

**CREDIBILITY, HYPOTHESIS TESTING AND
REGRESSION SOFTWARE**

Greg Taylor

Taylor Fry Consulting Actuaries
Level 8, 30 Clarence Street
Sydney NSW 2000
Australia

Professorial Associate
Centre for Actuarial Studies
Faculty of Economics and
Commerce
University of Melbourne
Parkville VIC 3052
Australia

Adjunct Professor in Actuarial
Studies
Faculty of Commerce and
Economics
University of New South Wales
Kensington NSW 2033
Australia

Phone: 61 2 9249 2901
Fax: 61 2 9249 2999
greg.taylor@taylorfry.com.au

May 2007

Table of Contents

1. Introduction	1
2. Basic credibility framework	2
3. Reducible credibility models	4
4. Estimation of credibility parameter	6
5. Numerical evaluation of credibility parameter	13
6. Hierarchical credibility	14
7. Conclusion.....	19
References	20

Summary. It has been known since Zehnwirth (1977) that a scalar credibility coefficient is closely related to the F-statistic of an analysis of variance between and within risk clauses. The F-statistic may also be viewed as testing a certain regression structure, associated with credibility framework, against the null hypothesis of a simpler structure.

This viewpoint is extended to multi-dimensional credibility frameworks in which the credibility coefficient is a matrix (Sections 3 and 4), and to hierarchical regression credibility frameworks (Section 6). In each case the credibility coefficient is expressed in terms of the F-statistic that tests the significance of a defined regression structure against a simpler one.

Section 5 points out how the computation may be implemented by means of regression software.

Keywords. Credibility, hierarchical credibility, hypothesis testing, regression.

1. Introduction

A credibility estimate is a linearised Bayes estimate, consisting of a convex combination of a prior quantity and a data-based estimate.

The credibility “coefficient” (which may be a matrix) defining the convex combination also requires estimation from data. Historically, therefore, each new credibility application has tended to be accompanied by an additional analysis indicating how the credibility coefficient may be estimated (see eg Bühlmann and Straub, 1970; De Vylder, 1978, 1985).

These analyses have usually been *ad hoc*. For complex credibility models, such as the hierarchical regression models considered in Section 6, the determination of the form of credibility coefficient is correspondingly complex, and possibly exceedingly tedious.

The purpose of the present paper is to construct a defined procedure by which the estimation of a credibility coefficient may be automated. This is done for a general non-hierarchical credibility framework in Sections 3 and 4, and extended to a hierarchical (regression) framework in Section 6.

Section 5 indicates how regression software may be used to carry out the estimation.

2. Basic credibility framework

Consider the basic framework of regression credibility, as introduced by Hachemeister (1975). Let β be a p -vector randomly drawn from a distribution with

$$E[\beta] = b, \quad \text{Var}[\beta] = \Gamma \quad (2.1)$$

where $\text{Var}[\cdot]$ is used here to denote the variance-covariance matrix of its argument.

Let Y be an observable n -vector satisfying

$$Y = X\beta + \varepsilon \quad (2.2)$$

with X an $n \times p$ ($n \geq p$) design matrix, assumed to be of full rank, and ε a centred stochastic error vector, independent of β and with

$$\text{Var}[\varepsilon] = V \quad (2.3)$$

The generalised linear regression estimate of β for model (2.2) and (2.3) is

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (2.4)$$

Note that

$$\text{Var}[\hat{\beta}] = (X^T V^{-1} X)^{-1} \quad (2.5)$$

Let \tilde{Y} denote the following linear combination of Xb (prior estimate of Y) and $X\hat{\beta}$ (linear regression fitted value for Y):

$$\tilde{Y} = (1 - Z) Xb + Z X\hat{\beta} \quad (2.6)$$

where Z is an as yet unspecified $n \times n$ (non-stochastic) matrix.

Define L to be the mean square loss function

$$L = E \left[(Y - \tilde{Y})^T V^{-1} (Y - \tilde{Y}) \right] \quad (2.7)$$

where the expectation is taken over the joint distribution of β and Y . Choose Z so as to minimise L .

This optimisation of \tilde{Y} is in fact no less general than if \tilde{Y} is defined as an affine function of Y (Hachemeister, 1975; Taylor 1977).

Substitute (2.6) into $Y - \tilde{Y}$ and rearrange to obtain

$$Y - \tilde{Y} = (Y - X\beta) + (1 - Z)X(\beta - b) - ZX(\hat{\beta} - \beta) \quad (2.8)$$

Substitute this into (2.7) and make use of the fact that the three members of (2.8) are stochastically independent, to obtain

$$L = E \left[(Y - X\beta)^T V^{-1} (Y - X\beta) + (\beta - b)^T X^T (1 - Z)^T V^{-1} (1 - Z) X(\beta - b) + (\hat{\beta} - \beta)^T X^T Z^T V^{-1} ZX(\hat{\beta} - \beta) \right] \quad (2.9)$$

Note that, for any square (non-stochastic) matrix M and dimensionally compatible centred stochastic vector v ,

$$E[v^T M v] = \text{Tr}\{\text{Var}[v]M\} \quad (2.10)$$

Application of this result to each of the three members on the right side of (2.9), with substitution of (2.1), (2.3) and (2.5) for the covariance matrices yields

$$L = \text{Tr} \left[1 + X\Gamma X^T (1 - Z)^T V^{-1} (1 - Z) + X(X^T V^{-1} X)^{-1} X^T Z^T V^{-1} Z \right] \quad (2.11)$$

Note that, for any $n \times n$ matrices M, N ,

$$\partial \text{Tr}(MZN) / \partial Z_{ij} = (NM)_{ji} \quad (2.12)$$

where the subscript denotes the relevant element of the named matrix.

