CREDIBILITY IN THE REGRESSION CASE REVISITED
(A LATE TRIBUTE TO CHARLES A. HACHEMEISTER)

H. BÜHLMANN
ETH Zurich

A. GISLER
Winterhur-Versicherungen

ABSTRACT

Many authors have observed that Hachemeisters Regression Model for Credibility – if applied to simple linear regression – leads to unsatisfactory credibility matrices: they typically 'mix up' the regression parameters and in particular lead to regression lines that seem 'out of range' compared with both individual and collective regression lines. We propose to amend these shortcomings by an appropriate definition of the regression parameters:
– intercept
– slope
Contrary to standard practice the intercept should however not be defined as the value at time zero but as the value of the regression line at the barycenter of time. With these definitions regression parameters which are uncorrelated in the collective can be estimated separately by standard one dimensional credibility techniques.

A similar convenient reparametrization can also be achieved in the general regression case. The good choice for the regression parameters is such as to turn the design matrix into an array with orthogonal columns.

1. THE GENERAL MODEL

In his pioneering paper presented at the Berkeley Credibility Conference 1974, Charlie Hachemeister introduced the following General Regression Case Credibility Model:
a) Description of individual risk r:
risk quality: θ_r
observations (random vector):

$$\begin{pmatrix} X_{1r} \\ X_{2r} \\ \vdots \\ X_{nr} \end{pmatrix} = X_r$$

with distribution $dP(X_r/\theta_r)$ and where X_{ir} = observation of risk r at time i

b) Description of collective:
$\{\theta_r;(r = 1, 2, \ldots, N)\}$ are i.i.d. with structure function $U(\theta)$

We are interested in the (unknown)
individually correct pure premiums $\mu_i(\theta_r) = E[X_{ir}/\theta_r]$ $(i = 1, 2, \ldots, n)$

$$\begin{pmatrix} \mu_1(\theta_r) \\ \mu_2(\theta_r) \\ \vdots \\ \mu_n(\theta_r) \end{pmatrix} = \mu(\theta_r) \quad \text{where } \mu_i(\theta_r) = \text{individual pure premium at time } i$$

and we suppose that these individual pure premiums follow a regression pattern

R) $$\mu(\theta_r) = Y_r\beta(\theta_r),$$

where $\mu(\theta_r) \sim n$-vector, $\beta(\theta_r) \sim p$-vector and $Y_r \sim n * p$-matrix (= design matrix).

Remark:

The model is usually applied for $p < n$ and maximal rank of Y_r; in practice p is much smaller than n (e.g. $p = 2$).

The goal is to have credibility estimator $\hat{\beta}(\theta_r)$ for $\beta(\theta_r)$
which by linearity leads to the credibility estimator $\hat{\mu}(\theta_r)$ for $\mu(\theta_r)$.

2. THE ESTIMATION PROBLEM AND ITS RELEVANT PARAMETERS AND SOLUTION (GENERAL CASE)

We look for

$$\hat{\beta}(\theta_r) = \mathbf{a} + AX_r$$
$$\mathbf{a} \sim p - \text{vector}$$
$$A \sim p * n \text{ matrix}$$

The following quantities are the "relevant parameters" for finding this estimator

$$E[Cov[X_r, X'_r / \theta_r]] = \Phi_r \quad \Phi_r \sim n * n \text{ matrix (regular)} \tag{1}$$
$$Cov[\beta(\theta_r), \beta'(\theta_r)] = \Lambda \quad \Lambda \sim p * p \text{ matrix (regular)} \tag{2}$$
$$E[\beta(\theta_r)] = \mathbf{b} \quad \mathbf{b} \sim p - \text{vector} \tag{3}$$

We find the credibility formula

$$\hat{\beta}(\theta_r) = (I - Z_r)\mathbf{b} + Z_r\mathbf{b}_r^X \tag{4}$$

where

$$Z_r = (I - W_r^{-1}\Lambda^{-1})^{-1} = (W_r + \Lambda^{-1})^{-1} W_r = \Lambda(\Lambda + W_r^{-1})^{-1} \quad \sim \text{credibility matrix}(p * p) \tag{5}$$

$$W_r = Y_r' \Phi_r^{-1} Y_r \quad \sim \text{auxiliary matrix} \quad (p * p) \tag{6}$$

$$\mathbf{b}_r^X = W_r^{-1} Y_r' \Phi_r^{-1} X_r \quad \sim \text{individual estimate} \quad (p * 1) \tag{7}$$

Discussion:

The generality under which formula (4) can be proved is impressiv, but this generality is also its weakness. Only by specialisation it is possible to understand how the formula can be used for practical applications. Following the route of Hachemeisters original paper we hence use it now for the special case of simple linear regression.

