OPTIMAL ESTIMATION UNDER LINEAR CONSTRAINTS

WALTHER NEUHAUS

University of Copenhagen and UNI Storebrand ASA (Oslo)

ABSTRACT

This paper shows how a multivariate Bayes estimator can be adjusted to satisfy a set of linear constraints. In the direct approach, the constraint is enforced by a restriction on the class of admissible estimators. In an alternative approach, the constraint is merely encouraged by a mixed risk function which penalises misbalance between the estimator and the constraint. The adjustment to the optimal unconstrained estimator is shown to depend on the risk function and the linear constraints only, not on the probability model underlying the Bayes estimator. Two practical examples are given, one of which involves reconciliation of independently assessed share values with current market values.

KEYWORDS

Bayes Estimation, Linear Bayes Estimation, Credibility Theory, Share Valuation

1. INTRODUCTION

Actuaries often need to reconcile the estimates they have arrived at, with the data used to calculate the estimates. In the estimation of pure premiums, for instance, the actuary would always check that the total premiums calculated are sufficient to cover the total cost of claims.

The concept of balanced linear estimators was introduced by Neuhaus (1995). In that paper, a linear estimator is called balanced if it satisfies certain linear constraints involving the original data; furthermore, the optimal linear estimator is called the credibility estimator, and the optimal balanced linear estimator is called the balanced credibility estimator.

This paper generalises the balancing concept in two directions. The first generalisation, presented in Section 2, involves balancing arbitrary estimators, i.e. estimators which are not necessarily linear in the data. By this approach one arrives at an optimal balanced estimator. As a by-product one obtains a much shorter derivation of the balanced credibility estimator than in Neuhaus (1995). A simple example of the calculations needed is given in Section 3.

Section 4 provides a more practical example. Given independent valuations of the different shares in a market, the actuary could wish to reconcile these values to the market value of the portfolio currently held, as well as the overall value of the market. We apply the balancing formula to calculate the optimal adjustment of individually assessed share values.

In Section 5 we compare the balanced credibility estimator with the homogeneous unbiased credibility estimator in the Bühlmann-Straub model. The latter estimator has long been known to be balanced, see e.g. Gisler (1987).

The second generalisation, briefly presented in Section 6, involves the use of a mixed risk function, where balancing is not enforced but misbalance is penalised. The mixed risk function is similar to the optimisation criterion used in Whittaker-Henderson graduation (see e.g. Taylor, 1992). A mixed risk function has also been used by Sundt (1992) as a way of smoothing a sequence of credibility estimators.

In both generalisations, the necessary adjustment to the unconstrained Bayes estimator turns out to be independent of the probability model underlying the Bayes estimator. This allows the actuary to balance the estimates after the unconstrained Bayes estimator has been calculated, and without reference to the model used.

Finally, a few words on terminology: Actuaries' frequent need to balance a set of estimators against a set of data, is the prime motivation for studying linearly constrained estimators. Since in the general context it is easy, however, to construct linear constraints that do not comply with any sensible notion of balancing, we will simply talk of constrained estimators in the balance of this paper.

2. Optimal estimation under binding linear constraints

We assume the existence of a latent random vector (the estimand),

$$\mathbf{b}^{p \times 1} := (b_1, \dots, b_p)', \tag{2.1}$$

as well as the existence of an observed the random vector (the statistic)

$$\mathbf{X}^{n \times 1} := (\mathbf{X}_1, \dots, \mathbf{X}_n)'. \tag{2.2}$$

Assume that $\mathbf{b}$ and $\mathbf{X}$ are defined over the same probability space and square integrable, and assume also that the joint distribution of $(\mathbf{b},\mathbf{X})$ is known.

An estimator $\mathbf{b}$ is any measurable function

$$\breve{\mathbf{b}} : \mathbb{R}^n \to \mathbb{R}^p : \mathbf{x} \to \mathbf{b}(\mathbf{x}), \tag{2.3}$$

such that $\breve{\mathbf{b}}(\mathbf{X})$ is square integrable. The criterion (risk function) we use to measure the performance of a given estimator is generalised mean squared error,

$$r(\breve{\mathbf{b}}) := \mathrm{E}[(\breve{\mathbf{b}}(\mathbf{X}) - \mathbf{b})'\mathbf{W}(\breve{\mathbf{b}}(\mathbf{X}) - \mathbf{b})], \tag{2.4}$$

with $\mathbf{W}^{p \times p}$ an some fixed, positive definite risk weighting matrix.

