PRACTICAL CREDIBILITY THEORY WITH EMPHASIS ON OPTIMAL PARAMETER ESTIMATION

F. De Vylder*

We develop Hachemeister's regression model in credibility theory (without proofs) and indicate how the involved structural parameters can be estimated from the observable variables (with proofs for the simple results and those not yet published).

Large families of unbiased estimators are available. From the practical viewpoint this is rather a handicap because it creates the problem to decide what estimators actually to use. In order to fix optimal estimators, we adopt the small-sample criterion of minimum-variance. But in the research for general solutions three kinds of difficulties arise.

(i) The calculations become too lengthy.
(ii) The optimal estimators depend on some of the parameters to be estimated. (Then we call them pseudo-estimators).
(iii) The optimal estimators depend on new structural parameters defined in terms of fourth-order moments.

Only a compromise allows to cope with this reality. Situation (iii) creates new estimation problems. They can only be avoided at the cost of the introduction of special assumptions or approximations. Then problem (i) is more or less automatically solved. By an obvious method of successive approximations pseudo-estimators can serve as true estimators. Thus (ii) is no real problem.

1. General notations and definitions

1.1. Matrices

The same main symbol is used to denote a matrix c and its elements c_i^j. Here c_i^j is the element at the intersection of row i and column j. The row i is denoted by c_i and the column j by c^j. When the number of rows is m and the number of columns n, we say that the matrix has dimensions $\binom{n}{m}$.

These conventions are most convenient for automatical calculation with matrices. For instance, if

(1)
$$a = b \quad c \quad d$$
$$\binom{n}{m} \quad \binom{p}{m} \binom{q}{p} \binom{n}{q}$$

then

$$a_i = b_i c d, \ a^j = b c d^j, \ a_i^j = b_i c d^j, \ a = \Sigma b^r c_r d, \ a_i = \sum_{r,s} b_i^r c_r^s d_s, \ \ldots$$

* The author is thankful to the referee who improved this paper.

Hence the following general rules in case of a matrix relation such as (1), with any number of factors in the last member:

(i)   The same lower index can be placed at the matrix in the first member and at the first matrix in the last member.

(ii) The same upper index can be placed at the matrix in the first member and the the last matrix in the last member.

(iii) In the last member, an upper index can be placed at any matrix and the same lower index at the following matrix, provided that index be summed out.

Of course, according to the same rules applied in the opposite direction, indices can also be dropped.

We recall that the trace of a square matrix c is the sum of its diagonal elements: $\operatorname{tr} c = \Sigma\, c_i^i$. When a product, say $abcd$, of any number of any matrices is a square matrix, then any cyclical permutation, say $cdab$, is also a square matrix (of possibly different dimensions) and $\operatorname{tr}(abcd) = \operatorname{tr}(cdab)$.

1.2. Random matrices

The expectation EA of a matrix A with random elements A_i^j is the matrix with elements EA_i^j. The operators tr and E commute.

To the $\binom{1}{n}$ random column X is associated the $\binom{1}{n}$ covariance matrix $\operatorname{Cov} X$ with elements

$$(\operatorname{Cov} X)_i^j = \operatorname{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j).$$

Using the accent for transposition, the last member can be displayed as $E(X_i X'^j) - E(X_i)E(X'^j)$. Therefore

$$\operatorname{Cov} X = E(XX') - E(X)E(X'), \quad \binom{n}{n}.$$

When X is $\binom{1}{1}$, then $\operatorname{Cov} X$ is in fact $\operatorname{Var} X$.

If t is $\binom{n}{m}$ and X $\binom{1}{n}$, then tX is $\binom{1}{m}$ and $\operatorname{Cov}(tX) = t(\operatorname{Cov} X)t'$

Similarly, for the conditional covariance matrix:

$$\operatorname{Cov}(X/\Theta) = E(XX'/\Theta) - E(X/\Theta)E(X'/\Theta), \quad \binom{n}{n}.$$

The following relation is easy:

$$\operatorname{Cov} X = E\operatorname{Cov}(X/\Theta) + \operatorname{Cov} E(X/\Theta), \quad \binom{n}{n}.$$

We define the *scalar variance* of any random matrix A by

$$\operatorname{ScaVar} A = \sum_{i,j} \operatorname{Var} A_j^i = \sum_{i,j} E(A_i^j A'^i_j) - \sum_{i,j} E(A)E(A'^i_j)$$

$$= \sum_i E(A_i A'^i) - \sum_i E(A_i)E(A'^i) = \operatorname{tr} E(AA') - \operatorname{tr}(EA.EA').$$

Thus, when A is the $\binom{1}{n}$ column X, then ScaVar $X = $ tr Cov X.

To say that the $\binom{1}{n}$ random vector X is $N(m, v)$ means that X is normally distributed with $E\,X = m$, Cov $X = v$. This implies that m is $\binom{1}{n}$ and that v is $\binom{n}{n}$. Similarly, "the conditional vector X, given Θ, is $N(m(\Theta), v(\Theta))$" means that this vector is normally distributed with $E(X/\Theta) = m(\Theta)$, Cov$(X/\Theta) = v(\Theta)$.

1.3. Miscellaneous

In case of multiple indices, the matrix rules are applied to the last indices only. For instance, if the scalars x^q_{jp} are defined for variable j, p, q, then, for each fixed j, x_j is the matrix with elements x^q_{jp}, x_{jp} is the p-th row of that matrix and x^q_j is its q-th column. These notations are unambiguous if the reader keeps in mind the number and positions of the subscripts on the initially considered scalars. Thus, in case of the just considered $x_{jp}, x_j, x_{jp}, x^q_j$ it must be remembered that the scalars x^q_{jp} have two subscripts and one superscript.