3. SIMPLE LINEAR REGRESSION

Let

$$Y_r = Y = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \\ 1 & n \end{pmatrix}$$

and

$$\beta(\theta_r) = \begin{pmatrix} \beta_0(\theta_r) \\ \beta_1(\theta_r) \end{pmatrix}$$

hence R) becomes

$$\mu_i(\theta_r) = \beta_0(\theta_r) + i \cdot \beta_1(\theta_r) \tag{8}$$

which is one of the most frequently applied regression cases. Assume further that Φ_r is diagonal, i.e. that observations X_{ir}, X_{jr} given θ_r are uncorrelated for $i \neq j$.
To simplify notation, we drop in the following the index r, i.e. we write Φ instead of Φ_r, W instead of W_r and Z instead of Z_r.

Hence

$$\Phi = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{pmatrix} \tag{9}$$

e.g. $\sigma_i^2 = \dfrac{\sigma_2}{V_i}$; V_i = "volume" of observation at time i.

Let $$\Lambda = \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{pmatrix} \quad \tau_{01} = \tau_{10}$$

https://doi.org/10.2143/AST.27.1.563207 Published online by Cambridge University Press

We find

$$W = Y'\Phi^{-1}Y = \begin{pmatrix} \sum_{k=1}^{n}\frac{1}{\sigma_k^2} & \sum_{k=1}^{n}\frac{k}{\sigma_k^2} \\ \sum_{k=1}^{n}\frac{k}{\sigma_k^2} & \sum_{k=1}^{n}\frac{k^2}{\sigma_k^2} \end{pmatrix} \tag{10}$$

It is convenient to write

$$\sigma_k^2 = \frac{\sigma^2}{V_k}, \quad V. = \sum_{k=1}^{n} V_k$$

(which is always possible for diagonal Φ). Hence we have

$$W = \frac{V.}{\sigma^2}\begin{pmatrix} \sum_k \frac{V_k}{V.} & \sum_k k\frac{V_k}{V.} \\ \sum_k k\frac{V_k}{V.} & \sum_k k^2\frac{V_k}{V.} \end{pmatrix}$$

Think of $\frac{V_k}{V.}$ as <u>sampling weights</u>, then we have

$$W = \frac{V.}{\sigma^2}\begin{pmatrix} 1 & E^{(s)}[k] \\ E^{(s)}[k] & E^{(s)}[k^2] \end{pmatrix} \tag{11}$$

where $E^{(s)}$, $Var^{(s)}$ denote the <u>moments with respect to the sampling distribution</u>.

One then also finds (see (7))

$$\begin{aligned} \mathbf{b}_r^X &= W^{-1}Y'\Phi^{-1}X_r \\ &= \frac{1}{Var^{(s)}[k]}\begin{pmatrix} E^{(s)}[k^2]\cdot E^{s}[X_{kr}] - E^{(s)}[k]\ E^{(s)}[kX_{kr}] \\ E^{s}[kX_{kr}] - E^{(s)}[k]\ E^{(s)}[X_{kr}] \end{pmatrix} \end{aligned} \tag{12}$$

where $\quad E^{(s)}[kX_{kr}] = \sum_k \frac{V_k}{V.}kX_{kr}, E^{(s)}[X_{kr}] = \sum_k \frac{V_k}{V.}X_{kr}$

Remark:

It is instructive to verify by direct calculation that the values given by (12) to b_{0r}^X, b_{1r}^X are identical with those obtained from

$$\sum_{k=1}^{n} V_k(X_{kr} - b_{0r}^X - kb_{1r}^X)^2 = \min!$$

The calculations to obtain the credibility matrix Z (see (5)) are as follows:

$$\Lambda^{-1} = \frac{1}{\tau_0^2\tau_1^2 - \tau_{01}^2}\begin{pmatrix} \tau_1^2 & -\tau_{01} \\ -\tau_{01} & \tau_0^2 \end{pmatrix} = \begin{pmatrix} \rho_0^2 & +\rho_{01} \\ +\rho_{01} & \rho_1^2 \end{pmatrix}$$

Abbreviate

$$\begin{aligned} \rho_0^2 \cdot \frac{\sigma^2}{V.} &= h_0 \\ \rho_1^2 \cdot \frac{\sigma^2}{V.} &= h_1 \\ \rho_{01} \cdot \frac{\sigma^2}{V.} &= h_{01} \end{aligned} \tag{13}$$

