

# Credibility Theory

Ragnar Norberg, Department of Statistics, London School of Economics,  
London WC2A 2AE, United Kingdom

**Abstract:** This survey of actuarial credibility theory traces its origins, describes its evolutionary history, gives an account of its main issues and results, and assesses its merits with a view to related work in statistical science.

*Key-words:* Experience rating, limited fluctuation theory, greatest accuracy theory, Bayes, linear Bayes, empirical Bayes, hierarchical models, Hilbert space methods, recursive computation, history of actuarial science.

**A. Credibility - what it was and what it is.** In actuarial parlance the term *credibility* was originally attached to experience rating formulas that were convex combinations (weighted averages) of individual and class estimates of the individual risk premium. *Credibility theory*, thus, was the branch of insurance mathematics that explored model-based principles for construction of such formulas. The development of the theory brought it far beyond the original scope so that in today's usage credibility covers more broadly linear estimation and prediction in latent variable models.

**B. The origins.** The advent of credibility dates back to Whitney [46], who in 1918 addressed the problem of assessing the risk premium  $m$ , defined as the expected claims expenses per unit of risk exposed, for an individual risk selected from a portfolio (class) of similar risks. Advocating the combined use of individual risk experience and class risk experience, he proposed that the premium rate be a weighted average of the form

$$\bar{m} = z \hat{m} + (1 - z) \mu, \tag{1}$$

where  $\hat{m}$  is the observed mean claim amount per unit of risk exposed for the individual contract and  $\mu$  is the corresponding overall mean in the insurance portfolio. Whitney viewed the risk premium as a random variable. In the language of modern credibility theory, it is a function  $m(\Theta)$  of a random element  $\Theta$  representing the unobservable characteristics of the individual risk. The random nature of  $\Theta$  expresses the notion of heterogeneity; the individual risk is a random selection from a portfolio of similar but not identical risks, and the distribution of  $\Theta$  describes the variation of individual risk characteristics across the portfolio.

The weight  $z$  in (1) was soon to be named *credibility (factor)* since it measures the amount of credence attached to the individual experience, and  $\bar{m}$  was called the *credibility premium*.

Attempts to lay a theoretical foundation for rating by credibility formulas bifurcated into two streams usually referred to as *limited fluctuation credibility theory* and *greatest accuracy credibility theory*. In more descriptive statistical terms they could appropriately be called the “fixed effect” and the “random effect” theories of credibility.

**C. The limited fluctuation approach.** The genealogy of the limited fluctuations approach takes us back to 1914, when Mowbray [29] suggested how to determine the amount of individual risk exposure needed for  $\hat{m}$  to be a fully reliable estimate of  $m$ . He worked with annual claim amounts  $X_1, \dots, X_n$ , assumed to be i.i.d. (independent and identically distributed) selections from a probability distribution with density  $f(x|\theta)$ , mean  $m(\theta)$ , and variance  $s^2(\theta)$ . The parameter  $\theta$  was viewed as non-random. Taking  $\hat{m} = \frac{1}{n} \sum_{i=1}^n X_j$ , he sought to determine how many years  $n$  of observation are needed to make  $\mathbb{P}_\theta [|\hat{m} - m(\theta)| \leq k m(\theta)] \geq 1 - \epsilon$  for some given (small)  $k$  and  $\epsilon$ . Using the normal approximation  $\hat{m} \sim N(m(\theta), \frac{s(\theta)}{\sqrt{n}})$ , he deduced the criterion  $k m(\theta) \geq z_{1-\epsilon/2} \frac{s(\theta)}{\sqrt{n}}$ , where  $z_{1-\epsilon/2}$  is the upper  $\epsilon/2$  fractile in the standard normal distribution. Plugging in the empirical estimates  $\hat{m}$  and  $\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{m})^2$  for the unknown parameters, he arrived at

$$n \geq \frac{z_{1-\epsilon/2}^2 \hat{s}^2}{k^2 \hat{m}^2}. \quad (2)$$

Whitney’s and Mowbray’s immediate successors adopted Whitney’s appealing formula and, replacing his random effect model with Mowbray’s fixed effect model, they saw Mowbray’s result (2) as a criterion for *full credibility* of  $\hat{m}$ , which means setting  $z = 1$  in (1). The issue of *partial credibility* was raised: how to choose  $z$  when  $n$  does not satisfy (2)? The plethora of papers that followed brought many tentative answers, but never settled on a unifying principle that covered all special cases and that opened for significant generalizations. Therefore, the limited fluctuation approach, despite its grand scale, does not really constitute a theory in the usual sense. A survey of the area is given in [27].