For any finite sequence x_i of square matrices, we define

$$x_\Sigma = \Sigma x_i, \quad \bar{x}_i = x_\Sigma^{-1} x_i,$$

provided x_Σ^{-1} exists, for the last relation. The sequence $\bar{x}_i$ is the *pre-normed sequence* x_i. In summations, all indices, matricial or not, must be completely summed out, unless stated otherwise.

If A is a random matrix, we denote by A° the *centered matrix* $A^\circ = A - E\,A$. No confusions can arise between $\circ$ and a matrix index, because $\circ$ shall never be used as a matrix index.

2. Hachemeister's regression model. (Hachemeister, 1975)

2.1. Definition of the model

We consider the array of *observable random variables*

$$
\begin{array}{cccc}
X_{11} & X_{21} & \ldots & X_{k1} \\
X_{12} & X_{22} & \ldots & X_{k2} \\
\ldots & \ldots & \ldots & \ldots \\
X_{1t} & X_{2t} & \ldots & X_{kt}
\end{array}
$$

The *class j* or *risk j* is the $\binom{1}{t}$ column X^j in that array.

The index s in X_{js} is interpreted as a time index. To X_j is associated the *structure variable* Θ_j (possibly multi-dimensional).

The possible values for i, j are always $i, j = 1, 2, \ldots, k$. It is assumed that all quantities further displayed actually exist and are finite.

The following assumptions are made.

(i) *Independence of completed classes*: the couples (X_1, Θ_1), (X_2, Θ_2), ..., (X_k, Θ_k) are independent.

(ii) *Equidistribution of structure variables*: the random variables Θ_1, Θ_2, . . ., Θ_k are identically distributed.

(iii) *Regression assumption*: $E(X_j/\Theta_j) = y\beta(\Theta_j)$, $\binom{1}{t}$, where
y is a given $\binom{g}{t}$ matrix of full rank $g < t$,
$\beta(.)$ is a $\binom{1}{g}$ vector of unknown functions $\beta_p(.)$.

(iv) *Covariance assumption*: $\text{Cov}(X_j/\Theta_j) = \sigma^2(\Theta_j)v_j$, $\binom{t}{t}$, where
$\sigma^2(.)$ is an unknown scalar function,
v_j is a given positive definite $\binom{t}{t}$ matrix.

2.2. Interpretations and problem

The meaning of the observable random variables X_{js} and of the matrices y, v_j is exemplified in Hachemeister (1975).

At a first stage, it is imagined that the distribution of X_j depends on a parameter θ_j. Because it is unknown, this parameter is interpreted as a realization of some hidden random variable Θ_j.

The actuary is interested in $E(X_j/\Theta_j)$ or equivalently, because of the regression assumption, in $\beta_p(\Theta_j)$, $(p = 1, 2, . . ., g)$. Usually however, $\beta_p(.)$ is unknown and Θ_j cannot be observed. Therefore, the actuary replaces his problem by a simpler one. He approximates $\beta_p(\Theta_j)$ by a linear expression.

$$B_{jp} = a_o + \sum_{is} a_{is} X_{is}$$

of the observable variables X_{js}. The unknown coefficients a_o, a_{is} (depending on j, p) are fixed in such a way that

$$E(\beta_p(\Theta_j) - B_{jp})^2$$

becomes minimum. The so obtained $\binom{1}{g}$ vector B_j is the *credibility estimator* for $\beta(\Theta_j)$.

2.3. Solution of the problem (SEE HACHEMEISTER 1975 OR DE VYLDER 1976)

The credibility estimator B_j for $\beta(\Theta_j)$ equals

(2)
$$B_j = (1\text{-}z_j)\, b + z_j\, \hat{b}_j, \ \binom{1}{g},$$

where

(3)
$$z_j = a(a + s^2 w_j)^{-1}\ \binom{g}{g}, \quad w_j = (y'v_j^{-1}y)^{-1}\ \binom{g}{g},$$

(4)
$$\hat{b}_j = (y'v_j^{-1}y)^{-1}y'v_j^{-1}X_j, \ \binom{1}{g},$$

and where

(5)
$$a = \text{Cov}\beta(\Theta_j)\ \binom{g}{g}, \quad b = E\beta(\Theta_j)\ \binom{1}{g}, \quad s^2 = E\sigma^2(\Theta_j).$$

With NORBERG (1979), we call a, b, s^2 *structural parameters* of the model. Briefly we shall call a, b, s^2 the *covariance matrix*, the *mean vector*, the *variance*

respectively. The matrix z_j is the *credibility matrix* for class j. When $g = 1$, all these quantities are scalars.

3. THE GENERAL PROBLEM OF STRUCTURAL PARAMETER ESTIMATION

3.1. Small-sample properties of the estimators

Unbiasedness and minimum-variance shall be considered as most wanted properties of the estimators for the structural parameters. These are small-sample properties.

Asymptotical properties of some estimators can also be proved. For instance see NORBERG (1979, 1980).

Such properties become only interesting in case of large samples of observed random variables. How large? This certainly depends on the complexity of the involved model. In case of Hachemeister's model, no precise answer to the question can be expected soon.

3.2. Families of unbiased estimators

In DE VYLDER (1978), we proposed families of unbiased estimators for the structural parameters. Here we shall propose the same families for b and s^2 but a different one for a.

3.3. Minimum-variance estimators

The *minimum-variance estimator* in a family of estimators $\hat{a}$ ($\hat{b}, \hat{s}^2$) for a (b, s^2) is that one making ScaVar $\hat{a}$ (ScaVar $\hat{b}$, Var $\hat{s}^2$) minimum. When looking for such minimum-variance estimators, one may have the sad surprise that the estimators contain some of the structural parameters to be estimated. However, that situation is far from hopeless as we shall see in 3.4.