Hence

$$W + \Lambda^{-1} = \frac{V.}{\sigma^2} \begin{pmatrix} 1 + h_0 & E^{(s)}[k] + h_{01} \\ E^{(s)}[k] + h_{01} & E^{(s)}[k^2] + h_1 \end{pmatrix}$$

$$(W + \Lambda^{-1})^{-1} = \frac{\sigma^2}{V.} \underbrace{\frac{1}{(1+h_0)(E^{(s)}[k^2]+h_1) - (E^{(s)}[k]+h_{01})^2}}_{N} \begin{pmatrix} E^{(s)}[k^2]+h_1 & -(E^{(s)}[k]+h_{01}) \\ -(E^{(s)}[k]+h_{01}) & 1+h_0 \end{pmatrix}$$

$$Z = (W + \Lambda^{-1})^{-1} \cdot W \tag{14}$$

$$Z = \frac{1}{N} \begin{pmatrix} Var^{(s)}[k] + h_1 - h_{01}E^{(s)}[k] & E^{(s)}[k]h_1 - E^{(s)}[k^2]h_{01} \\ h_0 E^{(s)}[k] - h_{01} & Var^{(s)}[k] + h_0 E^{(s)}[k^2] - h_{01}E^{(s)}[k] \end{pmatrix}$$

Discussion:

The credibility matrix obtained is <u>not</u> satisfactory from a practical point of view:

a) individual weights are <u>not</u> always between zero and one:

b) both intercept $\hat{\beta}_0(\theta_r)$ of the credibility line and slope $\hat{\beta}_1(\theta_r)$ of the credibility line may not lie between intercept and slope of individual line and collective line.

Numerical examples:

$n = 5 \quad V_k \equiv 1$

collective regression line: $b_0 = 100 \quad b_1 = 10$

individual regression line: $b_0^X = 70 \quad b_1^X = 7$

Example 1: $\sigma = 20 \quad \tau_0 = 10 \quad \tau_1 = 5 \quad \tau_{10} = 0$

resulting credibility line: $\hat{\beta}_0(\theta_r) = 88.8 \quad \hat{\beta}_1(\theta_r) = 3.7$

Example 2: $\sigma = 20 \quad \tau_0 = 100'000 \quad \tau_1 = 5 \quad \tau_{10} = 0$

resulting credibility line: $\hat{\beta}_0(\theta_r) = 64.5 \quad \hat{\beta}_1(\theta_r) = 8.8$

Example 3: $\sigma = 20 \quad \tau_0 = 10 \quad \tau_1 = 100'000 \quad \tau_{10} = 0$

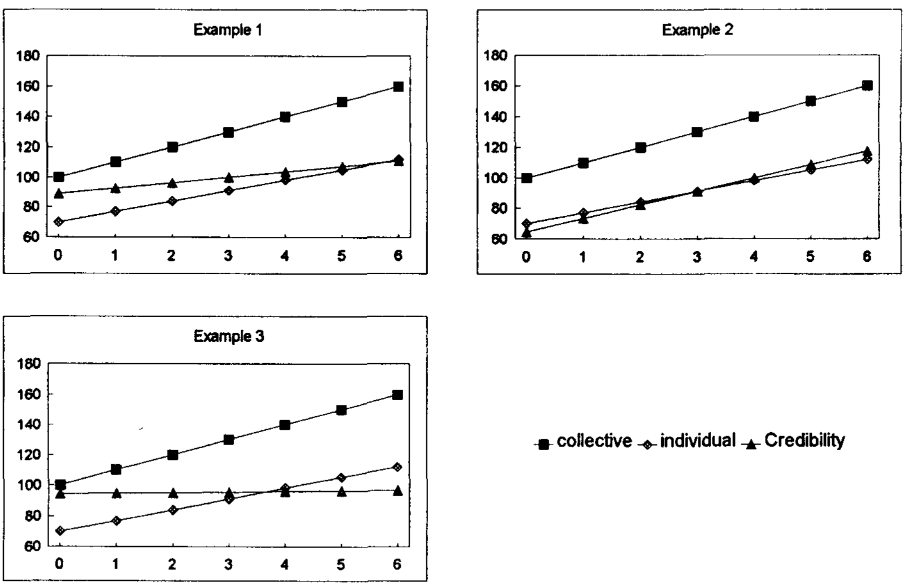
resulting credibility line: $\hat{\beta}_0(\theta_r) = 94.7 \quad \hat{\beta}_1(\theta_r) = 0.3$

https://doi.org/10.2143/AST.27.1.563207 Published online by Cambridge University Press

Comments:

In none of the 3 examples do both, intercept <u>and</u> slope of the credibility line, lie between the collective and the individual values. In example 2 there is a great prior uncertainty about the intercept (τ_0 very big). One would expect that the credibility estimator gives full weight to the intercept of the individual regression line and that $\hat{\beta}_0(\theta_r)$ nearly coincides with b_0^X. But $\hat{\beta}_0(\theta_r)$ is even smaller than b_0 and b_0^X. In example 3 there is a great prior uncertainty about the slope and one would expect, that $\hat{\beta}_1(\theta_r) \cong b_1^X$. But $\hat{\beta}_1(\theta_r)$ is much smaller than b_1 and b_1^X.