Write $\partial / \partial Z$ to denote the matrix of derivatives $\partial / \partial Z_{ij}$, so that (2.12) becomes

$$\partial \text{Tr}(MZN) / \partial Z = (NM)^T \quad (2.13)$$

Apply (2.13) to (2.11) and set the derivative to zero so as to minimize L , giving

$$-X\Gamma X^T (1 - Z)^T + X(X^T V^{-1} X)^{-1} X^T Z^T = 0 \quad (2.14)$$

Transposition of this, followed by post-multiplication by $V^{-1}X$ yields

$$Z(1 + X\Gamma X^T V^{-1})X = X\Gamma X^T V^{-1}X$$

which is solved by

$$Z = \left[1 + (X\Gamma X^T V^{-1})^{-1} \right]^{-1} \quad (2.15)$$

This is the classic result obtained by Hachemeister (1975). The matrix Z will be referred to henceforth as the **credibility matrix**.

3. Decomposition of credibility models

Consider the model described by (2.1) – (2.3), but now with X and Γ written as X_1 and Γ_1 . Express Γ_1 in the following block form:

$$\Gamma_1 = \begin{bmatrix} \Gamma_0 & * \\ * & \Gamma_+ \end{bmatrix} \quad (3.1)$$

where Γ_0 is of dimension $p_0 \times p_0$.

Partition X_1 in a dimensionally consistent manner:

$$X_1 = [X_0 \quad X_+] \quad (3.2)$$

where X_0 is of dimension $n \times p_0$.

No assumption is made about V other than the standard one of positive definiteness.

The notation here is that X_0, X_1 are the designs representing a regression null hypothesis and alternative hypothesis, and X_+ is the augmenting matrix connecting the two designs.

Example 3.1 (Bühlmann-Straub).

Let

$$X_0 = [u_n] \quad X_+ = \begin{bmatrix} 0 & \dots & 0 \\ u_{n_2} & & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & u_{n_r} \end{bmatrix} \quad (3.3)$$

where u_n denotes the n -dimensional column vector with all entries unity and

$$\sum_{i=1}^r n_i = n \quad (3.4)$$

where n_1 is the dimension of the top left zero vector in X_+ .

The interpretation of this model is as follows. If the r columns of X_1 are associated with r risk classes, then the model represents n_i observations on risk class i , $i=1, \dots, r$, which is assumed characterised by parameter β_1 for $i=1$, and by $\beta_1 + \beta_i$ otherwise. This is the data set-up of the model of Bühlmann and Straub (1970).

The model represented by the design matrix X_0 treats all risk classes as subject to the same parameter. \square

Example 3.2 (Hachemeister regression). Consider the case in which

$$X_0 = \begin{bmatrix} u_m & t_m \\ \vdots & \vdots \\ u_m & t_m \end{bmatrix} \quad X_+ = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ u_m & t_m & & & \\ \vdots & \vdots & \ddots & & \\ 0 & 0 & \dots & u_m & t_m \end{bmatrix} \quad (3.5)$$

where t_m is the time trend covariate vector $[1, 2, \dots, m]$, $n = rm$, and X_+ is of dimension $n \times 2r$.

Here the $2r$ parameters consist of r pairs and each pair may be regarded as defining a time trend for one of r risk classes. This corresponds to the model of Hachemeister (1975), in which the risk classes were states.

The model represented by the design matrix X_0 treats all states as subject to the same pair of risk parameters. \square

Example 3.3. A further example may be constructed by merging a model of the Hachemeister type with the sort of econometric model found in the workers compensation literature. For example, Butler (1994) found that statistical significant explanatory variables for real indemnity costs per employee included:

- wage replacement ratio;
- risky employment measure (proportion of workforce employed in manufacturing and construction);
- waiting period.

One might therefore define a model in which X_0 decomposes further:

$$X_0 = \begin{bmatrix} x^{(1)} & x^{(2)} & x^{(3)} & X_0^{(H)} \end{bmatrix}, \quad X_+ = X_+^{(H)} \quad (3.6)$$

where $x^{(1)}, x^{(2)}, x^{(3)}$ are the n -vectors of log (replacement ratio), log (risky employment measure) and waiting period respectively, and $X_0^{(H)}$ and $X_+^{(H)}$ are the Hachemeister versions of X_0 and X_+ defined by (3.5).

Moreover, one might choose

$$\Gamma_0 = \begin{bmatrix} 0 & 0 \\ 0 & \Gamma_0^{(H)} \end{bmatrix}, \quad \Gamma_+ = \Gamma_+^{(H)} \quad (3.7)$$

where the top left block of Γ_0 corresponds to $x^{(1)}, x^{(2)}, x^{(3)}$ and $\Gamma_0^{(H)}, \Gamma_+^{(H)}$ are as for the Hachemeister model.

The credibility model (2.6) then reduces to:

$$\tilde{Y} = \sum_{i=1}^3 \beta^{(i)} x^{(i)} + (1-Z) X^{(H)} b^{(H)} + ZX^{(H)} \hat{\beta}^{(H)} \quad (3.8)$$

where the terms with superscript (H) all relate to the Hachemeister portion of the model, ie the model with the $x^{(i)}$ components deleted.