3.4. Pseudo-statistics and pseudo-estimators

We call *pseudo-statistic* any known function of the observable random variables and of the structural parameters. Similarly, *pseudo-estimators* may depend on the parameters to be estimated. Such pseudo-estimators may nevertheless work as true estimators. For instance, consider the pseudo-estimators

(6) $$\hat{a} = f(X_{11}, \ldots, X_{kt}, a, b), \ \hat{b} = g(X_{11}, \ldots, X_{kt}, a, b)$$

for a, b. Suppose that the last members in (6) are not too sensitive to small variations of a, b. Then rather arbitrary initial estimates $a(0)$, $b(0)$ for a, b can be used in (6), furnishing first approximations $a(1)$, $b(1)$. The latter can be used again in (6), furnishing second approximations $a(2)$, $b(2)$ and so on. When practical convergence of the sequences $a(n)$, $b(n)$ is observed, the practical limits can be considered as final estimates for a, b.

Of course it amounts to the same to solve the system of equations

(7) $a = f(X_{11}, \ldots, X_{kt}, a, b), \; b = g(X_{11}, \ldots, X_{kt}, a, b)$

in the unknown quantities a, b.

Pseudo-estimators are used since several years by Dr. F. Bichsel in un-published practical work.

4. ESTIMATION OF THE MEAN VECTOR b

4.1. *Fixed-class estimator*

From the regression assumption results: $E\, X_j = yb$. Then $E\, \hat{b}_j = b$ by (4). This means that $\hat{b}_j$ is an unbiased estimator for b. In fact b_j is the classical least squares estimator, in class j, for b.

4.2. *Family of unbiased pseudo-estimators*

It follows that

(8) $\hat{b} = \Sigma\, x_j\, \hat{b}_j\, \binom{1}{g}, \; (x_j\, \binom{g}{g}), \; x_\Sigma = 1)$

is an unbiased estimator of b. Since we do not exclude that the x_j contain some structural parameters, relation (8) defines in fact a family of pseudo-estimators for b.

4.3. *Minimum-variance pseudo-estimator*

In DE VYLDER (1978), it is proved that the minimum-variance pseudo-estimator in the family (8) is obtained when x_j is the credibility matrix z_j pre-normed, i.e. $x_j = \bar{z}_j$. This minimum-variance estimator is denoted by

(9) $\hat{\bar{b}} = \Sigma\, \bar{z}_j\, \hat{b}_j$

It is noteworthy that TAYLOR (1977) and, in a particular case, BÜHLMANN and STRAUB (1970) obtained this optimal estimator automatically after the introduction of suitable constraints.

5. ESTIMATION OF THE VARIANCE s^2

5.1. *Fixed-class estimator*

The classical estimator for s^2, in class j, is

(10) $\hat{s}^2{}_j = \dfrac{1}{t\text{-}g}\, (X_j - y\hat{b}_j)'\, v_j^{-1}(X_j - y\hat{b}_j).$

It is unbiased. See DE VYLDER (1978).

5.2. *Family of unbiased estimators*

It follows that

(11) $\hat{s}^2 = \Sigma\, x_j\, \hat{s}_j^2, \; (x_\Sigma = 1)$

defines a family of unbiased estimators for s^2. We shall not consider the case where the scalars x_j depend on the structural parameters.

5.3. *Minimum-variance estimator under normal assumptions*

5.3.1. Normal assumption

We look for the minimum-variance estimator for s^2 in the family defined by (11). Unless new structural parameters are introduced, this problem is seen to have no solution under the general assumptions made hitherto. But it has a very simple one under the assumption that we introduce here.

Normal assumption

The conditional vector X_j, for Θ_j fixed, is normally distributed. Then it results from the general assumptions that this vector is $N(y\beta(\Theta_j),\ \sigma^2(\Theta_j)v_j)$.

The following simple lemma is needed in section 6.

Lemma.

Under the normal assumption, the conditional vector $\hat{b}_j$, for Θ_j fixed, is $N(\beta(\Theta_j),\ \sigma^2(\Theta_j)w_j)$.

Demonstration

By (4), $\hat{b}_j = u_j X_j$, with the obvious definition of $u_j \binom{t}{g}$.

Then

$$\text{Cov}\,(\hat{b}_j/\Theta_j) = u_j\,(\text{Cov}(X_j/\Theta_j)\,)\,u_j' = \sigma^2(\Theta_j)\,u_j\,v_j\,u_j'$$
$$= \sigma^2(\Theta_j)w_j.$$

Note also that the relation $E(\hat{b}_j/\Theta_j) = \beta(\Theta_j)$ results from the regression assumption and (4).

Now the lemma results from the fact that a linearly transformed normal vector remains normal.

5.3.2. Calculation of Var $\hat{s}^2$.

Theorem

Under the normal assumption,

$$(12) \qquad \text{Var}\,\hat{s}^2 = (\Sigma\,x_j^2)\,(s^{(4)}\,(1 + \frac{2}{t-g}) - s^4),$$

where $\hat{s}^2$ is defined by (11) and where

$$(13) \qquad s^4 = (s^2)^2,\ s^{(4)} = E\,\sigma^4(\Theta_j).$$

Demonstration

By the independence of the classes and the unbiasedness of $\hat{s}_j^2$,

$$(14) \qquad \text{Var}\,\hat{s}^2 = \Sigma\,x_j^2\,\text{Var}\,\hat{s}_j^2 = \Sigma\,x_j^2(E\,\hat{s}_j^4 - E^2\,\hat{s}_j^2)$$
$$= \Sigma\,x_j^2(E\,\hat{s}_j^4 - s^4)$$

By (4) and (10),

$$(t\text{-}g)\hat{s}_j^2 = X_j' \, r_j \, X_j,$$

where

$$r_j = v_j^{-1} - v_j^{-1}y(y' \, v_j^{-1}y)^{-1}y' \, v_j^{-1}, \; \binom{t}{t}.$$

Then

$$r_j y = 0, \; r_j \, v_j = 1^\circ - v_j^{-1}y(y' \, v_j^{-1}y)^{-1}y', \; (r_j \, v_j)^2 = r_j \, v_j,$$

$$\mathrm{tr}(r_j \, v_j) = t - \mathrm{tr}(y' \, v_j^{-1}y(y' \, v_j^{-1}y)^{-1}) = t - \mathrm{tr} \, 1^{\circ\circ} = t - g,$$

where 1° and $1^{\circ\circ}$ are respectively the $\binom{t}{t}$ and $\binom{g}{g}$ unit matrices and where we used the cyclical property of the trace.