For this reason many actuaries have either considered Hachemeisters regression model as not usable or have tried to impose artificially additional constraints (e.g. De Vylder (1981) or De Vylder (1985)). Dannenburg (1996) discusses the effects of such constraints and shows that they have serious drawbacks. This paper shows that by an appropriate reparametrization the defects of the Hachemeister model can be made to disappear and that hence no additional constraints are needed.



4. SIMPLE LINEAR REGRESSION WITH BARYCENTRIC INTERCEPT

The idea, that choosing the time scale in such a way as to have the intercept at the barycenter of time, is already mentioned in Hachemeisters paper, although it is then not used to make the appropriate model assumptions. Choosing the intercept at the barycenter of the time scale means formally that our design matrix is chosen as

$$Y = \begin{pmatrix} 1 & 1 - E^{(s)}[k] \\ 1 & 2 - E^{(s)}[k] \\ \vdots & \\ 1 & n - E^{(s)}[k] \end{pmatrix}$$

Remark:

It is well known, that any linear transformation of the time scale (or more generally of the covariates) does not change the credibility estimates. But what we do in the following changes the original model by assuming that the matrix Λ is now the covariance matrix of the 'new' vector $\beta(\theta_r), \beta_0(\theta_r)$ now being the intercept at the barycenter of time instead of the intercept at the time zero.
In our general formulae obtained in section 3 we have to replace

$$E^{(s)}[k] \leftarrow 0 \quad E^{(s)}[k^2] \leftarrow Var^{(s)}[k]$$

It is also important that sample variances and covariances are <u>not</u> changed by this shift of time scale.

We immediately obtain

$$b_{0r}^x = E^{(s)}[X_{kr}]$$
$$b_{1r}^x = \frac{Cov^{(s)}(k, X_{kr})}{Var(s)[k]} \qquad (12_{bar})$$

and

$$Z = \frac{1}{(1+h_0)(Var^{(s)}[k]+h_1) - h_{01}^2} \begin{pmatrix} Var^{(s)}[k]+h_1 & -Var^{(s)}[k]h_{01} \\ -h_{01} & Var^{(s)}[k](1+h_0) \end{pmatrix} \qquad (14_{bar})$$

These formulae are now becoming very well understandable, in particular the crosseffect between the credibility formulae for intercept and slope is only due to their correlation in the collective (off diagonal elements in the matrix Λ). In case of no <u>correlation between regression parameters in the collective</u> we have

$$Z = \frac{1}{(1+h_0)(Var^{(s)}[k]+h_1)} \begin{pmatrix} Var^{(s)}[k]+h_1 & 0 \\ 0 & Var^{(s)}[k](1+h_0) \end{pmatrix} \qquad (14_{sep})$$

which separates our credibility matrix into two separate one-dimensional credibility formulae with credibility weights

$$Z_{11} = \frac{1}{1+h_0} = \frac{1}{1+\frac{\sigma^2}{\tau_0^2 V.}} = \frac{V.}{V.+\frac{\sigma^2}{\tau_0^2}} \qquad (15)$$

$$Z_{22} = \frac{Var^{(s)}[k]}{Var^{(s)}[k]+h_1} = \frac{Var^{(s)}[k]}{Var^{(s)}[k]+\frac{\sigma^2}{\tau_1^2 V.}} = \frac{V.Var^{(s)}[k]}{V.Var^{(s)}[k]+\frac{\sigma^2}{\tau_1^2}}$$

Remark:

Observe the classical form of the credibility weights in (15) with volumes $V.$ for Z_{11} and $V.Var^{(s)}[k]$ for Z_{22}.

Numerical examples

The model assumptions of the following three examples numbered 4 – 6 are exactly the same as in the examples numbered 1 – 3 of the previous section with the only difference that the first element of the vector $\beta(\theta_r)$ now represents the intercept at the barycenter. Thus we have:

collective regression line: $b_0 = 130 \quad b_1 = 10$

individual regression line: $b_0^X = 91 \quad b_1^X = 7$

The resulting credibility lines are:

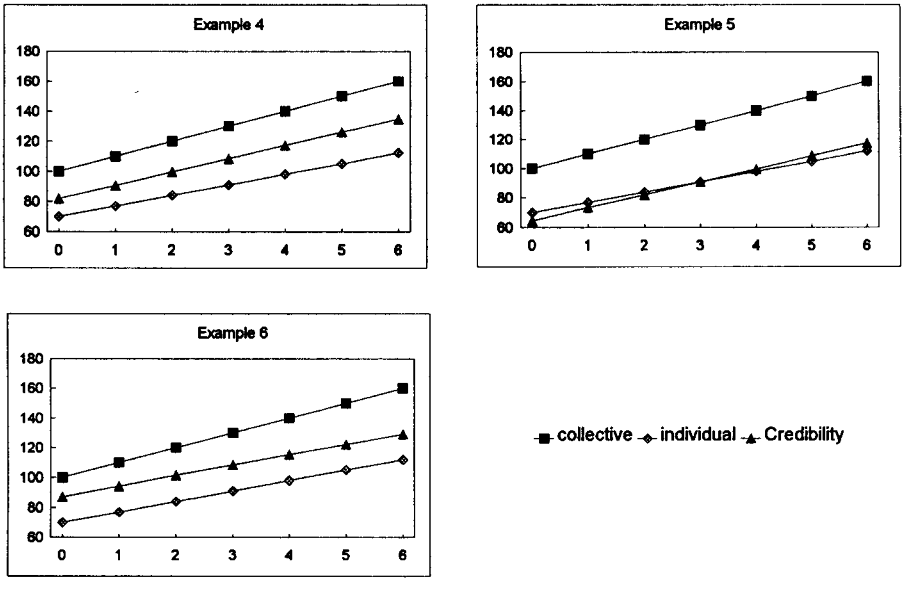
Example 4: $\hat{\beta}_0(\theta_r) = 108.3 \quad \hat{\beta}_1(\theta_r) = 8.8$

Example 5: $\hat{\beta}_0(\theta_r) = 91.0 \quad \hat{\beta}_1(\theta_r) = 8.8$

Example 6: $\hat{\beta}_0(\theta_r) = 108.3 \quad \hat{\beta}_1(\theta_r) = 7.0$

Comments:

Intercept and slope of the credibility lines are always lying between the values of the individual and of the collective regression line. In example 5 (respectively in example 6) the intercept $\hat{\beta}_0(\theta_r)$ (respectively the slope $\hat{\beta}_1(\theta_r)$) coincides with b_0^X (resp. b_1^X). It is also interesting to note that the credibility line of example 5 is exactly the same as the one of example 2.



5. HOW TO CHOOSE THE BARYCENTER?

Unfortunately the barycenter for each risk is shifting depending on the individual sampling distribution. There is usually no way to bring – simultaneously for all risks – the matrices *Y, W, Z* into the convenient form as discussed in the last section. This discussion however suggests that the most reasonable parametrization is the one using the intercept at the barycenter of the collective. This has two advantages: it is the point to which individual barycenters are (in the sum of least square sense) closest and the orthogonality property of parameters still holds for the collective.
In the following we work with this parametrization and assume that the regression parameters in this parametrization are uncorrelated.

Hence we work from now on with the regression line

$$\alpha_0(\theta_r)+(k-K)\alpha_1(\theta_r),$$

where K is the barycenter of the collective i.e. $K=\sum_{i=1}^{n}\frac{V_{i.}}{V_{..}}i.$

We assume also that the collective parameters are uncorrelated, i.e.

$$\Lambda^{(\alpha)}=\begin{pmatrix}\tau_0^2 & 0\\ 0 & \tau_1^2\end{pmatrix}$$

If we shift to the individual barycenter $E^{(s)}[k]$ we obtain the line:

$$\beta_0(\theta_r)+(k-E^{(s)}[k])\beta_1(\theta_r)$$

Hence

$$\beta_1(\theta_r)=\alpha_1(\theta_r)$$

$$\beta_0(\theta_r)=\alpha_0(\theta_r)+\alpha_1(\theta_r)\overbrace{(E^{(s)}[k]-K)}^{\Delta} \tag{16}$$

$$\Lambda(\beta)=\begin{pmatrix}\tau_0^2+\tau_1^2(E^{(s)}[k]-K)^2 & \tau_1^2(E^{(s)}[k]-K)\\ \tau_1^2(E^{(s)}[k]-K) & \tau_1^2\end{pmatrix}=\begin{pmatrix}\tau_0^2+\Delta^2\tau_1^2 & \Delta\tau_1^2\\ \Delta\tau_1^2 & \tau_1^2\end{pmatrix}$$

For the β-line we have further

$$\rho_0^2=\frac{1}{\tau_0^2};\qquad h_0^{(\beta)}=\frac{\sigma^2}{\tau_0^2\cdot V.}=h_0^{(\alpha)}$$

$$\rho_1^2=\frac{1}{\tau_1^2}+\Delta^2\frac{1}{\tau_0^2};\quad h_1^{(\beta)}=\frac{\sigma^2}{\tau_1^2\cdot V.}+\Delta^2\frac{\sigma^2}{\tau_0^2\cdot V.}=h_1^{(\alpha)}+\Delta^2 h_0^{(\alpha)}$$

$$\rho_{01}=-\Delta\frac{1}{\tau_0^2};\qquad h_{01}^{(\beta)}=\Delta\frac{\sigma^2}{\tau_0^2\cdot V.}=-\Delta h_0^{(\alpha)}$$

https://doi.org/10.2143/AST.27.1.563207 Published online by Cambridge University Press

