TWO CREDIBILITY REGRESSION APPROACHES FOR THE CLASSIFICATION OF PASSENGER CARS IN A MULTIPLICATIVE TARIFF[1]

By Bjørn Sundt

Storebrand Insurance Company Ltd, Oslo

Abstract

In the present paper we present two credibility regression models for the classification of passenger cars. As regressors we use technical variables like price, weight, etc. In both models we derive credibility estimators and find expressions for their estimation errors. Estimators for structure parameters are proposed. A numerical example based on real data is given. The second model is hierarchical with a level for make of car.

1. Background

In Norway there is no common passenger car tariff for all insurance companies, and thus there are several different tarriffs in the market. However, most of them seem to have about the same structure as the one used by Storebrand to be described below, but with different parameter values.

In this paper we are going to discuss vehicle damage insurance for passenger cars. The tariff structure is multiplicative with factors for bonus-malus, mileage/district, deductibles, age of car, and car model. We shall concentrate on the factor for car model. There are 65 classes numbered from 30 to 94, and the factor for class c is equal to 1.04^{c-30}.

Until the present research was started, the classification of individual car models was performed rather subjectively. There was one person classifying new car models. When a new car model appeared on the market, he looked at its specifications and tried to find out to which cars it was comparable. Then he looked at the factors for these cars, both by Storebrand and by the competing companies. When the car had been in the market for some time and claims statistics had become available, the rating factor was reassessed, taking into account the observed claims ratio, the observed volume of exposure (premium), the old factor, and the premiums of the competing companies. This reassessment was also performed in a rather subjective way, but not by the same person who had made the initial classification of the car.

The procedure described in the previous paragraph has obvious advantages compared to an objectively based statistical procedure. It would be impossible to build into a mathematical model all the experience, knowledge, and intuition of a skilled person. How could the model incorporate, say, the person's opinion of

[1] Paper presented at the XIXth ASTIN Colloquium, Tel Aviv, 20–24 September 1986.

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

the importance of the shape of the car (a limousine and a coupé are bought by quite different sorts of people)? And even if one should succeed in creating a model which to a great extent incorporated the knowledge of the skilled person, this model would probably be too complicated for practical use. However, the advantage of the subjective procedure is also a disadvantage. The procedure is too dependent on the person performing it. As it is impossible to build the knowledge of the skilled person into a statistical model, it is also impossible to give an adequate documentation of the procedure. And what then if the person leaves the company?

This was the background that motivated the present research. One wanted an objective method for classification of cars, and in this paper we are going to describe the models and methods that were considered. We are also going to comment upon the difficulties that occurred during the work. As should be well known to everyone who has worked on modelling insurance data, these data are very seldom what you want them to be.

When the project was started, it was decided that this time we should concentrate only on the determination of the factor for car model. Ideally, one should of course have developed models and methods for simultaneous determination of all the factors in the multiplicative model, but that would have been a much more ambitious and time-consuming project. It was discussed whether one should concentrate only on the classification of new car models, for which we have no claims data, but in the present author's opinion, classification of new models is only a special case of reclassification (i.e. the case with exposure volume equal to zero). It would therefore be unnatural not to treat these two situations together, and it was decided to follow this line.

As was argued above, the subjective approach has great advantages compared to a statistically based procedure, and it would be wrong to throw this system away completely. It is the author's intention that the methods presented in this paper should not replace the skilled person, but rather be an aid to him. The system proposes a class to the person, but he should himself decide whether to follow this proposal or not. In particular, this is important for reclassification of cars that have already been in the market for some time, and for which we know the rating of the competitors. It would be too ambitious to build a model that also incorporates the premiums of competing companies. For marketing reasons, it could also be desirable to make smaller changes by the reclassification than those proposed by the statistical procedure.

Furthermore, in the statistical investigations it became clear that some car models behaved so strangely, relative to the model studied, that they ought to be considered as outliers in the present context. For such cars one should not use the factor suggested by the system, and perhaps even more important, these cars should be left out when estimating the model parameters. The most striking example in our investigations was Volkswagen Golf GTI, and the parameter estimates changed considerably when this car was taken out of the estimation procedures. It is important that the person doing the classification identifies such cars and sees to it that they are left out of the statistical analysis. One could of course argue that the model assumptions should also embrace such cars, but it is the opinion of the present author that it is preferable to have a relatively simple model giving satisfactory results for "normal" cars, than a complicated model that could be used for all cars. In particular, as he believes that in practice the "outliers" would usually be easy to identify.

For the numerical computations we used the program package SAS, which in particular was very convenient for the matrix calculus.

2. PRELIMINARIES

2.1. *Optimality Criterion for Estimators*

Let m be a random variable. We shall say that an estimator $m^{(1)}$ of m is better than another estimator $m^{(2)}$ if

$$E(m^{(1)}-m)^2 < E(m^{(2)}-m)^2,$$

that is, we use the quadratic loss function.

Let $\boldsymbol{m}=(m_1, \ldots, m_s)'$ be an unknown random vector and $\boldsymbol{m}^{(1)}=(m_1^{(1)}, \ldots, m_s^{(1)})'$ and $\boldsymbol{m}^{(2)}=(m_1^{(2)}, \ldots, m_s^{(2)})'$ two estimators of $\boldsymbol{m}$. Then we shall say that $\boldsymbol{m}^{(1)}$ is a better estimator of $\boldsymbol{m}$ than $\boldsymbol{m}^{(2)}$ if

$$E(m_i^{(1)}-m_i)^2 \leqslant E(m_i^{(2)}-m_i)^2 \quad i=1, \ldots, s$$

with strict inequality for at least one i.

We implicitly assume that second-order moments exist for all random variables to be considered.

2.2. *Credibility Estimators*

Let $\boldsymbol{x}$ and $\boldsymbol{m}$ be random vectors, $\boldsymbol{x}$ observable and $\boldsymbol{m}$ unknown. We shall call $\dot{\boldsymbol{m}}$ a linear estimator of $\boldsymbol{m}$ (based on $\boldsymbol{x}$) if $\dot{\boldsymbol{m}}$ may be written in the form $\dot{\boldsymbol{m}}=\boldsymbol{a}+\boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{a}$ is a non-random vector and $\boldsymbol{A}$ a non-random matrix. The credibility estimator of $\boldsymbol{m}$ (based on $\boldsymbol{x}$) is defined as the best linear estimator of $\boldsymbol{m}$. We summarize some results about credibility estimators in the following theorem.

THEOREM 2.1. (i) *There always exists a unique credibility estimator of* $\boldsymbol{m}$.

(ii) *Let* $\dot{\boldsymbol{m}}$ *be a linear estimator of* $\boldsymbol{m}$. *Then* $\dot{\boldsymbol{m}}$ *is a credibility estimator of* $\boldsymbol{m}$ *if and only if* $\dot{\boldsymbol{m}}$ *satisfies the two conditions*

$$E\dot{\boldsymbol{m}}=E\boldsymbol{m} \tag{2.1}$$

$$\operatorname{Cov}(\dot{\boldsymbol{m}},\boldsymbol{x}')=\operatorname{Cov}(\boldsymbol{m},\boldsymbol{x}'). \tag{2.2}$$

(iii) *Let* $\tilde{\boldsymbol{m}}$ *be the credibility estimator of* $\boldsymbol{m}$. *Then we have*

$$\operatorname{Cov}(\boldsymbol{m},\tilde{\boldsymbol{m}}')=\operatorname{Cov}(\tilde{\boldsymbol{m}},\boldsymbol{m}')=\operatorname{Cov}(\tilde{\boldsymbol{m}})=\operatorname{Cov}(\boldsymbol{m})-\operatorname{Cov}(\boldsymbol{m}-\tilde{\boldsymbol{m}}). \tag{2.3}$$

For proof of (i) we refer to DE VYLDER (1976), for proof of (ii) to SUNDT (1980), and for proof of (iii) to SUNDT (1981).

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

3. A NON-HIERARCHICAL APPROACH

3.1. *Model*

Consider a group of K different car models. These could be all passenger cars (station wagons included) that are rated in Storebrand, or a well-defined subgroup of these (e.g. diesel cars, cars with four-wheel drive, all Volkswagen models, all cars produced after 1982, etc.). For the parameter estimation described in Subsection 3.3 it could be reasonable to take a representative sample from the group considered.

For car model k we have observed I_k risk units (policies). Let X_{ki} be the total claim amount in the exposure time for unit i of model k, and let p_{ki} be the earned premium. We want to use earned premium as a measure of risk volume, but this premium also contains the car model factor which we are going to reassess, and this old value should not be included in the risk measure. Hence, let

$$w_{ki} = p_{ki}/f_k, \tag{3.1}$$

where f_k is the old factor, be our measure of risk volume. We assume that for fixed k, the X_{ki}'s are independent of the corresponding data from other car models, and that $X_{k1}, \ldots, X_{kI_k}$ are conditionally independent given Θ_k, a random parameter characterizing car model k. It is assumed that $\Theta_1, \ldots, \Theta_K$ are independent and identically distributed.

Let

$$Y_{ki} = X_{ki}/w_{ki}.$$

It is assumed that

$$E[Y_{ki} \mid \Theta_k] = m_k(\Theta_k)$$

$$\text{Var } m_k(\Theta_k) = \lambda \tag{3.2}$$

$$\text{Var}[Y_{ki} \mid \Theta_k] = s^2(\Theta_{ki})/v_{ki} \tag{3.3}$$

with $v_{ki} = w_{ki}$ (the reason for introducing v_{ki} will become clear in subsection 3.6, where we modify the present assumptions), and

$$\mu_k = Em_k(\Theta_k) = \boldsymbol{x}_k'\boldsymbol{\beta},$$

where $\boldsymbol{x}_k$ is a known $q \times 1$ design vector based on the technical data of the car and $\boldsymbol{\beta}$ is an unknown $q \times 1$ regression vector. We further introduce

$$\phi = Es^2(\Theta_k) \qquad \varkappa = \phi/\lambda$$

$$X_k = \sum_{i=1}^{I_k} X_{ki} \qquad v_k = \sum_{i=1}^{I_k} v_{ki} \qquad w_k = \sum_{i=1}^{I_k} w_{ki}$$

$$Y_k = X_k/w_k. \tag{3.4}$$

We note that in the special case when

$$Em_k(\Theta_k) = \mu,$$

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

independent of k, the conditions of the Bühlmann–Straub model (BÜHLMANN and STRAUB (1970)) are satisfied.