Then, for Θ_j fixed, we have by (A4) of the Appendix:

$$(t - g)^2 \, E(\hat{s}_j^4/\Theta_j) = E(X_j' \, r_j \, X_j \, X_j' \, r_j X_j/\Theta_j) =$$

$$\sigma^4(\Theta_j) \, (\mathrm{tr}^2(r_j v_j) + 2\mathrm{tr}(r_j \, v_j)^2) = \sigma^4(\Theta_j) \, ((t-g)^2 + 2(t-g)).$$

Applying E and using (14), relation (12) follows.

5.3.3. Minimum-variance-estimator
Theorem

Under the normal assumption, the minimum-variance estimator in the family

(11) is

(15)
$$\hat{\hat{s}}^2 = \frac{1}{k} \sum \hat{s}_j^2.$$

Demonstration

Results from (12) because the minimum of $\sum x_j^2$ under the constraint $x_\Sigma = 1$ is obtained for $x_1 = x_2 = \ldots = x_k = 1/k$.

6. ESTIMATION OF THE COVARIANCE MATRIX

6.1. Fixed-class estimator
Theorem

(16)
$$\hat{a}_j = (\hat{b}_j - b) \, (\hat{b}_j - b)' - s^2 w_j, \; \binom{g}{g}$$

is an unbiased pseudo-estimator for the covariance matrix a.

Demonstration

The demonstration of the lemma in 5.3.1. shows that

(17)
$$\mathrm{Cov}(\hat{b}_j/\Theta_j) = \sigma^2(\Theta_j)w_j, \; E(\hat{b}_j/\Theta_j) = \beta(\Theta_j).$$

(The normal assumption is not used at that stage). Then, because $E \, \hat{b}_j = b$:

$$E(\, (\hat{b}_j - b) \, (\hat{b}_j - b)') = \mathrm{Cov} \, \hat{b}_j = \mathrm{Cov} \, E(\hat{b}_j/\Theta_j) + E \, \mathrm{Cov}(\hat{b}_j/\Theta_j)$$

(18) $\qquad = \text{Cov } \beta(\Theta_j) + E \, \sigma^2(\Theta_j) w_j = a + s^2 w_j.$

The theorem is clear from this.

6.2. Family of unbiased pseudo-estimators

From the theorem results that

(19) $\qquad \hat{a} = \Sigma \, x_j \, \hat{a}_j, \quad (x_j \left(\genfrac{}{}{0pt}{}{g}{g}\right), x_\Sigma = 1)$

defines a family of unbiased pseudo-estimators for the covariance matrix a.

The matrix a is symmetrical, whereas $\hat{a}$ furnished by (19) is not. Therefore
NORBERG (1979) replaces $\hat{a}$ by

(20) $\qquad\qquad\qquad \frac{1}{2}(\hat{a} + \hat{a}')$

(and also b, s^2 by estimators for these structural parameters).

We shall not perform the symmetrization at this stage. Of course, in practical
work, any estimate for a must finally be symmetrized.

6.3. Minimum-variance pseudo-estimator under normal and other assumptions

6.3.1. Notation.
We define

$$a^{(2)} = E(\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)).$$

Note that

$$a = E(\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)\,).$$

The notation $s^{(4)}$ already introduced in (13) is also used further.
$a^{(2)}$ and $s^{(4)}$ are fourth-order structural parameters.

6.3.2. Independence assumption

We shall use the assumption of independence of $\sigma^2(\Theta_j)$ and $\beta(\Theta_j)$. Briefly,
we call it the *independence assumption.*

6.3.3. Calculation of ScaVar $\hat{a}$.

Lemma

Under the normal assumption (5.3.1.):

(21) $\quad E(\hat{b}_j^{\circ'} x_j' x_j \hat{b}_j^\circ \; \hat{b}_j^{\circ'} \hat{b}_j^\circ / \Theta_j) =$

$\qquad \sigma^2(\Theta_j)\text{tr}(x_j'x_jw_j)\text{tr}(w_j) + 2\sigma^4(\Theta_j)\text{tr}(x_j'x_jw_jw_j)$

$+ \; 4 \; \sigma^2(\Theta_j)\text{tr}(x_j'x_jw_j\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)\,)$

$+ \quad \sigma^2(\Theta_j)\text{tr}(x_j'x_j\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)\,)\text{tr}(w_j)$

$+ \quad \sigma^2(\Theta_j)\text{tr}(x_j'x_jw_j)\text{tr}(\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)\,)$

$+ \qquad \text{tr}(x_j'x_j\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)\beta^\circ(\Theta_j)\beta^{\circ'}(\Theta_j)\,).$

Demonstration

By the lemma in 5.3.1, the conditional vector $\hat{b}_j^\circ$, for fixed Θ_j is $N(\beta^\circ(\Theta_j),$ $\sigma^2(\Theta_j)w_j)$. Then (21) follows from (A4) in the Appendix with

$$a = x_j'x_j, \; b = 1, \; \mathrm{m} = \beta^\circ(\Theta_j), \; v = \sigma^2(\Theta_j)w_j.$$