In this example, the regression coefficients associated with the first three regressors, which are not state-specific (the regressors themselves may be, but their coefficients are not), are given full credibility in (3.8). However, the state-specific trends comprising the full extension model are credibility-weighted.

The distinction between the $x^{(i)}$ and the $X_0^{(H)}$ is the distinction between fixed and random effects in the regression model (see eg Ohlsson and Johansson, 2006).

4. Estimation of credibility parameter

Application of the credibility formula (2.6) and (2.15) requires a knowledge, or estimate, of some properties of Γ_1 and V . Full estimation of these matrices is a substantial task. Hachemeister (1975) shows how to estimate the former.

The present paper will be concerned with the restricted case in which the structure of each of the two matrices is known, ie each is known up to a multiplier,

$$V = \sigma^2 W, \quad \Gamma_1 = \tau^2 G_1 \quad (4.1)$$

with W, G_I known and σ^2, τ^2 unknown.

Substitution of (4.1) in (2.15) yields

$$Z = \left[1 + \left(v X_1 G_1 X_1^T W^{-1} \right)^{-1} \right]^{-1} \quad (4.2)$$

with

$$v = \tau^2 / \sigma^2 \quad (4.3)$$

and all other terms in (4.2) known. An estimate of only the ratio (4.3), rather than of its separate components, is required for computation of Z . The ratio v will be referred to as the **credibility parameter** in view of its central role. The following paragraphs address its estimation. They require the following elementary results.

Result 4.1. For any conformable non-stochastic matrix A and centred stochastic vector x ,

$$E \left[x^T A x \right] = Tr \{ A Var [x] \} \quad (4.4)$$

□

Result 4.2 For any matrices A, B of dimensions $m \times n$ and $n \times m$ respectively

$$Tr (AB) = Tr (BA) \quad (4.5)$$

It follows that the trace of a matrix product (of any number of factors) is invariant under cyclic permutation of the matrix factors. □

Consider the following two regression models:

$$\text{Model 0: } Y = X_0 \beta_0 + \varepsilon$$

$$\text{Model 1: } Y = X_1 \beta_1 + \varepsilon$$

subject to (2.3) and (3.7). Generally in the following a subscript 0 or 1 will be used to indicate which of the models is under discussion.

The fitted value of Y , denoted \hat{Y}_i , $i = 0, 1$, is

$$\hat{Y}_i = P_i Y \quad (4.6)$$

with

$$P_i = X_i \left(X_i^T W^{-1} X_i \right)^{-1} X_i^T W^{-1} \quad (4.7)$$

Define the **residual sum of squares** for Model 1 as

$$\begin{aligned} RSS &= (Y - \hat{Y}_1)^T W^{-1} (Y - \hat{Y}_1) \\ &= \sigma^2 \text{Tr} \left\{ W^{-1} (1 - P_1) W (1 - P_1)^T \right\} \quad [\text{using Results 4.1 and 4.2}] \end{aligned}$$

which, after some minor manipulation, yields

$$E[RSS] = \sigma^2 \text{Tr}(1 - P_1) \quad (4.8)$$

where the matrices on the right are of dimension $n \times n$.

By Result 4.2, $\text{Tr} P_1$ reduces to the trace of the $p_1 \times p_1$ identity matrix, and so (4.8) gives

$$E[RSS] = (n - p_1) \sigma^2 \quad (4.9)$$

Now define the **regression sum of squares**

$$\begin{aligned} SS_{reg} &= (\hat{Y}_1 - \hat{Y}_0) W^{-1} (\hat{Y}_1 - \hat{Y}_0) \\ &= Y^T (P_1 - P_0)^T W^{-1} (P_1 - P_0) Y \end{aligned} \quad (4.10)$$

Consider the matrix $(P_1 - P_0)^T W^{-1} (P_1 - P_0)$. It is elementary to check that

$$P_i^T W^{-1} P_i = W^{-1} P_i, \quad i = 0, 1 \quad (4.11)$$

Now consider the cross terms in $(P_1 - P_0)^T W^{-1} (P_1 - P_0)$. By (4.7),

$$P_1^T W^{-1} P_0 = W^{-1} X_1 \left[(X_1^T W^{-1} X_1)^{-1} X_1^T W^{-1} X_0 \right] (X_0^T W^{-1} X_0)^{-1} X_0^T W^{-1} \quad (4.12)$$

Expand X_1 according to (3.7) within the square bracket:

$$\begin{aligned} (X_1^T W^{-1} X_1)^{-1} X_1^T W^{-1} X_0 &= \begin{bmatrix} X_0^T W^{-1} X_0 & X_0^T W^{-1} X_+ \\ X_+^T W^{-1} X_0 & X_+^T W^{-1} X_+ \end{bmatrix}^{-1} \begin{bmatrix} X_0^T W^{-1} X_0 \\ X_+^T W^{-1} X_0 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned} \quad (4.13)$$

Substitution of (4.13) in (4.12) yields

$$P_1^T W^{-1} P_0 = W^{-1} [X_0 \ X_+] \begin{bmatrix} 1 \\ 0 \end{bmatrix} (X_0^T W^{-1} X_0)^{-1} X_0^T W^{-1} = W^{-1} P_0 \quad (4.14)$$