It is also interesting to relate our present model to HACHEMEISTER'S (1975) regression model. In that model one assumes that

$$E[Y_{ki} \mid \Theta_k] = \mathbf{x}_{ki}'\mathbf{b}(\Theta_k),$$

where $\mathbf{x}_{ki}$ is a known $q \times 1$ design vector and $\mathbf{b}$ is a $q \times 1$ vector function. To correspond to our present set-up we assume that $\mathbf{x}_{ki} = \mathbf{x}_k$ independent of i. We introduce

$$\mathbf{\Lambda} = \text{Cov } \mathbf{b}(\Theta_k) \qquad \boldsymbol{\beta} = E\mathbf{b}(\Theta_k)$$

and get

$$\text{Var } E[Y_{ki} \mid \Theta_k] = \mathbf{x}_k'\mathbf{\Lambda}\mathbf{x}_k.$$

Thus this variance would typically vary between car models whereas in (3.2) we have assumed it to be constant. Let us now assume that the first element of $\mathbf{x}_k$ is equal to one, which will usually be the case. Then we obtain our present model by assuming that only the first element of $\mathbf{b}(\Theta_k)$ is random, which makes all elements of $\mathbf{\Lambda}$ except the (1,1) element equal to zero. We note that this $\mathbf{\Lambda}$ is not positive definite.

3.2. *Credibility Estimation of $m_s(\Theta_s)$*

Let $\tilde{m}_s$ be the credibility estimator of $m_s(\Theta_s)$ based on the observed Y_{ki}'s. We also introduce the estimation error

$$\psi_s = E(m_s(\Theta_s) - \tilde{m}_s)^2$$

of $\tilde{m}_s$. From Theorem 2.1 we get

$$\tilde{m}_s = \zeta_s Y_s + (1 - \zeta_s)\mu_s \tag{3.5}$$

$$\psi_s = \phi/(v_s + \varkappa) = \lambda(1 - \zeta_s)$$

with

$$\zeta_s = v_s/(v_s + \varkappa).$$

3.3. *Parameter Estimation*

The structure parameters ϕ, λ, and $\boldsymbol{\beta}$ will in practical applications be unknown and have to be estimated.

We have that

$$\phi_k^* = (I_k - 1)^{-1} \sum_{i=1}^{I_k} v_{ki}(Y_{ki} - Y_k)^2 \tag{3.6}$$

satisfies $E[\phi_k^* \mid \Theta_k] = s^2(\Theta_k)$, and thus

$$\phi^* = \sum_{k=1}^{K} u_k \phi_k^*$$

is an unbiased estimator of ϕ for all weights u_k ($\sum_{k=1}^{K} u_k = 1$). In an earlier version of the paper we suggested that one should simply apply $u_k = K^{-1}$. This choice has been criticized by Ragnar Norberg, who suggests that one should apply

$$u_k = (I_k - 1) \Big/ \left(\sum_{r=1}^{K} I_r - K \right).$$

An optimal choice of weights is difficult, involving fourth-order moments (cf. e.g. NORBERG (1982)), and it was not within the scope of the present research to perform a profound analysis of this problem. Both our choice and Norberg's choice can be criticized; our choice because it gives too much weight to cars with low exposure; Norberg's choice because if I_s is much greater than the other I_k's for some s, then the value of Θ_s will have a too dominant influence on the estimate of ϕ. The present discussion also applies to the analogously weighted estimators in subsections 3.6 and 4.3. We note that in the special case $I_1 = I_2 = \ldots = I_K$ Norberg's choice and our choice are equal, and in this case ϕ^* is equal to the estimator proposed by BÜHLMANN and STRAUB (1970) for the Bühlmann–Straub model.

We introduce

$$Y = (Y_1, \ldots, Y_K)' \qquad X = (x_1, \ldots, x_K)'$$

$$v = \sum_{k=1}^{K} v_k \qquad D = \mathrm{diag}(v_1/v, \ldots, v_K/v)$$

and get

$$EY = X\beta$$

$$\text{Cov } Y = (\phi/v)D^{-1} + \lambda I_K \tag{3.7}$$

with I_K denoting the $K \times K$ identity matrix. It is assumed that X has rank q.

We trivially have that

$$\hat{\beta} = (X'DX)^{-1}X'DY$$

is an unbiased estimator of β. It seems reasonable to base an estimator of λ on the statistic

$$Q = (Y - X\hat{\beta})'D(Y - X\hat{\beta}), \tag{3.8}$$

and we therefore want to find the expectation of Q. In the deduction we use that for an $r \times s$ matrix A and an $s \times r$ matrix B we have

$$\mathrm{tr}(AB) = \mathrm{tr}(BA), \tag{3.9}$$

where "tr" denotes the trace of a quadratic matrix (i.e. the sum of its diagonal elements); this result is easily proved. We have

$$EQ = E(Y - X\hat{\beta})'D(Y - X\hat{\beta}) = \mathrm{tr}\{D\ E(Y - X\hat{\beta})(Y - X\hat{\beta})'\}$$
$$= \mathrm{tr}\{D\ \mathrm{Cov}(Y - X\hat{\beta})\}$$

as

$$E(Y - X\hat{\beta}) = \mathbf{0}.$$

We further get

$$EQ = \mathrm{tr}(D \operatorname{Cov}[\{I_K - X(X'DX)^{-1}X'D\}Y]) = \mathrm{tr}(D\{I_K - X(X'DX)^{-1}X'D\}(\operatorname{Cov} Y)\{I_K - DX(X'DX)^{-1}X\}),$$

and insertion of (3.7) gives

$$EQ = \lambda\tau_1 + (\phi/v)\tau_2 \tag{3.10}$$

with

$$\tau_1 = \mathrm{tr}(D\{I_K - X(X'DX)^{-1}X'D\}\{I_K - DX(X'DX)^{-1}X'\}) \tag{3.11}$$

and

$$\tau_2 = \mathrm{tr}(D\{I_K - X(X'DX)^{-1}X'D\}D^{-1}\{I_K - DX(X'DX)^{-1}X'\}). \tag{3.12}$$

From (3.11) we get

$$\tau_1 = \mathrm{tr}\ D - \mathrm{tr}\{DX(X'DX)^{-1}X'D\} - \mathrm{tr}\{D^2X(X'DX)^{-1}X'\} + \mathrm{tr}\{DX(X'DX)^{-1}X'D^2X(X'DX)^{-1}X'\}. \tag{3.13}$$

By repeated use of (3.9) we see that the three last terms in (3.13) are all equal to $\mathrm{tr}\{(X'DX)^{-1}X'D^2X\}$, and as in addition $\mathrm{tr}\ D = 1$, we get

$$\tau_1 = 1 - \mathrm{tr}\{(X'DX)^{-1}X'D^2X\}. \tag{3.14}$$

From (3.12) we obtain

$$\tau_2 = \mathrm{tr}(\{I_K - DX(X'DX)^{-1}X'\}\{I_K - DX(X'DX)^{-1}X'\}) = \mathrm{tr}\ I_K - \mathrm{tr}\{DX(X'DX)^{-1}X'\} = \mathrm{tr} I_K - \mathrm{tr}\{(X'DX)^{-1}X'DX\} = \mathrm{tr}\ I_K - \mathrm{tr}\ I_q,$$

and as the trace of an identity matrix is equal to its dimension, we get

$$\tau_2 = K - q. \tag{3.15}$$

From (3.8), (3.10), (3.14), and (3.15) we get that

$$\hat{\lambda} = \{(Y - X\hat{\beta})'D(Y - X\hat{\beta}) - (K - q)\phi^*/v\}/[1 - \mathrm{tr}\{(X'DX)^{-1}X'D^2X\}]$$

is an unbiased estimator of λ. It has, however, the disadvantage that it can take negative values whereas λ is always non-negative. Therefore we replace it by

$$\lambda^* = \max(0, \hat{\lambda}).$$

However, by this adjustment we lose the unbiasedness. For simplicity, in the following we proceed as if $\lambda^* > 0$; the adaption to the case $\lambda^* = 0$ is trivial. To avoid having to take special care of the case $\lambda^* = 0$, one could instead of putting λ^* equal to zero when $\hat{\lambda} \leqslant 0$, put λ^* equal to some small positive number; one possible choice would be ε/K for some small ε, as we would then have asymptotic unbiasedness when K goes to infinity.

If Cov Y were known, the best linear unbiased estimator of $\boldsymbol{\beta}$ would be

$$\ddot{\boldsymbol{\beta}} = \{X'(\text{Cov } Y)^{-1}X\}^{-1}X'(\text{cov } Y)^{-1}Y,$$

and as $Y_1, \ldots, Y_K$ are independent,

$$\ddot{\boldsymbol{\beta}} = \left(\sum_{k=1}^{K} \boldsymbol{x}_k(\text{Var } Y_k)^{-1}\boldsymbol{x}_k'\right)^{-1} \sum_{k=1}^{K} \boldsymbol{x}_k(\text{Var } Y_k)^{-1}Y_k$$

$$= \left(\sum_{k=1}^{K} \boldsymbol{x}_k v_k(\phi + \lambda v_k)^{-1}\boldsymbol{x}_k'\right)^{-1} \sum_{k=1}^{K} \boldsymbol{x}_k v_k(\phi + \lambda v_k)^{-1}Y_k$$

$$= \left(\sum_{k=1}^{K} \zeta_k \boldsymbol{x}_k \boldsymbol{x}_k'\right)^{-1} \sum_{k=1}^{K} \zeta_k \boldsymbol{x}_k Y_k,$$

and we propose to estimate $\boldsymbol{\beta}$ by

$$\boldsymbol{\beta}^* = \left(\sum_{k=1}^{K} \zeta_k^* \boldsymbol{x}_k \boldsymbol{x}_k'\right)^{-1} \sum_{k=1}^{K} \zeta_k^* \boldsymbol{x}_k Y_k$$

with

$$\zeta_k^* = v_k/(v_k + \varkappa^*) \qquad \varkappa^* = \phi^*/\lambda^*.$$

It should be noted that in the Bühlmann–Straub model the estimators λ^* and $\boldsymbol{\beta}^*$ reduce to the estimators studied by BÜHLMANN and STRAUB (1970).

3.4. *Determination of the Tariff Class*

By inserting the estimators ζ_k^* and $\boldsymbol{\beta}^*$ in (3.5) we get the empirical credibility estimator

$$\tilde{m}_s^* = \zeta_s^* Y_s + (1 - \zeta_s^*)\mu_s^*$$

with

$$\mu_s^* = \boldsymbol{x}_s' \boldsymbol{\beta}^*.$$

The estimation error ψ_s is estimated by

$$\psi_s^* = \lambda^*(1 - \zeta_s^*).$$

The estimator $\tilde{m}_s^*$ cannot yet be used as the proposed rating factor for car model s; it still needs to be adjusted by some scaling factor. The approach used in our numerical investigations was to determine the scaling factor such that the total premium for the portfolio used for the estimations would be the same with the new values of the model factor as with the old ones.

