ACTUARIAL RESEARCH CLEARING HOUSE
1985 VOL. 2

USING MODEL-BASED SMOOTHING AND EMPIRICAL BAYES ESTIMATION TECHNIQUES TO PREDICT AUTOMOBILE ACCIDENT FREQUENCIES FOR CALIFORNIA DRIVERS IN A TWO-WAY CLASSIFICATION

DRAFT

Thomas J. Tomberlin
Department of Quantitative Methods
Concordia University
Montréal, Québec

ABSTRACT

For predicting accident frequencies, a succession of log-linear models for Poisson data, some of which include nested random effects, is introduced. By applying maximum likelihood and empirical Bayes estimation techniques to these models, the notions of risk classification, model-based smoothing, credibility theory and experience rating are incorporated under a unified statistical approach to loss prediction. The performance of these methods is evaluated using accident data from California.

October 4, 1984
DRAFT

1. INTRODUCTION

Insurance serves the purpose of protecting individuals from the possibility of a large loss by exchanging that possibility for the certainty of a small loss in the form of a premium paid to an insurer. In theory, the consumer should be willing to pay the insurer an amount equal to the expected loss for the policy period, plus some additional premium to cover administrative expenses and company profits. Usually, that part of the premium equal to the expected loss, often referred to as the "pure premium", dominates the overall cost of insurance.

In general, the expected loss over the policy period is not known. For insuring against losses which are relatively common, such as automobile accidents, insurance companies generally make use of previous accident or claims experience to estimate expected future losses. However, the experience of any single individual is not sufficiently stable to yield reliable pure premium estimates. For this reason, actuaries generally place individuals into classes of similar risks, pooling their experience to obtain more stable estimates of pure premiums.

Ideally, individuals within these classes should be as homogeneous as possible with regard to their "true" expected losses. In statistical terms, the classification system should be a good discriminator for pure premiums.

For captive markets, such as those of state owned insurance companies or heavily regulated private companies, accurate risk assessment (pure premium estimation) is necessary in order for the system to be fair and equitable. For example, low risk drivers who find themselves in high risk classes can arguably feel unfairly treated by the system. (See for example, Shayer, 1978.) In competitive markets, an "unfair" classification system of one company can be exploited by another company, endangering the profitability of the first. Low risk individuals improperly classified into high risk categories by one company can be attracted to a second company with a "better" classification system, leaving the original company with a class of individuals who, on average, have larger expected losses than indicated by previous class experience. (See Doherty, 1981 and Tryfos, 1980.)

Companies are faced with conflicting requirements. For purposes of maintaining profitability and/or fairness, one would want as fine a classification system as possible. However, having too many classes would result in small numbers of individuals in any single class, making their pooled experience unreliable for estimating pure premiums.

One way to resolve this dilemma is to use a relatively fine classification system, but to impose structure on the system via a statistical model. Such a structure allows for pooling of information across categories yielding more stable estimates of pure premiums for relatively homogeneous groups of individuals.

Actuaries have long used such methods for estimating pure premiums for individuals in cells of a cross-classification of risks, but have only recently begun to formulate these methods in terms of statistical models. Bailey (1963), for example, described methods for obtaining estimates of pure premiums based on a multi-way classification. These are made up of products of factors associated with levels of each of the classification variables. In a two-way classification, such estimates take the form:

(1.1) $P_{ijk} = X_i Y_j$,

where P_{ijk} is the estimated pure premium for the kth individual in the ijth cell, and X_i and Y_j are factors associated with the ith and jth levels of the first and second classification variables respectively. Bailey also considered additive procedures and described methods for deriving estimates from data.

Recently, substantial effort has been invested in the reformulation of some of these traditional estimation techniques into model based estimates, and the exploration of the consequences of more complex models. See for example, Seal (1969), Harris and Van Slyke (1978), Chang and Fairley (1979), Massachusetts Automobile Rating and Accident Prevention Bureau (1981), Chamberlain (1980), Weisberg and Tomberlin (1982), Tomberlin (1982), and Weisberg, Tomberlin and Chatterjee (1984). Some of these methods, as well as the more traditional actuarial methods are described in Section 2.

In Section 3, a simple model for the claims generation process is introduced. It allows for separation of the two components of loss--the frequency component, how many claims are made; and the severity component, how much does each claim cost. In this paper, we concentrate on the prediction of the frequency component of loss. The severity component can be treated as a continuous variable, and standard linear model analysis employed, perhaps after a suitable data transformation.

Data corresponding to the frequency component of loss are discrete. Furthermore, for individuals, the range of observed numbers of accidents is quite limited, and analyses tailored for count data are required. It is common to approximate the distribution of the number of claims associated with any single individual by a Poisson. In Section 3, standard maximum likelihood methods for fitting models to Poisson data in a cross-classification are reviewed and empirical Bayes estimates based on nested random effects models are developed. It is shown that these models incorporate the actuarial notions of risk classification, credibility theory and experience rating discussed in Section 2, into a unified, model based estimation procedure. A more complete description of these estimates and associated algorithms can be found in Tomberlin (1982).

An evaluation of the proposed estimates based on accident data from California is presented in Section 4. These data are taken from the Longitudinal Study of California Driver Accident Frequencies conducted by the California Department of Motor Vehicles (Kwong, Kuan and Peck, 1976). Although not insurance data, this data set is especially useful for two reasons, it contains data on individuals, and it is longitudinal. These two characteristics allow for an evaluation of the year to year prediction capabilities of the detailed model based estimates of claim frequency described in Section 3.

Finally, conclusions drawn from this study, as well as directions for further research are reported in Section 5.

2. TRADITIONAL ACTUARIAL APPROACH

Much of traditional actuarial practice in estimating pure premiums is not based on any explicitly stated model or set of estimation principles. (See for example, Stern, 1965.) It is therefore not clear under what conditions such methods provide adequate information for setting insurance premiums.

Generally, data for estimation consists of the claims experience for a population over a given period. Pure premium estimates are obtained by calculating the observed per policy average claim cost, i.e.,

$$\text{Estimated pure premium} = \frac{\text{Total Losses}}{\text{Total Exposures}} \quad ,$$

where an exposure represents an insured risk over the time period of the insurance contract. For most practical purposes, the number of exposures can be thought of as the number of individuals insured.

2.1 Risk Classification

If there is reason to believe that pure premiums for individuals in the population are quite homogeneous, or if the cost of insurance is to be kept constant across the population, as for example is the case for many compulsory health insurance schemes, then the simple pure premium estimate described above should be adequate. However, it is more common that pure premiums vary substantially across the population, and estimates which reflect this diversity are required.

Generally, these estimates are obtained for subgroups of the population by imposing some kind of classification system. In automobile insurance, this usually consists of a cross-classification by two or more variables. For example, Massachusetts in 1978 used a driver class variable with 11 categories formed from a complex combination of individual driver characteristics as one dimension of the classification. For the other dimension, policies were divided into discrete territories according to where the automobile was principally garaged. These geographical subdivisions were intended to reflect variations in traffic conditions and other risk factors related to location. (Shayer, 1978)

2.2 Premium Estimation for Cross-Classifications of Risks

One of the most common methods of premium estimation for such a cross-classification is to assume that the effect of being in a particular driver class is to modify the pure premium by a constant multiplicative factor, and the effect of being in a particular territory is another constant factor. Such a multiplicative method was proposed by Almer (1957). A pure premium estimate for the ijth cell takes the form:

$$(2.2.1) \quad \pi_{ij} = x_i \, y_j \, ,$$

where the factors x_i and y_j are referred to as relativities.

In 1963, Bailey suggested a criterion for determining the relativities in a multi-way classification. He specified that the multiplicative method should yield cell estimates which were "balanced" for each of the classification variables. By "balanced", Bailey meant that single variable marginal estimated pure premiums should be equal to the corresponding marginal observed pure premiums. In a two-way classification, the requirement is that

$$(2.2.2) \quad \sum_j x_i y_j n_{ij} / n_{i\cdot} = p_{i\cdot}$$

$$\sum_i x_i y_j n_{ij} / n_{\cdot j} = p_{\cdot j}$$

where,

$$(2.2.3) \quad n_{ij} = \text{number of exposures in cell } ij$$

$$p_{ij} = \text{observed pure premium in cell } ij$$

$$p_{i\cdot} = \sum_j n_{ij} \, p_{ij} / n_{i\cdot} = \text{marginal observed pure premium in the } i\text{th row}$$

$$p_{\cdot j} = \sum_i n_{ij} \, p_{ij} / n_{\cdot j} = \text{marginal observed pure premium in the } j\text{th column.}$$

The set of equations (2.2.2) can be solved iteratively for x_i and y_j. In Section 3, it is shown that the same estimates can be obtained via the iterative proportional fitting (IPF) algorithm described by Deming and Stephan (1940). Bailey termed estimates obtained via

(2.2.2), "minimum bias multiplicative" estimates, and went on to propose additive estimates satisfying similar balancing constraints.

Bailey's multiplicative method seems to have been widely adopted by the insurance industry. Indeed, the latest reference to the subject of pure premium estimation in the 1981 curriculum studied by casualty actuarial students in preparation for their examination was still the 1963 Bailey paper.

Recently, various investigators have proposed additive and multiplicative statistical models for pure premium data in cross-classifications. Chang and Fairley (1979) and Fairley, Tomberlin and Weisberg (1981) used least squares techniques for estimation based on ANOVA type additive models. Sant (1980) proposed a multiplicative model, and Derrig and Conger (Massachusetts Automobile Rating and Accident Prevention Bureau, 1981) proposed a hybrid model which contains the additive and multiplicative models as special cases. For all three models, the authors used least squares to estimate expected losses or pure premiums.

2.3 Credibility Theory: Experience Rating For Classes.

Even before adopting the structure based estimation methods discussed in Section 2.2, actuaries were using Bayesian, or Bayesian-like premium estimates which they called credibility estimates. In their simplest form for a one-way classification of risks, credibility estimates can be expressed as,

$$(2.3.1)\quad \pi_i = Z\, p_i + (1-Z)\, P,$$

where p_i is the observed pure premium in the ith class, P is referred to as the "manual" rate, based on a large number of exposures and thus regarded as a constant, and Z is a credibility weight chosen so as to reflect the amount of information in the observed pure premium, p_i, relevant to the ith risk class.