Theorem

Under the normal (5.3.1) and independence (6.3.2) assumptions, for $\hat{a}$ defined by (19):

$$
\begin{aligned}
(22) \quad \text{ScaVar } \hat{a} = \; & \Sigma s^{(4)} \mathrm{tr}(x_j'x_jw_j)\mathrm{tr}(w_j) + 2\Sigma s^{(4)}\mathrm{tr}(x_j'x_jw_jw_j) \\
& + 4\Sigma s^2 \, \mathrm{tr}(x_j'x_jw_ja) + \Sigma \, s^2 \mathrm{tr}(x_j'x_ja)\mathrm{tr}(w_j) \\
& + \Sigma \, s^2 \, \mathrm{tr}(x_j'x_jw_j) \, \mathrm{tr} \, a + \Sigma \mathrm{tr}(x_j'x_j(a)^2) \\
& - \Sigma \, \mathrm{tr}(x_j'x_j(a+s^2w_j)\,(a+s^2w_j)\,).
\end{aligned}
$$

Demonstration

Let $T\binom{g}{g}$ denote the pseudo-statistic

$$(23) \qquad\qquad T = \Sigma \, x_j\hat{b}_j\hat{b}_j' = \Sigma \, x_j(\hat{b}_j-b)\,(\hat{b}_j-b)', \; \binom{g}{g}$$

Then

$$(24) \qquad\qquad\qquad \hat{a} = T - s^2 \, \Sigma \, x_jw_j.$$

Because the last term is non-random,

$$\text{ScaVar } \hat{a} = \text{ScaVar } T = \mathrm{tr} \, E(TT') - \mathrm{tr}(ET.ET').$$

Using the independence of classes,

$$
\begin{aligned}
(25) \quad E(TT') = \; & \Sigma_{ij} E(x_i\hat{b}_i^\circ\hat{b}_i^{\circ\prime}\hat{b}_j^\circ\hat{b}_j^{\circ\prime}x_j') \\
= \; & \Sigma_j E(x_j\hat{b}_j^\circ\hat{b}_j^{\circ\prime}\hat{b}_j^\circ\hat{b}_j^{\circ\prime}x_j') \; + \Sigma_{i\neq j} E(x_i\hat{b}_i^\circ\hat{b}_i^{\circ\prime})E(\hat{b}_j^\circ\hat{b}_j^{\circ\prime}x_j').
\end{aligned}
$$

Similarly,

$$(26) \quad ET.ET' = \Sigma_j E(x_j\hat{b}_j^\circ\hat{b}_j^{\circ\prime})E(\hat{b}_j^\circ\hat{b}_j^{\circ\prime}x_j') \; + \Sigma_{i\neq j} E(x_i\hat{b}_i^\circ\hat{b}_i^{\circ\prime})E(\hat{b}_j^\circ\hat{b}_j^{\circ\prime}x_j').$$

Thus, in the difference $E(TT') - ET.ET'$, the sum in $i \neq j$ cancels.

By the cyclical property of the trace and the general relation $E(\cdot) = EE(\cdot /\Theta_j)$, the trace of the first sum in the last member of (25) equals

$$\Sigma \, E \, E(\hat{b}_j^{\circ\prime}x_j'x_j\hat{b}_j^\circ \; \hat{b}_j^{\circ\prime}\hat{b}_j^\circ/\Theta_j).$$

The positive terms in (22) result from this expression and from the lemma. The negative sum in (22) comes from the first sum in (26) after an application of (18).

6.3.4. Minimum-variance pseudo-estimator.

Lemma

For $j = 1, 2, \ldots, k$ and $n = 1, 2, \ldots, r$, let x_j be variable $\binom{g}{g}$ matrices, let c_{nj} be fixed $\binom{g}{g}$ matrices and let s_{nj} be fixed scalars. Then the extremum of

$$(27) \qquad \sum_{nj} s_{nj} \operatorname{tr}(x_j' x_j c_{nj}),$$

under the constraint $x_{\Sigma} = 1$, is obtained for

$$(28) \qquad x_j = \bar{y}_j,$$

where

$$(29) \qquad y_j = \left(\sum_n s_{nj}(c_{nj} + c_{nj}') \right)^{-1}.$$

Demonstration

Applying Lagrange's method, we consider g^2 multipliers $- \lambda_p^q$ corresponding to the constraints $x_{\Sigma} = 1$. Expanding (27), we have to put equal to zero the partial derivatives in the variables of

$$L = \sum_{nj\alpha\beta\gamma} s_{nj} x_{j\beta}^{\alpha} x_{j\beta}^{\gamma} c_{nj\gamma}^{\alpha} - \sum_{jpq} \lambda_p^q x_{jp}^q.$$

Derivation in x_{jp}^q gives the equation

$$\sum_{n\gamma} s_{nj} x_{jp}^{\gamma} c_{nj\gamma}^q + \sum_{n\alpha} s_{nj} x_{jp}^{\alpha} c_{njq}^{\alpha} = \lambda_p^q,$$

or, in matrix form

$$(30)\ \sum_n s_{nj} x_j c_{nj} + \sum_n s_{nj} x_j c_{nj}' = \lambda, \ x_j \left(\sum_n s_{nj}(c_{nj} + c_{nj}') \right) = \lambda, \ x_j = \lambda\, y_j,$$

where y_j is defined by (29). Summing over j, we obtain $1 = \lambda\, y_j$. From this relation results λ. Substitution in (30) gives (28).

Theorem

Under the normal (5.3.1) and independence (6.3.2) assumptions, the minimum-variance pseudo-estimator in the family (19) is

$$(31) \qquad \hat{a} = \sum \bar{x}_j \hat{a}_j,$$

where

$$(32) \qquad x_j = a^2 \left[\tfrac{3}{2}(s^{(4)} - s^4)(w_j)^2 - \tfrac{1}{2}(3a^2 - a^{(2)}) + (a + s^2 w_j)^2 \right]^{-1}.$$

For $g > 1$, that result is based on the approximations (33).