Let γ^* be our scaling factor. Then the new model factor will be

$$\rho_s^* = \gamma^* \tilde{m}_s^*,$$

and thus the total "new" premium will be $\gamma^* \sum_{k=1}^{K} w_k \tilde{m}_k^*$ whereas the "old" premium is $\sum_{k=1}^{K} p_k$ with

$$p_k = \sum_{i=1}^{I_k} p_{ki}.$$

As these two premiums should be equal, we get

$$\gamma^* = \left(\sum_{k=1}^{K} p_k\right) \Big/ \sum_{k=1}^{K} w_k \tilde{m}_k^*.$$

In addition to our estimator for the class, we also want a confidence interval for the "correct" factor, by which we mean $\gamma m_s(\Theta_s)$, where γ denotes the mean of γ^*. To get such a confidence interval we need some additional assumption, and for simplicity we assume that the conditional distribution of $m_s(\Theta_s)$ given the observations is normal with mean $\tilde{m}_s$ and variance ψ_s. This assumption seems highly unrealistic, in particular for cars with low exposure, but we really did not need an exact confidence interval, only some measure of the uncertainty of the estimator, and for this purpose the assumption seems adequate. As a 1-ε confidence interval (in the Bayesian sense, cf. e.g. DEGROOT (1970, subsections 11.5–6)) for the factor we obtain $\tilde{m}_s \pm g_{1-\varepsilon/2}\sqrt{\psi_s}$, where $g_{1-\varepsilon/2}$ denotes the $1-\varepsilon/2$ fractile in the standard normal distribution $N(0, 1)$ and by insertion of estimators for unknown parameters we finally get the estimated confidence interval $\rho_s^* \pm \gamma^* g_{1-\varepsilon/2}\sqrt{\psi_s^*}$.

From the estimator and the confidence interval of the model factor, we can trivially derive an estimator and a confidence interval for the model class (cf. Section 1).

When a new car model t, for which we have no data, is to be classified, we have $v_t = \zeta_t^* = 0$, which gives

$$\rho_t^* = \gamma^* \mu_t^* = \gamma x_t' \boldsymbol{\beta}^*,$$

that is,

$$\rho_t^* = x_t' \boldsymbol{\alpha}^* = \sum_{j=1}^{q} \alpha_j^* x_{tj} \tag{3.16}$$

with

$$\boldsymbol{\alpha}^* = (\alpha_1^*, \ldots, \alpha_q^*)' = \gamma^* \boldsymbol{\beta}^*.$$

Thus, (3.16) is the formula to be used to find the model factor for car model t.

Let us for a moment call ρ_t^*, given by (3.16), $\rho_t^*(0)$ to stress that this is the factor estimate without exposure. When we get an observed exposure, we get the factor

$$\rho_t^*(1) = \zeta_t^* \gamma^* Y_t + (1 - \zeta_t^*)\rho_t^*(0),$$

that is, a weighted mean of the initial factor estimate and the empirical factor $\gamma^* Y_t$. We also note that with no exposure we have $\psi_t = \lambda$ and $\psi_t^* = \lambda^*$.

3.5. *Choice of Regressors*

In subsection 3.1 we said that x_k should be a design vector based on the technical data of car model k without giving any further indication of which regressors one should use. In our numerical investigations we registered for each car model in our test sample the four basic variables engine power, cylinder volume, weight, and price. Diesel cars and cars with four-wheel drive were not included in our sample; otherwise it would have been appropriate to include (0,1)-variables for these characteristics. As interesting regressors we concentrated on the four basic variables and ratios between them.

It should be noted that the estimator ϕ^* of ϕ does not depend on the chosen regressors. For λ^* and $\boldsymbol{\beta}^*$ we made several computations using different regressors. In each design we of course included a constant term, that is, the first element of $\boldsymbol{x}_k$ being equal to 1.

As

$$\lambda = E(m_k(\Theta_k) - \boldsymbol{x}_k'\boldsymbol{\beta})^2,$$

λ measures how close the prior mean is to $m_k(\Theta_k)$, and it was therefore felt that one should use a set of regressors making λ^* small. This is also consistent with our choice of the quadratic loss function; one could think of all the possible regressors being studied as included in a huge design, but that for most of them we estimate the corresponding element of $\boldsymbol{\beta}$ by zero.

An important point when choosing regressors is that we know something about monotonicity. To motivate this, let us look at an example. At an early stage of our research we wanted to classify some new car models for which the prices were still unknown. A design giving small λ^* under these circumstances was (1 power/weight weight/power)'. For two of the cars we got the following results:

Car	Weight	Power	Class
1	1200 kg	63 HP	59
2	1227 kg	86 HP	42

This seems of course very unreasonable. Car 2 has a slightly higher weight and a much higher power than Car 1, but should be rated lower!

In accordance with our opinion about monotonicity, several sets of regressors were rejected when looking at $\boldsymbol{\beta}^*$. It should be noted that the more regressors we include, the more difficult it would be to control that our opinion about monotonicity is satisfied as the different regressors could be strongly correlated; even if we mean that the factor should be increasing with cylinder volume, it need not be disturbing to get a negative coefficient for this regressor if engine power has a positive coefficient, as cylinder volume and engine power are strongly correlated. Under these considerations we conclude that q should not be too large, say, at most 4–5.

It should be noted that the monotonicity secured by the choice of regressors is not necessarily satisfied for the posterior estimates $\tilde{m}_s^*$ with positive exposures. This is reasonable as we then have more information apart from the technical data; the monotonicity is important only when we base the factor on only the technical data.

One should be aware that in one respect price is different from the other basic variables considered, as the price may change whereas the car model is still the same.

We conclude this subsection by briefly recapitulating the criteria that should be taken into consideration by the choice of regressors:

(i) small λ^*;
(ii) monotonicity;
(iii) small q.

3.6. *Some Practical Modifications*

In subsection 3.3 we described how we would have estimated ϕ if we had had the necessary data. Unfortunately, we did not have them. From (3.6) we see that for each policy we had to match the exposure with the total amount of the claims occurred during the exposure period. At present, the data of Storebrand are organized such that for each calendar year we have one claims file and one policy file. The claims file contains data for all claims *reported* during the year. As stated above, we really wanted the claims *occurred* during the year, but this does not seem to be a serious problem. The policy file contains data from the middle of the year. The registered premium is the premium at the latest renewal prior to the middle of the year, which means that these renewals range from the middle of the previous year until the middle of the present year. Thus a match between claims and policies would be awkward. We also have the problem that the total registered premium for a fixed car model is not really the premium we wanted it to be, but we decided to use it as an approximation. If the premium volume of the car model is relatively stable over time, this approximation should be acceptable. However, if the premium volume is growing, we would register too low a value for the exposure volume. This will in particular be the case when a new car model is introduced, most extremely for cars introduced in the second half of the year, for which we may have claims, but no premium. Such cars should not be included in the parameter estimation.

The following additional model assumptions and estimation method were applied. Let N_{ki} be the number of claims from risk unit i of car model k, and let Z_{kij} denote the claim amount of the jth of these claims. Then

$$X_{ki} = \sum_{j=1}^{N_{ki}} Z_{kij}.$$

We assume that given Θ_k, the Z_{kij}'s are conditionally independent and identically distributed and conditionally independent of the N_{ki}'s. It is further assumed that N_{ki} is conditionally Poisson distributed with parameter $w_{ki}r_k(\Theta_k)$ given Θ_k. It is well known that under these conditions

$$\mathrm{Var}[X_{ki} \mid \Theta_k] = w_{ki}r_k(\Theta_k)t_k(\Theta_k)$$

with

$$t_k(\Theta_k) = E[Z_{kij}^2 \mid \Theta_k],$$

and by using (3.3) we obtain

$$s^2(\Theta_k) = r_k(\Theta_k)t_k(\Theta_k). \tag{3.17}$$

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

We now have that

$$\phi_k^* = \left(\sum_{i=1}^{I_k} \sum_{j=1}^{N_{ki}} Z_{kij}^2\right) \bigg/ w_k$$

satisfies $E[\phi_k^* \mid \Theta_k] = s^2(\Theta_k)$, and thus

$$\phi^* = \sum_{k=1}^{K} u_k \phi_k^*$$

is an unbiased estimator of ϕ for all weights u_k. (We stress that the quantities ϕ_k^* and ϕ^* defined in the present subsection are not the same as the quantities defined in subsection 3.3, hoping that this abuse of notation will not present any problems to the reader.)

The author is not quite happy with the introduction of the present compound Poisson assumption in our model. From (3.17) we see that the functions r_k and t_k depend on k whereas their product is independent of k. And r_k really should depend on k as an independence assumption would imply that technical data have no influence on the number of claims, which seems very unrealistic.

The fact that in our test sample ϕ_k^* was strongly correlated with the four basic technical variables, could be a consequence of the issue discussed in the previous paragraph. Let a_k denote the engine power of car model k. From our test sample consisting of 62492 policies distributed on 90 different car models, we found the correlations displayed in Table 3.1 by using a correlation procedure in SAS. As is seen from the table, the correlations become considerably lower if we divide ϕ_k^* by a_k. Therefore we replace assumption (3.3) by

$$\mathrm{Var}[Y_{ki} \mid \Theta_k] = a_k s^2(\Theta_k)/w_{ki} = s^2(\Theta_k)/v_{ki}$$

with $v_{ki} = w_{ki}/a_k$. Under this assumption (3.17) should be replaced by

$$s^2(\Theta_k) = r_k(\Theta_k) t_k(\Theta_k)/a_k.$$

We get that

$$\phi_k^* = \left(\sum_{i=1}^{I_k} \sum_{j=1}^{N_{ki}} Z_{kij}^2\right) \bigg/ (a_k w_{ki})$$

satisfies $E[\phi_k^* \mid \Theta_k] = s^2(\Theta_k)$, and thus

$$\phi^* = \sum_{k=1}^{K} u_k \phi_k^*$$

TABLE 3.1

CORRELATION OF ϕ_k^* AND ϕ_k^*/a_k WITH THE FOUR BASIC RISK VARIABLES

	ϕ_k^*		ϕ_k^*/a_k	
	unweighted	weight w_k	unweighted	weight v_k
Weight	0.249	0.445	−0.009	0.078
Power	0.337	0.499	0.041	0.083
Cylinder volume	0.346	0.509	0.063	0.106
Price	0.286	0.486	0.060	0.187

is an unbiased estimator of $\phi = Es^2(\Theta_k)$ for all weights u_k; in the numerical example in Section 6 we applied $u_k = K^{-1}$.

The reason that we introduced a v_{ki} in subsection 3.1 should now be clear: The derivations made in the previous subsections are still valid under our revised assumptions; we have only changed the definition of some of the quantities.