Formula (14_{bar}) then leads to

$$Z = \frac{1}{(Var^{(s)}[k] + h_1^{(\alpha)})(1 + h_0^{(\alpha)}) + \Delta^2 h_0^{(\alpha)}} \cdot \begin{pmatrix} Var^{(s)}[k] + h_1^{(\alpha)} + \Delta^2 h_0^{(\alpha)} & Var^{(s)}[k] \cdot \Delta h_0^{(\alpha)} \\ \Delta h_0^{(\alpha)} & Var^{(s)}[k](1 + h_0^{(\alpha)}) \end{pmatrix} \qquad (17)$$

Remarks:

a) Obviously formula (17) is not as attractive as (14_{sep}) but the two are similar

b) In the case Δ small (or $h_0^{(\alpha)}$ small) the numbers differ by little. The case of Δ small is very often encountered in practice. For instance in the data used by Hachemeister with 5 groups (states) and an observation period of 12 time units (quarters), $|\Delta|_{max}$ is only 0.17 time units.

c) An interesting case is $h_0^{(\alpha)} = 0 \Leftrightarrow \tau_0^2 = \infty$ (great prior uncertainty about the intercept). Then for arbitrary Δ we have

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{Var^{(s)}[k]}{Var^{(s)}[k] + h_1^{(\alpha)}} \end{pmatrix}$$

This case is also studied by Hachemeister under the heading "mixed model". It also explains why the credibility lines coincide in our numerical examples 2 and 5.

6. BARYCENTRIC CREDIBILITY – GENERAL CASE

6.1 Motivation

The basic idea which we have pursued in the case of simple linear regression can be summarized as follows:

Step 1: Reparametrize the regression problem conveniently

Step 2: In the reparametrized model assume that regression parameters are uncorrelated.

It turns out that for simple linear regression the good choice of the parameters is as follows:

– intercept at barycenter of time

– slope

You should observe that for this choice the design matrix

$$Y = \begin{pmatrix} 1 & 1 - E^{(s)}[k] \\ 1 & 2 - E^{(s)}[k] \\ \vdots & \vdots \\ 1 & n - E^{(s)}[k] \end{pmatrix}$$

has two orthogonal columns (using the weights of the sampling distribution). This is the clue for the general regression case. The good choice of the regression parameters is such as to render the design matrix into an array with orthogonal columns.

6.2 The Barycentric Model

Let
$$Y = \begin{pmatrix} Y_{11} & Y_{12} & Y_{1p} \\ Y_{21} & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ Y_{n1} & Y_{n2} & Y_{np} \end{pmatrix}$$

and assume volumes $V_1, V_2, \ldots, V_n$ and let be $V. = \sum_{k=1}^{n} V_k$.

We think of column j in Y as a random variable Y_j which assumes Y_{jk} with sampling weight $\frac{V_k}{V.}$ in short $P^{(s)}[Y_j = Y_{jk}] = \frac{V_k}{V.}$ where $P^{(s)}$ stands for the sampling distribution. As in the case of simple linear regression it turns out that also in the general case this sampling distribution allows a concise and convenient notation.
We have from (9)

$$\Phi^{-1} = \frac{V.}{\sigma_2} \begin{pmatrix} \frac{V_1}{V.} & & \\ & \frac{V_2}{V.} & \\ & & \frac{V_n}{V.} \end{pmatrix}$$

and from (10)
$$W = Y' \Phi^{-1} Y = (w_{ij})$$

where
$$w_{ij} = \frac{V.}{\sigma_2} E^{(s)}[Y_i \cdot Y_j].$$

Under the barycentric condition we find

$$W = \frac{V.}{\sigma^2} \begin{pmatrix} E^{(s)}[Y_1^2] & & 0 \\ & E^{(s)}[Y_2^2] & \\ 0 & & E^{(s)}[Y_p^2] \end{pmatrix} \tag{18}$$

i.e. a matrix of diagonal form.