It may also be noted that

$$\begin{aligned} P_0^T W^{-1} P_1 &= (P_1^T W^{-1} P_0)^T = (W^{-1} P_0)^T \quad [\text{by (4.14)}] \\ &= W^{-1} P_0 \quad [\text{by (4.7)}] \end{aligned} \quad (4.15)$$

Substitution of (4.11), (4.14) and (4.15) in (4.10) yields

$$SS_{reg} = Y^T W^{-1} (P_1 - P_0) Y \quad (4.16)$$

and then, by Result 4.1,

$$E[SS_{reg}] = Tr\{W^{-1} (P_1 - P_0) Var[Y]\} \quad (4.17)$$

To evaluate this quantity, write

$$Y = X_1 b + X_1 (\beta - b) + (Y - X_1 \beta)$$

and note that covariances between all three terms on the right are zero. Therefore, by (4.1), (2.1) and (2.3),

$$\begin{aligned} Var[Y] &= X_1 \Gamma_1 X_1^T + V \\ &= \tau^2 X_1 G_1 X_1^T + \sigma^2 W \end{aligned} \quad (4.18)$$

where (4.1) has been used again.

By substitution of (4.18) into (4.17),

$$E[SS_{reg}] = Tr\{W^{-1} (P_1 - P_0) (\tau^2 X_1 G_1 X_1^T + \sigma^2 W)\} \quad (4.19)$$

Now

$$Tr(W^{-1} P_i W) = Tr 1_{p_i \times p_i} \quad (4.20)$$

by (4.7) and Result 4.2, where X_i is of dimension $n \times p_i$. Thus

$$Tr\{W^{-1} (P_1 - P_0) \sigma^2 W\} = (p_1 - p_0) \sigma^2 \quad (4.21)$$

Note that, by Result 4.2,

$$Tr[W^{-1} (P_1 - P_0) X_1 G_1 X_1^T] = Tr[X_1^T W^{-1} (P_1 - P_0) X_1 G_1] \quad (4.22)$$

To evaluate this, consider

$$X_1^T W^{-1} (P_1 - P_0) X_1 = X_1^T W^{-1} X_1 - X_1^T W^{-1} P_0 X_1 \quad (4.23)$$

by (4.7).

Now

$$\begin{aligned}
 X_1^T W^{-1} P_0 X_1 &= (X_1^T W^{-1} X_0) (X_0^T W^{-1} X_0)^{-1} (X_0^T W^{-1} X_1) \\
 &= \begin{bmatrix} X_0^T W^{-1} X_0 \\ X_+^T W^{-1} X_0 \end{bmatrix} (X_0^T W^{-1} X_0)^{-1} \begin{bmatrix} X_0^T W^{-1} X_0 & X_0^T W^{-1} X_+ \end{bmatrix} \quad (4.24) \\
 &= \begin{bmatrix} X_0^T W^{-1} X_0 & X_0^T W^{-1} X_+ \\ X_+^T W^{-1} X_0 & (X_+^T W^{-1} X_0) (X_0^T W^{-1} X_0)^{-1} (X_0^T W^{-1} X_+) \end{bmatrix}
 \end{aligned}$$

Substitute (4.24) into (4.23) and expand $X_1^T W^{-1} X_1$ in a similar block form to obtain

$$X_1^T W^{-1} (P_1 - P_0) X_1 = \begin{bmatrix} 0 & 0 \\ 0 & X_+^T W^{-1} (1 - P_0) X_+ \end{bmatrix} \quad (4.25)$$

Substitution of (4.25) in (4.22) gives

$$Tr[W^{-1} (P_1 - P_0) X_1 G_1 X_1^T] = Tr[X_+^T W^{-1} (1 - P_0) X_+ G_+] \quad (4.26)$$

Combining (4.19), (4.21) and (4.26)

$$E[SS_{reg}] = \sigma^2 (p_1 - p_0) + \tau^2 t(D) \quad (4.27)$$

where D is the regression design that recognises X_0, X_+ , in addition to W and G_+ , and

$$t(D) = tr[X_+^T W^{-1} (1 - P_0) X_+ G_+] \quad (4.28)$$

Combine (4.27) with (4.9) to obtain

$$\frac{E(SS_{reg}) / (p_1 - p_0)}{E[RSS] / (n - p_1)} = 1 + \nu (p_1 - p_0)^{-1} t(D) \quad (4.29)$$

where ν was defined in (4.3).

Thus, ν is estimated by

$$\hat{\nu} = \max[0, (F - 1)(p_1 - p_0) / t(D)] \quad (4.30)$$

where F is the conventional regression F -statistic for testing Model 1 against Model 0, ie

$$F = [SS_{reg} / (p_1 - p_0)] / [RSS / (n - p_1)] \quad (4.31)$$

The credibility matrix Z in (4.2) is thus estimated by replacing ν with $\hat{\nu}$.

The dependence of Z on the F -statistic was demonstrated by Zehnwirth (1977) in the simple case of 1-dimensional credibility. It may also be remarked that the estimator of Z derived here is different from Hachemeister's (1975) estimate because he did not make the reducibility assumption (3.3) and assumed no prior knowledge of Γ_1 . He therefore estimated Γ_1 in its entirety rather than just the scaling parameter τ^2 in (4.1).