3.7. *Introduction of Subjective Assessment*

Classification of individual car models by credibility has also been treated by CAMPBELL (1986). He computes a model factor by a pure Bühlmann–Straub model, that is, he makes no regression assumption about the technical attributes of the car. However, before performing the credibility analysis, he divides the cars by using cluster analysis into groups of cars that are similar with respect to technical attributes. The credibility analysis is then performed within each group of car models. Roughly speaking, one could say that in our set-up the regression assumption plays the role of the cluster analysis in Campbell's set-up.

After the Bühlmann–Straub analysis has been performed, Campbell lets the final value of the model factor be a weighted mean of the value found by the Bühlmann–Straub analysis and a subjective estimate based on a technical assessment of the car.

Let us now see how one could incorporate a subjective estimator in our model. We assume that when car model k is initially classified, a skilled person proposes a class C_k. His proposal is based on a technical assessment of the car. From the class C_k we find the factor

$$F_k = 1.04^{C_k - 30}.$$

This factor is not yet comparable to $m_k(\Theta_k)$ as it is differently scaled (cf. subsection 3.4). From (3.1) and (3.4) we get

$$Y_k = (X_k / p_k) f_k,$$

which motivates the scaling factor

$$N = \left(\sum_{k=1}^{K} X_k\right) \Big/ \sum_{k=1}^{K} p_k,$$

and we introduce the rescaled model factor

$$A_k = N F_k.$$

We now assume that A_k is independent of the data from the other car models, that it is conditionally independent of $Y_{k1}, \ldots, Y_{kI_k}$ given Θ_k, and that

$$E[A_k \mid \Theta_k] = m_k(\Theta_k) \qquad E\,\mathrm{Var}[A_k \mid \Theta_k] = \tau.$$

Now let $\tilde{m}_s$ be the credibility estimator of $m(\Theta_s)$ based on $Y_{s1}, \ldots, Y_{sI_s}$, and

A_s and let

$$\psi_s = E(m_s(\Theta_s) - \tilde{m}_s)^2.$$

From Theorem 2.1 we get

$$\tilde{m}_s = (v_s Y_s + \varepsilon A_s + \varkappa \mu_s)/(v_s + \varepsilon + \varkappa)$$
$$\psi_s = \phi/(v_s + \varepsilon + \varkappa)$$

with

$$\varepsilon = \phi/\tau.$$

We have that

$$\hat{\tau} = \sum_{k=1}^{K} u_k\{(Y_k - A_k)^2 - \phi^*/v_k\}$$

is an unbiased estimator of τ for all weights u_k, but as $\hat{\tau}$ can take negative values, we propose to estimate τ by $\tau^* = \max(\hat{\tau}, 0)$.

We can of course still estimate λ and $\boldsymbol{\beta}$ by the estimators previously found, but if we also want to include the A_k's in the estimation, we can easily modify the estimators presented in subsection 3.3 by using the following trick: We simply transform the subjective estimator A_k to an artificial risk unit $I_k + 1$ with "risk volume"

$$v_{k,I_k+1} = \varepsilon \tag{3.18}$$

and "claim amount"

$$X_{k,I_k+1} = \varepsilon A_k. \tag{3.19}$$

By adding the new risk units $X_{1,I_1+1}, \ldots, X_{K,I_K+1}$ to the statistics data, we can estimate λ and $\boldsymbol{\beta}$ in exactly the same way as in subsection 3.3. In (3.18) and (3.19) we estimate ε by

$$\varepsilon^* = \phi^*/\tau^*.$$

This author is for two reasons a bit reluctant about the introduction of the subjective estimator A_s in the credibility estimator $\tilde{m}_s$. Both reasons really have as a consequence that the model assumptions made about the A_k's are not fulfilled in practice.

Firstly, the person performing the assessment will probably gradually adapt himself to the statistical model. He will get a feeling of what class the statistical model will propose, and thus his assessment is no longer independent. This does not seem to be an important objection, but it means that after a while the attitude of the person is apt to change, and thus one should frequently update the estimate of τ.

The second objection is more serious. In a competitive market like the Norwegian one, not only the risk level of the car will influence the person performing the assessment, but also the classification of similar cars, not only by Storebrand, but also by the competing companies.

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

Thus this author is more attracted by the opinion expressed in Section 1, that the subjective assessment should be influenced by the statistical method instead of influencing the method itself.

4. A HIERARCHICAL APPROACH

4.1. *Model*

The make of the car is a characteristic that we have not mentioned yet, but it could contain valuable information about the risk of a car; the information that the car is a Mercedes Benz, may contain information about both the car and its driver that is not contained in characteristics like price, power, etc. It should be noted that make differs from the characteristics studied in subsection 3.5 in one important respect; whereas those characteristics were *quantitative,* make is *qualitative,* and thus we cannot directly include make in the set-up of Section 3. One possible approach would be to extend the regression analysis of Section 3 to a covariance analysis. Instead of following that line we are going to extend the non-hierarchical regression model of Section 3 to a hierarchical model with a new level representing the make of the car.

Consider a group of N different makes. For make n we have observed K_n different car models, and for model k of these we have observed I_{nk} risk units. Let X_{nki} denote the total claim amount in the exposure time for unit i of model k of make n, and let p_{nki} be the earned premium. We introduce

$$w_{nki} = p_{nki}/f_{nk},$$

where f_{nk} denotes the old factor for make of car.

We assume that claim amounts from cars of different makes are independent, and that from within one make n, claim amounts from different car models are conditionally independent given a random parameter H_n (capital Greek eta) characterizing make n. Within car model k of make n, the claim amounts from different risk units are assumed to be conditionally independent given $(\Theta_{nk}, \mathrm{H}_n)$, where Θ_{nk} is a random parameter characterizing car model k of make n. It is assumed that $\Theta_{n1}, \ldots, \Theta_{nK_n}$ are conditionally independent and identically distributed given H_n, and that their common conditional distribution depends on the make only through the value of H_n. We further assume that $\mathrm{H}_1, \ldots, \mathrm{H}_N$ are independent and identically distributed.

Let

$$Y_{nki} = X_{nki}/w_{nki}.$$

It is assumed that

$$E[Y_{nki} \mid \Theta_{nk}, \mathrm{H}_n] = m_{nk}(\Theta_{nk}, \mathrm{H}_n)$$

$$E\operatorname{Var}[m_{nk}(\Theta_{nk}, \mathrm{H}_n) \mid \mathrm{H}_n] = \lambda$$

$$\operatorname{Var}[Y_{nki} \mid \Theta_{nk}, \mathrm{H}_n] = s^2(\Theta_{nk}, \mathrm{H}_n)/v_{nki}$$

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

with $v_{nki} = w_{nki}/a_{nk}$, where a_{nk} is a known quantity which could be equal to one, engine power (cf. subsection 3.6), or something else. We further assume that

$$E[m_{nk}(\Theta_{nk}, \mathrm{H}_n)|\,\mathrm{H}_n] = \boldsymbol{x}'_{nk}\boldsymbol{b}(\mathrm{H}_n), \tag{4.1}$$

where $\boldsymbol{x}_{nk}$ is a known, non-random $q \times 1$ design vector based on the technical data of the car and $\boldsymbol{b}$ is a $q \times 1$ vector function. We introduce

$$\phi = Es^2(\Theta_{nk}, \mathrm{H}_n) \qquad \boldsymbol{\beta} = E\boldsymbol{b}(\mathrm{H}_n)$$
$$\boldsymbol{\Xi} = \mathrm{Cov}\; \boldsymbol{b}(\mathrm{H}_n) \tag{4.2}$$
$$\varkappa = \phi/\lambda$$
$$X_{nk} = \sum_{i=1}^{I_{nk}} X_{nki} \qquad v_{nk} = \sum_{i=1}^{I_{nk}} v_{nki} \qquad w_{nk} = \sum_{i=1}^{I_{nk}} w_{nki}$$
$$Y_{nk} = X_{nk}/w_{nk}.$$

We note that for $\Xi = \mathbf{0}$, the model reduces to the non-hierarchical model studied in Section 3.

4.2. *Credibility Estimators of $m_r(\Theta_{rs}, \mathrm{H}_r)$*

Let $\tilde{m}_{rs}$ and $\tilde{\boldsymbol{b}}_r$ denote the credibility estimators of $m_{rs}(\Theta_{rs}, \mathrm{H}_r)$ and $\boldsymbol{b}(\mathrm{H}_r)$ based on the observed Y_{nki}'s. We introduce the estimation errors

$$\psi_{rs} = \mathrm{Var}(m_{rs}(\Theta_{rs}, \mathrm{H}_r) - \tilde{m}_{rs}) \qquad \boldsymbol{\Pi}_r = \mathrm{Cov}(\boldsymbol{b}(\mathrm{H}_r) - \tilde{\boldsymbol{b}}_r).$$

Then we have the following result.

THEOREM 4.1. *We have*

$$\tilde{m}_{rs} = \zeta_{rs} Y_{rs} + (1 - \zeta_{rs})\boldsymbol{x}'_{rs}\tilde{\boldsymbol{b}}_r \tag{4.3}$$

$$\psi_{rs} = (1 - \zeta_{rs})\{\lambda + (1 - \zeta_{rs})\boldsymbol{x}'_{rs}\boldsymbol{\Pi}_r\boldsymbol{x}_{rs}\} \tag{4.4}$$

with

$$\zeta_{rs} = v_{rs}/(v_{rs} + \varkappa).$$

PROOF. As the coefficients of credibility estimators depend on only first- and second-order moments, it is sufficient to prove the result for a special case having the same first- and second-order moments as the general case. It is convenient to consider multinormal distributions as it is well-known that in that case the Bayes estimators are linear, and hence they are equal to the credibility estimators.