It is well known that the optimal estimator in the absence of any constraints, which we denote by $\bar{\mathbf{b}}$, is the conditional mean of $\mathbf{b}$ given $\mathbf{X}$:

$$\bar{\mathbf{b}}(\mathbf{X}) = \mathrm{E}[\mathbf{b} \mid \mathbf{X}]. \tag{2.5}$$

The risk of that estimator is

$$r(\bar{\mathbf{b}}) = \mathrm{tr}[\mathbf{W} \cdot \mathrm{ECov}[\mathbf{b} \mid \mathbf{X}]]. \tag{2.6}$$

The optimality of $\bar{\mathbf{b}}$ and its risk (2.6) are a direct consequence of the decomposition

$$\begin{aligned} r(\breve{\mathbf{b}}) &= \mathrm{EE}[(\breve{\mathbf{b}} - \mathbf{b})'\mathbf{W}(\breve{\mathbf{b}} - \mathbf{b}) \mid \mathbf{X}] \\ &= \mathrm{tr}[\mathbf{W} \cdot \mathrm{ECov}(\mathbf{b} \mid \mathbf{X})] + \mathrm{E}[(\breve{\mathbf{b}} - \bar{\mathbf{b}})'\mathbf{W}(\breve{\mathbf{b}} - \bar{\mathbf{b}})] \\ &= r(\bar{\mathbf{b}}) + \mathrm{EE}[(\breve{\mathbf{b}} - \bar{\mathbf{b}})'\mathbf{W}(\breve{\mathbf{b}} - \bar{\mathbf{b}}) \mid \mathbf{X}]. \end{aligned} \tag{2.7}$$

We refer to the last term in the above expression as the excess risk of the estimator $\breve{\mathbf{b}}$.

Assume now that a constraint of the following general form has been imposed on the class of admissible estimators:

$$f(\breve{\mathbf{b}}(\mathbf{X})) = g(\mathbf{X}) \quad a.s., \tag{2.8}$$

where $f : \mathbb{R}^p \to \mathbb{R}^q$ and $g : \mathbb{R}^n \to \mathbb{R}^q$ are known, fixed functions. The constraint may equivalently be stated as

$$\breve{\mathbf{b}}(\mathbf{X}) \in f^{-1}\{g(\mathbf{X})\}, \tag{2.9}$$

provided that $f^{-1}\{g(\mathbf{X})\} \neq \emptyset$ a.s. An estimator will be called constrained if it satisfies (2.8).

From the decomposition (2.7) and the observation that both $\bar{\mathbf{b}}$ and $\breve{\mathbf{b}}$ are functions of the statistic $\mathbf{X}$, it is evident that the optimal constrained estimator, which we denote by $\tilde{\bar{\mathbf{b}}}(\mathbf{X})$, can be found by pointwise minimisation for each possible realisation $\mathbf{X}=\mathbf{x}$:

$$\tilde{\bar{\mathbf{b}}}(\mathbf{x}) = \mathrm{pro}_{\mathbf{W})}(\bar{\mathbf{b}}(\mathbf{x}) \mid f^{-1}\{g(\mathbf{x})\}), \tag{2.10}$$

where $\mathrm{pro}_{\mathbf{W}}(\mathbf{a} \mid B)$ denotes a projection of a vector $\mathbf{a}$ into a set B, with respect to the metric derived from the inner product $<\mathbf{a}, \mathbf{b}> = \mathbf{a}'\mathbf{W}\mathbf{b}$. For general functions f, g, this projection need not be unique and may not even exist; however, the projection does exist if f is a continuous function. In particular, if f and g are linear functions, the projection has the explicit formula given in the next theorem.

Theorem

Assume that the two constraining functions are linear,

$$f(\breve{\mathbf{b}}) = \mathbf{L}\breve{\mathbf{b}}, \quad g(\mathbf{x}) = \mathbf{P}\mathbf{x}, \tag{2.11}$$

where $\mathbf{L}^{qxp}$ and $\mathbf{P}^{qxn}$ are known, fixed matrices, and $\mathbf{L}$ is of full row rank $q \leq p$; then the optimal constrained estimator is

$$\tilde{\mathbf{b}}(\mathbf{X}) = \bar{\mathbf{b}}(\mathbf{X}) + \mathbf{W}^{-1}\mathbf{L}'\mathbf{Q}^{-1}(\mathbf{PX} - \mathbf{L}\bar{\mathbf{b}}(\mathbf{X})), \tag{2.12}$$

with $\mathbf{Q} = \mathbf{LW}^{-1}\mathbf{L}'$.

Proof

Consider a fixed value of the statistic ($\mathbf{X} = \mathbf{x}$) and note that both $\breve{\mathbf{b}}$ and $\bar{\mathbf{b}}$ are functions of $\mathbf{x}$. Define the Lagrange functional

$$F = \frac{1}{2}(\breve{\mathbf{b}} - \bar{\mathbf{b}})'\mathbf{W}(\breve{\mathbf{b}} - \bar{\mathbf{b}}) - \lambda'(\mathbf{L}\breve{\mathbf{b}} - \mathbf{Px}), \tag{2.13}$$

with $\lambda^{q\times 1}$ a vector of Lagrange multipliers. Solving the equation

$$\frac{\partial F}{\partial \breve{\mathbf{b}}} = (\breve{\mathbf{b}} - \bar{\mathbf{b}})'\mathbf{W} - \lambda'\mathbf{L} = \mathbf{0} \tag{2.14}$$

yields

$$\breve{\mathbf{b}}(\lambda) = \bar{\mathbf{b}} + \mathbf{W}^{-1}\mathbf{L}'\lambda. \tag{2.15}$$

Now use the constraint (2.11) to determine $\lambda = \mathbf{Q}^{-1}(\mathbf{Px}-\mathbf{L}\bar{\mathbf{b}})$; substitute this expression in (2.15) to find (2.12).