The regression whose statistics appear on the left side of (4.29) may have a number of equivalent designs. Example 3.1, for example, might have been formulated with X_+ taking any of the block diagonal forms

$$X_+ = \text{diag}(u_{n_1}, \dots, u_{n_{s-1}}, 0, u_{n_{s+1}}, \dots, u_{n_r}) \quad (4.32)$$

However, changing from one design D to another would not change the left side of (4.29), and so $t(D)$ is invariant over $D \in \mathbf{D}$, where \mathbf{D} denotes the set of all regression designs equivalent to (and including) the one of interest.

Example 3.1 (continued). Suppose that

$$\text{Var}[\beta] = \tau^2 1, \quad W^{-1} = \text{diag}(w_1, \dots, w_n) \quad \text{with} \quad \sum_{i=1}^n w_i = w.$$

Note that, since the parameter vector for module H_1 is represented in the form $(\beta_1, \beta_1 + \beta_2, \dots, \beta_1 + \beta_r)^T$ instead of just $(\beta_1, \dots, \beta_r)$, it is necessary for the former vector to have covariance matrix $\tau^2 1$. That is

$$\text{Var}[M\beta] = \tau^2 1$$

with

$$M = \begin{bmatrix} 1 & 0 \\ u_{r-1} & 1 \end{bmatrix}$$

It may be checked that, in this representation,

$$\text{Var}[\beta] = \tau^2 M^{-1} (M^{-1})^T = \tau^2 \begin{bmatrix} 1 & -u_{r-1}^T \\ -u_{r-1} & 1 + u_{r-1} u_{r-1}^T \end{bmatrix}$$

and so

$$G_+ = 1 + u_{r-1} u_{r-1}^T \quad (4.33)$$

Now

$$X_0^T W^{-1} X_0 = u_n^T W^{-1} u_n = w \quad (4.34)$$

$$P_0 = w^{-1} u_n u_n^T W^{-1} \quad (4.35)$$

$$X_+^T W^{-1} X_+ = \begin{bmatrix} u_{n_2}^T W_{(2)}^{-1} u_{n_2} & & \\ & \ddots & \\ & & u_{n_r}^T W_{(r)}^{-1} u_{n_r} \end{bmatrix} \quad (4.36)$$

where W^{-1} is written in block diagonal form $W^{-1} = \text{diag}(W_{(1)}^{-1}, \dots, W_{(r)}^{-1})$ with blocks corresponding to those in X_1 .

Thus

$$X_+^T W^{-1} X_+ = \text{diag}(w_{(2)}, \dots, w_{(r)})$$

where $w_{(j)} = \text{Tr} W_{(j)}^{-1}$, and so

$$\text{Tr}(X_+^T W^{-1} X_+ G_+) = 2(w_{(2)} + \dots + w_{(r)}) = 2(w - w_{(1)}) \quad (4.37)$$

Also

$$X_+^T W^{-1} P_0 X_+ G_+ = w^{-1} (X_+^T W^{-1} u_n) (X_+^T W^{-1} u_n)^T G_+ \quad (4.38)$$

and

$$X_+^T W^{-1} u_n = \begin{bmatrix} w_{(2)} \\ \vdots \\ w_{(r)} \end{bmatrix} \quad (4.39)$$

Therefore, by substitution of (4.39) in (4.38) and use of (4.33) and Result 4.2,

$$\begin{aligned} \text{Tr}(X_+^T W^{-1} P_0 X_+ G_+) &= w^{-1} \text{Tr} \left[(X_+^T W^{-1} u_n)^T (1 + u_{r-1} u_{r-1}^T) (X_+^T W^{-1} u_n) \right] \\ &= w^{-1} \left\{ \sum_{j=2}^r w_{(j)}^2 + \left[\sum_{j=2}^r w_{(j)} \right]^2 \right\} \\ &= w^{-1} \left[\sum_{j=1}^r w_{(j)}^2 - w_{(1)}^2 + (w - w_{(1)})^2 \right] \end{aligned} \quad (4.40)$$

It then follows from (4.27) and (4.28) that

$$\begin{aligned} E[SS_{reg}] &= (r-1)\sigma^2 + \tau^2 w^{-1} \left[2w(w - w_{(1)}) - (w - w_{(1)})^2 + w_{(1)}^2 - \sum_{j=1}^r w_{(j)}^2 \right] \\ &= (r-1)\sigma^2 + \tau^2 \left[w - w^{-1} \sum_{j=1}^r w_{(j)}^2 \right] \end{aligned} \quad (4.41)$$

Comparison with (4.27) indicates that

$$t(D) = w - w^{-1} \sum_{j=1}^r w_{(j)}^2 \quad (4.42)$$

Note that $t(D)$ is indeed independent of D , as predicted earlier.

The estimator (4.30) now takes the explicit form

$$\hat{\nu} = \max \left\{ 0, (F-1)(r-1) / \left[w - w^{-1} \sum_{j=1}^r w_{(j)}^2 \right] \right\} \quad (4.43)$$

with

$$F = [SS_{reg} / (r-1)] / [RSS / (n-r)] \quad (4.44)$$

Here

$$SS_{reg} = \sum_{j=1}^r w_{(j)} (\bar{Y}_{(j)} - \bar{Y})^2 \quad (4.45)$$

where $\bar{Y}_{(j)}$ is the weighted mean of Y over the j -th risk class with weight vector $w_{(j)}$ and \bar{Y} is the grand weighted mean.