Let

$$W_{nki} = v_{nki}^{1/2}\{Y_{nki} - m_{nk}(\Theta_{nk}, \mathrm{H}_n)\}$$
$$U_{nk} = m_{nk}(\Theta_{nk}, \mathrm{H}_n) - \boldsymbol{x}'_{nk}\boldsymbol{b}(\mathrm{H}_n).$$

We assume that the W_{nki}'s are independent and identically distributed $N(0, \phi)$, the U_{nk}'s are independent and identically distributed $N(0, \lambda)$, the $\boldsymbol{b}(\mathrm{H}_n)$'s are independent and identically distributed $N(\boldsymbol{\beta}, \Xi)$, and that the W_{nki}'s, the U_{nk}'s, and the $b(\mathrm{H}_n)$'s are independent. It is obvious that we have the same first- and second-order moments as in the distribution-free model. Furthermore, we have

$$E[m_{rs}(\Theta_{rs}, \mathrm{H}_r) \mid b(\mathrm{H}_r), Y_{rki} \forall (k, i)] = \zeta_{rs} Y_{rs} + (1 - \zeta_{rs}) x'_{rs} b(\mathrm{H}_r)$$

as under the conditional probability measure given $b(\mathrm{H}_r)$ we have the same first- and second-order moment structure for make r as in subsection 3.1 (cf. formula (3.5)). We get

$$\begin{aligned} \tilde{m}_{rs} &= E[m_{rs}(\Theta_{rs}, \mathrm{H}_r) \mid Y_{nki} \forall (n, k, i)] \\ &= E[E[m_{rs}(\Theta_{rs}, \mathrm{H}_r) \mid b(\mathrm{H}_r), Y_{nki} \forall (n, k, i)] \mid Y_{nki} \forall (n, k, i)] \\ &= E[E[m_{rs}(\Theta_{rs}, \mathrm{H}_r) \mid b(\mathrm{H}_r), Y_{rki} \forall (k, i)] \mid Y_{nki} \forall (n, k, i)] \end{aligned}$$

as different makes are independent, and thus

$$\begin{aligned} \tilde{m}_{rs} &= \zeta_{rs} Y_{rs} + (1 - \zeta_{rs}) x'_{rs} E[b(\mathrm{H}_r) \mid Y_{nki} \forall (n, k, i)] \\ &= \zeta_{rs} Y_{rs} + (1 - \zeta_{rs}) x'_{rs} \tilde{b}_r, \end{aligned}$$

which proves (4.3)

For ψ_{rs} we apply the same way of reasoning and get

$$\begin{aligned} \psi_{rs} &= E \operatorname{Var}[m_{rs}(\Theta_{rs}, \mathrm{H}_r) \mid Y_{nki} \forall (n, k, i)] \\ &= E \operatorname{Var}[m_{rs}(\Theta_{rs}, \mathrm{H}_r) \mid b(\mathrm{H}_r), Y_{nki} \forall (n, k, i)] \\ &\quad + E \operatorname{Var}[E[m_{rs}(\Theta_{rs}, \mathrm{H}_r) \mid b(\mathrm{H}_r), Y_{nki} \forall (n, k, i)] \mid Y_{nki} \forall (n, k, i)] \\ &= \lambda(1 - \zeta_{rs}) + (1 - \zeta_{rs})^2 x'_{rs} (\operatorname{Cov}[b(\mathrm{H}_r) \mid Y_{nki} \forall (n, k, i)] \, x_{rs} \\ &= (1 - \zeta_{rs}) [\lambda + (1 - \zeta_{rs}) x'_{rs} \Pi_r x_{rs}], \end{aligned}$$

which proves (4.4).

This completes the proof of Theorem 4.1. Q.E.D.

We now want an expression for $\tilde{b}_r$. To reduce the dimension of the problem we first prove the following lemma.

LEMMA 4.1. *The credibility estimator $\tilde{b}_r$ depends on the Y_{nki}'s only through $Y_{r1}, \ldots, Y_{rK_r}$.*

PROOF. Let $\tilde{b}_r^{(1)}$ be credibility estimator of $b(\mathrm{H}_r)$ based on $Y_{r1}, \ldots, Y_{rK_r}$. Then by Theorem 2.1(ii)

$$\begin{aligned} E\tilde{b}_r^{(1)} &= Eb(\mathrm{H}_r) \\ \operatorname{Cov}(\tilde{b}_r^{(1)}, Y_{rk}) &= \operatorname{Cov}(b(\mathrm{H}_r), Y_{rk}). \qquad k = 1, \ldots, K_r \end{aligned}$$

As

$$\begin{aligned} \operatorname{Cov}(Y_{rs}, Y_{rki}) &= \operatorname{Cov}(Y_{rs}, Y_{rk}) \\ \operatorname{Cov}(b(\mathrm{H}_r), Y_{rki}) &= \operatorname{Cov}(b(\mathrm{H}_r), Y_{rk}) \end{aligned}$$

for all (k, s, i), we get

$$\operatorname{Cov}(\tilde{b}_r^{(1)}, Y_{rki}) = \operatorname{Cov}(b(\mathrm{H}_r), Y_{rki}).$$

Furthermore, as different makes are independent, we have

$$\mathrm{Cov}(\tilde{\boldsymbol{b}}_r^{(1)}, Y_{nki}) = \mathrm{Cov}(\boldsymbol{b}(\mathrm{H}_r), Y_{nki}) = \mathbf{0}$$

for all $n \neq r$, and thus Lemma 4.1 follows from Theorem 2.1(ii). Q.E.D.

We want to find a matrix expression for $\tilde{\boldsymbol{b}}_r$ and introduce

$$\mathbf{Z}_r = \mathrm{diag}(\zeta_{r1}, \ldots, \zeta_{rK_r})$$

$$X_r = (\boldsymbol{x}_{r1}, \ldots, \boldsymbol{x}_{rK_r})' \qquad Y_r = (Y_{r1}, \ldots, Y_{rK_r})'.$$

We write $\tilde{\boldsymbol{b}}_r$ as

$$\tilde{\boldsymbol{b}}_r = \boldsymbol{\gamma}_r + \boldsymbol{\Gamma}_r Y_r.$$

From (2.1) we get

$$\boldsymbol{\gamma}_r = \boldsymbol{\beta} - \boldsymbol{\Gamma}_r X_r \boldsymbol{\beta},$$

that is,

$$\tilde{\boldsymbol{b}}_r = \boldsymbol{\Gamma}_r(Y_r - X_r\boldsymbol{\beta}) + \boldsymbol{\beta}.$$

From (2.2) we obtain

$$\boldsymbol{\Gamma}_r \,\mathrm{Cov}\, Y_r = \mathrm{Cov}(\boldsymbol{b}(\mathrm{H}_r), Y_r'). \tag{4.5}$$

We easily get

$$\mathrm{Cov}\; Y_r = \lambda \mathbf{Z}_r^{-1} + X_r \boldsymbol{\Xi} X_r' \qquad \mathrm{Cov}(\boldsymbol{b}(\mathrm{H}_r), Y_r') = \boldsymbol{\Xi} X_r',$$

and insertion in (4.5) gives

$$\boldsymbol{\Gamma}_r(\lambda \mathbf{Z}_r^{-1} + X_r \boldsymbol{\Xi} X_r') = \boldsymbol{\Xi} X_r'. \tag{4.6}$$

We multiply (4.6) by $\mathbf{Z}_r X_r$ from the right to obtain

$$\boldsymbol{\Gamma}_r X_r(\lambda I_q + \boldsymbol{\Xi} X_r' \mathbf{Z}_r X_r) = \boldsymbol{\Xi} X_r' \mathbf{Z}_r X_r,$$

which gives

$$\boldsymbol{\Gamma}_r X_r = \boldsymbol{\Xi} X_r' \mathbf{Z}_r X_r(\lambda I_q + \boldsymbol{\Xi} X_r' \mathbf{Z}_r X_r)^{-1}. \tag{4.7}$$

From (4.6) we get

$$\boldsymbol{\Gamma}_r \lambda \mathbf{Z}_r^{-1} = (I_q - \boldsymbol{\Gamma}_r X_r)\boldsymbol{\Xi} X_r', \tag{4.8}$$

that is,

$$\boldsymbol{\Gamma}_r = \lambda^{-1}(I_q - \boldsymbol{\Gamma}_r X_r)\boldsymbol{\Xi} X_r' \mathbf{Z}_r.$$

Insertion of (4.7) gives

$$\boldsymbol{\Gamma}_r = (\lambda I_q + \boldsymbol{\Xi} X_r' \mathbf{Z}_r X_r)^{-1} \boldsymbol{\Xi} X_r' \mathbf{Z}_r.$$

If $X_r' \mathbf{Z}_r X_r$ is non-singular, we introduce

$$\hat{\boldsymbol{b}}_r = (X_r' \mathbf{Z}_r X_r)^{-1} X_r' \mathbf{Z}_r Y_r.$$

Then we have

$$\Gamma_r Y_r = (\lambda I_q + \Xi X_r' Z_r X_r)^{-1} \Xi X_r' Z_r Y_r = (\lambda I_q + \Xi X_r' Z_r X_r)^{-1} \Xi X_r' Z_r X_r \hat{b}_r$$
$$= \Xi X_r' Z_r X_r (\lambda I_q + \Xi X_r' Z_r X_r)^{-1} \hat{b}_r = \Delta_r \hat{b}_r,$$

where we have introduced the credibility matrix

$$\Delta_r = \Xi X_r' Z_r X_r (\lambda I_q + \Xi X_r' Z_r X_r)^{-1},$$

and we get

$$\tilde{b}_r = \Delta_r \hat{b}_r + (I_q - \Delta_r)\beta.$$

We still have to find an expression for the estimation error matrix Π_r. By Theorem 2.1(iii) we get

$$\Pi_r = \text{Cov } b(\text{H}_r) - \text{Cov } \tilde{b}_r = \Xi - \Gamma_r (\text{Cov } Y_r) \Gamma_r' = \Xi - \Gamma_r (\lambda Z_r^{-1} + X_r \Xi X_r') \Gamma_r'.$$

We insert (4.6) and obtain

$$\Pi_r = \Xi - \Xi X_r' \Gamma_r' = \Xi (I_q - X_r' \Gamma_r'),$$

that is,

$$\Pi_r = \Xi (I_q - \Delta_r') = (I_q - \Delta_r) \Xi,$$

the last equality because Π_r and Ξ are symmetric.

We now have expressions for all the quantities that we need for the computation of $\tilde{m}_{rs}$ and ψ_{rs}.

4.3. *Parameter Estimation*

Corresponding to (3.6) we introduce

$$\phi_{nk}^* = (I_{nk} - 1)^{-1} \sum_{i=1}^{I_{nk}} v_{nki} (Y_{nki} - Y_{nk})^2,$$

for which we have

$$E[\phi_{nk}^* \mid \Theta_{nk}, \text{H}_n] = s^2(\Theta_{nk}, \text{H}_n)$$

if $v_{nki} = w_{nki}$, and in that case

$$\phi^* = \sum_{n=1}^{N} \sum_{k=1}^{K_n} u_{nk} \phi_{nk}^*$$

is an unbiased estimator of ϕ for all weights $u_{nk} (\sum_{n=1}^{N} \sum_{k=1}^{K_n} u_{nk}) = 1$.

It should be obvious how one could generalize the assumptions and estimators introduced in subsection 3.6 to the hierarchical model, and we shall not go any further into details on that matter.

In the following we just assume that we have got an unbiased estimator ϕ^* of

ϕ, and the following derivations do not depend on whether $v_{nki} = w_{nki}$ or not.

For the estimation of λ, $\boldsymbol{\beta}$, and $\boldsymbol{\Xi}$ we shall also assume that X_n has full rank q for all n. In practice, this will mean that we exclude data from makes for which we have observed only a few car models, from the estimation procedures. It is of course questionable not to utilize these data, but the estimation procedures become much simpler.