Remarks

Two remarks on the hypotheses of the theorem are in order. Firstly, the full rank assumption on $\mathbf{L}$ is needed in order to ensure that the equation (2.8) is consistent. Secondly, the assumption that g is linear has not been used at all; thus the estimator (2.12) may easily be extended to more general functions g; however, the assumption that g is linear will allow us to derive transparent formulas for the excess risk, which is our next stopping point.

Using (2.12), the excess risk of the optimal constrained estimator over the unconstrained Bayes estimator is easily seen to be

$$\mathrm{E}[(\tilde{\mathbf{b}} - \bar{\mathbf{b}})'\mathbf{W}(\tilde{\mathbf{b}} - \bar{\mathbf{b}})] = \mathrm{tr}[\mathbf{Q}^{-1}\mathrm{E}(\mathbf{PX} - \mathbf{L}\bar{\mathbf{b}})(\mathbf{PX} - \mathbf{L}\bar{\mathbf{b}})']. \tag{2.16}$$

One can write

$$\begin{aligned}
\mathrm{E}(\mathbf{PX} - \mathbf{L}\bar{\mathbf{b}})(\mathbf{PX} - \mathbf{L}\bar{\mathbf{b}})' &= \mathrm{E}[\mathrm{E}(\mathbf{PX} - \mathbf{Lb} \mid \mathbf{X}) \cdot \mathrm{E}'(\mathbf{PX} - \mathbf{Lb} \mid \mathbf{X})] \\
&= \mathrm{Cov}\mathrm{E}[\mathbf{PX} - \mathbf{Lb} \mid \mathbf{X}] + \mathrm{E}(\mathbf{PX} - \mathbf{Lb}) \cdot \mathrm{E}'(\mathbf{PX} - \mathbf{Lb}) \\
&= \mathrm{Cov}(\mathbf{PX} - \mathbf{Lb}) - \mathrm{E}\mathrm{Cov}[\mathbf{PX} - \mathbf{Lb} \mid \mathbf{X}] \\
&\quad + \mathrm{E}(\mathbf{PX} - \mathbf{Lb}) \cdot \mathrm{E}'(\mathbf{PX} - \mathbf{Lb}) \\
&= \mathrm{E}(\mathbf{PX} - \mathbf{Lb})(\mathbf{PX} - \mathbf{Lb})' - \mathbf{L} \cdot \mathrm{E}\mathrm{Cov}[\mathbf{b} \mid \mathbf{X}] \cdot \mathbf{L}' \\
&= \mathbf{P} \cdot \mathrm{E}\mathrm{Cov}[\mathbf{X} \mid \mathbf{b}] \cdot \mathbf{P}' - \mathbf{L} \cdot \mathrm{E}\mathrm{Cov}[\mathbf{b} \mid \mathbf{X}] \cdot \mathbf{L}' \\
&\quad + \mathrm{E}[(\mathbf{P} \cdot \mathrm{E}(\mathbf{X} \mid \mathbf{b}) - \mathbf{Lb})(\mathbf{P} \cdot \mathrm{E}(\mathbf{X} \mid \mathbf{b}) - \mathbf{Lb})'].
\end{aligned} \tag{2.17}$$

In the important special case where $n = p$ and $\mathrm{E}(\mathbf{X} \mid \mathbf{b}) = \mathbf{b}$ and $\mathbf{L} = \mathbf{P}$, the last term drops out and the expression is reduced to $\mathbf{L}(\mathrm{ECov}[\mathbf{X} \mid \mathbf{b}] - \mathrm{ECov}[\mathbf{b} \mid \mathbf{X}])\, \mathbf{L}'$.

Note that the adjustment matrix in (2.12), namely

$$\mathbf{J} = \mathbf{W}^{-1}\mathbf{L}'\mathbf{Q}^{-1}, \tag{2.18}$$

does not depend on the probability distribution of $(\mathbf{b}, \mathbf{X})$. As a consequence, one can calculate the constraining adjustment after the unconstrained Bayes estimator has been calculated, and without reference to the model used to derive that estimator. Unlike the unconstrained Bayes estimator, however, the optimal constrained estimator $\tilde{\mathbf{b}}$ depends on the risk weighting matrix $\mathbf{W}$ and, of course, the constraints.