As originally proposed by Whitney (1918), the credibility weight Z was obtained using Bayesian reasoning. He assumed that claim severity was constant, so that premiums could be determined solely on the basis of the number or frequency of claims. Further, he modeled the number of claims in the ith class, f_i, as binomial with parameters q_i and n_i. The parameter q_i was assumed to be a realization of a normal random variable C having mean μ and variance σ^2. He showed that the mode of the posterior distribution of q_i given the data f_i could be approximated by

$$(2.3.2) \quad \pi_i = \frac{f_i \vartheta^2 + \mu^2}{n_i \vartheta^2 + \mu}$$

Further, if each claim cost one unit, then (2.3.2) can be re-expressed in the form of (2.3.1) as

$$(2.3.3) \quad \pi_i = p_i \frac{n_i \vartheta^2}{n_i \vartheta^2 + \mu} + \mu \frac{\mu}{n_i \vartheta^2 + \mu}$$

where the "manual" rate, P, is set to the mean of the prior distribution of C, and the credibility weight is given by

$$(2.3.4) \quad Z = \frac{n_i \vartheta^2}{n_i \vartheta^2 + \mu} = \frac{\vartheta^2}{\vartheta^2 + \mu / n_i}$$

Whitney went on to conduct a kind of empirical Bayes analysis, estimating the parameters of his normal prior from the data. Later, A. L. Bailey (1950) obtained similar credibility factors for Beta-Binomial models, and Seal (1969) for Normal-Poisson and Gamma-Poisson models.

Subsequent to Whitney's landmark work, practicing actuaries seem to have lost sight of the underlying Bayesian nature of credibility theory. Instead of attempting to estimate the scale parameters of the prior distribution from the data, they focused on developing simple, easy to use approximations of Z as a function of the number of exposures or number of claims in the data. They noted from (2.3.4) that for a large number of exposures, n_i, Z approaches one. They then set about identifying minimum requirements for setting Z to one, a condition they called "full credibility". The criterion for attaining full credibility was sometimes based on the number of exposures, and later more commonly on the number of claims. For example, Stern (1965) in his classical review article recommended that a minimum of 1084 claims be observed for establishing full credibility.

Having set a standard for full credibility, there remained the need for a means of determining Z when the criterion for full credibility was not met. Various functions relating the number of exposures or number of claims to Z for "partial credibility" were considered in the literature. Norberg (1979), in an extensive review of credibility theory termed this type of credibility "limited fluctuation theory". He traced its origins to a paper by Mowbray (1914) which was aimed at determining the amount of experience necessary for obtaining a "dependable" pure premium estimate.

The original Bayesian nature of credibility theory was re-introduced to practicing actuaries by A. L. Bailey (1946, 1950). It has since received much attention. See for example, Norberg (1979), Kahn (1975), and Buhlmann (1970). Obviously, Bayesian credibility does away with determining minimum experience for "full credibility", since that is only achieved in the limit. Furthermore, approximate formulas for partial credibility are not needed. On the other hand, a prior distribution must be specified, and techniques for estimating its parameters to allow for empirical Bayes estimation must be devised. One such method based on normal priors for observed pure premiums was suggested by Morris and Van Slyke (1978). Another based on log-normal priors for expected claims frequency is proposed in Section 3 of this paper.

2.4 Merit Rating: Experience Rating for Individuals

A Bayesian or empirical Bayesian analysis can be applied to individual risks in a manner similar to Bayesian credibility for classes. This has been recognized for some time. (Seal, 1969 and Buhlmann, 1970.) However, in practice, merit rating is generally accomplished by establishing another classification determined by previous loss experience. (In addition to previous loss experience, the use of information on traffic violations is also referred to as merit rating. Such systems are not considered here.)

One common merit rating scheme consists of forming a classification variable determined by the number of full years since the individual's latest accident or since licensing. (Bailey and Simon, 1959 and Ferreira, 1978.) This classification variable is then treated as any other and class relativities are computed in the usual manner. Note that a number of years of data is required for such a system. Several years are needed to establish class membership. Relativities are computed based on the actual loss experience of these drivers in a subsequent year.

There are some obvious shortcomings in this type of merit rating scheme. First, data required for setting up a merit classification variable cannot be used for calculating pure premiums and it is not clear that this is the most efficient use of the wealth of data that is available. Second, no use is made of the inherent ordering in such a classification variable. More reliable pure premium estimates could probably be obtained by imposing more structure via a model. Kwong, Kuan and Peck (1976) did just that using linear regression models in their analysis of accident rates in California. Finally, when merit rating surcharges are applied independently of other classification variables, the interaction between the two operations--model-based smoothing and merit rating--cannot be predicted. Such is the case in the Massachusetts system described by Ferreira (1978).

Recently, there has been a move to apply the tools of Bayesian credibility theory to the problem of individual experience rating. Ralph (1981) describes an empirical Bayes experience rating scheme for medical malpractice insurance which is based on gamma-Poisson models for frequency within rating classes. Zehnwirth (1979) and Hackman (1980) considered methods for incorporating the two stages of credibility estimation, described here as class and individual experience rating, into a single comprehensive methodology. Zehnwirth proposed a kind of empirical Bayes estimator for individual expected claim frequencies based on a hierarchical set of priors for individuals within classes. In each case, estimates of pure premiums for individuals are obtained via credibility theory at appropriate levels of aggregation.

In the following section, we propose a method for setting pure premiums which is similar in spirit to those proposed by Zehnwirth and Hackman. There are two principal differences. Hackman used pure premium data where here, data on claims frequency and claim cost are treated separately. (Actually, Hackman's models are based on loss ratios which are defined to be the ratio of total losses to total collected premiums. This data format is commonly used by actuaries, but seems to this author to be unnecessarily complicated.) The approach proposed here also incorporates model estimates using the structure of cross-classification of risks where both Zehnwirth and Hackman's models are aimed solely at deriving credibility estimates in a one-way classification.

October 5, 1984
DRAFT

3. MODEL BASED PREDICTION OF CLAIMS FREQUENCY

3.1 Stochastic Models for Insurance Data

Models for the claims generation process generally focus on the total loss associated with claims generated by an individual or group of individuals over a certain period of time. (See Buhlmann, 1970 and Seal, 1969.) The process is generally modelled as two independent processes, the discrete occurrence of claims, and the claim amount associated with each individual claim.

Let f be the number of claims made over a certain period, Y_i be the claim amount associated with the ith claim, and

$$(3.1.1)\quad L = \sum_{i}^{f} Y_i$$

be the total loss for the period. Then assuming the two processes are independent, the cumulative distribution function (c.d.f.) of L can be expressed as,

$$(3.1.2)\quad P(L) = \sum_{f=0}^{\infty} p(f)\, P_f^{*}(L)\,,$$

where P_f^{*} is the f-fold convolution of the c.d.f. of the claim amount variable Y.

Provided that the p.d.f's of f and Y are reasonably simple, the moment generating function of L can be easily derived from (3.1.2). However, it is very difficult to get exact results or even good approximations for the distribution function itself. (Seal, 1969 and Beard, Pentikäinen and Pesonen, 1977.) Generally, simulation studies are required for studying the effects of changing frequency and severity distributions.

For purposes of pure premium estimation, one need not be concerned with the complex form of P(L). As long as the two processes, frequency and severity, are independent, the parameters of the two can be independently estimated.

Rather than summarizing the data using (3.1.1), by adding losses over the period, let the data for the ith insured individual be given by,

$$(3.1.3)\quad (f_i\,, x_i) = (f_i\,, x_i(1)\,, x_i(2)\,, \ldots)\,,$$

where f_i is the number of claims and $x_i(j)$ is the cost associated with the jth claim of the ith individual. For j greater than f_i, $x_i(j)$ is defined to be zero. Then, if there are n individuals, the likelihood of the data is given by,

$$(3.1.4) \quad \prod_{i=1}^{n} p(f_i, x_i) = \prod_{i=1}^{n} p(x_i \mid f_i)\, p(f_i)$$

Further, if the claim amounts $x_i(j)$, $j \le f_i$ are assumed independent identically distributed and independent of f_i, this can be simplified to

$$(3.1.5) \quad \prod_{i=1}^{n} p(f_i) \prod_{j=1}^{f_i} p(x_i(j)) = [\prod_{i=1}^{n} p(f_i)]\,[\prod_{i=1}^{n} \prod_{j=1}^{f_i} p(x_i(j))].$$

From (3.1.5), it can be seen that so long as the distributions of frequency and amount have no parameters in common, and the parameters of the two distributions are a priori independent, they can be estimated separately.

The assumption that the parameters of the frequency and severity distributions should be independent is non-trivial. Jewell (1980) notes that correlations between frequency and average claim size have been observed in studies of automobile insurance claims. Such a phenomena could occur if the parameters of the two distributions were related. There are two ways to address such a dependence. First, the relationship between the two sets of parameters could be modelled, and included in the likelihood. However, the analysis of data based on such models could be quite complicated. Second, the dependence can be reduced, if not eliminated, by conditioning on other variables.

In this paper, the problem of possible relationships between frequency and severity is not addressed. Instead, we only consider the prediction of the frequency component of loss. However, estimation conditional on categorical variables defining a cross-classification of risks is considered, and thus, the methodology for reducing the relationship by conditioning is developed. Weisberg, Tomberlin, and Chatterjee (1984) describe an empirical investigation into the consequences of treating the two components independently. They conclude that such a methodology does not adversely affect the quality of the estimates. This observation, coupled with the more extensive set of methodologies available justifies the treatment of the two components separately

3.2 Models for Predicting Frequency

Adopting the common assumption that accidents and associated claims occur via a Poisson process, the frequency, f, associated with the an individual is a Poisson random variable with parameter ζ,

$$(3.2.1) \quad p(f) = e^{-\zeta} \zeta^{f} / f!$$

We propose to predict these rates using a series of log-linear models, incorporating, in succession, the three actuarial notions of risk classification, credibility and experience rating in a uniform statistical theory of estimation.

For purposes of illustration, the population of policy holders is assumed to be classified by two categorical variables. The indices (i,j,k) refer to the kth individual within the ijth cell of the cross-classification. Let n_{ij} be the number of individuals within the ijth cell, and let f_{ijk} represent the number of events occuring to the kth individual within the ijth cell during a specified period of time, and f represent the complete set of frequencies f_{ijk}.