We introduce

$$v_n = \sum_{k=1}^{K_n} v_{nk} \qquad v = \sum_{n=1}^{N} v_n$$

$$D_n = v_n^{-1}\text{diag}(v_{n1}, \ldots, v_{nK_n})$$

$$\ddot{b}_n = (X_n' D_n X_n)^{-1} X_n' D_n Y_n.$$

Analogous to what we did in subsection 3.3, we get

$$E(Y_n - X_n\ddot{b}_n)' D_n (Y_n - X_n\ddot{b}_n) = \lambda[1 - \text{tr}\{(X_n'D_nX_n)^{-1}X_n'D_n^2X_n\}] + (K_n - q)\phi/v_n,$$

and thus

$$\hat{\lambda} = \frac{1}{v}\sum_{i=1}^{N} \frac{v_n(Y_n - X_n\ddot{b}_n)'D_n(Y_n - X_n\ddot{b}_n) - (K_n - q)\hat{\phi}}{1 - \text{tr}\{(X_n'D_nX_n)^{-1}X_n'D_n^2X_n\}}$$

is an unbiased estimator of λ. As $\hat{\lambda}$ may take negative values, we proceed as in subsection 3.3 to construct a modified estimator λ^* which is non-negative or positive. In the following, we assume for simplicity that λ^* is positive.

Let

$$W_n = \left(\sum_{r=1}^{N} X_r' \mathbf{Z}_r X_r\right)^{-1} X_n' \mathbf{Z}_n X_n$$

$$\hat{\boldsymbol{\beta}} = \sum_{n=1}^{N} W_n \hat{b}_n = \left(\sum_{n=1}^{N} X_n' \mathbf{Z}_n X_n\right)^{-1} \sum_{n=1}^{N} X_n' \mathbf{Z}_n Y_n.$$

It seems reasonable to base our estimator of $\boldsymbol{\Xi}$ on

$$Q = \sum_{n=1}^{N} W_n(\hat{b}_n - \hat{\boldsymbol{\beta}})(\hat{b}_n - \hat{\boldsymbol{\beta}})'.$$

We have

$$\begin{aligned} EQ &= \sum_{n=1}^{N} W_n\, E(\hat{b}_n - \hat{\boldsymbol{\beta}})(\hat{b}_n - \hat{\boldsymbol{\beta}})' = \sum_{n=1}^{N} W_n \,\text{Cov}(\hat{b}_n - \hat{\boldsymbol{\beta}}) \\ &= \sum_{n=1}^{N} W_n[\text{Cov}\ \hat{b}_n - \text{Cov}(\hat{b}_n, \hat{\boldsymbol{\beta}}') - \text{Cov}(\hat{\boldsymbol{\beta}}, \hat{b}_n') + \text{Cov}\ \hat{\boldsymbol{\beta}}] \\ &= \sum_{n=1}^{N} W_n[\text{Cov}\ \hat{b}_n - \text{Cov}(\hat{\boldsymbol{\beta}}, \hat{b}_n')], \end{aligned}$$

that is,

$$EQ = \sum_{n=1}^{N} W_n(I_q - W_n)\ \text{Cov}\ \hat{b}_n. \tag{4.9}$$

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

For all n we have

$$\text{Cov } \hat{b}_n = (X_n' \mathbf{Z}_n X_n)^{-1} X_n' \mathbf{Z}_n (\text{Cov } Y_n) \mathbf{Z}_n X_n (X_n' \mathbf{Z}_n X_n)^{-1}$$
$$= (X_n' \mathbf{Z}_n X_n)^{-1} X_n' \mathbf{Z}_n (\lambda \mathbf{Z}_n^{-1} + X_n' \Xi X_n) \mathbf{Z}_n X_n (X_n' \mathbf{Z}_n X_n)^{-1} = \lambda (X_n' \mathbf{Z}_n X_n)^{-1} + \Xi,$$

and insertion in (4.9) gives

$$EQ = \mathbf{T}\lambda + \left(\sum_{n=1}^{N} W_n (I_q - W_n)\right) \Xi$$

with

$$\mathbf{T} = \sum_{n=1}^{N} W_n (I_q - W_n)(X_n' \mathbf{Z}_n X_n)^{-1} = \sum_{n=1}^{N} (I_q - W_n) W_n (X_n' \mathbf{Z}_n X_n)^{-1}$$
$$= \left(\sum_{n=1}^{N} (I_q - W_n)\right) \left(\sum_{r=1}^{N} X_r' \mathbf{Z}_r X_r\right)^{-1} = (N-1) \left(\sum_{r=1}^{N} X_r' \mathbf{Z}_r X_r\right)^{-1}$$

that is,

$$EQ = (N-1) \left(\sum_{r=1}^{N} X_r' \mathbf{Z} X_r\right)^{-1} \lambda + \left(I_q - \sum_{r=1}^{N} W_n^2\right) \Xi,$$

and

$$\ddot{\Xi} = \left(I_q - \sum_{n=1}^{N} W_n^2\right)^{-1} \left\{\sum_{n=1}^{N} W_n (\hat{b}_n - Y\hat{\beta})(\hat{b}_n - Y\hat{\beta})' - (N-1) \left(\sum_{r=1}^{N} X_r' \mathbf{Z}_r X_r\right)^{-1} \hat{\lambda}\right\}$$

is an unbiased estimator of Ξ. However, as Ξ is symmetric whereas $\ddot{\Xi}$ does not in general have this property, we replace $\ddot{\Xi}$ by

$$\hat{\Xi} = (\ddot{\Xi} + \ddot{\Xi}')/2.$$

When estimating λ, we had the problem that $\hat{\lambda}$ was not necessarily positive. The analogous problem when estimating Ξ is that $\hat{\Xi}$ is not necessarily positive semi-definite. As $\hat{\Xi}$ is symmetric, it can be written as

$$\hat{\Xi} = A' \mathbf{T} A,$$

where A is an orthonormal $q \times q$ matrix (i.e. $A'A = I_q$) and

$$\mathbf{T} = \text{diag}(\tau_1, \ldots, \tau_q),$$

where $\tau_1, \ldots, \tau_q$ denote the eigenvalues of $\hat{\Xi}$. Let

$$\tau_i^0 = \max(\tau_i, 0) \qquad i = 1, \ldots, q$$
$$\mathbf{T}^0 = \text{diag}(\tau_1^0, \ldots, \tau_q^0).$$

It can be shown (cf. BUNKE and GLADITZ (1974), RAO (1965)) that

$$\Xi^* = A' \mathbf{T}^0 A$$

satisfies

$$v' (\Xi^* - \hat{\Xi}) v \leqslant v' (P - \hat{\Xi}) v$$

for all $q \times 1$ vectors v and all positive semi-definite $q \times q$ matrices P, and hence it seems reasonable to replace $\hat{\Xi}$ by Ξ^* to get a positive semi-definite estimator. To avoid having to take special care of the case when Ξ^* is not strictly positive definite, one could instead of replacing negative eigenvalues by zero, replace them by some small positive number; one possible choice would be ε/N for some ε as we would than have asymptotic unbiasedness when N goes to infinity.

The computation of Ξ^* from $\hat{\Xi}$, involving the construction of A and $\mathbf{T}$, may seem complicated. However, in SAS we had standard procedures for the computation of A and $\mathbf{T}$.

The procedure for estimation of Ξ depends on the parameters ϕ and λ, which were assumed to be unknown, and we therefore insert the estimators ϕ^* and λ^* for these parameters.

We have that

$$\beta^* = \left(\sum_{n=1}^{N} \Delta_n\right)^{-1} \sum_{n=1}^{N} \Delta_n \hat{b}_n \tag{4.10}$$

is the best linear unbiased estimator of β. As β^* depends on the unknown parameters ϕ, λ, and Ξ, we insert the estimators ϕ^*, λ^*, and Ξ^* for these parameters in (4.10).

We have now found estimators for all the unknown parameters involved in the credibility estimators presented in subsection 4.2, and we are therefore able to construct empirical versions of these credibility estimators.

4.4. *A Disadvantage of the Hierarchical Model*

For a new car model s of make r (i.e. $w_{rs} = 0$) we have

$$\tilde{m}_{rs} = x'_{rs}\tilde{b}_r.$$

In the non-hierarchical model the corresponding formula was

$$\tilde{m}_{rs} = x'_{rs}\beta.$$

In subsection 3.5 on the choice of regressors, we said that we have some prior opinion on monotonicity, and that the regressors should be chosen such that this monotonicity was preserved. This was not too complicated in the non-hierarchical model. In the hierarchical model it is much more difficult. Whereas in the non-hierarchical model we could just look at the sign of the elements of β, in the hierarchical model we have to look at the elements of $\tilde{b}_r$ for all r.

In a parametric empirical Bayes analysis we could solve the problem by restricting the support of the distribution of $b(\mathrm{H}_r)$ to a set $\mathscr{B}$ for which the monotonicity is preserved. Then of course also the posterior mean of $b(\mathrm{H}_r)$ would be contained in $\mathscr{B}$. However, such a parametric model would presumably be complicated to handle, and we would probably have to leave the linearity of the estimators.

If our statistical models should be used as proposed in Section 1, that is, not as giving the final answer, but as an aid for the person who finally makes the decision, this author would recommend that this person receives the estimates from both the hierarchical and the non-hierarchical models, using the same regressors in both models. In his decision he should be aware that the hierarchical model utilizes information about the make of the car, information that is not used in the non-hierarchical model. On the other hand, for the assessment of new car models, the non-hierarchical model will preserve some monotonicity properties, which may be violated in the hierarchical model.

4.5. *A Modified Approach*

When the author gave a seminar on the present research, Ragnar Norberg suggested a modified approach that avoids the monotonicity problem discussed in the previous subsection. We replace the assumptions (4.1) and (4.2) by

$$E[m_{nk}(\Theta_{nk}, \mathrm{H}_n) | \mathrm{H}_n] = M_{nk}(\mathrm{H}_n) \qquad EM_{nk}(\mathrm{H}_n) = \boldsymbol{x}'_{nk}\boldsymbol{\beta} \qquad \mathrm{Var}\, M_{nk}(\mathrm{H}_n) = \xi.$$

One could say that these assumptions are more consistent with the assumptions made in the non-hierarchical model whereas (4.1) and (4.2) are more in line with HACHEMEISTER's (1975) regression model.

Under these new assumptions we obtain

$$\tilde{m}_{rs} = \zeta_{rs} Y_{rs} + (1 - \zeta_{rs})(\boldsymbol{x}'_{rs}\boldsymbol{\beta} + D_r) \tag{4.11}$$

with

$$D_r = \left\{\xi \sum_{p=1}^{K_r} \zeta_{rp}(Y_{rp} - \boldsymbol{x}'_{rp}\boldsymbol{\beta})\right\} \Bigg/ \left(\lambda + \xi \sum_{p=1}^{K_r} \zeta_{rp}\right).$$

It is interesting to compare (4.11) to (3.5). We see that the only formal difference is that we have added a correction term D_r to the prior mean $\boldsymbol{x}'_{rs}\boldsymbol{\beta}$. For the case with no exposure for car model s this property is very attractive. We then get

$$\tilde{m}_{rs} = \boldsymbol{x}'_{rs}\boldsymbol{\beta} + D_r,$$

that is, we compute the prior mean $\boldsymbol{x}'_{rs}\boldsymbol{\beta}$ based on the technical data and add a correction term D_r as the car is of make r.