Demonstration

By the lemma, the minimum of (22) is obtained for the matrices $\bar{x}_j$, where

$$x_j^{-1} = 2s^{(4)} w_j \operatorname{tr} w_j + 4s^{(4)} w_j w_j + 4s^2 w_j a + 4s^2 a w_j$$
$$+ 2s^2 a \operatorname{tr} w_j + 2s^2 w_j \operatorname{tr} a + 2\, a^{(2)} - 2(a + s^2 w_j)(a + s^2 w_j).$$

Using the approximations (if $g > 1$)

$$(33) \quad \begin{cases} w_j \operatorname{tr} w_j + 2 \, w_j w_j \cong 3 \, w_j w_j \\ a \operatorname{tr} w_j + 2 \, a \, w_j \cong 3 \, a \, w_j \\ w_j \operatorname{tr} a \; + 2 \, w_j a \; \cong 3 \, w_j \, a \end{cases}$$

one obtains

$$x_j^{-1} \cong 6(s^{(4)} - s^4) \, (w_j)^2 - 2(3a^2 - a^{(2)}) + 4(a + s^2 w_j)^2.$$

From this results (32), except for a constant (not depending on j) pre-factor. But such a pre-factor cancels in the formation of $\bar{x}_j$.

Comments

1. In any practical case, the quality of the approximations (33) can be tested on basis of the numerical data, at least with a replaced by its estimate.

2. Let $q_j \binom{g}{g}$ be defined by

$$(34) \quad \tfrac{3}{2}(s^{(4)} - s^4) \, (w_j)^2 - \tfrac{1}{2}(3a^2 - a^{(2)}) + (a + s^2 w_j)^2 = (a + s^2 w_j)^2 \, q_j.$$

Then, if $q_j = q$, independent of j, we are rid of the fourth-order structural parameters $s^{(4)}$ and $a^{(2)}$. Indeed, the inverse of the last member of (34) equals, in that case, $q^{-1} \, (a + s^2 w_j)^{-2}$ and the constant pre-factor q^{-1} cancels in the pre-norming of these matrices. Then we obtain the following unbiased pseudo-estimator for a:

$$(35) \qquad\qquad \hat{\hat{a}} = \Sigma \, \bar{z}_j^{(2)} \, \hat{a}_j,$$

where

$$(36) \qquad\qquad z_j^{(2)} = a^2(a + s^2 w_j)^{-2}.$$

The irrelevant constant pre-factor a^2 is introduced again, because so $z_j^{(2)} = (z_j)^2$ for $g = 1$.

3. Les us now examine some cases where $q_j = q$, constant.

  (i) Of course $q_j = q$ in the trivial case $w_j = w$, constant.

  (ii) Let $\beta(\Theta_j) = b$ be degenerated. Then $a = (a)^2 = 0$ and in (34) each remaining term can be simplified by the factor $(w_j)^2$. Then $q_j = q$, constant.

(iii) Let $a^{(2)} = 3a^2$. For $g = 1$, this means that the random variable $\beta(\Theta_j)$ has a coefficient of excess equal to zero. That is the case if it is normally distributed. Furthermore, assume the degeneracy $\sigma^2(\Theta_j) = s^2$. Then $s^{(4)} = s^4$ and $q_j = 1$.

4. We summarize the arguments collected in favour of the pseudo-estimator $\hat{\hat{a}}$ (35)

  — It is unbiased, because $\bar{z}_\Sigma^{(2)} = 1$.

  — It does not depend on fourth-order structural parameters.

  — It is optimal in degenerated situations. It will remain approximately optimal in cases close to such situations. It must be noted that it is

precisely in case of the degeneracy described in (ii) that the estimation of a is most delicate.

— For $g = 1$, the weights $z_j^{(2)}$ are simply the squared credibility weights.

6.4. Practical estimation of the covariance matrix

6.4.1. Problem.

The pseudo-estimator $\mathring{a}$ (35) and more generally, the pseudo-estimators (19), have a great inconvenience. Although they are unbiased when b receives its true value in (16), they no longer have that interesting quality when b is replaced, say by its optimal pseudo-estimator $\mathring{b}$ (9). But that is precisely what should be done in practice! The bias appears because b does not figure linearly in the last member of (16). For s^2, there is no problem. Its replacement by $\mathring{s}^2$ (15) causes no bias. For the structural parameters a, s^2 appearing in $z_j^{(2)}$ (36) the situation is different. Even when they are replaced by wrong estimates, $\mathring{a}$ (35) is unbiased because $\bar{z}_\Sigma^{(2)} = 1$.

Thus, we have only to correct the situation for b. We shall start now from the pseudo-statistic

$$(37) \qquad \acute{T} = \sum_j x_j (\hat{b}_j - \mathring{b}) (\hat{b}_j - \mathring{b})' \; \binom{g}{g}, \; (x_j \binom{g}{g}, \; x_\Sigma = 1).$$

A look at (23) shows that $\hat{T}$ is T wherein b has been replaced by its optimal pseudo-estimator $\mathring{b}$.

In case of $\hat{T}$, the research for optimal weights x_j is based on so lengthy combinatorial calculations that it loses its interest.

In that respect, we shall simply transpose to the case of $\hat{T}$, the results obtained in section 6.3.4 in the case of T. At least for $g = 1$, lengthy arguments show that this is a good approximation.

6.4.2. A less "pseudo" family of unbiased pseudo-estimators for a.

Theorem

The relation

$$(38) \qquad \tilde{a} = \hat{T} - s^2 \sum_j x_j w_j + z_\Sigma^{-1} a, \; (x_j \binom{g}{g}, \; x_\Sigma = 1),$$

defines a family of unbiased pseudo-estimators for a.