For the first three models considered below, it is assumed that the number of events, f_{ijk}, is distributed as a Poisson random variable with intensity parameter ζ_{ij}. That is, within a cell, individuals are assumed to have a common intensity parameter. For these models, let

$$(3.2.2) \quad f_{ij} = \sum_{k} f_{ijk}$$

represent the total number of events occurring to members of the ijth cell. Then f_{ij} is distributed as a Poisson random variable with parameter $n_{ij} \zeta_{ij}$. Prediction of accident rates can be made on the basis of estimates of the accident propensity parameters ζ_{ij}.

The simplest estimate for members of the ijth cell is the within cell maximum likelihood estimate of ζ_{ij}, f_{ij}/n_{ij}. This procedure incorporates the rudiments of a risk classification system. Individuals who are thought to have similar accident propensities on the basis of some set of classification variables are grouped together, and their pooled accident experience used to predict future accident rates.

The simple estimate described above can be obtained from the completely saturated log-linear model:

$$(3.2.3)\quad \text{MODEL 0:}\ \ln C_{ij} = l_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

As such, the u-terms are not unique, but can be made so by adopting the usual analysis of variance model constraints,

$$(3.2.4)\quad \sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_{ij} u_{12(ij)} = 0 .$$

See Bishop, Fienberg and Holland (1975) or Haberman (1974) for a more complete discussion of log-linear models for count data in general, and for Poisson data specifically.

Let u_{ij} refer to the u-terms for the ijth cell, and u the complete set of u-terms for all cells in the cross-classification. Knowledge of u is the same as knowledge of the C_{ij}. Let this relationship be denoted by

$$(3.2.5)\quad C_{ij}(u) = \exp[u + u_{1(i)} + u_{2(j)} + u_{12(ij)}] .$$

Then, the likelihood of the data is given by

$$(3.2.6)\quad p(f \mid u) = \frac{\exp[-\Sigma\, n_{ij}\, C_{ij}(u)]\ \Pi\, C_{ij}(u)^{f_{ij}}}{\Pi\, f_{ij}!}$$

Model 0 is "completely saturated" and corresponding estimates of expected cell frequencies are simply the observed cell frequencies. There is no pooling of information between cells. For a very large population or sample with sufficient number of individuals within each cell, Model 0 estimates may be acceptable. However, it is more common for at least some cells to be too small to produce useful estimates. In such cases, information from cells within the same category of at least one classification variable can be used to produce more stable estimates.

As described earlier, Bailey and Simon (1960) and Bailey (1963) accomplished this for pure premium data by various additive and multiplicative estimation methods. For frequency data, a similar effect can be achieved with a log-linear model having no interaction term.

(3.2.7) MODEL I: $l_{ij} = u + u_{1(i)} + u_{2(j)}$

In fact, it is shown in Section 3.3.1 that maximum likelihood estimates based on Model I for frequency data are numerically equivalent to Bailey's "minimum bias multiplicative method" estimates for pure premium data.

Many times some cells of a classification contain enough data so that no pooling across categories is necessary, and Model 0 would seem appropriate. However, other cells within the same classification may be small so that some pooling of data across categories, as with Model I, is required. In fact Models 0 and I are two extremes, and some compromise is appropriate using individual cell data to the extent that it is reliable, while allowing for some pooling of information across categories, as required.

From a statistical perspective, it makes sense to view the estimation of several Poisson rates as a multiple parameter estimation problem. Stein (1955) considered such a problem in estimating a vector of normal means. The estimator he developed "shrinks" all individual estimates towards some common point and can be shown to dominate the classical unbiased estimator under several loss functions, including squared error loss. The decision theoretic work of Stein has been extended to include the simultaneous estimation of parameters of non-normal populations, in particular, Poisson means. See for example, James and Stein (1961), Clevenson and Zidek (1975), Peng (1975), Hudson and Tsui (1981) and Albert (1981). For the present application, these developments share the disadvantage of requiring that the common "shrinking point" be specified prior to analysis.

Efron and Morris (1972, 1973) used Bayes and empirical Bayes justifications for their James-Stein type estimators. Under the empirical Bayes approach, shrinkage toward an unspecified common mean is allowed. Using variance stabilizing transformations, Carter and Rolph (1974) applied these normal theory empirical Bayes estimators to binomial data for estimating probabilities of false fire alarms.

Here, log-linear models with fixed and random effects are employed for the purpose of multiple parameter estimation under an empirical Bayes framework. Similar methods have been employed in the past in the context of mixed linear models, though their connection with Bayesian reasoning has not been explicit. For example, Henderson derived estimates for individual random effects based on the assumption of a known function of the variance components. His work is reported in a paper by Henderson, Kempthorne, Searle and von Krosigk

(1969). Estimates developed here are similar in spirit though the variance components are parameters of prior distributions.

We adopt the empirical Bayes approach termed by Maritz (1970) "smooth empirical Bayes procedures". A common prior distribution is specified for a group of parameters. The parameters of the prior are then estimated from the data. Finally, Bayes estimates are formed using the estimated prior parameters.

For Poisson data in categories, the conjugate prior is the multivariate gamma distribution. (See Fienberg and Holland, 1973.) However this prior does not easily allow for the exploitation of the structure in a contingency table via the use of log-linear models. On the other hand, the multivariate normal distribution adapts readily to these linear, ANOVA type models, and has received some attention.

Laird (1975, 1978) employed normal priors for log-linear models with fixed and random effects to obtain empirical Bayes estimates for cell probabilities in a multiway contingency table. The development of frequency predictors based on the following two models draws heavily on this work by Laird.

One can obtain empirical Bayes estimates for count data in a log-linear model by specifying some of the parameters as random. Instead of dropping interaction terms, we specify them as random effects, and assume a normal prior distribution for them. Consider,

(3.2.8) $\quad$ MODEL II: $l_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$

(3.2.9) $\quad$ $u_{12(ij)} \sim NORMAL(0, \sigma^2)$

In effect, this specification means no a priori information exists indicating systematic departures from Model I. For small cells, the prior distribution (3.2.9) dominates the posterior of the $u_{12(ij)}$ terms, yielding posterior estimates which are close to zero. Thus, the estimated accident rates, $\hat{C}_{ij}$, from such cells are close to those which would be obtained under Model I. Data from large cells dominate prior information in the posterior distribution of the interaction terms associated with these cells so that the corresponding estimated frequency rates are close to observed rates. That is, under Model II, large cells produce estimates close to those obtained under Model 0. For moderate size cells, Model II yields estimates which are a compromise between Models 0 and I.

Estimates which combine information from individual experience and pooled group experience are classically referred to as credibility estimates by actuaries. Combining model estimates and individual cell estimates is somewhat different but the idea is the same. Actuaries have noted that weights for combining individual and group experience can be obtained via Bayesian analysis. Morris and Van Slyke (1978) have demonstrated empirical Bayes estimates for insurance premiums in a one-way classification. Model II demonstrates that methods for combining pooled cell estimates and model estimates for Poisson frequencies can also be obtained via empirical Bayes techniques.

Laird (1976) showed that specifying u, a_1, and a_2 as *fixed effects* is equivalent to placing a flat prior on these terms. Thus, we have the joint distribution of f and a given by

$$(3.2.10)\quad p(f,a \mid \sigma^2) = p(f \mid a,\sigma^2)\, p(a \mid \sigma^2)$$

$$= \frac{\exp[-\Sigma n_{ij}\, \xi_{ij}(a)]\ \Pi\, \xi_{ij}(a)^{f_{ij}}}{\Pi\, f_{ijk}!}\ [\sqrt{2\pi}\,\sigma]^{-RC} \exp[\Sigma a_{12(ij)}{}^2/2\sigma^2]$$

Each of Models 0, I, and II share the assumption that within a cell, individuals have a common underlying propensity to accidents, as quantified by their expected number of accidents, ξ_{ij}. Several studies have shown that substantial variation in expected accident rates exists among individuals within cells of a cross-classification based on commonly available variables. See for example, Casey, Pezier and Spetzler (1976). In fact the individual intensity parameters, ξ_{ij} have a distribution within cells.

In using a common cell average rate for each individual within a class, those individuals having lower than average accident propensities would be penalized for the higher propensities of their fellow drivers. As noted in Section 2, such penalties not only call into question the fairness of the system, but also can endanger the profitability of a company.

The use of individual claims histories in setting rates is referred to as "experience rating". (See Felk and Bolnick, 1975.) It is usually applied to group policies in health and workman's compensation insurance where the experience of the group may well be substantial. For automobile insurance, the use of individual claim experience, together with other individual information such as records of traffic violations, in setting rates is referred to as "merit rating". (See Ferreira, 1978.) Usually these merit rating systems are set up separately from the classification system, and involve a surcharge or bonus applied to the basic rate for individuals in a cell.

By adding a term for individuals within classes to Model II, the notion of experience rating can be incorporated within a unified approach for predicting accident rates based on a statistical model.

$$(3.2.11) \quad \text{MODEL III:} \; f_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} + u_{123(ijk)}$$

$$(3.2.12) \quad u_{12(ij)} \sim \text{NORMAL}(0, \delta_1^2)$$

$$(3.2.13) \quad u_{123(ijk)} \sim \text{NORMAL}(0, \delta_2^2)$$

Again, as with Model II, specifying u, u_1, and u_2 as fixed effects is equivalent to placing flat priors on them, and the joint distribution of f and u is given by,

$$(3.2.14) \quad p(f, u \mid \delta_1^2, \delta_2^2) = p(f \mid u, \delta_1^2, \delta_2^2)\, p(u \mid \delta_1^2, \delta_2^2)$$

$$= \frac{\exp[-\Sigma\, C_{ijk}(u)]\; \Pi\, C_{ijk}(u)^{f_{ijk}}}{\Pi\, f_{ijk}!} [\sqrt{2\pi}\,\delta_1]^{-RC} \exp[-\Sigma u_{12(ij)}{}^2/2\delta_1{}^2]$$

$$\cdot [\sqrt{2\pi}\,\delta_2]^{-n} \exp[-\Sigma u_{123(ijk)}{}^2/2\delta_2{}^2]$$

Thus, by adding terms to models, the actuarial notions of risk classification, credibility and experience rating can be incorporated into a single, unified statistical approach. Now, let us turn our attention to deriving estimators for these models.