Assuming non-correlation for the corresponding parametrization we have

$$\Lambda = \begin{pmatrix} \tau_1^2 & & 0 \\ & \tau_2^2 & \\ 0 & & \tau_p^2 \end{pmatrix} \quad \Lambda^{-1} = \frac{V.}{\sigma^2} \begin{pmatrix} h_1 & & 0 \\ & h_2 & \\ 0 & & h_p \end{pmatrix}$$

with $$h_j = \frac{1}{\tau_j^2} \cdot \frac{\sigma^2}{V.}$$

Hence

$$(W + \Lambda^{-1}) = \frac{V.}{\sigma^2} \begin{pmatrix} E^{(s)}[Y_1^2] + h_1 & & 0 \\ & E^{(s)}[Y_2^2] + h_2 & \\ 0 & & E^{(s)}[Y_p^2] + h_p \end{pmatrix}$$

and finally

$$Z = (W + \Lambda^{-1})W = \begin{pmatrix} \frac{E^{(s)}[Y_1^2]}{E^{(s)}[Y_1^2] + h_1} & 0 \\ 0 & \frac{E^{(s)}[Y_p^2]}{E^{(s)}[Y_p^2] + h_p} \end{pmatrix} \tag{19}$$

(19) shows that our credibility matrix is of diagonal form. Hence the multidimensional credibility formula breaks down into p one dimensional formulae with credibility weights:

$$Z_{jj} = \frac{V.E^{(s)}[Y_j^2]}{V.E^{(s)}[Y_j^2] + \frac{\sigma^2}{\tau_j^2}} \tag{20}$$

Observe the "volume" $V.E^{(s)}[Y_j^2]$ for the j-th component.

6.3 The Summary Statistics for the Barycentric Model

From (7) we have

$$\mathbf{b}_r^x = W^{-1}Y^{-1}\Phi^{-1}\mathbf{X}_r = C\mathbf{X}_r$$

where the elements of C are

$$c_{ij} = \frac{1}{E^{(s)}[Y_i^2]} \cdot Y_{ij} \frac{V_j}{V.} \tag{21}$$

hence $$b_{ir}^x = \frac{1}{E^{(s)}[Y_i^2]} \sum_{j=1}^{n} X_{ij} Y_{jr} \frac{V_j}{V.}$$

or

$$b_{ir}^{x} = \frac{E^{(s)}[Y_i \cdot X_r]}{E^{(s)}[Y_i^2]} \quad i = 1, 2, \ldots p \tag{22}$$

6.4 How to find the Barycentric Reparametrization

We start with the design matrix
Y and its column vectors $Y_1, Y_2, \ldots, Y_p$
and want to find the new design matrix
Y^* with orthogonal column vectors $Y_1^*, Y_2^*, \ldots, Y_p^*$
The construction of the vectors Y_k^* is obtained recursively:

i) Start with $Y_1^* = Y_1$
ii) If you have constructed $Y_1^*, Y_2^*, \ldots, Y_{k-1}^*$, you find Y_k^* as follows
   a) Solve $E^{(s)}[(Y_k - a_1^* Y_1^* - a_2^* Y_2^* - \ldots - a_{k-1}^* Y_{k-1}^*)^2] = \min!$
      over all values of $a_1, a_2, \ldots, a_{k-1}$
   b) Define $Y_k^* := Y_k - a_1^* Y_1^* - a_2^* Y_2^* - \ldots - a_{k-1}^* Y_{k-1}^*$

Remarks:

i) obviously this leads to Y_k^* such that

$$E^{(s)}[Y_k^* \cdot Y_l^*] = 0 \quad \text{for all} \quad l < k \tag{23}$$

ii) The procedure of orthogonalisation is called weighted Gram-Schmitt in Numerical Analysis.
iii) The result of this procedure depends on the order of the colums of the original matrix. Hence there might be several feasible solutions.

With the new design matrix Y^* we can now also find the new parameters
$\beta_j^*(\theta_r) \quad j = 1, 2, \ldots p$. The regression equation becomes

R) $$\mu(\theta_r) = Y^* \beta^*(\theta_r)$$

which reads componentwise

$$\mu_j(\theta_r) = \sum_{j=1}^{p} Y_{ij}^* \beta_j^*(\theta_r).$$

Multiply both sides by $Y_{ik}^* \cdot \frac{V_i}{V_\cdot}$ and sum over i

$$\sum_{i=1}^{n} Y_{ik}^* \mu_i(\theta_r) \cdot \frac{V_i}{V_\cdot} = \sum_{j=1}^{p} \sum_{i=1}^{n} Y_{ik}^* Y_{ij}^* \beta_j^*(\theta_r) \cdot \frac{V_i}{V_\cdot}$$

leading to

$$E^{(s)}[Y_k^* \mu(\theta_r)] = E^{(s)}[(Y_k^*)^2] \cdot \beta_k^*(\theta_r) \tag{24}$$

https://doi.org/10.2143/AST.27.1.563207 Published online by Cambridge University Press

where, on the right hand side, we have used the orthogonality of Y_k^* and Y_j^* for $j \neq k$.
Hence

$$\beta_k^*(\theta_r) = \frac{E^{(s)}[Y_k^* \mu(\theta_r)]}{E^{(s)}[(Y_k^*)^2]} \quad k = 1,2,\ldots,p \tag{25}$$

which defines our new parameters in the barycentric model.
You should observe that this transformation of the regression parameters $\beta_j(\theta_r)$ may lead to new parameters $\beta_j^*(\theta_r)$ which are sometimes difficult to interprete. In each application one has therefore to decide whether the orthogonality property of the design matrix or the interpretability of the regression parameters is more important.
Luckily – as we have seen – there is no problem with the interpretation in the case of simple linear regression and interpretability is also not decisive if we are interested in prediction only.