An important special case is where $\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_p)$ is a diagonal matrix and there is only one constraint ($q = 1$). In that case one can write $\mathbf{L} = (l_1, \ldots, l_p)$ and $\mathbf{P} = (p_1, \ldots, p_n)$. Inserted into (2.12), this gives the following formula for the ith component of the optimal constrained estimator:

$$\begin{aligned}\tilde{b}_i &= \bar{b}_i + \frac{l_i}{w_i}\left(\sum_{j=1}^{p}\frac{l_j^2}{w_j}\right)^{-1}\left(\sum_{j=1}^{n} p_j X_j - \sum_{j=1}^{p} l_j \bar{b}_j\right)\\ &=: \bar{b}_i + \frac{l_i}{w_i}\left(\sum_{j=1}^{p}\frac{l_j^2}{w_j}\right)^{-1}\cdot\Delta,\end{aligned} \tag{2.19}$$

where Δ is the 'amount of misbalance' exhibited by the unconstrained Bayes estimator. The excess risk in this case becomes

$$r(\tilde{\mathbf{b}}) - r(\bar{\mathbf{b}}) = \left(\sum_{j=1}^{p}\frac{l_j^2}{w_j}\right)^{-1}\cdot\mathrm{E}(\Delta^2). \tag{2.20}$$

Neuhaus (1995) treated the case where, on top of linear constraints of the form (2.11), one imposes the additional constraint that the estimator $\breve{\mathbf{b}}$ be a linear function of the statistic $\mathbf{X}$ (i.e. a 'credibility' estimator). Using Lagrange minimisation, the resulting 'balanced credibility estimator' was shown to be of the same form as (2.12), with $\bar{\mathbf{b}}$ the credibility estimator rather than the Bayes estimator. Given the present result, a simple reasoning leading to that result goes as follows: since the risk $r(\breve{\mathbf{b}})$ of a linear estimator $\breve{\mathbf{b}}$ depends on the distribution of $(\mathbf{b}, \mathbf{X})$ only through its moments of first and second order, the balanced credibility estimator can only depend on those moments, too. But those moments could have been generated by a Gaussian distribution, in which case even the optimal constrained estimator is a linear function of $\mathbf{X}$. Being selected from a wider class of admissible estimators, the optimal constrained estimator in the Gaussian case must coincide with the balanced credibility estimator; which in turn coincides with the balanced credibility estimator in any other model that generates the same first and second order moments.

The excess risk (2.16) measures the average cost of constraining the estimator in the long run by repeated independent estimation situations. One could argue that the expectation should be dropped and that the pointwise increase of loss is what matters. If one takes this view, the pointwise increase at $\mathbf{X} = \mathbf{x}$ can be easily calculated by the formula

$$(\breve{\mathbf{b}}(\mathbf{x}) - \bar{\mathbf{b}}(\mathbf{x}))\,'\mathrm{W}(\breve{\mathbf{b}}(\mathbf{x}) - \bar{\mathbf{b}}(\mathbf{x})) = (\mathbf{Px} - \mathbf{L}\bar{\mathbf{b}}(\mathbf{x}))\,'\mathbf{Q}^{-1}(\mathbf{Px} - \mathbf{L}\bar{\mathbf{b}}(\mathbf{x})), \qquad (2.21)$$

being simply the distance between $\bar{\mathbf{b}}$ and its projection onto $f^{-1}\{g(\mathbf{x})\}$.

3. Example: Optimal constrained estimation of lognormal means

As we have seen, the optimal constrained estimator (2.12) always has a component that is linear in the statistic **X**. Let us now consider a model in which the optimal unconstrained estimator is non-linear in X, and quantify the excess risk generated by the constraint.

Assume that the portfolio under consideration consists of stochastically independent policies labelled by $i = 1, \ldots, p$. For policy no. i, assume we have observed claim amounts $X_{ij} : j = 1, \ldots, n_i$, with n_i fixed. Now assume that the X_{ij} are conditionally independent, given the value θ_i of a hidden random parameter Θ_i, and that under the same conditional distribution,

$$Y_{ij} := \log(X_{ij}) \sim \text{Normal } (\theta_i, \phi), \qquad (3.1)$$

with a fixed value of ϕ. Assume that the hidden risk parameters Θ_i are independent with

$$\Theta_i \sim \text{Normal}(\mu, \lambda). \qquad (3.2)$$

The properties of the lognormal distribution are summarised in e.g. Hogg & Klugman (1984). In particular, the conditional mean of X_{ij}, given Θ_i, is

$$b_i = \mathrm{E}[X_{ij} \mid \Theta_i] = \mathrm{e}^{\Theta_i + \frac{\phi}{2}}. \qquad (3.3)$$

Now assume that it is our intention to estimate the vector of lognormal means,

$$\mathbf{b} := (b_1, \ldots, b_p)' \qquad (3.4)$$

under the constraint that the weighted sum of the estimates must equal the sum of claims:

$$\sum_{i=1}^{p} n_i \breve{b}_i = \sum_{i=1}^{p} \sum_{j=1}^{n_i} X_{ij} = \sum_{i=1}^{p} n_i X_i, \qquad (3.5)$$

where $X_i := n_i^{-1} \sum_j X_{ij}$ is the average of claims against policy no i.