3.3 Model Based Estimates for Predicting Frequency

3.3.1 Models 0 and 1

Model 0 describes the data, f, as arising from R C independent Poisson processes, with intensity parameter $n_{ij} \zeta_{ij}$ within the ijth cell. No relationship between cells is implied by the model, and maximum likelihood estimates are obtained in the usual manner as the average rate within each cell. That is, under Model 0, the estimated individual expected accident rates are

(3.3.1) $\zeta_{ij} = f_{ij}/n_{ij}$

Model 1, on the other hand, sets up a structure relating the individual cell rates. The likelihood of the data is given by (3.2.6) where

(3.3.2) $\zeta_{ij}(u) = \exp(u + u_{1(i)} + u_{2(j)})$

Maximum likelihood estimates of the ζ_{ij} are the solutions of the normal equations:

(3.3.3) $$\frac{\partial}{\partial u} \ln p(f|u) = \sum_{ij}[\, -n_{ij}\, \zeta_{ij}(u) + f_{ij}\,] = 0$$

(3.3.4) $$\frac{\partial}{\partial u_{1(i)}} \ln p(f|u) = \sum_{j}[\, -n_{ij}\, \zeta_{ij}(u) + f_{ij}\,] = 0$$

(3.3.5) $$\frac{\partial}{\partial u_{2(j)}} \ln p(f|u) = \sum_{i}[\, -n_{ij}\, \zeta_{ij}(u) + f_{ij}\,] = 0 .$$

Note that (3.3.3) is implied by (3.3.4) and (3.3.6). Furthermore, defining f^{*}_{ij} to be the expected frequency in the ijth cell, and using the usual "dot notation", the normal equations simplify to

(3.3.6) $f^{*}_{i\cdot} = f_{i\cdot}$

(3.3.7) $f^{*}_{\cdot j} = f_{\cdot j}$

That is, the normal equations imply that the maximum likelihood estimates of the ζ_{ij} should be such that the resultant fitted frequencies have row and column margins which match the observed row and column margins.

Haberman (1974) shows that the IPF algorithm can be adapted for solving these normal equations. The procedure starts with a table made up of the exposures, n_{ij}. Entries in the table are then rescaled successively so that first, estimated row margins match observed row margins, and second, estimated column margins match observed column margins. Iterations continue until equations (3.3.6) and (3.3.7) are satisfied.

Note that these constraints for estimating frequencies are the same as Bailey's (1963) balancing constraints (2.2.2) for estimating pure premiums. Thus, it can easily be seen that Bailey's minimum bias multiplicative estimates are maximum likelihood in the event that claim severity is constant. In addition, this yields an alternative method for obtaining Bailey's estimates. By starting with a table of exposures, n_{ij}, and using the IPF algorithm to force the estimated total marginal losses to match observed marginal losses, the balancing constraints on pure premium estimates can be satisfied.

3.3.2 Models II and III: Bayes estimates.

A traditional Bayesian analysis of Model II requires the specification of the prior variance of the $u_{12(ij)}$ terms, σ^2, in advance. Estimation then is based on the posterior distribution of the u-terms given σ^2 and the observed frequencies, f:

$$(3.3.8) \quad p(u|f, \sigma^2) = \frac{p(f,u|\sigma^2)}{p(f|\sigma^2)}.$$

This expression is difficult to work with analytically because of the intractable integration required to obtain the denominator.

Following suggestions by Laird (1978) and Leonard (1975), the posterior distribution in (3.3.8) can be approximated by a normal distribution having its mean at the mode of $p(u|f,\sigma^2)$ and its covariance matrix equal to the inverse of the second derivative matrix of $p(u|f,\sigma^2)$ evaluated at the mode. In effect, this is an approximation of the log posterior distribution at its mode by a quadratic in u. In calculating the mode and second derivative matrix, the denominator of the right hand side of (3.3.8) can be ignored since it does not depend on u.

The modal values of the u-terms can be determined by setting the corresponding partial derivatives of the posterior distribution to zero and solving. For main effects this yields equations (3.3.3-5), where here,

$$(3.3.9)\quad \mathcal{E}_{ij}(u) = \exp(u + u_{1(i)} + u_{2(j)} + u_{12(ij)}).$$

Consequently, the mean for the normal approximation of the posterior distribution under Model II must yield expected marginal frequencies which match observed marginal frequencies. That is, equations (3.3.6) and (3.3.7) must be satisfied.

In addition, the partial derivative with respect to $u_{12(ij)}$ must be zero,

$$(3.3.10)\quad \frac{\partial}{\partial u_{12(ij)}} \ln p(f,u \mid \sigma^2)$$

$$= -n_{ij}\exp(u + u_{1(i)} + u_{2(j)} + u_{12(ij)}) + f_{ij} - \frac{u_{12(ij)}}{\sigma^2} = 0.$$

In the same way that the IPF algorithm can be used to preserve the factor n_{ij} in the final fitted frequencies associated with Model I, it can be used to preserve an arbitrary, pre-specified interaction structure. By starting the algorithm with the value $n_{ij}\exp(u_{12(ij)})$ in the ijth cell, the model

$$(3.3.11)\quad \ln \mathcal{E}_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

can be fit, where u, u_1, and u_2 are unspecified.

Using these properties, the set of equations (3.3.6), (3.3.7), and (3.3.10) can be solved with the following iterative procedure. First, set the $u_{12(ij)}$ equal to zero and solve (3.3.6-7) using the IPF algorithm starting with the value n_{ij} in the ijth cell. That is, fit Model I. Second, using the values for u, $u_{1(i)}$, and $u_{2(j)}$ obtained from the IPF step, solve (3.3.10) for $u_{12(ij)}$ using a Newton-Raphson algorithm. Cycle back to the IPF step, starting the ijth cell at $n_{ij}\exp(u_{12(ij)})$ and solve for u, $u_{1(i)}$, and $u_{2(j)}$. That is, fit model (3.3.11), fixing $u_{12(ij)}$ at the value obtained at the previous step. Continue to cycle back and forth until the algorithm converges.

The approximate posterior covariance matrix, Σ can be obtained by inverting the second derivative matrix of the log of the right hand side of (3.3.8) evaluated at u^*, the mode. That is,

$$(3.3.12)\ \Sigma = -\left(\frac{\partial^2}{\partial u \partial u^T}\left[\ln p(f, u \mid \delta^2)\right]\right)^{-1}.$$

evaluated at $u = u^*$.

By listing the two-way tables of numbers of exposures and lambda parameters, row by row, into column vectors n and λ, model II can be restated in matrix form as,

$$(3.3.13)\ \ln(\lambda) = 1 \cdot Du = [D_1 \quad I_{RC}]u\ ,$$

where the logarithm of a vector is defined to be the vector of logarithms, u is a vector of log-linear parameters and D_1 is the usual two-way classification design matrix. Using this notation, (3.3.12) can be re-expressed as the following partitioned matrix,

$$(3.3.14)\ \begin{vmatrix} D_1^T \operatorname{diag}(n \cdot \lambda(u^*))\, D_1 & D_1^T \operatorname{diag}(n \cdot \lambda(u^*)) \\ \operatorname{diag}(n \cdot \lambda(u^*))\, D_1 & \operatorname{diag}(n \cdot \lambda(u^*)) + I_{RC} / \delta^2 \end{vmatrix}$$

Here, "·" represents an element by element product. In order to insure non-singularity of the matrix in (3.3.14), those rows of the design matrix D corresponding to the last row and column of the table have been deleted. The approximate posterior covariance matrix of the estimated model parameters, Σ, is then obtained by inverting (3.3.14). As the matrix is partitioned, and the lower right corner is diagonal, this only requires the inversion of a matrix of order R + C - 1.

Model III, described in (3.2.13-14), contains an additional set of random effects corresponding to individuals within cells of the classification. Bayesian analysis, which requires the specification of the two variance components, δ_1^2 and δ_2^2, proceeds in a manner similar to that for Model II. Finding the mode u^* of the posterior distribution requires solving the following system of equations:

$$(3.3.15)\ f^*_{i\cdot} = f_{i\cdot}$$

$$(3.3.16)\ f^*_{\cdot j} = f_{\cdot j}$$

$$(3.3.17)\ f_{ij\cdot} - f^{*}_{ij} = \frac{u^{*}_{12(ij)}}{\lambda_1^2}$$

$$(3.3.18)\ f_{ijk} - f^{*}_{ijk} = \frac{u^{*}_{123(ijk)}}{\lambda_2^2}$$

This set of equations can be solved using using an algorithm similar to that used for Bayesian analysis of Model II, an iterative combination of an IPF step and a Newton–Raphson step. The algorithm begins with all interaction and individual terms set to zero. The fixed effects, u, $u_{1(i)}$, and $u_{2(j)}$ are obtained via application of IPF, fitting the observed margins, $f_{i\cdot\cdot}$ and $f_{\cdot j\cdot}$ to the table of exposures n_{ij}. The fixed effects are set to these solutions, and a multivariate Newton–Raphson method is applied to solve for $u_{12(ij)}$ and $u_{123(ijk)}$. Cycle back to the IPF step, starting the algorithm with

$$(3.3.19)\ \sum_{k} \exp[u_{12(ij)} + u_{123(ijk)}]$$

in the ijth cell, where the $u_{12(ij)}$ and $u_{123(ijk)}$ terms are those obtained in the previous Newton–Raphson step. Continue to cycle back and forth between these two steps until the combined algorithm converges.

October 5, 1984
DRAFT

3.3.3 Models II and III: Empirical Bayes estimates

A Bayesian analysis of Model II requires that the prior variance of the $u_{12(ij)}$ terms, σ^2, be specified. Here there are RC $u_{12(ij)}$ terms, and we are in a good position to estimate this parameter from the data. That is, this is a multiple parameter estimation problem which lends itself to an empirical Bayes treatment.