The estimator (4.43) is the same as that obtained by Bühlmann and Straub (1970). It is emphasised, however, that, despite its algebraic development for illustrative purposes here, it could have been derived numerically (with no algebra) as described in Section 5 below.

5. Numerical evaluation of credibility parameter

Suppose one is faced with the credibility regression design represented by (3.2), (3.3) and (4.1). One wishes to apply the credibility formulas (3.5) and (4.2), and needs an estimate of ν in order to do so.

One may proceed by means of the following steps.

- 5.1 Choose a regression design D and evaluate $t(D)$ according to (4.28). This may be done algebraically or numerically.
- 5.2 Evaluate the F -statistic for testing the regression Model 1 against the Model 0 null hypothesis. This may be done by applying regression software to the data, or by direct calculation. The latter would amount to a re-creation of the regression software.
- 5.3 Assemble the result into the estimator \hat{v} given by (4.30), and hence an estimator of Z .

Step 5.2 provides a quick and systematic way of evaluating the credibility matrix in cases of complicated design. An example is given in the next section.

6. Hierarchical credibility

6.1 Definition of hierarchical model

A general hierarchical credibility model (Taylor 1979, Sundt 1979, 1980) can be extensive, and the procedure set out in Section 5 may be helpful in the evaluation of the various credibility matrices.

The standard hierarchical credibility framework, as defined by Taylor, is one in which risk classes consist of sub-classes, and sub-sub-classes, and so on.

If a risk class is labelled j_1 , it will be composed of sub-classes j_11, j_12, \dots , and generally j_1j_2 . This will consist of sub-sub-classes $j_1j_2j_3$, and ultimately $j_1j_2\dots j_q$.

The nodes $(j_1\dots j_k)$, $k = 1, 2, \dots, q$ form a tree. A regression structure may be placed at each node.

This sort of structure is studied in full generality by Sundt (1979, 1980) who sets up a framework involving an observable vector $Y = [Y_1^T, \dots, Y_q^T]^T$, in which the q components represent the q levels of the hierarchy. Latent parameters $\theta_1, \dots, \theta_q$ are associated with the different levels of the hierarchy.

A slightly simplified version of Sundt's assumptions, sufficient for present purposes, is as follows:

- (i) Y_j and θ_k are conditionally independent given $\theta_1, \dots, \theta_j$ if $k > j$.
- (ii) $(Y_1^T, \dots, Y_j^T)^T$ and $(Y_{j+1}^T, \dots, Y_q^T)^T$ are conditionally independent given $\theta_1, \dots, \theta_j$, $j = 1, \dots, q - 1$.
- (iii) $E[Y_j | \theta_1, \dots, \theta_j] = X_j \beta_j(\theta_1, \dots, \theta_j)$, $j = 1, 2, \dots, q$ (6.1)

for non-stochastic matrix X_j and vector function $\beta_j(\cdot)$,

$$(iv) \quad E[\beta_j(\theta_1, \dots, \theta_j) | \theta_1, \dots, \theta_{j-1}] = \beta_{j-1}(\theta_1, \dots, \theta_{j-1}), j = 2, \dots, q \quad (6.2)$$

and

$$E[\beta_1(\theta_1)] = \beta_0 \quad (6.3)$$

Condition (i) is actually stronger than Sundt, who assumed only that Y_j and θ_k were unconditionally independent.

Let

$$\begin{aligned} G_j &= EVar[\beta_j(\theta_1, \dots, \theta_j) | \theta_1, \dots, \theta_{j-1}], j = 2, \dots, q \\ &= Var[\beta_1(\theta_1)] \text{ for } j = 1 \end{aligned} \quad (6.4)$$

where the expectation operator is tacitly assumed taken over all conditioning variables of its operand.

Similarly, let

$$W_j = EVar[Y_j | \theta_1, \dots, \theta_j], j = 1, \dots, q \quad (6.5)$$

Define

$$Y_{(j)} = [Y_j^T, \dots, Y_q^T]^T, j = 1, \dots, q \quad (6.6)$$

$$X_{(j)} = [X_j^T, \dots, X_q^T]^T, j = 1, \dots, q \quad (6.7)$$

$$W_{(j)} = EVar[Y_{(j)} | \theta_1, \dots, \theta_j], j = 1, \dots, q \quad (6.8)$$

$$\begin{aligned} N_{(j)} &= EVar[Y_{(j)} | \theta_1, \dots, \theta_{j-1}], j = 2, \dots, q \\ &= Var[Y_{(1)}] \text{ for } j = 1 \end{aligned} \quad (6.9)$$

Then $W_{(j)}$ and $N_{(j)}$ may be calculated recursively as follows (Sundt, 1980):

$$\begin{aligned} W_{(j)} &= W_j \text{ for } j = q \\ &= \begin{bmatrix} W_j & 0 \\ 0 & N_{(j+1)} \end{bmatrix}, j = q-1, q-2, \dots, 1 \end{aligned} \quad (6.10)$$

$$N_{(j)} = W_{(j)} + X_{(j)} G_j X_{(j)}^T, j = q, q-1, \dots, 1 \quad (6.11)$$

Finally the credibility estimator of $E[\beta_j(\theta_1, \dots, \theta_j) | \theta_1, \dots, \theta_{j-1}]$ based on data Y (ie least squares estimator, linear in Y), denoted $\tilde{\beta}_j$, is calculated recursively as follows (Sundt, 1980).