**D. The greatest accuracy point of view.** After three decades dominated by limited fluctuation studies, the post World War II era saw the

revival of Whitney's random effect idea. Combined with suitable elements of statistical decision theory developed meanwhile, it rapidly developed into a huge body of models and methods - the greatest accuracy theory. The experience rating problem was now seen as a matter of estimating the random variable  $m(\Theta)$  with some function  $\tilde{m}(X)$  of the individual data  $X$ , the objective being to minimize the *mean squared error* (MSE)

$$\rho(\tilde{m}) = \mathbb{E} [m(\Theta) - \tilde{m}(X)]^2. \quad (3)$$

The calculation

$$\mathbb{E} [m(\Theta) - \tilde{m}(X)]^2 = \mathbb{E} [m(\Theta) - \mathbb{E}[m|X]]^2 + \mathbb{E} [\mathbb{E}[m|X] - \tilde{m}(X)]^2 \quad (4)$$

shows that the optimal estimator is the conditional mean,

$$\tilde{m}(X) = \mathbb{E}[m|X], \quad (5)$$

and that its MSE is

$$\tilde{\rho} = \mathbb{E} \text{Var}[m(\Theta)|X] = \text{Var } m - \text{Var } \tilde{m}.$$

In statistical terminology  $\tilde{m}$  is the *Bayes estimator* under squared loss and  $\tilde{\rho}$  is the *Bayes risk*.

Assuming that the data is a vector  $X = (X_1, \dots, X_n)$  with density  $f(x_1, \dots, x_n|\theta)$  conditional on  $\Theta = \theta$ , and denoting the distribution of  $\Theta$  by  $G$ , we have

$$\tilde{m}(X) = \int m(\theta) dG(\theta|X_1, \dots, X_n), \quad (6)$$

where  $G(\cdot | x_1, \dots, x_n)$  is the conditional distribution of  $\Theta$ , given the data:

$$dG(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta) dG(\theta)}{\int f(x_1, \dots, x_n|\theta') dG(\theta')}. \quad (7)$$

In certain well structured models, with  $f(x_1, \dots, x_n|\theta)$  parametric (i.e.  $\theta$  finite-dimensional) and  $G$  conveniently chosen, the integrals appearing in (6) and (7) are closed form expressions. The quest for such models is a major enterprise in Bayesian statistics, where  $f(\cdot | \theta)$  is called the *likelihood* (function),  $G$  is called the *prior* (distribution) since it expresses subjective beliefs prior to data, and  $G(\cdot | x_1, \dots, x_n)$  is called the *posterior* (distribution) accordingly. A family of priors is said to be *conjugate* to a given likelihood if the posterior stays within the same family. If the conjugate prior is mathematically tractable, then so is the posterior. The Bayes theory boasts a

huge body of results on conjugate priors for well structured parametric likelihoods that possess finite-dimensional sufficient statistics. For an overview see e.g. [8]. In 1972 Ferguson [12] launched the Dirichlet process (a gamma process normed to a probability) and showed that it is a conjugate prior to the non-parametric family of distributions in the case of i.i.d. observations. Conjugate analysis has had limited impact on credibility theory, the reason being that in insurance applications it is typically not appropriate to impose much structure on the distributions. In quest for a nice estimator it is better to impose structure on the class of admitted estimators and seek the optimal solution in the restricted class. This is what the insurance mathematicians did, and the programme of greatest accuracy credibility thus became to find the best estimator of the linear form

$$\check{m}(X) = a + b \hat{m}(X),$$

where  $\hat{m}$  is some natural estimator based on the individual data. The MSE of such an estimator is just a quadratic form in  $a$  and  $b$ , which is straightforwardly minimized. One arrives at the *linear Bayes* (LB) estimator

$$\bar{m} = \mathbb{E} m(\Theta) + \frac{\text{Cov}[m, \hat{m}]}{\text{Var} \hat{m}} (\hat{m} - \mathbb{E} \hat{m}), \quad (8)$$

and the LB risk

$$\bar{\rho} = \text{Var} m - \frac{\text{Cov}^2[m, \hat{m}]}{\text{Var} \hat{m}}. \quad (9)$$