The prior variance can be estimated using an EM algorithm as described by Dempster, Laird and Rubin (1977). In their terminology, the u terms are treated as missing data. The E-step of a single iteration of the algorithm consists of deriving the expected value of the log-likelihood, $\ln p(f, u \mid \sigma^2)$, given the data, f, and a current value of σ^2. At the M-step, σ^2 is estimated by maximizing the resultant likelihood function. That is, the M-step consists of updating σ^2 with the maximum likelihood estimate based on the expected log-likelihood obtained at the previous M-step. For regular exponential densities like that in (3.2.10), one need only calculate the expected value of the sufficient statistic at the E-step, in this case $\Sigma u^2_{12(ij)}$. At the subsequent M-step, σ^2 is updated to the maximum likelihood function of this expected value, in this case,

$$(3.3.20) \quad E\{\Sigma u^2_{12(ij)}\} / RC$$

Thus, at the (t+1) iteration of the EM algorithm, one requires the conditional distribution of the $u_{12(ij)}$ given the observed frequencies f and the estimate of σ^2 obtained from the previous M-step, σ^2_t.

$$(3.3.21) \quad p(u_{12} \mid f, \sigma^2_t)$$

As described in Section 3.3.2, we approximate this by a multivariate normal distribution. At iteration (t+1), the updated value of the prior variance, σ^2_{t+1}, is given by,

$$(3.3.22) \quad \sigma_{t+1} = \Sigma\{(u^*_{12(ij)}(t))^2 + Var_t(u_{12(ij)})\} / RC$$

where $u^*_{12(ij)}(t)$ and $Var_t(u_{12(ij)})$ are the mean and variance obtained from the normal approximation at step t of the EM algorithm.

Iterations of the EM algorithm are continued until it converges. Dempster, Laird and

Rubin (1977) show that the likelihood increases at each cycle of the algorithm. Thus, assuming that the likelihood, $p(f \mid \sigma^2)$, is convex, the algorithm will eventually converge.

It is also possible to approximate the likelihood of σ^2 at a set of points using $p(u, f \mid \sigma^2)$ and the normal approximation of u, together with equation (3.3.8). Substituting the normal approximation for $\ln p(u \mid f, \sigma^2)$ at u^*, we have

$$(3.3.22) \quad \ln p(f \mid \sigma^2) = K(f) - \Sigma n_{ij} C_{ij}(u^*) + \Sigma f_{ij} \ln[C_{ij}(u^*)]$$

$$- \{\ln(2\pi\sigma^2) - \Sigma[u^*_{12(ij)}]^2 / \sigma^2 - \ln|\Sigma^{-1}|\} / 2$$

The only heavy computation required to evaluate (3.3.22) is the determinant of Σ^{-1}, a matrix of order RC + R + C - 1, with an RC X RC diagonal matrix in the lower right hand corner. The prior variance can then be estimated by plotting the log-likelihood (3.3.22) at a set of points and approximating the maximum of the resulting graph. This can be used as a final estimate, or can be used to obtain a starting point for the EM algorithm. In addition, the shape of the likelihood gives information regarding the stability of the estimated variance.

Empirical Bayes estimates for Model III are obtained in a manner similar to that used for Model II. There are two sets of random effects in this model, and consequently two prior variances to estimate. The EM algorithm can be adapted to this problem, though the computational problems are much more difficult. For example, the covariance matrix for the normal approximation of the posterior distribution of the u-terms is of the order RC + R + C - 1 + n.., where n.. is the number of observations in the data set. For the California data set, there are over 100,000 observations. One cannot produce the whole matrix. However, for the EM algorithm, one requires only the trace of this matrix, and this can be computed. The details of these calculations as well as those required for Model III can be found in Tomberlin (1982).

4. PREDICTING ACCIDENT RATES IN CALIFORNIA

Accident frequencies were predicted using data obtained from the Longitudinal Study of California Driver Accident Frequencies. (See Kwang, Kuan and Peck, 1976.) For purposes of illustration, drivers were classified by two variables--place of residence and driver class. Driver class is actually a one-dimensional classification based on three variables--age, sex and marital class. Driver class definitions are given in Table I. The place of residence variable divided drivers up into 16 territories by zip codes defined in such a way as to separate central cities, suburbs and rural areas. The distribution of drivers in this cross-classification is given in Table II and observed 1972 accident rates per 100 drivers is given in Table III.

All calculations were carried out using the MATRIX procedure in SAS. (See SAS Institute, Inc., 1979.) A program listing is available in Tomberlin (1982).

4.1 Models I and II

First, the data were smoothed by fitting Model I, the fixed effects log-linear model. This result is obtained from the first iteration of the EM algorithm for Model II estimates with the interaction terms set to zero. The results are presented in Table IV, expressed in expected numbers of accidents per 100 drivers.

For Model II, a starting value for the EM algorithm was obtained by calculating the approximate likelihood (3.3.22) at a set of values for γ^2, and then graphically approximating the maximum. The final estimate was 0.0067. The corresponding empirical Bayes estimates of cell accident rates are given in Table V.

A comparison of the three tables indicates that Model II estimates perform as anticipated. Where there are substantial numbers of drivers in a cell (e.g., the 9862 middle aged, married males in territory 1), the empirical Bayes estimates are relatively closer to the observed accident rate than when the number of drivers is small (e.g., the 53 young married males in Territory 12). Thus, the empirical Bayes estimates allow the observed cell experience to be used when it is reliable (or credible) and the smoothed model estimates to be used when only a small amount of data are available. In effect, the Model II empirical Bayes estimator is a kind of credibility estimator.

Measures comparing the fit of the Models I and II estimates are given in Table VI under the heading "Fit to 1972 rates". These measures are based on the differences between observed and estimated accident rates within cells. The mean absolute error per 100 is given by

$$(4.1.1) \quad 100 \times \sum | f_{ij} - \hat{f}_{ij} | / n \quad ,$$

the root mean square error is given by

$$(4.1.2) \quad 100 \times [\sum (f_{ij} - \hat{f}_{ij})^2 / n_{ij} n \ ]^{1/2}$$

and finally, the Chi-square measure is given by

$$(4.1.3) \quad \sum (f_{ij} - \hat{f}_{ij})^2 / f_{ij} \quad .$$

Here, f_{ij} refers to an estimated frequency of accidents for the ijth cell.

It should come as no surprise that the empirical Bayes estimates fit the data better than the fixed effect Model I estimates. They are based on a model with an extra parameter, the prior variance, and in fact are a compromise between the observed rates and multiplicative model estimates. For example, the mean absolute error for all 128 cells of 0.6365 per cent for Model I is reduced to 0.4569 per cent for Model II estimates. An even larger reduction is achieved in the root mean square error.

The Chi-square associated with Model I is that which would be used to test the fit of that model. Under the null hypothesis, it is distributed as a Chi-square with 105 degrees of freedom. This corresponds to a p-value of about 0.001, indicating a statistically significant departure from Model I. The Chi-square measure, like the other two measures of fit, is reduced substantially by the empirical Bayes estimates.

Measures of fit were also computed separately for the 27 cells with more than 1000 exposures, the 34 cells with between 500 and 1000 exposures, and the remaining 101 cells with fewer than 500 exposures. As expected, the difference in fit between the two procedures is largest for cells with large numbers of exposures. For example, the root mean square error is 0.4430 per cent for Model I estimates and 0.2660 per cent for the empirical Bayes estimates. For cells with fewer exposures, the differences are not as great. For these cells, the empirical

Bayes estimates are very close to the Model I estimates.

A more interesting evaluation of the two methods is to compare their ability to predict accident rates for 1973 using estimates derived from 1972 data. Rather than predicting actual accident rates for 1973, the observed rates were adjusted so that the overall adjusted number of accidents in 1973 would be the same as the number of accidents in 1972. This was done on the premise that these methods are useful for predicting cell rates relative to the overall rate. They are not designed for predicting changes in the overall accident rates. Adjusted rates for 1973, $r_{ij}(73)$, were obtained by the following formula,

$$(4.1.4) \quad r_{ij}(73) = f_{ij}(73)\left[\Sigma f_{ij}(72) / \Sigma f_{ij}(73)\right]$$

where, $f_{ij}(72)$ and $f_{ij}(73)$ are the observed numbers of accidents occurring to members of the ijth cell in 1972 and 1973 respectively. The overall accident rate in 1973 was 8.60 per cent lower than in 1972, probably reflecting the beginning of the 1973-74 gasoline shortage. Thus, observed accident rates in 1973 were adjusted by a factor of 1.094.

Summary measures of predictions were calculated for estimates based on Models I and II as well as for estimates based on the observed 1972 rates, that is, Model 0 estimates. The results of these comparisons are shown in Table VI under the heading "Prediction of 1973 Rates".

For all 128 cells, the Model II empirical Bayes estimates outperform Model I estimates on all three measures, though the differences are not dramatic. Both empirical Bayes and the Model I estimates show substantial improvements over the unsmoothed cell estimates. For example, the mean absolute error is reduced from 0.8523 per cent for the 1972 observed cell estimates to 0.7762 per cent for the empirical Bayes estimates.

For cells with large numbers of exposures, the advantage of the empirical Bayes estimates is quite evident. The experience available for these cells is such that the simple observed cell rates outperform the fixed effect model estimates in terms of mean ablsolute error. The situation is reversed for root mean square error and the Chi-square measure which emphasize larger prediction errors. For all three measures, Model II estimates offer substantial improvements over the fixed effect model estimates. The mean absolute errors of the empirical Bayes estimates and the 1972 observed rates are virtually the same. However, in terms of mean square error and Chi-square, the empirical Bayes estimates are definitely superior.

For cells of moderate size, between 600 and 1000 exposures, the observed rates outperform the fixed effect model estimates while the empirical Bayes estimates are clearly superior to both.

Finally, for cells with fewer than 500 exposures, the two smoothed estimates dramatically outperform the 1972 observed rates. These cell sizes are too small to achieve any reliability for prediction, and are precisely those cells for which model estimates or smoothing techniques were designed. The differences between the two sets of model based estimates are not large. The estimates themselves are very close since the empirical Bayes estimator relies on the fixed effect model when cell sizes are small. In fact, for these cells, the simple fixed effects model estimates perform marginally better than the empirical Bayes estimates.

4.2 Model III

The raw data for Model III analysis consists of a classification of individuals according to territory, driver class, and number of accidents (in 1972). Although in theory one must produce a separate estimate of $u_{123(ijk)}$ for each individual, in practice, because of symmetry, one only need produce predictions corresponding to observed numbers of accidents within cells. For example, all middle aged, married male drivers living in Territory 3 who had two accidents in 1972 will have the same Model III estimated accident rate. In this sample, there were no drivers with more than four accidents in 1972, so that at most, only five unique estimates need be produced for each cell.