Define

$$M_j = G_j X_{(j)}^T W_{(j)}^{-1} X_{(j)} \quad (6.12)$$

$$Z_j = M_j (1 + M_j)^{-1} \quad (6.13)$$

$$\hat{\beta}_j = \left[X_{(j)}^T N_{(j)}^{-1} X_{(j)} \right]^{-1} X_{(j)}^T N_{(j)}^{-1} Y_{(j)} \quad (6.14)$$

which may be recognized as a generalised least squares regression estimator of $E[\beta_j(\theta_1, \dots, \theta_j) | \theta_1, \dots, \theta_{j-1}]$.

Then

$$\tilde{\beta}_j = (1 - Z_j) \tilde{\beta}_{j-1} + Z_j \hat{\beta}_j, \quad j = 1, 2, \dots, q \quad (6.15)$$

with $\tilde{\beta}_0 = \beta_0$.

It is of interest to consider the special case in which the covariance matrices G_j and W_j are independent of $\theta_1, \dots, \theta_j$ and are known up to scaling constants, and the relation between the different W_j also known:

$$G_j = \tau_j^2 \bar{G}_j, \quad W_j = \sigma^2 \bar{W}_j \quad (6.16)$$

where \bar{G}_j and \bar{W}_j are known and τ_j^2 and σ^2 are to be estimated from data.

Define

$$\bar{W}_{(j)} = W_{(j)} / \sigma^2, \quad \bar{N}_{(j)} = N_{(j)} / \sigma^2 \quad (6.17)$$

It may be checked from (6.10) and (6.11) that

$$\begin{aligned} \bar{W}_{(j)} &= \bar{W}_j, \quad j = q \\ &= \begin{bmatrix} \bar{W}_j & 0 \\ 0 & \bar{N}_{(j+1)} \end{bmatrix}, \quad j = q-1, q-2, \dots, 1 \end{aligned} \quad (6.18)$$

$$\bar{N}_{(j)} = \bar{W}_{(j)} + \nu_j X_{(j)} \bar{G}_j X_{(j)}^T, \quad j = q, q-1, \dots, 1 \quad (6.19)$$

where

$$\nu_j = \tau_j^2 / \sigma^2 \quad (6.20)$$

By (6.12), (6.17) and (6.20),

$$M_j = \nu_j \bar{G}_j X_{(j)}^T \bar{W}_{(j)}^{-1} X_{(j)} \quad (6.21)$$

and (6.13) then yields Z_j in a form parallel to (4.2) with the ν_j now referred to as the credibility parameters in parallel with ν in (4.3).

6.2 Estimation of the credibility parameters

The credibility parameters of the hierarchical regression model may be estimated by the approach established in Sections 4 and 5. the details of the application will depend on the parametric details of the model. An example follows.

Begin with the regression

$$Y_j = X_j \beta_j (\theta_1, \dots, \theta_j) + \varepsilon_j \quad (6.22)$$

where, by (6.1), (6.8) and (6.16),

$$E[\varepsilon_j] = 0 \quad (6.23)$$

Example 6.1. Consider the model defined in Section 6.1, and note that, for $j =$ fixed k , the required credibility parameter ν_k arises from τ_k^2 and σ^2 . If the former is set to zero, then (6.4) and (6.16) imply that

$$\text{Var}[\beta_k(\theta_1, \dots, \theta_k) | \theta_1, \dots, \theta_k] = 0 \quad (6.24)$$

and so, by (6.2),

$$\beta_k(\theta_1, \dots, \theta_k) = \beta_{k-1}(\theta_1, \dots, \theta_{k-1}) \quad (6.25)$$

By (6.1), (6.2), (6.6) and (6.7),

$$E[Y_{(j)} | \theta_1, \dots, \theta_j] = X_{(j)} \beta_j(\theta_1, \dots, \theta_j), \quad j = 1, 2, \dots, q \quad (6.26)$$

By (6.1) with $j = k - 1$ and (6.26) with $j = k$,

$$\begin{aligned} E \begin{bmatrix} Y_{k-1} | \theta_1, \dots, \theta_{k-1} \\ Y_{(k)} | \theta_1, \dots, \theta_k \end{bmatrix} &= \begin{bmatrix} X_{k-1} \beta_{k-1}(\theta_1, \dots, \theta_{k-1}) \\ X_{(k)} \beta_k(\theta_1, \dots, \theta_k) \end{bmatrix} \\ &= \begin{bmatrix} X_{k-1} & -X_{k-1} \\ X_{(k)} & 0 \end{bmatrix} \begin{bmatrix} \beta_k(\theta_1, \dots, \theta_k) \\ \eta_{k-1} \end{bmatrix} \end{aligned} \quad (6.27)$$

where

$$\eta_{k-1} = \beta_k(\theta_1, \dots, \theta_k) - \beta_{k-1}(\theta_1, \dots, \theta_{k-1}) \quad (6.28)$$

Let Model 1 denote (6.27), and Model 0 denote the same subject to (6.25), ie with $\eta_{k-1} = 0$. The situation is now parallel to Example 3.2, with

$$X_0 = \begin{bmatrix} X_{k-1} \\ X_{(k)} \end{bmatrix} = X_{(k-1)}, \quad X_+ = \begin{bmatrix} -X_{k-1} \\ 0 \end{bmatrix} \quad (6.29)$$