The LB risk measures the accuracy of the LB estimator. For linear estimation to make sense, the LB risk ought to approach 0 with increasing amounts of data  $X$ . Since

$$\bar{\rho} \leq \mathbb{E}[m - \hat{m}]^2,$$

a sufficient condition for  $\bar{\rho}$  to tend to 0 is

$$\mathbb{E}[m - \hat{m}]^2 \rightarrow 0. \quad (10)$$

From the decomposition

$$\mathbb{E} [m(\Theta) - \hat{m}]^2 = \mathbb{E} [m(\Theta) - \mathbb{E}[\hat{m}|\Theta]]^2 + \mathbb{E} \text{Var}[\hat{m}|\Theta]$$

it is seen that a pair of sufficient conditions for (10) to hold true are asymptotic conditional unbiasedness in the sense

$$\mathbb{E} [m(\Theta) - \mathbb{E}[\hat{m}|\Theta]]^2 \rightarrow 0$$

and asymptotic conditional consistency in the sense

$$\mathbb{E} \operatorname{Var}[\hat{m}|\Theta] \rightarrow 0.$$

Usually  $\hat{m}$  is conditionally unbiased, not only asymptotically:

$$\mathbb{E}[\hat{m}|\Theta] = m(\Theta).$$

If this condition is in place, then

$$\begin{aligned} \mathbb{E} \hat{m} &= \mathbb{E} m, \\ \operatorname{Cov}[m, \hat{m}] &= \operatorname{Var} m, \\ \operatorname{Var} \hat{m} &= \operatorname{Var} m + \mathbb{E} \operatorname{Var}[\hat{m}|\Theta], \end{aligned}$$

and (8) assumes the form (1) with

$$\begin{aligned} \mu &= \mathbb{E} m(\Theta), \\ z &= \frac{\operatorname{Var} m(\Theta)}{\operatorname{Var} m(\Theta) + \mathbb{E} \operatorname{Var}[\hat{m}|\Theta]}. \end{aligned}$$

This is the greatest accuracy justification of the credibility approach.

**E. The greatest accuracy break-through.** The programme of the theory was set out clearly in the late 1960-es by Bühlmann [4] [5]. He emphasized that the optimization problem is simple (a matter of elementary algebra) and that the optimal estimator and its MSE depend only on first and second moments that are usually easy to estimate from statistical data. The greatest accuracy resolution to the credibility problem had essentially been set out already two decades earlier by Bailey [1] [2], but like many other scientific works ahead of their time, they did not receive wide recognition. They came prior to, and could not benefit from, modern statistical decision theory, and the audience was not prepared to collect the message.

Bühlmann considered a non-parametric model specifying only that, conditional on  $\Theta$ , the annual claim amounts  $X_1, \dots, X_n$  are i.i.d. with mean  $m(\Theta)$  and variance  $s^2(\Theta)$ . Taking

$$\hat{m} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_j, \tag{11}$$

which is best linear unbiased estimator (BLUE) of  $m(\theta)$  in the conditional model, given  $\Theta = \theta$ , he arrived at the credibility formula (1) with

$$\mu = \mathbb{E} m(\Theta) = \mathbb{E} X_j, \tag{12}$$

$$z = \frac{\lambda n}{\lambda n + \phi}, \quad (13)$$

$$\lambda = \text{Var}[m(\Theta)], \quad \phi = \mathbb{E} s^2(\Theta). \quad (14)$$

The credibility factor  $z$  behaves as it ought to do: it increases and tends to 1 with increasing number of observations  $n$ ; it increases with  $\lambda$ , which means that great uncertainty about the value of the true risk premium will give much weight to the individual risk experience; it decreases with  $\phi$ , which measures the purely erratic variation in the observations. The LB risk (9) becomes

$$\bar{\rho} = \frac{\phi \lambda}{\lambda n + \phi} = (1 - z)\lambda, \quad (15)$$

which depends in a natural way on  $n$  and the parameters.