Demonstration

We have

$$\hat{T} = \sum_i x_i \left(\hat{b}_i^\circ - \sum_j \bar{z}_j \hat{b}_j^\circ \right) \left(\hat{b}_i^\circ - \sum_k \bar{z}_k \hat{b}_k^\circ \right)'$$

because it follows from $\bar{z}_\Sigma = 1$ that the replacement of $\hat{b}_i$ by $\hat{b}_i^\circ$ is allowed. Then

$$\hat{T} = \sum_i x_i \hat{b}_i^\circ \hat{b}_i^{\circ\prime} + \sum_{ijk} x_i \bar{z}_j \hat{b}_j^\circ \hat{b}_k^{\circ\prime} \bar{z}_k^\prime - \sum_{ik} x_i z_j \hat{b}_j^\circ \hat{b}_i^{\circ\prime} - \sum_{ik} x_i \hat{b}_i^\circ \hat{b}_k^{\circ\prime} \bar{z}_k^\prime.$$

Because $\Sigma x = 1$, the x_i can be summed out in the sum in ijk.

When we apply E, the dubble sums must be replaced by single sums and the indices must be egalized, say replaced by j. This follows from the fact that the $\hat{b}_j^\circ$ have zero expectation and are independent. By (18) we then obtain

$$E\,\hat{T} = \sum_i x_j (a + s^2 w_j) + \sum \bar{z}_j (a + s^2 w_j) \bar{z}_j^\prime - \sum x_j \bar{z}_j (a + s^2 w_j) - \sum x_j (a + s^2 w_j) \bar{z}_j^\prime$$

The first term in the last member equals $a + s^2 \sum x_i w_i$.

From these consideration it is clear that

$$\tilde{a} = \hat{T} - s^2 \sum x_j w_j - \sum \bar{z}_j (a + s^2 w_j) \bar{z}_j$$

(39)
$$+ \sum x_j \bar{z}_j (a + s^2 w_j) + \sum x_j (a + s^2 w_j) \bar{z}_j^\prime,$$

is an unbiased pseudo-estimator for a. From the relations

$$\bar{z}_j = z_\Sigma^{-1} z_j = z_\Sigma^{-1} a (a + s^2 w_j)^{-1}$$

we obtain

$$\sum \bar{z}_j (a + s^2 w_j) \bar{z}_j^\prime = \sum z_\Sigma^{-1} a \bar{z}_j^\prime = z_\Sigma^{-1} a$$

(showing in particular that $z_\Sigma^{-1} a$ is symmetrical) and similar relations allowing to simplify the last member of (39). Finally (38) is obtained.

6.4.3. Practical estimation of a.

Because a is symmetric, we adopt for it the pseudo-estimator (38) symmetrized in the obvious way. As matrical weights x_j in (38) we suggest to take the pre-normed $z_j^{(2)}$ defined by (36), i.e. the same weights $\bar{z}_j^{(2)}$ as in (35).

7. SUMMARY OF ALL STEPS TO BE EXECUTED FOR THE CALCULATION OF THE CREDIBILITY ESTIMATORS

Steps depending on j must be executed for $j = 1, 2, \ldots, k$.

0) Given: $X_j \binom{1}{t}$, $v_j \binom{t}{t}$, $y \binom{g}{t}$

    1) Calculate $w_j = (y' v_j^{-1} y)^{-1} \binom{g}{g}$

    2) Calculate $\hat{b}_j = w_j y' v_j^{-} X_j \binom{1}{g}$

    3) Calculate the scalar $s^2 = \dfrac{1}{k(t-g)} \sum (X_j - y\hat{b}_j)' v_j^{-1} (X_j - y\hat{b}_j)$

    4) Fix initial $a = \text{diag}(c, c, \ldots, c) \binom{g}{g}$, ($c$ large)

5) Calculate $z_j = a(a + s^2 w_j)^{-1} \binom{g}{g}$, $z_\Sigma \binom{g}{g}$, $\bar{z}_j \binom{g}{g}$,

   $z_j^{(2)} = a^2(a + s^2 w_j)^{-2} \binom{g}{g}$, $z_\Sigma^{(2)} \binom{g}{g}$, $x_j = \bar{z}_j^{(2)} \binom{g}{g}$

6) Calculate $b = \Sigma \, \bar{z}_j \, \hat{b}_j \, \binom{1}{g}$

7) Calculate $S = \Sigma \, x_j(\hat{b}_j - b)\,(\hat{b}_j - b)' - s^2 \, \Sigma \, x_j w_j + z_\Sigma^{-1} a \, \binom{g}{g}$

8) Calculate New $a = \frac{1}{2}(S + S') \binom{g}{g}$

9) Return to step 5) with a replaced by New a, as many times as necessary, until a is stabilized.

10) Calculate $B_j = (1 - z_j)b + z_j \hat{b}_j$.

APPENDIX

Expectation of products of quadratic forms in normal variables

Theorem

Let a, b, v be symmetrical $\binom{t}{t}$ matrices, v positive definite. If the $\binom{1}{t}$ vector X is $N(o, v)$, then

(A1) $$E(X'aX) = \mathrm{tr}(av),$$

(A2) $$E(X'aX \; X'bX) = \mathrm{tr}(av)\mathrm{tr}(bv) + 2 \, \mathrm{tr}(avbv)$$

If X is $N(m, v)$, then

(A3) $$E(X'aX) = \mathrm{tr}(av) + \mathrm{tr}(amm'),$$

(A4) $$\begin{aligned} E(X'aX \; X'bX) = \; &\mathrm{tr}(av)\mathrm{tr}(bv) + 2 \, \mathrm{tr}(avbv) + 4 \, \mathrm{tr}(avbmm') \\ &+ \mathrm{tr}(amm')\mathrm{tr}(bv) + \mathrm{tr}(bmm')\mathrm{tr}(av) + \mathrm{tr}(amm'bmm'). \end{aligned}$$