It is well known that the conditional distribution of Θ_i, given $X_{i1}, \ldots, X_{i,n_i}$, is again a normal with conditional mean

$$\mu_i = \frac{n_i \lambda}{n_i \lambda + \phi} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \log(X_{ij}) \right\} + \frac{\phi}{n_i \lambda + \phi} \mu =: z_i \bar{Y}_i + (1 - z_i)\mu \qquad (3.6)$$

and conditional variance

$$\lambda_i = \frac{\lambda\phi}{n_i\lambda + \phi} = (1 - z_i)\lambda. \tag{3.7}$$

The Bayes estimator of b_i is its conditional mean

$$\bar{b}_i = \mathrm{E}[b_i \mid X_{i1}, \ldots, X_{i,n_i}] = \mathrm{e}^{\mu_i + \frac{\lambda_i + \phi}{2}}, \tag{3.8}$$

and its conditional variance is

$$\mathrm{Var}[b_i \mid X_{i1}, \ldots, X_{i,n_i}] = \mathrm{e}^{2\mu_i + \phi + \lambda_i}(\mathrm{e}^{\lambda_i} - 1). \tag{3.9}$$

Using (3.9) and the marginal distribution of $\bar{Y}_i$, which is Normal$(\mu, \lambda + \phi/n_i)$, we find after some tedious manipulation the Bayes risk for estimating b_i:

$$\mathrm{EVar}[b_i \mid X_{i1}, \ldots, X_{i,n_i}] = \mathrm{e}^{2\mu + 2\lambda + \phi}(1 - \mathrm{e}^{-\lambda_i}). \tag{3.10}$$

Let us assume that the matrix $\mathbf{W}$ is diagonal. Inserting $l_i = p_i = n_i (i = 1, \ldots, p)$ and $\bar{b}_i$ given by (3.8) into (2.19), the optimal constrained estimator can be read off directly.

In order to calculate the excess risk using (2.16) and (2.17), we must also find

$$\mathrm{EVar}[X_i \mid b_i] = \frac{1}{n_i}\mathrm{e}^{2\mu + 2\lambda + \phi}(\mathrm{e}^{\phi} - 1). \tag{3.11}$$

Now inserting the expressions (3.11) and (3.10) into (2.17), we find the excess risk generated by the constraint:

$$r(\tilde{\mathbf{b}}) - r(\bar{\mathbf{b}}) = \left(\sum_{i=1}^{p} \frac{n_i^2}{w_i}\right)^{-1} \mathrm{e}^{2\mu + 2\lambda + \phi} \sum_{i=1}^{p} n_i[(\mathrm{e}^{\phi} - 1) - n_i(1 - \mathrm{e}^{-\lambda_i})]. \tag{3.12}$$

Using Jensen's inequality one can check that each of the summands on the right hand side is non-negative.

The weighted Bayes risk is

$$r(\bar{\mathbf{b}}) = \mathrm{e}^{2\mu + 2\lambda + \phi} \sum_{i=1}^{p} w_i(1 - \mathrm{e}^{-\lambda_i}). \tag{3.13}$$

In the case where $n_1 = \ldots = n_p = n$ and $\mathbf{W} = \mathbf{I}$, it is easy to calculate the relative excess risk,

$$\frac{r(\tilde{\mathbf{b}}) - r(\bar{\mathbf{b}})}{r(\bar{\mathbf{b}})} = \frac{1}{p}\left[\frac{1}{n} \cdot \frac{\mathrm{e}^{\phi} - 1}{1 - \mathrm{e}^{-\lambda_1}} - 1\right], \tag{3.14}$$

with $\lambda_1 = \lambda\phi/(n\lambda + \phi) = (1 - z_1)\lambda$. The relative excess risk can become arbitrarily large. One can also show that

$$\lim_{n\to\infty} \frac{r(\tilde{\mathbf{b}}) - r(\bar{\mathbf{b}})}{r(\bar{\mathbf{b}})} = \frac{1}{p}\left[\frac{e^{\phi} - 1}{\phi} - 1\right] > 0, \tag{3.15}$$

independent of λ. Thus constraints should never be applied uncritically, and special care must be taken when the original observations have a heavy-tailed distribution.

4. Example: Combining share valuations with market values

Consider an actuary who has been charged with an analysis of a portfolio of p shares. Assume that in addition to the market values at any time t, which we denote by

$$\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))', \tag{4.1}$$

the actuary has access to an individual valuation of the shares, based on an analysis of the economic fundamentals. Let us assume that the individually assessed share values reflect the analyst's conditional expectation, given all the information available to him, and denote the analyst's best bet by

$$\bar{\mathbf{b}}(t) = (\bar{b}_1(t), \ldots, \bar{b}_p(t))'. \tag{4.2}$$

Of course it is possible that $\bar{b}_i(t) = X_i(t)$ for the major stocks and those stocks that have not had the attention of the analyst.