We hope to return to the present model in a subsequent paper.

5. SOME CARS ARE MORE EQUAL THAN OTHERS

As is well known, there are often several variants of one car model. In a Norwegian price list from 1984 (OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1984)) we found for instance 9 entries for Volkswagen Golf and 28 for Opel Ascona. The technical differences between such variants may be number of doors, engine, shape (coupé/sedan), etc. Such differences will of course in most cases also influence the price. In our investigations we have considered each variant as a separate model. However, variants of a car model usually have very much in common, and it is tempting to try to utilize this informaton in the estimation of the model factors.

One possible solution would be to extend our two-level hierarchical model (make, model) to three levels (make, model, variant) (for multi-level hierarchical models, cf. e.g. SUNDT (1980), NORBERG (1985)). This would be a more complicated model, and we would have to estimate more parameters.

Another possibility would be to drop the make level in the three-level hierarchical model to obtain a two-level model with levels for model and variant. For this model we could make the same assumptions as in Section 4, but the grouping of the cars would be different.

A third approach would be simply to consider different variants as one model. Then we have the difficulty that the different variants do not have the same technical specifications, but as design vector we could use a weighted mean of the design vectors for the different variants with weights proportional to the observed exposures. In this set-up, possible differences in risk characteristics of the variants now pooled together would be incorporated in $s^2(\Theta_k)$ (to use the notation of the non-hierarchical model). The present approach should be used with care as there exist variants with risk characteristics so different from other variants of the same model that they should definitely not be pooled together; a striking example is Volkswagen Golf GTI. Usually, one would be able to identify such "outliers" already before one obtains the risk statistics. However, this need not always be the case, and one should therefore, even if the variants are pooled together, always register the variant of each car in the statistics data so that one is able to detect an "outlier" and revise the pooling if necessary.

6. NUMERICAL EXAMPLE

6.1. *The Data*

We have already mentioned our numerical studies a couple of times. Our first studies were based on data from Storebrand for the year 1983, and in subsection 3.6 we presented some results based on these data. When our first studies had been performed, data from 1984 became available, and in our investigations on these data, we included a greater number of makes and car models than in our 1983 studies. In the present section we shall display figures found in our 1984 study; the 1983 data were analysed in the same way.

For each car model included in the study, we registered the technical variables weight, engine power, cylinder volume, and price. The price was the price given in a list from April 1984 (OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1984)), and we only included car models that were found in this list. This implies that we excluded car models that were no longer produced or imported to Norway. If one should also include older car models, one would have had to use older prices, which would have had to be adjusted to the price level of 1984. At the present stage of the development of models and methods, we decided to leave out this problem, but it is further discussed in SUNDT (1986). As already mentioned, for simplicity we also excluded diesel cars and cars with four-wheel drive.

In the following presentation we use the codes of Storebrand for make and

TABLE 6.1

Code	Name	K_n	Risk units	v_n
11	Audi	7	1112	9050
14	BMW	11	2754	15117
15	Citroën	10	1190	10429
16	Fiat	7	782	9541
17	Ford, British	7	2322	22326
18	Ford, German	24	13107	115557
24	Lancia	1	58	942
25	Mercedes Benz	7	1561	7444
31	Opel	33	8860	67880
33	Peugeot	14	1467	13017
34	Renault	1	950	13771
37	SAAB	6	3382	21261
39	Skoda	2	248	2549
45	Volkswagen	11	3145	35722
46	Volvo	19	3946	29881
47	Daihatsu	2	105	1555
53	Subaru	6	349	3076
54	Mitsubishi	14	1844	17962
66	Talbot	6	507	4976
93	Lada	5	3490	28800
94	Honda	9	3823	29963
96	Toyota	16	4034	35365
97	Nissan	14	3653	33067
98	Mazda	21	8069	69041
	Total	253	70758	598290

model. In Table 6.1 we give some summary policy data for our sample. For the headings of the table we have used the notation of Section 4, and in the following we use $v_{nk} = w_{nk}/$ (engine power). As we see from the table, we have applied data from in all 253 different car models distributed on 24 different makes. We applied no such pooling of car models as described in Section 5.

It would obviously be too much to present the results for all 253 car models, and we therefore restrict ourselves to give more detailed data for a representative sample of 25 car models found by including each tenth model from our total sample, ordered by the codes for make and model. In Table 6.2 we display the exposure and the technical variables engine power, weight, price, and price/weight. Prices are given in NOK and weights in kg.

We estimated ϕ by the procedure described in subsection 3.6 and found $\phi^* = 651.1$.

6.2. *The Non-hierarchical Approach*

For the non-hierarchical model we computed from the 1983 data for several different sets of regressors the estimates λ^* and β^* as described in subsection 3.3. According to the criteria given in subsection 3.5, it seemed reasonable to use the two regressors cylinder volume and price/weight, giving $q = 3$. However, it was

TABLE 6.2

Make	Model	Name	Power	Weight	Price	Price/weight	Risk units	v_{nk}
14	541	BMW 320 I	125	1105	162540	147.10	173	928
15	313	Citroën Visa GT	80	830	79200	95.42	35	288
16	321	Fiat Panda 45	45	670	48400	72.24	179	3251
17	328	Ford Escort 1.6 L	79	880	87560	99.50	738	6437
18	451	Ford Sierra 2.0	105	1095	102100	93.24	77	628
18	741	Ford Sierra 1.6	75	1100	109260	99.33	19	206
25	504	Mercedes Benz 190 E	122	1100	199560	181.42	185	1043
31	327	Opel Corsa 1.2 ST Sedan	55	775	67270	86.80	144	1747
31	347	Opel Kadett 1.2 S Combi	60	870	72620	83.47	269	2770
31	421	Opel Rekord 2.0 S	100	1140	118290	103.76	2879	18532
33	354	Peugeot 305 GLS	74	930	88020	94.65	249	2146
33	892	Peugeot 505 Break	100	1295	146580	113.19	50	372
39	323	Skoda 120 GLS	58	910	50627	55.63	74	877
45	523	Volkswagen Santana 1.9 GX	115	1100	138810	126.19	17	120
46	506	Volvo 240 GLT B23A	129	1330	178900	134.51	21	110
46	907	Volvo 240 GLE B23	129	1300	178400	137.23	10	52
53	349	Subaru 1600 GL Swing-Back	71	885	74800	84.52	48	374
54	396	Mitsubishi Galant 1600 GL	75	1065	99900	93.80	206	2188
93	411	Lada 1600 S	78	1040	54570	52.47	716	5669
94	417	Honda Prelude EX	106	985	181400	184.16	36	306
96	433	Toyota Carina Coupé	75	1060	94000	88.68	321	2790
97	321	Nissan Stanza 1.6 GL	81	970	93800	96.70	49	408
97	832	Nissan Bluebird 1.8 GL	88	1150	108300	94.17	327	2804
98	353	Mazda 626 1.6 GLX Sedan	81	1035	93900	90.72	153	1351
98	474	Mazda 929 2.0 DX St.Wagon	90	1200	108400	90.33	350	2835

argued that cylinder volume and engine power were strongly correlated, and that diesel cars and petrol cars were more comparable with respect to engine power than with respect to cylinder volume. Therefore it was felt that if we should later include also diesel cars in the analysis, it would be better to replace the regressor cylinder volume by engine power. We did this and got only a slightly higher value of λ^*. With the 1984 data we therefore concentrated on the design (1 power price/weight). We obtained

$$\lambda^* = 0.2063$$

$$\beta^* = (-0.4183\ 0.01238\ 0.01007)',$$

and from the values of ϕ^* and λ^* we found

$$\varkappa^* = \phi^*/\lambda^* = 3156.$$

In Table 6.3 we have displayed the observed Y_k, the estimated prior mean μ_k^*, the empirical credibility weight ζ_k^*, and the estimated estimation error ψ_k^* for each of the car models.

We see that Volkswagen Santana 1.9 GX and Volvo 240 GLE B23 have rather extreme values of Y_k. However, as these cars also have low exposure, $\tilde{m}_k^*$ does not differ much from μ_k^*.

We also computed estimates for tariff classes as described in subsection 3.4.

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

TABLE 6.3

Make	Model	Y_k	μ_k^*	$\tilde{m}_k^*$	ζ_k^*	ψ_k^*
14	541	3.336	2.610	2.775	0.2272	0.1595
15	313	0.502	1.533	1.447	0.0836	0.1891
16	321	2.283	0.866	1.585	0.5075	0.1016
17	328	1.465	1.561	1.497	0.6711	0.0679
18	451	5.628	1.820	2.452	0.1660	0.1721
18	741	0.147	1.510	1.426	0.0614	0.1937
25	504	2.075	2.919	2.709	0.2485	0.1551
31	327	0.844	1.136	1.032	0.3563	0.1328
31	347	1.135	1.165	1.151	0.4675	0.1099
31	421	1.644	1.864	1.676	0.8545	0.0300
33	354	0.954	1.451	1.250	0.4048	0.1228
33	892	3.043	1.959	2.073	0.1054	0.1846
39	323	1.301	0.860	0.956	0.2175	0.1615
45	523	10.904	2.276	2.591	0.0365	0.1988
46	506	1.556	2.533	2.500	0.0337	0.1994
46	907	0.000	2.560	2.519	0.0161	0.2030
53	349	1.307	1.312	1.311	0.1060	0.1845
54	396	1.230	1.455	1.363	0.4094	0.1219
93	411	1.133	1.076	1.112	0.6424	0.0738
94	417	3.046	2.748	2.774	0.0883	0.1881
96	433	1.811	1.403	1.594	0.4693	0.1095
97	321	0.862	1.558	1.478	0.1145	0.1827
97	832	1.749	1.619	1.680	0.4705	0.1092
98	353	1.243	1.498	1.422	0.2998	0.1445
98	474	1.186	1.605	1.407	0.4732	0.1087

After having computed γ^*, we computed estimates for the classes based on both the credibility estimates and based on the prior means. For the estimates based on prior means, the deviations from the classes that were actually used in 1984, were in most cases quite small; for the estimates based on the credibility estimates, the deviations were somewhat larger. The explanation is probably that one has been a bit reluctant to alter the class of a car model. For the actual rating, one might feel that the procedure is too sensitive to the random variable Y_k, and one should pay attention to this in the final subjective determination of the class; the *statistical* procedures do not make *political* considerations.