**F. The Bühlmann-Straub model.** The greatest accuracy paradigm, a merger of a sophisticated model concept and a constructive optimization criterion, had great potential for extensions and generalizations. This was demonstrated in a much cited paper by Bühlmann and Straub [6], henceforth abbreviated B-S, where the i.i.d. assumption in Bühlmann's model was relaxed by letting the conditional variances be of the form  $\text{Var}[X_j|\Theta] = s^2(\Theta)/p_j$ ,  $j = 1, \dots, n$ . The motivation was that  $X_j$  is the loss ratio in year  $j$ , which is the total claim amount divided by the amount of risk exposed,  $p_j$ . The volumes  $(p_1, \dots, p_n)$  constitute the observational *design*, a piece of statistical terminology that has been adopted in insurance mathematics despite its connotation of planned experiments. The admitted estimators were taken to be of the linear form

$$\tilde{m} = g_0 + g_1 X_1 + \dots + g_n X_n,$$

with constant coefficients  $g_j$ . Minimization of the MSE (3) is just another exercise in differentiation of a quadratic form and solving the resulting set of linear equations. The LB estimator is of the credibility form (1), now with

$$\hat{m} = \frac{\sum_{j=1}^n p_j X_j}{\sum_{j=1}^n p_j} \quad (16)$$

(the BLUE of  $m(\theta)$  in the conditional or fixed effects model), and

$$z = \frac{\sum_{j=1}^n p_j \lambda}{\sum_{j=1}^n p_j \lambda + \phi}. \quad (17)$$

The LB risk is

$$\bar{\rho} = (1 - z)\lambda. \quad (18)$$

In retrospect this was a humble extension of the results (11) – (15) for the i.i.d. case, but of great importance in its time since it manifestly showed the wide applicability of the greatest accuracy construction. The way was now paved to more elaborate models, and new results followed in rapid succession.

**G. Multi-dimensional credibility.** Jewell [20] introduced a multi-dimensional model in which the data is a random  $n \times 1$  vector  $\mathbf{x}$ , the estimand  $\mathbf{m}$  is a random  $s \times 1$  vector, and the admitted estimators are of the linear form

$$\check{\mathbf{m}} = \mathbf{g} + \mathbf{G}\mathbf{x},$$

with constant coefficients  $\mathbf{g}$  ( $s \times 1$ ) and  $\mathbf{G}$  ( $s \times n$ ). The objective is to minimize the MSE

$$\rho(\check{\mathbf{m}}) = (\mathbf{m} - \check{\mathbf{m}})' \mathbf{A}(\mathbf{m} - \check{\mathbf{m}}),$$

where  $\mathbf{A}$  is a fixed positive definite  $s \times s$  matrix. Again one needs only to minimize a quadratic form. The LB solution is a transparent multi-dimensional extension of the one-dimensional formulas (8) - (9):

$$\bar{\mathbf{m}} = \mathbb{E} \mathbf{m} + \text{Cov}[\mathbf{m}, \mathbf{x}'] [\text{Var} \mathbf{x}]^{-1} (\mathbf{x} - \mathbb{E} \mathbf{x}), \quad (19)$$

$$\bar{\rho} = \text{tr}(\mathbf{A}\bar{\mathbf{R}}),$$

where  $\text{tr}$  denotes the trace operator and  $\mathbf{R}$  is the *LB risk matrix*

$$\bar{\mathbf{R}} = \text{Var} \mathbf{m} - \text{Cov}[\mathbf{m}, \mathbf{x}'] [\text{Var} \mathbf{x}]^{-1} \text{Cov}[\mathbf{x}, \mathbf{m}']. \quad (20)$$