In order not to stray too far from the market value of the shares, the actuary now wishes to ensure that at least the overall value of the shares coincides with their overall market value. Thus assume that the number of shares listed is

$$\mathbf{n}(t) = (n_1(t), \ldots, n_p(t))', \tag{4.3}$$

while the number of shares held by the company (or pension fund) is

$$\mathbf{m}(t) = (m_1(t), \ldots, m_p(t))', \tag{4.4}$$

Assume that $\mathbf{n}(t)$ and $\mathbf{m}(t)$ are linearly independent. We must also assume that shares have a common nominal value.

There could conceivably be two different constraints the actuary wishes to obey:

$$\begin{aligned} (1) &: \sum_{i=1}^{p} n_i(t)\breve{b}_i(t) = \sum_{i=1}^{p} n_i(t)X_i(t), \\ (2) &: \sum_{i=1}^{p} m_i(t)\breve{b}_i(t) = \sum_{i=1}^{p} m_i(t)X_i(t). \end{aligned} \tag{4.5}$$

The first constraint prevents any deviation from a market weighted index, while the second constraint ensures that the values used in the actuary's analysis add up to the total value the company must show in its books. There could be additional constraints, for example a constraint to prevent deviation from a major sub-index like the All Industrials.

The two constraints in (4.5) are formalised by the matrices

$$\mathbf{L}^{2xp}(t) = \mathbf{P}^{2xp}(t) = \begin{pmatrix} \mathbf{n}'\ (t) \\ \mathbf{m}'\ (t) \end{pmatrix}. \tag{4.6}$$

We suppress the argument (t) in the rest of this section.

The optimal adjustment to make to the analyst's set of estimates, follows directly from (2.12):

$$\tilde{\mathbf{b}} - \bar{\mathbf{b}} = \mathbf{W}^{-1}\mathbf{L}'\mathbf{Q}^{-1}\Delta, \tag{4.7}$$

where the vector

$$\Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} \sum_i n_i(X_i - \bar{b}_i) \\ \sum_i m_i(X_i - \bar{b}_i) \end{pmatrix} \tag{4.8}$$

contains the misbalance of the analyst's estimates against each of the two constraints.

If $\mathbf{W}$ is diagonal, one can derive the following expression for the adjustment to vector of share values:

$$\tilde{\mathbf{b}} - \bar{\mathbf{b}} = \left(\frac{\frac{n_i}{w_i}\left[\sum_j \frac{m_j^2}{w_j}\right]\Delta_1 + \frac{m_i}{w_i}\left[\sum_j \frac{n_j^2}{w_j}\right]\Delta_2 - \sum_j \frac{n_j m_j}{w_j}\left[\frac{m_i}{w_i}\Delta_1 + \frac{n_i}{w_i}\Delta_2\right]}{\left[\sum_j \frac{n_j^2}{w_j}\right]\cdot\left[\sum_j \frac{m_j^2}{w_j}\right] - \left[\sum_j \frac{m_j n_j}{w_j}\right]^2} \right)_{i=1,\ldots,p} \tag{4.9}$$

In particular if mean squared error is weighted by the number of shares listed ($w_i = n_i$), we obtain the simple adjustment

$$\tilde{\mathbf{b}} - \bar{\mathbf{b}} = \left(\frac{[\bar{s} - s_i]\delta_1 + [s_i - S]\delta_2}{\bar{s} - S} \right)_{i=1,\ldots,p}, \tag{4.10}$$

with $N := \Sigma_j n_j$ the total number of shares listed, $M := \Sigma_j m_j$ the total number of shares held, $s_i := m_i/n_i$ the stake held in stock i, $\bar{s} := M^{-1}\Sigma_j m_j s_j$ the average stake held and $S := M/N$ the overall stake in the stock market; $\delta_1 := \Delta_1/N$ and $\delta_2 := \Delta_2/M$ denote the relative (per share) deviations between the market values and the analyst's values.

From (4.10) one sees how the optimal adjustment depends on the relative misbalances and the company's relative exposure to the different stocks. In particular, if $\delta_2 > \delta_1$, then the adjustment to be made to share price no. i, is an increasing function of s_i, the stake held in stock no. i. Roughly speaking, $\delta_2 > \delta_1$ means that the market likes the portfolio held by the company better than the analyst; in that case it seems reasonable that the value of stocks of which the company has a major holding, should be assessed more highly than the value of stocks of which the company only has a small holding. The opposite holds if $\delta_2 < \delta_1$; if $\delta_2 = \delta_1 = \delta$, then each share price is adjusted with the same amount δ.

5. Balanced vs. homogeneous credibility estimation in the Bühlmann-Straub model

For the purpose of this section only, we revert to using the term 'balanced estimators', since the constraint applied here neatly fits in with the intuitive notion of balancing. Moreover, the estimators analysed in this section satisfy several different constraints, so that the simple term 'constrained estimator' without a number of qualifiers would be highly ambiguous.