Then $\nu_k = \tau_k^2 / \sigma^2$ is estimated by (4.30). The detail is as follows. First note that, by condition (i) of the hierarchical regression model, Y_{k-1} and θ_k are conditionally independent given $\theta_1, \dots, \theta_{k-1}$, and it then follows that $Y_{k-1} | \theta_1, \dots, \theta_{k-1}$ and $Y_{k-1} | \theta_1, \dots, \theta_k$ are equal in distribution. Then (6.27) becomes

$$E \begin{bmatrix} Y_{k-1} \\ Y_{(k)} \end{bmatrix} \Big| \theta_1, \dots, \theta_k = \begin{bmatrix} X_{k-1} & -X_{k-1} \\ X_{(k)} & 0 \end{bmatrix} \begin{bmatrix} \beta(\theta_1, \dots, \theta_k) \\ \eta_{k-1} \end{bmatrix} \quad (6.30)$$

In the notation associated with (4.32),

$$G_+ = \text{Var}[\eta_{k-1} | \theta_1, \dots, \theta_{k-1}] = \text{Var}[\beta_k(\theta_1, \dots, \theta_k) | \theta_1, \dots, \theta_{k-1}] \quad (6.31)$$

$$W = \text{Var} \begin{bmatrix} Y_{k-1} \\ Y_{(k)} \end{bmatrix} \Big| \theta_1, \dots, \theta_k \quad (6.32)$$

Now consider the special case in which the covariance matrices in (6.31) and (6.32) do **not** depend on the conditioning parameters shown. Then they may be written as

$$G_+ = E \text{Var}[\beta_k(\theta_1, \dots, \theta_k) | \theta_1, \dots, \theta_{k-1}] = G_{k-1} \quad (6.33)$$

$$W = E \text{Var} \begin{bmatrix} Y_{k-1} \\ Y_{(k)} \end{bmatrix} \Big| \theta_1, \dots, \theta_k = \begin{bmatrix} W_{k-1} & 0 \\ 0 & W_{(k)} \end{bmatrix} \quad (6.34)$$

where the entries in the matrix are justified as follows. The bottom right entry follows directly from (6.8). The top left follows from (6.5) and the fact that $Y_{k-1} | \theta_1, \dots, \theta_{k-1}$ and $Y_{k-1} | \theta_1, \dots, \theta_k$ are equal in distribution. The off-diagonal zeros are justified by observing that $Y_{k-1} | \theta_1, \dots, \theta_{k-1}$ and $Y_{(k)} | \theta_1, \dots, \theta_{k-1}$ are independent, by condition (ii) of the model. Also $Y_{k-1} | \theta_1, \dots, \theta_{k-1}$ and $\theta_k | \theta_1, \dots, \theta_{k-1}$ are independent, from condition (i). It follows that $Y_{k-1} | \theta_1, \dots, \theta_k$ and $Y_{(k)} | \theta_1, \dots, \theta_k$ are independent. This is just an application of the general result that independence of $U|A$ and $V|A$, and of $U|A$ and $B|A$, implies independence of $U|AB$ and $V|AB$, whose proof is left to the reader.

The credibility parameter ν_k is now estimated by $\hat{\nu}_k$ given by (4.30).

7. Conclusion

Traditionally, credibility coefficients have been estimated by manipulation of squared error terms of some sort. The manipulations have been *ad hoc*, and each new credibility model has required a new exercise in estimation.

This paper has developed an estimate of a credibility coefficient on the basis of an analysis of variance for the testing of one regression model against another. The credibility coefficient is expressed in terms of the F test statistic.

Sections 4 and 6 give examples of the procedure's application. Some of these, especially Example 6.2, are complex, and the *ad hoc* algebra involved in manipulating squared error terms in order to arrive at estimators for the credibility coefficients would be laborious and possibly error-prone.

The suggested procedure reduces this to an algorithm once null and alternative hypotheses have been formulated.

Further, as pointed out in Section 5, the reduction of the estimation to hypothesis testing of regression models means that the required F -statistic may be computed by standard regression software.

References

Bühlmann H and Straub E (1970). Glaubwürdigkeit für Schadensätze. **Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker**, 70.

Butler RJ (1994). Economic determinants of workers' compensation trends. **Journal of Risk and Insurance**, 61, 383 – 401.

De Vylder F (1978). Parameter estimation in credibility theory. **Astin Bulletin**, 10, 99 – 112.

De Vylder F (1985). Non-linear regression in credibility theory. **Insurance: mathematics and economics**, 4, 163 – 172.

Hachemeister CA (1975). Credibility for regression with application to trend. Appears in Kahn (1975).

Kahn PM (1975) (ed). **Credibility: theory and applications**. Academic Press, New York, NY.

Ohlsson E and Johansson B (2006). Exact credibility and Tweedie models. **Astin Bulletin**, 36, 121 – 133.

Sundt B (1979). A hierarchical credibility regression model. **Scandinavian Actuarial Journal**, 107 – 114.

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

Taylor GC (1977). Abstract credibility. **Scandinavian Actuarial Journal**, 149 – 168.

Taylor GC (1979). Credibility analysis of a general hierarchical model. **Scandinavian Actuarial Journal**, 1 – 12.

Zehnwirth B (1977). The credible distribution function is an admissible Bayes rule. **Scandinavian Actuarial Journal**, 121 – 127.