6.3. *The Hierarchical Approach*

Also for the hierarchical model we used the design (1 power price/weight)′. The parameters λ, $\boldsymbol{\Xi}$, and $\boldsymbol{\beta}$ were estimated as described in subsection 4.3.

For λ we found the estimate

$$\lambda^* = 0.1913,$$

from which we obtained

$$\varkappa^* = \phi^*/\lambda^* = 3404.$$

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

It is reasonable that the value of λ^* is lower than in the non-hierarchical model.

When estimating Ξ, we obtained

$$\hat{\Xi} \cdot 10^5 = \begin{pmatrix} -299650 & -3113 & 4724 \\ -3113 & 75 & 18 \\ 4724 & -18 & -33 \end{pmatrix}.$$

This matrix is obviously not positive definite. It has one negative eigenvalue, and by replacing this eigenvalue by $10^{-6}/N = 4 \cdot 10^{-8}$ as described in subsection 4.3, we arrived at

$$\Xi^* \cdot 10^5 = \begin{pmatrix} 0.0488 & -2.1678 & 1.351 \\ -2.1678 & 107.385 & 66.820 \\ 1.3511 & 66.820 & 41.709 \end{pmatrix}.$$

As the value of λ^* was only slightly lower than in the non-hierarchical model whereas the difference between $\hat{\Xi}$ and Ξ^* is considerable, we presume that for practical purposes we would choose the non-hierarchical model, but we shall go on presenting some results for the hierarchical model for illustrative purposes. We mention that computations made on the same data with the modified model described in subsection 4.5, gave much more reasonable results.

For $\boldsymbol{\beta}$ we found

$$\boldsymbol{\beta}^* = (-0.0587 \ 0.01228 \ 0.00687)'.$$

TABLE 6.4

Make	$\tilde{b}_n^{*\prime}$		
11	−0.05086	0.01078	0.00762
14	−0.05181	0.05778	−0.02127
15	−0.05134	0.03461	−0.00683
16	−0.05132	0.03384	−0.00638
17	−0.05091	0.01332	0.00631
18	−0.05102	0.01853	0.00276
24	−0.05090	0.01263	0.00662
25	−0.05058	−0.00309	0.01623
31	−0.05065	0.00014	0.01424
33	−0.05059	−0.00233	0.01643
34	−0.05100	0.01753	0.00355
37	−0.05078	0.00707	0.01019
39	−0.05093	0.01407	0.00577
45	−0.05106	0.02029	0.00170
46	−0.05043	−0.01066	0.02107
47	−0.05086	0.01094	0.00775
53	−0.05093	0.01400	0.00575
54	−0.05071	0.00327	0.01236
66	−0.05091	0.01307	0.00630
93	−0.05082	0.00877	0.00905
94	−0.05070	0.00284	0.01290
96	−0.05083	0.00927	0.00862
97	−0.05079	0.00733	0.01003
98	−0.05078	0.00682	0.01037

https://doi.org/10.2143/AST.17.1.2014983 Published online by Cambridge University Press

In Table 6.4 we have displayed the empirical credibility estimate $\tilde{b}_n^*$ for the 24 makes included in the study. The table illustrates the problem discussed in subsection 4.4; we see that for makes 25, 33, and 46 $x'_{nk}\tilde{b}_n^*$ will be decreasing in engine power, and for makes 14, 15, and 16 it will be decreasing in price/weight.

As examples of the values found for Π_n^* we display the value for one make with low exposure (Skoda) and one with high exposure (Opel). We found

$$\Pi_{39}^* \cdot 10^5 = \begin{pmatrix} 0.0211 & -7.9151 & 4.979 \\ -7.9151 & 39.063 & 24.483 \\ 4.9790 & 24.483 & 15.463 \end{pmatrix}$$

$$\Pi_{31}^* \cdot 10^5 = \begin{pmatrix} 0.0071 & -0.095 & 0.068 \\ -0.0954 & 4.459 & 3.114 \\ 0.0678 & 3.114 & 2.255 \end{pmatrix}.$$

Table 6.5 is the hierarchical analogue to Table 6.3. The quantities displayed in the last three columns are the estimates of the quantities

$$\psi_{nk} = (1 - \zeta_{nk})[\lambda + (1 - \zeta_{nk})x'_{nk}\Pi_n x_{nk}]$$

$$\psi_{nk}^{(\text{make})} = \lambda + x'_{nk}\Pi_n x_{nk}$$

$$\psi_{nk}^{(0)} = \lambda + x'_{nk}\Xi X_{nk}.$$

TABLE 6.5

Make	Model	Y_{nk}	$x'_{nk}\tilde{b}_n^*$	$\tilde{m}_{nk}^*$	ζ_{nk}^*	$x'_{nk}\beta^*$	ψ_{nk}^*	$\psi_{nk}^{*(\text{make})}$	$\psi_{nk}^{*(0)}$
14	541	3.336	4.042	3.891	0.2141	2.495	0.1911	0.2573	1.4167
15	313	0.502	2.066	1.944	0.0779	1.587	0.2366	0.2621	0.6572
16	321	2.283	1.010	1.632	0.4885	0.998	0.0994	0.1973	0.1975
17	328	1.465	1.629	1.522	0.6541	1.603	0.0798	0.3050	0.5152
18	451	5.628	2.152	2.693	0.1558	1.880	0.1882	0.2288	2.5677
18	741	0.147	1.613	1.529	0.0572	1.553	0.1871	0.1989	0.3889
25	504	2.075	2.517	2.413	0.2346	2.694	0.1666	0.2258	0.3189
31	327	0.844	1.193	1.075	0.3391	1.221	0.1296	0.1987	0.2013
31	347	1.135	1.146	1.141	0.4486	1.260	0.1072	0.1969	0.2688
31	421	1.644	1.441	1.612	0.8448	1.890	0.0307	0.2336	1.5493
33	354	0.954	1.332	1.186	0.3867	1.508	0.1294	0.2235	0.4459
33	892	3.043	1.576	1.721	0.0985	1.955	0.2666	0.3072	1.1428
39	323	1.301	1.086	1.130	0.2049	1.044	0.2863	0.4036	0.7804
45	523	10.904	2.497	2.782	0.0339	2.229	0.5712	0.6053	1.6357
46	506	1.556	1.408	1.413	0.0313	2.458	0.2641	0.2753	2.4120
46	907	0.000	1.466	1.444	0.0150	2.477	0.2641	0.2692	2.2515
53	349	1.307	1.430	1.417	0.0990	1.402	0.3058	0.3557	0.5623
54	396	1.230	1.354	1.306	0.3912	1.515	0.1290	0.2252	0.4975
93	411	1.133	1.108	1.124	0.6248	1.268	0.0827	0.2688	2.3997
94	417	3.046	2.625	2.660	0.0824	2.517	0.2420	0.2703	0.3121
96	433	1.811	1.409	1.590	0.4505	1.480	0.1137	0.2199	0.6209
97	321	0.862	1.513	1.443	0.1071	1.609	0.2003	0.2283	0.6666
97	832	1.749	1.539	1.634	0.4517	1.677	0.1268	0.2641	1.1276
98	353	1.243	1.442	1.385	0.2841	1.567	0.1460	0.2089	0.8462
98	474	1.186	1.499	1.357	0.4544	1.675	0.1142	0.2244	1.4243

The quantity ψ_{nk} has already been defined as the estimation error of the credibility estimator $\tilde{m}_{nk}$. We have that $\psi_{nk}^{(\text{make})}$ would be the estimation error of $x'_{nk}\tilde{b}_n$ as estimator of $m_{nk'}(\Theta_{nk'}, \mathrm{H}_n)$ for a car model k' with the same technical specifications as car model k, but for which we have no exposure. (To say that $\psi^{(\text{make})}$ is the estimation error of $x'_{nk}\tilde{b}_n$ considered as estimator of $m_{nk}(\Theta_{nk}, \mathrm{H}_n)$ would be wrong as $\tilde{b}_n$ contains claims data from car model k.) Similarly, $\psi_{nk}^{(0)}$ would be the estimation error of $x'_{nk}\beta$ considered as estimator of $m_{n'k'}(\Theta_{n'k'}, \mathrm{H}_{n'})$ for a car model k' of make n', for which we have no exposure.

As a consequence of the fact that the value of λ^* was lower in the present model than in the non-hierarchical model, we see that the values of ζ_{nk}^* are also lower. This is intuitively reasonable as $\tilde{b}_{nk}$ in the hierarchical model would contain more information than β in the non-hierarchical model.

ACKNOWLEDGEMENTS

The present research project was performed by the present author, but of course it was discussed on several occasions in the actuarial department in Storebrand, and the author is grateful to his colleagues Walther Neuhaus, Johan Lindtvedt, and Hans Haugland for useful comments and suggestions. He is also grateful to Storebrand for letting him present this paper, and to Ragnar Norberg for useful suggestions and comments; the present proof of Theorem 4.1 was suggested by him.

REFERENCES

BUNKE, H. and GLADITZ, J. (1974) Empirical linear Bayes decision rules for a sequence of linear models with different regressor matrices. *Mathematische Operationsforschung und Statistik* **5**, 234–244.

BÜHLMANN, H. and STRAUB. E. (1970) Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung schweizerischer Versicherungs mathematiker* **70**, 111–133.

CAMPBELL, M. (1986) An integrated system for estimating the risk premium of individual car models in motor insurance. *ASTIN Bulletin* **16**, 165–183.

DEGROOT, M. H. (1970) *Optimal Statistical Decisions*. McGraw-Hill, New York.

DE VYLDER, F. (1976) Geometrical credibility. *Scandinavian Actuarial Journal*, 121–149.

HACHEMEISTER, C. A. (1975) Credibility for regression models with application to trend. In *Credibility: Theory and Applications* (ed. P. M. KAHN), pp. 129–163. Academic Press, New York.

NORBERG, R. (1982) On optimal parameter estimation in credibility. *Insurance* **1**, 73–89.

NORBERG, R. (1985) Hierarchical credibility: Analysis of a random effect linear model with nested classification. *Statistical Research Report*. Institute of Mathematics, University of Oslo.

OPPLYSNINGSRÅDET FOR VEITRAFIKKEN (1984) Bilpriser 1984 – Vår. Norsk Veiforening, Oslo.

RAO, C. R. (1965) *Linear Statistical Inference and Its Applications*. Wiley, New York.

SUNDT, B. (1980) A multi-level hierarchical credibility regression model. *Scandinavian Actuarial Journal*, 25–32.

SUNDT, B. (1981) Recursive credibility estimation. *Scandinavian Actuarial Journal*, 3–21.

SUNDT, B. (1986) Credibility and old estimates. *Storebrand-Norden Actuarial Research Report* **2**.

BJØRN SUNDT
Storebrand P.O. Box 1380 Vika, N-0114 Oslo 1, Norway.