Assume that the actuary is charged with estimating the pure premiums of n independent insurance policies. For policy no. i, what has been observed is a measure of exposure, denoted by p_i, and the total claims cost, denoted by S_i. The empirical pure premium per unit of exposure of policy no. i is then $X_i = S_i/p_i$.

Assume now that the probability distribution of X_i is governed by an unobserved random parameter θ_i coming from a distribution U, in such a way that

$$\begin{aligned} \mathrm{E}_{\theta_i}(X_i) &= b(\theta_i), \\ \mathrm{Var}_{\theta_i}(X_i) &= v(\theta_i)/p_i. \end{aligned} \tag{5.1}$$

Define the following structural parameters:

$$\beta = \mathrm{E}(b(\Theta)), \quad \phi = \mathrm{E}(v(\Theta)), \quad \lambda = \mathrm{Var}(b(\Theta)). \tag{5.2}$$

If a diagonal risk weighting matrix $\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_n)$ is used, the balanced credibility estimator under the constraint

$$\sum_{i=1}^{n} p_i \breve{b}_i = \sum_{i=1}^{n} p_i X_i \tag{5.3}$$

is given by

$$\tilde{b}_i = z_i X_i + (1 - z_i)\beta + \frac{p_i}{w_i}\left(\sum_{j=1}^{n} \frac{p_j^2}{w_j}\right)^{-1} \sum_{j=1}^{n} p_j(1 - z_j)(X_j - \beta), \tag{5.4}$$

$(i = 1, \ldots, n)$, and its risk is

$$r(\tilde{\mathbf{b}}) = \lambda \sum_{j=1}^{n} w_j(1 - z_j) + \phi\left(\sum_{j=1}^{n} \frac{p_j^2}{w_j}\right)^{-1} \sum_{j=1}^{n} p_j(1 - z_j), \tag{5.5}$$

where $z_j = \lambda p_j/(\lambda p_j + \phi)$ are the credibility factors. The results (5.4) and (5. 5) are proved in Neuhaus (1995).

Another linear estimator that is known to balance (i.e. satisfy the constraint (5.3)), is the optimal homogeneous and unbiased credibility estimator,

$$\hat{b}_i = z_i X_i + (1 - z_i)\left(\sum_{j=1}^{n} z_j\right)^{-1} \sum_{j=1}^{n} z_j X_j. \tag{5.6}$$

This fact has been noted by e.g. Gisler (1987).

Since the optimal homogeneous and unbiased estimator has two constraints to satisfy and just happens to be balanced as well, there is no prize for guessing that its risk will exceed that of the optimal balanced estimator. The question is just, by how much the risks differ.

Specialising equation (5.27) of Neuhaus (1995) or using the representation (5.6) directly, one easily shows that the risk of $\hat{\mathbf{b}}$ is

$$r(\hat{\mathbf{b}}) = \lambda \sum_{j=1}^{n} w_j(1 - z_j) + \lambda\left(\sum_{j=1}^{n} z_j\right)^{-1} \sum_{j=1}^{n} w_j(1 - z_j)^2. \tag{5.7}$$

After some manipulations one can write the difference in risk as

$$r(\hat{\mathbf{b}}) - r(\tilde{\mathbf{b}}) = \frac{\phi^2}{\lambda}\left[\left(\sum_{j=1}^{n} z_j\right)^{-1} \sum_{j=1}^{n} z_j \frac{z_j w_j}{p_j^2} - \left(\sum_{j=1}^{n} z_j \frac{p_j^2}{z_j w_j}\right)^{-1} \sum_{j=1}^{n} z_j\right] \geq 0. \tag{5.8}$$

The inequality is a consequence of Jensen's inequality applied to the convex function $x \longrightarrow x^{-1}$ $(x > 0)$, and will be strict unless all $z_j w_j / p_j^2$ are identical.

6. Optimal estimation under a penalty for misbalance

In the example of Section 3 we noted that the relative excess risk introduced by the constraint, can become arbitrarily large. Thus one must consider whether the benefit of constraining the estimator is worth the added risk.

A compromise approach would be not to enforce the constraint, but merely to 'encourage' it by a suitable modification of the risk function, so that the modified risk function reflects our preference of constrained estimators. This approach is similar to that taken in Whittaker-Henderson graduation, see e.g. Taylor (1992). It is also very similar to the smoothing approach proposed by Sundt (1992).

Let us therefore introduce the following, mixed risk function:

$$r_\alpha(\breve{\mathbf{b}}) = (1 - \alpha) \cdot \mathrm{E}(\breve{\mathbf{b}} - \mathbf{b})\,'\mathbf{W}(\breve{\mathbf{b}} - \mathbf{b}) + \alpha \cdot \mathrm{E}(\mathbf{L}\breve{\mathbf{b}} - \mathbf{PX})\,'\mathbf{V}(\mathbf{L}\breve{\mathbf{b}} - \mathbf{PX}), \tag{6.1}$$

with $\mathbf{W}$, $\mathbf{V}$ fixed, positive definite matrices, and $\alpha \in [0, 1)$ a parameter which quantifies the trade-off between estimator precision and estimator constraints.