As with Model II, an approximation of the likelihood function for the variance components, σ_1^2 and σ_2^2, was used to determine starting values for the EM algorithm. The likelihood was calculated for a grid of points, and its maximum approximated by eye. The algorithm converged to its maximum at $(\sigma_1^2, \sigma_2^2) = (0.0046, 0.46)$. The likelihood is fairly concentrated on the σ_2^2 dimension. If one defines a "confidence region" as the set of all points whose likelihood is no less than 1/20 of the maximum, then there is good evidence that σ_2^2 is between 0.35 and 0.60. On the other hand, in the σ_1^2 dimension, any such region would include σ_1^2 equal to zero and would go beyond σ_1^2 equal to 0.01. Thus with the additional variance component corresponding to individuals within cells, there is no strong evidence for a non-zero variance component corresponding to cell effects.

Table VII gives Model III predicted accident rates for drivers in a selection of territories for purposes of illustration. The complete table of predictions can be found in Tomberlin (1982). For example, middle aged, married males living in Territory 1 with no accidents in 1972 have a predicted accident rate of 0.68 per year. Note that the information in a single year's driving record is substantial. For example, if the average accident claim were $1000, drivers in this class having no accidents would be assessed a pure premium of $68. This would increase to $106 for those having a single accident, $161 for those with two accidents, and so on.

Further, note that the increase in expected accident frequencies is not a linear function of the observed numbers of accidents. Additional accidents increase the predicted rate by increasing amounts. This contrasts with the linear, multiple regression approach for predicting frequencies adopted by others. See, for example, Kwong, et al. (1975)

Data from the California study were again used to compare the predictions of Model III estimates with those of Models 0, I, and II. Estimates based on 1972 data were used to predict accident rates for 1973 and 1974 both separately and combined. Since Model III estimates are designed to predict individual accident rates on the basis of individual driver records, ideally one would compare the performance of each of the competing estimation techniques in making individual predictions. However, since the expected number of accidents for a single individual is so small, the observed rates are subject to so much variability that summary measures of prediction errors based on individual data are not very illuminating.

For this reason, drivers were classified according to their territory, driver class and number of accidents in 1972. Observed accident rates in 1973 and 1974 were obtained by pooling the observed rates within these categories. Again, observed rates in 1973 and 1974 were adjusted as described in equation (4.1.4) so that the overall adjusted rates for these years would match the observed 1972 rate. In 1974 there was a dramatic decrease in the accident rate due to the gasoline shortage caused by the Arab oil embargo. The overall accident rate in 1974 was 30% lower than in 1972! With such a large decrease one worries about structural changes in the relationship between accidents and predictor variables and the effect this might have on the relative performances of the various model predictions. Conclusions based on these comparisons should be made with this qualification in mind.

Mean absolute errors (MAE), root mean square errors (RMSE), and mean errors (ME) for the predictions of the four models are compared in Table VIII. The MAE for predicting 1973 rates, for example, is defined by

$$(4.2.1) \sum | r^{*}_{ij(k)}(73) - \hat{f}_{ij(k)}(72) | / n_{...}$$

where $r^{*}_{ij(k)}(73)$ is defined to be the adjusted number of accidents occurring in 1973 to the $n_{ij(k)}$ drivers in the ith territory, and jth driver class who had k accidents in 1972; and $\hat{f}_{ij(k)}(72)$ is the appropriate model-based estimate of the expected number of accidents for the same group based on 1972 data.

The results for all drivers comparing predictions for 1973 and 1974 individually and combined are given in the first two columns of Table VIII. In every case, the summary measures of prediction errors become smaller with increasing model complexity. For example, the root mean square error in predicting 1973 rates decreases from 0.0264 for Model 0, the observed territory-driver class cell rates, to 0.0251 for Model I, the fixed effects model, to 0.0249 for Model II estimates, to 0.0217 for Model III empirical Bayes estimates. Mean errors for all drivers are zero for each the four models, and are therefore not reported.

The remaining ten columns of Table VIII contain summary measures of prediction errors for drivers classified according to the number of accidents they experienced in 1972. Here again, with a few exceptions, summary prediction errors are smallest for Model III estimates. This is true, without exception, for drivers having at most one accident. The exceptions occur for predictions for drivers with three or four accidents. In each case, the MAE is largest for Model III estimates. However, the RMSE are smallest for these drivers except when predicting 1974 accidents.

Even more striking are the comparisons of mean (signed) errors. This measure reflects consistent errors in prediction for particular groups of drivers. In every case, the absolute value of the mean error is reduced dramatically when Model III estimates are used, whereas there is not much difference in the performance of Models 0, I, or II. For example in predicting 1973 accident rates for drivers with three or four accidents in 1972, the mean error is 0.2167 for Model 0 estimates and only -0.0262 for Model III estimates. On the average, the observed 1972 cell accident rates underestimate 1973 accident rates by more than 20 accidents per 100 drivers where Model III overestimates by less than 3 per 100 drivers. Relative to the overall observed 1972 accident rate of 6.9 per 100 drivers, the reduction in average error achieved by Model III estimates is indeed large!

October 5, 1984
DRAFT

V. CONCLUSION

A succession of models for claims frequency has been introduced which incorporates the actuarial notions of risk classification, model-based smoothing, credibility theory, and experience rating, cumulatively under a unified statistical approach to the problem. Model 0 incorporates the essential features of a risk classification system. Model I adds the notion of model-based smoothing. Model II employs a random term which allows for empirical Bayes estimation of class rates in a manner similar to actuarial credibility estimates. Finally, Model III incorporates an additional random effect term corresponding to individuals within cells of the cross-classification, allowing for empirical Bayes estimates for individuals in a manner similar to actuarial experience rating.

As is indicated in the review in Section 2, the use of Bayes and empirical Bayes reasoning for loss prediction is not new to actuaries. In that sense, the ideas presented in Section 3 are not new. What is new is the incorporation of all the notions into a unified approach to estimation. Usually, notions of credibility are applied separate from (and sometimes confused with) model based smoothing techniques for obtaining cell estimates for a cross-classification of risks. Model II incorporates both of the notions. Experience rating or merit rating for individuals within cells of a classification is usually achieved either by a completely separate mechanism or by establishing merit-rating classes within the classification system. By including nested random effects corresponding to individuals within cells of the cross-classification, Model III incorporates empirical Bayes (credibility theory) estimates for individuals along with all the features of Model II within a single overall estimation procedure.

For the California data, prediction errors associated with Model II estimates are reduced from those associated with the fixed effects Model I estimates for the population as a whole, though the reduction is not dramatic. When the data are broken down into size categories by cells, the results are more interesting. For large cells, Model II predictions are close to the individual observed accident rates, and the predictions of these are superior to those of the fixed effects Model I estimates. For small cells, Model II estimates are close to Model I estimates and both provide predictions superior to the observed cell accident rates. For moderate size cells, the empirical Bayes estimates combine information from the observed cell rates and the smoothed Model I rates to provide predictions which are superior to both. These findings are consistent with those reported by Weisberg, Tomberlin and Chatterjee (1984) based on an anlaysis of insurance data from Massachusetts.

Experience rating estimates for individuals obtained using Model III provide further reductions in prediction errors over those produced by Model II. More importantly, estimates which take no account of individual driving experience consistently overestimate accident rates for drivers with no previous accidents and correspondingly underestimate rates associated with drivers with at least one previous accident. The size of these consistent errors in prediction is reduced dramatically by the Model III estimates which take into account individual driving records.

In summary, the analysis provides evidence that the empirical Bayes estimates based on Models II and III provide predictions which are superior to the more conventional estimation techniques. Furthermore, they incorporate actuarial notions, which are many times employed in an ad hoc, piece-meal manner, into a unified approach to estimation.

October 5, 1984
DRAFT

REFERENCES

Albert, James (1981), "Simultaneous estimation of Poisson Means", Journal of Multivariate Analysis, 400-417.

Almer, B. (1957), "Risk analysis in theory and practical statistics", Transactions of the XV International Congress of Actuaries, New York, 2, 314-349.

Bailey, A. L. (1945), "A generalized theory of credibility", Proceedings of the Casualty Actuarial Society, 32, 13-20.

Bailey, A. L. (1950), "Credibility procedures, LaPlace's generalization of Bayes' rule, and the combination of collateral knowledge with observed data", Proceedings of the Casualty Actuarial Society, 37, 7-23.

Bailey, Robert A. (1963), "Insurance rates with minimum bias", Proceedings of the Casualty Actuarial Society, 50, 4-11.

Bailey, Robert A. and LeRoy J. Simon (1959), "An actuarial note on the credibility of experience of a single private passenger car", Proceedings of the Casualty Actuarial Society, 46, 159-164.

Bailey, Robert A., and LeRoy J. Simon (1960), "Two studies in automobile insurance ratemaking", Proceedings of the Casualty Actuarial Society, 47, Part I, 1-19.

Beard, R. E., T. Pentikäinen, and E. Pesonen (1977), Risk Theory, 2nd edition, Chapman and Hall Ltd.

Bishop, Y., S. Fienberg, and P. Holland (1975), Discrete Multivariate Analysis: Theory and Practice, Cambridge, MA: MIT Press.

Buhlmann, H. (1970), Mathematical Methods in Risk Theory, Berlin: Springer-Verlag.

Carter, Grace M. and John E. Rolph (1974), "Empirical Bayes methods applied to estimating fire alarm probabilities", Journal of the American Statistical Association, 69, 880-885.

Casey, B., J. Pezier, and C. Spetzler (1976), The Role of Classification in Property and Casualty Insurance: A Study of the Risk Assessment Process, Menlo Park, CA: Stanford Research Institute.

Chamberlain, C. (1980), "Relativity pricing through analysis of variance", Pricing Property and Casualty Insurance Products, a volume of discussion papers presented at the Spring meeting of the Casualty Actuarial Society, 4-24.

Chang, L. and W. Fairley (1979), "Pricing automobile insurance under a multivariate classification", The Journal of Risk and Insurance, 46, 75-93.

Clevenson, M.L. and Zidek, J.V. (1975), "Simultaneous estimation of means of independent

Poisson laws", *Journal of the American Statistical Association*, 70, 698-706.

Deming, W. E. and F. Stephan (1940), "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", *Annals of Mathematical Statistics*, 11, 427-444.

Dempster, A. P., N. Laird, and D. B. Rubin (1977), "Maximum likelihood from incomplete data via the EM algorithm" (With discussion), *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Doherty, N. W. (1981), "Is rate classification profitable?", *Journal of Risk and Insurance*, 68, 286-295.