Demonstration

(A1) $$E(X'aX) = \underset{i,j}{\Sigma} E(X_i a_i^j X_j) = \underset{i,j}{\Sigma} a_i^j E(X_i X_j) = \underset{i,j}{\Sigma} a_i^j v_j^i = \mathrm{tr}(av)$$

(A2) Let the $\binom{1}{t}$ vector Y be $N(o, 1)$. Then each component is normal with zero mean and unit variance. Therefore

$$E \, Y_i = o, \; E \, Y_i^2 = 1, \; E \, Y_i^3 = o, \; E \, Y_i^4 = 3.$$

Moreover, Y_1, Y_2, ..., Y_t are independent. From these facts follows the combinatorial formula

(A5) $$E(Y_i Y_j Y_k Y_l) = \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{lj} + \delta_{il}\delta_{jk}$$

where δ_{ij} is Kronecker's symbol.

Indeed, if m, n, p, q are 4 different indices, the possible patterns for i, j, k, l are

i	m	m	n	n	n	m	m	m	m	...	m
j	m	n	m	n	n	m	n	n	m	...	n
k	m	n	n	m	n	n	m	n	n	...	p
l	m	n	n	n	m	n	n	m	p	...	q

and for each pattern, the verification of (A5) is immediate. Then

$$E(Y'aYY'bY) = \sum_{ijkl} E(Y_i a_i^j Y_j Y_k b_k^l Y_1) = \sum_{ijkl} a_i^j b_k^l E(Y_i Y_j Y_k Y_l) =$$

$$\sum_{ijkl} a_i^j b_k^l (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{lj} + \delta_{il}\delta_{jk}) = \sum_{ik} a_i^i b_k^k + \sum_{il} a_i^l b_i^l + \sum_{ij} a_i^j b_j^i =$$

$$\sum_i a_i^i \sum b_k^k + \sum_{il} a_i^l b_l^i + \sum_{ij} a_i^j b_j^i = \text{tr. } a \cdot \text{tr } b + \text{tr}(ab) + \text{tr}(ab).$$

This proves (A2) for $v = 1$.

Now we prove (A2) for X supposed to be $N(o, v)$. It is classical that for some $\binom{t}{t}$ matrix s, we have $X = sY$ with Y $N(o, 1)$.

Then

$$v = \text{Cov } X = s(\text{Cov } Y)s' = ss'$$

and

$$\begin{aligned} E(X'aXX'bX) &= E(Y's'asYY's'bsY) \\ &= \text{tr}(s'as)\text{tr}(s'bs) + 2\,\text{tr}(s'ass'bs) \\ &= \text{tr}(ass')\text{tr}(bss') + 2\,\text{tr}(ass'bss') \\ &= \text{tr}(av)\text{tr}(bv) + 2\,\text{tr}(avbv), \end{aligned}$$

by the cyclical property of the trace.

(A3)   We decompose $X = Y + m$, where $Y = X - m$ is $N(o, v)$.

Then

(A6) $X'aX = Y'aY + 2\,Y'am + m'am$

and (A3) follows from (A1) and the relations

$$E(Y'am) = o, \quad m'am = \text{tr}(m'am) = \text{tr}(amm').$$

(A4)   For (A4) we use (A6) and the similar decomposition

(A7) $X'bX = Y'bY + 2\,m'bY + m'bm.$

The expectation of the product of (A6) and (A7) is best displayed in a table. Making use of (A1), (A2) we obtain the table

	$Y'bY$	$2\,m'bY$	$m'bm$
$Y'aY$	$\text{tr}(av)\text{tr}(bv)$ $+ 2\,\text{tr}(avbv)$	o	$\text{tr}(av)m'bm$
$2\,Y'am$	o	$4\,\text{tr}(amm,bv)$	o
$m'am$	$m'amtr(bv)$	o	$m'amm'bm$

After slight transformations (use cyclical property of trace, and also that a matrix and its transposed matrix have the same trace), formula (A4) results from this table. However, some supplementary explanations may be useful.

First, the matrix $amm'b$ is not necessarily symmetrical. The corresponding term $4 \, \text{tr}(amm'bv)$ is the right one because (A1) is valid even if a is not symmetrical (see demonstration). We have also to explain the zero's in the table. For instance, let us justify the relation

$$E(Y'aYm'bY) = o.$$

For some $\binom{t}{t}$ matrix s, Y can be expressed as $Y = sZ$ where the $\binom{1}{t}$ vector Z is $N(o, 1)$. Then the components Z_i of Z are independent and have zero expectation. The development of $Y'aYm'bY$ is an expression $\Sigma \, c_{ijk} \, Z_i Z_j Z_k$. Among the indices i, j, k, at least one must appear an odd number of times. Therefore $E(Z_i Z_j Z_k) = o$.

REFERENCES

BÜHLMANN, H. and STRAUB, E. (1970). Glaubwürdigkeit für Schadensätze. *Mitt. Ver. Schw. Vers.-Math.*, **79**, 111-133.

DE VYLDER, F. (1976). Geometrical Credibility. *Scandinavian Actuarial Journal*, 1976, 121-149.

DE VYLDER, F. (1978). Parameter estimation in credibility theory. *Astin Bulletin*, **10**, 99-112.

HACHEMEISTER, C. (1975). *Credibility for regression models with application to trend*. In Kahn (1975).

KAHN, P. M., Editor. (1975). *Credibility. Theory and Applications*. (Academic Press).

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

NORBERG, R. (1980). Emperical Bayes credibility. *Scandinavian Actuarial Journal*, 1980, 177-194.

TAYLOR, G. C. (1977). Abstract Credibility. *Scandinavian Actuarial Journal*, 1977 149-168.