One can write

$$r_\alpha(\breve{\mathbf{b}}) = \mathrm{EE}[(1 - \alpha) \cdot (\breve{\mathbf{b}} - \mathbf{b})\,'\mathbf{W}(\breve{\mathbf{b}} - \mathbf{b}) + \alpha \cdot (\mathbf{L}\breve{\mathbf{b}} - \mathbf{PX})\,'\mathbf{V}(\mathbf{L}\breve{\mathbf{b}} - \mathbf{PX}) \mid \mathbf{X}]. \tag{6.2}$$

Thus the optimal estimator may be determined pointwise for each possible realisation of $\mathbf{X} = \mathbf{x}$, and after having decomposed the first term in (6.2) as in (2.7), it is easy to verify that the optimal estimator is

$$\tilde{\mathbf{b}}_\alpha(\mathbf{x}) = ((1-\alpha)\mathbf{W} + \alpha\mathbf{L}'\mathbf{V}\mathbf{L})^{-1}[(1-\alpha)\mathbf{W}\bar{\mathbf{b}}(\mathbf{x}) + \alpha\mathbf{L}'\mathbf{V}\mathbf{P}\mathbf{x}]. \tag{6.3}$$

Using tedious but straightforward matrix transformations, we can derive an equivalent expression,

$$\tilde{\mathbf{b}}_\alpha(\mathbf{x}) = \bar{\mathbf{b}}(\mathbf{x}) + \mathbf{J}_\alpha(\mathbf{P}\mathbf{x} - \mathbf{L}\bar{\mathbf{b}}(\mathbf{x})), \tag{6.4}$$

with

$$\mathbf{J}_\alpha = \alpha\mathbf{W}^{-1}\mathbf{L}'((1-\alpha)\mathbf{V}^{-1} + \alpha\mathbf{Q})^{-1}. \tag{6.5}$$

It is plain to see that

$$\lim_{\alpha\uparrow 1} \tilde{\mathbf{b}}_\alpha(\mathbf{x}) = \tilde{\mathbf{b}}(\mathbf{x}), \tag{6.6}$$

i.e. the optimal estimator under the risk function r_α converges to the constrained estimator when the relative penalty for misbalance increases.

7. Conclusion

The results of the author's previous paper (Neuhaus, 1995) have been generalised.

Constrained estimators solve a practical problem faced by most actuaries and, as it turns out, the necessary adjustment is often very simple to compute. However, the warning about constraints creating an excess risk cannot be put too strongly; in unfavourable cases, an elaborate search for a realistic model and the optimal estimator may well have been in vain, if subsequent use of constraints greatly increases the risk of the estimator.

An interesting aspect concerns the use of constraints in empirical Bayes estimation and empirical credibility estimation. The normal procedure followed by actuaries is to estimate the distribution of $(\mathbf{b}, \mathbf{X})$ (or its first and second order moments if only a credibility estimator is sought), and then to act as if the estimated model was the true model. In that case the constraining adjustment is still appropriate because, as we have seen, it is independent of the model used to derive the Bayes (or credibility) estimator. However, the resulting constrained estimator will not be the optimal constrained estimator (or balanced credibility estimator), only an approximation of it.

The last argument may be extended to arbitrary estimators. The derivation in Section 2 rests fully on a pointwise minimum distance projection of the optimal estimator into the space of constrained estimators. Now if one redefines the optimisation problem from one of finding the optimal constrained estimator, to one of finding the constrained estimator with minimum distance from a given estimator, one still arrives at the same constraining adjustment. A similar argument is valid for the weighted estimator of Section 6.

ACKNOWLEDGEMENTS

This paper was written during the author's stay at the Laboratory of Actuarial Mathematics. Thanks go to the Laboratory staff for their hospitality and in particular to Ragnar Norberg for making my visit possible. Thanks also to Ragnar Norberg for the probing questions that lead to this paper, and to Thomas Mack (Münchener Rück) for posing the question that is answered in Section 5. Finally, thanks to two anonymous referees for their thoughtful comments.

References

HOGG, R.V. AND KLUGMAN, S.A. (1984) *Loss Distributions*. John Wiley & Sons.

GISLER, A. (1987) Einige Bemerkungen zum hierarchischen Credibility-Modell. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 1987, Heft 1: 91-98.

NEUHAUS, W. (1995) Balanced Credibility Estimation. *Applied Stochastic Models and Data Analysis* **11**, No. 3: 217-230.

SUNDT, B. (1992) On Greatest Accuracy Credibility with Limited Fluctuation. *Scandinavian Actuarial Journal* 1992, No. 2: 109-119.

TAYLOR, G.C. (1992) A Bayesian Interpretation of Whittaker-Henderson Graduation. *Insurance: Mathematics & Economics* **11**, No. 1: 7-16.

Address for correspondence
Walther Neuhaus,
UNI Storebrand ASA,
P.O. Box 1380
Vika, N-0114 Oslo,
Norway