6.5 An example

Suppose that we want to model $\mu_k(\theta_r)$ as depending on time in a quadratic manner, *i.e.*

$$\mu_k(\theta_r) = \beta_0(\theta_r) + k\beta_1(\theta_r) + k^2\beta_2(\theta_r)$$

Our design matrix is hence of the following form

$$Y = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \vdots & \vdots & \vdots \\ 1 & k & k^2 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{pmatrix}$$

Let us construct the design matrix Y^* with orthogonal columns.
Following the procedure as outlined in 6.4 we obviously have for the first two columns those obtained in the case of simple linear regression (measuring time from its barycenter) and we only have to construct Y_3^*
Formally:

$$Y^* = \begin{pmatrix} 1 & 1 - E^{(s)}[k] & Y_{13}^* \\ 1 & 2 - E^{(s)}[k] & Y_{23}^* \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & n - E^{(s)}[k] & Y_{n3}^* \end{pmatrix}$$

To find Y_3^* we must solve

$$\sum_{k=1}^{n}\left(k^2 - a_1^* - a_2^*(k - E^{(s)}[k])\right)^2 \cdot \frac{V_k}{V.} = \min!$$

Using relation (23) we obtian

$$a_1^* = E^{(s)}[k^2]$$
$$a_2^* = \frac{E^{(s)}[k^2(k - E^{(s)}[k])]}{Var^{(s)}[k]}$$

Hence we get

$$Y_{k3}^* = k^2 - E^{(s)}[k^2] - \frac{E^{(s)}[k^2(k - E^{(s)}[k])]}{Var^{(s)}[k]}(k - E^{(s)}[k]) \quad k = 1,2,\ldots,n \qquad (26)$$

and from

$$\mu_i(\theta) = \sum\nolimits_{j=1}^{3} Y_{ij}^* \beta_j^*(\theta_r) \qquad \text{R)}$$

we get both

– the interpretation of $\beta_j^*(\theta_r)$ (use (25))

– the prediction $\hat{\mu}_i(\theta) = \sum\nolimits_{j=1}^{3} Y_{ij}^* \hat{\beta}_j^*(\theta_r)$

where $\hat{\beta}_j^*(\theta_r)$ is the credibility estimator. Due to orthogonality of Y^* it can be obtained componentwise.

7. FINAL REMARKS

Our whole discussion of the general case is based on a particular fixed sampling distribution. As this distribution typically varies from risk to risk Y^*, β^* and Z^* depend on the risk r and we cannot achieve orthogonality of Y^* simultaneously for all risks r. This is the problem which we have already discussed in section 5. The observations made there apply also to the general case and the basic lesson is the same. You should construct the orthogonal Y^* for the sampling distribution of the whole collective which then will often lead to "nearly orthogonal" design matrices for the individual risks which again "nearly separates" the credibility formula into componentwise procedures.

The question not addressed in this paper is the one of choice of the number of regression parameters. In the case of simple linear regression this question would be: Should you use a linear regression function, a quadratic or a higher order polynominal? Generally the question is: How should one choose the design matrix to start with? We hope to address this question in a forthcoming paper.

ACKNOWLEDGEMENT

The authors wish to thank Peter Bühlmann, University of California, Berkeley for very stimulating ideas and discussions which have led to the formulation of the general case barycentric credibility.

REFERENCES

HACHEMEISTER, C. (1975) *Credibility for regression models with application to trend.* Credibility: Theory and Applications, (e.d. D.M. Kahn), 129-163 Academic Press, New York.

DANNENBURG, D. (1996) *Basic actuarial credibility models.* PhD Thesis University of Amsterdam.

DE VYLDER, F. (1981) *Regression model with scalar credibility weights.* Mitteilungen Vereinigung Schweizerischer Versicherungsmathematiker, Heft **1**, 27-39.

DE VYLDER, F. (1985) *Non-linear regression in credibility theory.* Insurance: Mathematics and Economics **4**, 163-172.

PROF. H BÜHLMANN
Mathematics Department
ETH Zurich
CH – 8092 Zurich

DR. A. GISLER
"Winterthur", Swiss Insurance Co.
P.O. Box 357
CH – 8401 Winterthur

https://doi.org/10.2143/AST.27.1.563207 Published online by Cambridge University Press