Efron, B. and C. Morris (1972), "Limiting the risk of Bayes and empirical Bayes estimators--Part II: the empirical Bayes case", *Journal of the American Statistical Association*, 67, 130-139.

Efron, B. and C. Morris (1973), "Stein's estimator and its competitors--an empirical Bayes approach", *Journal of the American Statistical Association*, 68, 117-130.

Felk, W. J. and H. J. Belnick (1975), "Credibility in group experience rating", *Credibility, Theory and Applications*, P.M. Kahn, Editor, New York: Academic Press.

Ferreira, J. (1978), "Identifying equitable insurance premiums for risk classes: An alternative to the classical approach", *Automobile Insurance Risk Classification: Equity and Accuracy*, Boston, MA: Massachusetts Division of Insurance.

Fienberg, S. E. and Holland, P. W. (1973), "Simultaneous estimation of multinomial cell probabilities", *Journal of the American Statistical Association*, 68, 683-691.

Haberman, S. J. (1979) *Analysis of Qualitative Data, Vols I and II.* New York: Academic Press.

Heckman, P. E. (1980), "Credibility and solvency", *Pricing Property and Casualty Insurance Products*, a volume of discussion papers presented at the Spring meeting of the Casualty Actuarial Society, 116-152.

Henderson, C. R., O. Kempthorne, S. R. Searle and Von Krosigk (1959), "The estimation of environmental and genetic trends from records subject to culling", *Biometrics*, 15, 192-218.

Hudson, H. M. and K. W. Tsui (1981), "Simultaneous Poisson estimators for a priori hypothesis about means", *Journal of the American Statistical Association*, 76, 182-187.

James, W. and Stein, C. (1961), "Estimation with quadratic loss", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 361-379.

Jewell, W. S. (1980), "Models in insurance: Paradigms, puzzles, communications and revolutions", presented to the 21st International Congress of Actuaries in Zurich and Lausanne.

Kahn, P. M. (1975), Credibility, Theory and Applications, New York: John Wiley.

Kwong, K. W., J. Kuan and R. Peck (1976), "Longitudinal study of California driver accident frequencies I: An exploratory multivariate analysis", Sacramento, CA: Department of Motor Vehicles, Office of Program Development and Evaluation.

Laird, N. (1975), "Log-linear models with random parameters: An empirical Bayes approach", unpublished Ph.D. Thesis, Cambridge, MA: Department of Statistics, Harvard University.

Laird, N. (1978), "Empirical Bayes for two-way tables", Biometrika, 65, 805-811.

Leonard, T. (1975), "Bayesian estimation methods for two-way contingency tables", Journal of the Royal Statistical Society, Series B, 37, 23-37.

Maritz, J. S. (1970), Empirical Bayes Methods, London: Methuen and Co., Ltd.

Massachusetts Automobile Rating and Accident Prevention Bureau (1981), "Automobile insurance rate filing, presented to the hearings before the Massachusetts Commissioner of Insurance, Boston, MA.

Morris, C. and L. Van Slyke (1978), "Empirical Bayes methods for pricing insurance classes", Proceedings of the Business and Economics Section, San Deigo, CA.

Mowbray, A. H. (1914), "How extensive a payroll exposure is necessary to give a dependable pure premium?", Proceedings of the Casualty Actuarial Society, 1, 24-30.

Norberg, R. (1979), "The credibility approach to experience rating", Scandinavian Actuarial Journal, 181-221.

Pang, S. (1975), "Simultaneous estimation of the parameters of independent Poisson distributions", Technical Report No 78, Stanford University, Department of Statistics.

Rolph, J. E. (1981), "Some statistical evidence on merit rating in medical malpractice insurance", Journal of Risk and Insurance, 48, 247-260.

Sant, D. T. (1980), "Estimating expected losses in auto insurance", Journal of Risk and Insurance, 47, 133-151.

SAS Institute, Inc. (1979), SAS User's Guide, Cary, NC.

Seal, H. (1969), Stochastic Theory of a Risk Business, New York: John Wiley.

Shayer, N. (1978), "Driver classification in automobile insurance", Automobile Insurance Risk Classification: Equity and Accuracy, Boston, MA: Massachusetts Division of Insurance.

Stern, P. K. (1965), "Ratemaking procedures for automobile liability insurance", Proceedings of the Casualty Actuarial Society, 52, 139-202.

October 5, 1984
DRAFT

Tomberlin, T. J. (1982), "A statistical perspective on predicting losses in automobile insurance", unpublished Ph.D. dissertation, Department of Statistics, Harvard University.

Tryfos, P. (1980), "On classification in automobile insurance", Journal of Risk and Insurance, 47, 331-337.

Weisberg, Herbert I. and Thomas J. Tomberlin (1982), "A Statistical Perspective on Actuarial Methods for Estimating Pure Premiums from Cross-Classified Data", Journal of Risk and Insurance.

Weisberg, Herbert I., Thomas J. Tomberlin and Sangit Chatterjee (1984), "Predicting Insurance Losses Under a Cross-Classification: A Comparison of Alternative Approaches", Journal of Business and Economic Statistics.

Whitney, A. W. (1918), "The theory of experience rating", Proceedings of the Casualty Actuarial Society, 4, 274-292.

Zehnwirth, B. (1979), "A hierarchical model for the estimation of claim rates in a motor car insurance portfolio", Scandinavian Actuarial Journal, 75-82.

TABLE I

DRIVER CLASS DEFINITIONS

Driver Class	
MD_MAR_M	Married males aged 26-65*
MD_SGL_M	Single males aged 26-65
MD_MAR_F	Married females aged 26-65
MD_SGL_F	Single females aged 26-65
OLD	All drivers over age 65
YG_MAR_M	Married males under age 26
YG_SGL_M	Single males under age 26
YG_F	females under age 26

*Age and marital status are as reported in 1974.

TABLE II

NUMBER OF DRIVERS BY TERRITORY
AND DRIVER CLASS

DRIVER CLASS

TERRITORY	MO_MAR_M	MO_SGL_M	MO_MAR_F	MO_SGL_F	OLD	YG_MAR_M	YG_SGL_M	YG_F
1	9852	4682	9830	4110	2243	588	2981	3103
2	2033	934	1980	813	552	152	593	644
3	1148	478	1030	365	408	92	354	287
4	1045	465	1018	428	414	125	340	358
5	700	319	716	299	171	71	222	243
6	743	263	740	222	226	51	217	233
7	647	286	567	247	213	40	229	218
8	717	310	654	247	186	46	226	234
9	1580	707	1436	585	294	88	438	463
10	881	325	840	324	288	77	297	299
11	1699	1164	1508	872	533	96	511	510
12	819	485	795	455	176	33	260	273
13	1523	680	1341	574	384	96	457	449
14	758	345	718	259	356	86	245	287
15	1725	1518	1394	1288	651	128	492	512
16	2309	1037	2231	906	577	132	648	644

Data taken from the 1974 California Drivers record study

TABLE III

OBSERVED ACCIDENT RATE PER
100 DRIVERS BY TERRITORY
AND DRIVER CLASS
IN 1972

TERRITORY	DRIVER CLASS MD_MAR_M	MD_SGL_M	MD_MAR_F	MD_SGL_F	OLD	YG_MAR_M	YG_SGL_M	YG_F
1	7.05	9.33	3.96	6.06	5.31	8.50	13.62	7.19
2	5.80	6.42	3.58	4.31	4.17	9.21	12.98	5.43
3	5.14	6.28	3.11	4.66	3.19	10.87	8.19	6.27
4	5.07	7.31	2.55	5.61	3.86	5.6	12.65	5.03
5	6.14	8.15	3.91	6.35	4.68	2.82	11.71	5.35
6	6.86	9.51	3.65	4.50	3.10	3.92	17.51	6.87
7	5.72	9.79	3.35	8.50	1.88	0	17.47	6.42
8	8.51	9.68	3.82	6.48	5.91	4.35	18.14	7.26
9	6.27	9.62	3.13	4.96	5.78	11.36	15.07	7.56
10	7.04	14.77	2.50	6.79	8.33	11.69	17.51	7.69
11	8.00	11.60	3.65	7.45	5.25	4.17	16.83	8.63
12	8.79	6.60	4.40	4.62	5.11	0	11.15	9.89
13	7.68	11.32	4.18	7.14	5.99	17.71	16.63	6.68
14	5.94	9.86	3.62	2.32	3.37	8.14	19.18	4.53
15	9.45	10.21	5.81	7.14	7.22	7.81	12.40	6.45
16	8.45	10.99	3.90	5.41	2.95	8.33	12.81	9.94

Data taken from the 1974 California Driver Record Study.

TABLE IV

MODEL 1 ESTIMATES
OF ACCIDENT RATES PER 100 DRIVERS
BY TERRITORY AND DRIVER CLASS
IN 1972

TERRITORY	MD_MAR_M	MD_SGL_M	MD_MAR_F	MD_SGL_F	OLD	YG_MAR_M	YG_SGL_M	YG_F
1	7.14	9.42	3.84	5.91	4.96	8.29	14.16	7.14
2	5.86	7.73	3.15	4.85	4.07	6.80	11.62	5.86
3	5.18	6.84	2.79	4.29	3.60	6.01	10.28	5.18
4	5.49	7.25	2.95	4.55	3.81	6.27	10.89	5.49
5	6.25	8.26	3.36	5.18	4.34	7.26	12.40	6.26
6	6.90	9.10	3.71	5.71	4.79	8.01	13.68	6.90
7	6.79	8.96	3.65	5.62	4.71	7.88	13.47	6.79
8	7.97	10.52	4.28	6.60	5.53	9.25	15.81	7.97
9	6.82	9.01	3.67	5.65	4.74	7.92	13.53	6.83
10	8.14	10.74	4.37	6.74	5.65	9.45	16.14	8.14
11	8.16	10.78	4.39	6.76	5.67	9.40	16.19	8.17
12	7.00	9.24	3.76	5.80	4.86	8.13	13.89	7.01
13	8.17	10.78	4.39	6.76	5.67	9.48	16.20	8.17
14	6.48	8.55	3.48	5.36	4.50	7.52	12.84	6.48
15	8.41	11.10	4.52	6.97	5.84	9.76	16.68	8.41
16	7.64	10.09	4.11	6.33	5.31	8.87	15.16	7.65

Data taken from the 1974 California driver record study.

TABLE V

MODEL II ESTIMATES OF ACCIDENT
RATES PER 100 DRIVERS BY
TERRITORY AND DRIVER CLASS
IN 1972

DRIVER CLASS

TERRITORY	MD_MAR_M	MD_SGL_M	MD_MAR_F	MD_SGL_F	OLD	YG_MAR_M	YG_SGL_M	YG_F
1	7.08	9.38	3.92	6.01	5.10	8.35	13.86	7.18
2	5.84	7.35	3.25	4.72	4.05	6.94	12.22	5.77
3	5.19	6.80	2.83	4.33	3.55	6.20	10.07	5.30
4	5.37	7.30	2.86	4.65	3.79	6.31	11.43	5.43
5	6.22	8.26	3.40	5.26	4.31	7.07	12.48	6.15
6	6.85	9.15	3.65	5.55	4.60	7.81	14.48	6.84
7	6.53	9.10	3.57	5.83	4.48	7.66	14.35	6.72
8	8.07	10.36	4.15	6.51	5.47	9.02	16.43	7.83
9	6.62	9.27	3.52	5.53	4.82	8.10	14.24	6.98
10	7.82	11.65	4.01	6.76	5.91	9.59	16.82	8.11
11	8.05	11.14	4.11	6.89	5.52	9.07	16.54	8.20
12	7.48	8.64	3.82	5.57	4.81	7.91	13.51	7.29
13	7.98	11.06	4.32	6.85	5.70	10.00	16.64	7.91
14	6.30	8.75	3.45	5.05	4.32	7.46	14.12	6.21
15	8.92	10.65	4.88	7.00	6.07	9.56	15.39	7.95
16	8.03	10.41	3.97	5.99	4.81	8.67	14.28	8.12

Data taken from the 1974 California driver record study.

TABLE VI

SUMMARY MEASURES OF FIT AND PREDICTION
OF CELL AVERAGE ACCIDENT RATES

	FIT TO 1972 RATES			PREDICTION OF 1973 RATES		
	Mean Absolute Error (Per 100 Drivers)	Root Mean Square Error (Per 100 Drivers)	Chi Square	Mean Absolute Error (Per 100 Drivers)	Root Mean Square Error (Per 100 Drivers)	Chi Square
	All 128 Cells					
1972 Observed Rates (Model 0)				.8523	1.4040	292.55
Fixed Effects Model I Estimates	.6365	1.6096	152.18	.7752	1.1256	178.35
Empirical Bayes Model II Estimates	.4569	.8564	100.90	.7303	1.0931	165.90
	27 Cells With > 1000 Exposures					
1972 Observed Rates (Model 0)				.4084	.6175	39.54
Fixed Effects Model I Estimates	.3167	.4433	20.79	.4494	.5054	36.31
Empirical Bayes Model II Estimates	.1760	.2660	8.18	.4053	.5625	30.91
	34 Cells With Between 500 and 1000 Exposures					
1972 Observed Rates (Model 0)				1.1911	1.5482	68.50
Fixed Effects Model I Estimates	.8557	1.0855	39.96	1.2875	1.5822	74.64
Empirical Bayes Model II Estimates	.6294	.8060	24.02	1.2100	1.5006	66.24
	67 Cells With < 500 Exposures					
1972 Observed Rates (Model 0)				2.0810	2.7725	184.51
Fixed Effects Model I Estimates	1.5546	2.1626	91.43	1.3293	1.7420	67.41
Empirical Bayes Model II Estimates	1.2901	1.8294	68.69	1.3245	1.7653	68.74

TABLE VII

MODEL III PREDICTIONS OF ACCIDENT RATES BY TERRITORY, DRIVER CLASS AND NUMBER OF ACCIDENTS IN 1972

Territory	Driver Class	Number of Accidents: Zero	One	Two	Three	Four
1	MD_MAR_M	0.068	0.105	0.161	0.243	0.361
1	MD_SCL_M	0.089	0.137	0.208	0.311	0.457
1	MD_MAR_F	0.038	0.059	0.092	0.141	0.214
1	MD_SCL_F	0.058	0.090	0.138	0.209	0.313
1	OLD	0.049	0.076	0.118	0.179	0.270
1	YG_MAR_M	0.080	0.123	0.187	0.281	0.415
1	YG_SCL_M	0.129	0.197	0.295	0.435	0.626
1	YG_F	0.069	0.106	0.163	0.246	0.365
2	MD_MAR_M	0.057	0.088	0.134	0.204	0.306
2	MD_SCL_M	0.071	0.110	0.168	0.254	0.377
2	MD_MAR_F	0.032	0.049	0.076	0.117	0.179
2	MD_SCL_F	0.046	0.072	0.111	0.169	0.255
2	OLD	0.040	0.062	0.095	0.146	0.221
2	YG_MAR_M	0.066	0.102	0.157	0.237	0.353
2	YG_SCL_M	0.113	0.173	0.260	0.386	0.560
2	YG_F	0.056	0.087	0.133	0.203	0.304
3	MD_MAR_M	0.050	0.078	0.120	0.183	0.276
3	MD_SCL_M	0.066	0.101	0.155	0.235	0.349
3	MD_MAR_F	0.028	0.043	0.067	0.103	0.158
3	MD_SCL_F	0.042	0.065	0.101	0.154	0.234
3	OLD	0.035	0.054	0.084	0.129	0.197
3	YG_MAR_M	0.059	0.092	0.141	0.214	0.319
3	YG_SCL_M	0.096	0.147	0.223	0.333	0.488
3	YG_F	0.051	0.079	0.122	0.186	0.279
4	MD_MAR_M	0.052	0.081	0.125	0.190	0.286
4	MD_SCL_M	0.070	0.108	0.165	0.249	0.370
4	MD_MAR_F	0.028	0.044	0.069	0.106	0.162
4	MD_SCL_F	0.045	0.070	0.108	0.164	0.248
4	OLD	0.037	0.058	0.089	0.137	0.208
4	YG_MAR_M	0.061	0.094	0.145	0.220	0.328
4	YG_SCL_M	0.106	0.162	0.245	0.364	0.530
4	YG_F	0.053	0.082	0.126	0.192	0.288
5	MD_MAR_M	0.060	0.093	0.143	0.217	0.324
5	MD_SCL_M	0.079	0.122	0.185	0.279	0.412
5	MD_MAR_F	0.033	0.052	0.080	0.123	0.188
5	MD_SCL_F	0.051	0.079	0.121	0.183	0.278
5	OLD	0.042	0.066	0.101	0.155	0.234
5	YG_MAR_M	0.069	0.106	0.162	0.245	0.364
5	YG_SCL_M	0.117	0.178	0.269	0.397	0.575
5	YG_F	0.060	0.092	0.142	0.215	0.322
6	MD_MAR_M	0.066	0.102	0.156	0.236	0.351
6	MD_SCL_M	0.087	0.133	0.203	0.304	0.447
6	MD_MAR_F	0.036	0.056	0.086	0.132	0.201
6	MD_SCL_F	0.054	0.084	0.129	0.196	0.294
6	OLD	0.045	0.070	0.108	0.166	0.250
6	YG_MAR_M	0.075	0.116	0.177	0.267	0.395
6	YG_SCL_M	0.132	0.201	0.301	0.443	0.636
6	YG_F	0.066	0.102	0.156	0.236	0.351
7	MD_MAR_M	0.064	0.098	0.150	0.228	0.340
7	MD_SCL_M	0.086	0.132	0.201	0.301	0.443
7	MD_MAR_F	0.035	0.055	0.084	0.130	0.197
7	MD_SCL_F	0.056	0.086	0.132	0.201	0.301
7	OLD	0.044	0.069	0.106	0.162	0.245
7	YG_MAR_M	0.074	0.114	0.174	0.262	0.388
7	YG_SCL_M	0.130	0.198	0.298	0.438	0.630
7	YG_F	0.065	0.100	0.153	0.232	0.346
8	MD_MAR_M	0.077	0.118	0.181	0.272	0.402

TABLE VIII

Summary Measures of Prediction of accident rates
within categories defined according to number of accidents in 1972

Year(s) Predicted	Model Used For Estimation	All drivers		Drivers with 0 Accident in 1972			Drivers with 1 Accident in 1972			Drivers with 2 Accidents in 1972			Drivers with 3 or 4 Accidents in 1973		
		MAE	RMSE	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RSME	ME
1973	0	.0118	.0264	.0087	.0142	-.0028	.0492	.0681	.0371	.1452	.2107	.0927	.3452	.4631	.2157
	I	.0111	.0251	.0080	.0114	-.0029	.0496	.0680	.0384	.1447	.2099	.0962	.3383	.4638	.2250
	II	.0107	.0249	.0076	.0111	-.0029	.0494	.0678	.0381	.1449	.2101	.0955	.3396	.4635	.2238
	III	.0095	.0217	.0070	.0103	-.0000	.0370	.0572	.0014	.1421	.1885	-.0144	.3589	.3778	-.0252
1974	0	.0143	.0295	.0117	.0168	-.0021	.0449	.0659	.0274	.1495	.2562	.0737	.3048	.5923	.1266
	I	.0130	.0288	.0112	.0155	-.0022	.0446	.0653	.0287	.1480	.2548	.0773	.3008	.6016	.1359
	II	.0135	.0285	.0108	.0149	-.0022	.0447	.0651	.0284	.1484	.2551	.0765	.2898	.5993	.1343
	III	.0127	.0275	.0101	.0145	.0007	.0407	.0613	-.0083	.1587	.2458	-.0334	.4536	.6116	-.1143
1973-74	0	.0113	.0231	.0088	.0135	-.0025	.0408	.0563	.0323	.1271	.1868	.0832	.2826	.3944	.1711
	I	.0105	.0219	.0080	.0112	-.0026	.0408	.0559	.0335	.1257	.1854	.0868	.2801	.4018	.1805
	II	.0101	.0216	.0076	.0106	-.0026	.0404	.0557	.0333	.1260	.1857	.0860	.2804	.3999	.1788
	III	.0087	.0191	.0068	.0099	.0003	.0300	.0468	-.0035	.1153	.1668	-.0239	.2843	.3623	-.0697
Number of Drivers		107,731		100,771			6,522			408			30		