**H. The random coefficient regression model** Hachemeister [16] introduced a regression extension of the B-S model specifying that

$$\mathbb{E}[X_j | \Theta] = \sum_{r=1}^s y_{jr} b_r(\Theta),$$

where the regressors  $y_{jr}$  are observable, and

$$\text{Var}[X_j | \Theta] = s^2(\Theta)/p_j.$$

The design now consists of the  $n \times q$  regressor matrix  $\mathbf{Y} = (Y_{jr})$  and the  $n \times n$  volume matrix  $\mathbf{P} = \text{Diag}(p_j)$  with the  $p_j$  placed down the principal

diagonal and all off-diagonal entries equal to 0. In matrix form, denoting the  $n \times 1$  vector of observations by  $\mathbf{x}$  and the  $s \times 1$  vector of regression coefficients by  $\mathbf{b}(\Theta)$ ,

$$\begin{aligned}\mathbb{E}[\mathbf{x}|\Theta] &= \mathbf{Y}\mathbf{b}(\Theta), \\ \text{Var}[\mathbf{x}|\Theta] &= s^2(\Theta)\mathbf{P}^{-1}.\end{aligned}$$

The problem is to estimate the regression coefficients  $\mathbf{b}(\Theta)$ . Introducing

$$\boldsymbol{\beta} = \mathbb{E}\mathbf{b}, \quad \boldsymbol{\Lambda} = \text{Var}\mathbf{b}, \quad \phi = \mathbb{E}s^2(\Theta),$$

the entities involved in (19) and (20) now become

$$\begin{aligned}\mathbb{E}\mathbf{x} &= \mathbf{Y}\boldsymbol{\beta}, \\ \text{Var}\mathbf{x} &= \mathbf{Y}\boldsymbol{\Lambda}\mathbf{Y}' + \phi\mathbf{P}^{-1}, \\ \text{Cov}[\mathbf{x}, \mathbf{b}'] &= \mathbf{Y}\boldsymbol{\Lambda}.\end{aligned}$$

If  $\mathbf{Y}$  has full rank  $s$ , some matrix algebra leads to the appealing formulas

$$\bar{\mathbf{b}} = \mathbf{Z}\hat{\mathbf{b}} + (\mathbf{I} - \mathbf{Z})\boldsymbol{\beta}, \quad (21)$$

$$\bar{\mathbf{R}} = (\mathbf{I} - \mathbf{Z})\boldsymbol{\Lambda}, \quad (22)$$

where

$$\hat{\mathbf{b}} = (\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{P}\mathbf{x}, \quad (23)$$

$$\mathbf{Z} = (\boldsymbol{\Lambda}\mathbf{Y}'\mathbf{P}\mathbf{Y} + \phi\mathbf{I})^{-1}\boldsymbol{\Lambda}\mathbf{Y}'\mathbf{P}\mathbf{Y}. \quad (24)$$

Formula (21) expresses the LB estimator as a credibility weighted average of the sample estimator  $\hat{\mathbf{b}}$ , which is BLUE in the conditional model, and the prior estimate  $\boldsymbol{\beta}$ . The matrix  $\mathbf{Z}$  is called the *credibility matrix*. The expressions in (23) and (24) are matrix extensions of (16) and (17), and their dependence on the design and the parameters follows along the same lines as in the univariate case in Paragraph E.

**I. Heterogeneity models and empirical Bayes.** Whitney's notion of heterogeneity (Paragraph B) was set out in precise terms in the cited works of Bailey and Bühlmann: The portfolio consists of  $N$  independent risks, the unobservable risk characteristics of risk No.  $i$  is denoted by  $\Theta_i$ , and the  $\Theta_i$  are i.i.d. selections from some distribution  $G$  called the *structural distribution*. The device clarifies the idea that the risks are different, but still have something in common that motivates pooling them into one risk class or portfolio.

Thus, in the B-S set-up the annual loss ratios of risk No.  $i$ ,  $X_{i1}, \dots, X_{in_i}$ , are conditionally independent with

$$\mathbb{E}[X_{ij}|\Theta] = m(\Theta_i), \quad \text{Var}[X_{ij}|\Theta] = s^2(\Theta_i)/p_{ij}.$$

Due to independence, the Bayes estimator and the linear Bayes estimator of each individual risk premium  $m(\Theta_i)$  remain as in Paragraphs E and F (add subscript  $i$  to all entities.) Thus, for the purpose of assessing  $m(\Theta_i)$ , the observations stemming from the collateral risks  $i' \neq i$  are irrelevant if the parameters in the model would be known. However, the parameters are unknown and are to be estimated from portfolio statistics. This is how data from collateral risks become useful in the assessment of the individual risk premium  $m(\Theta_i)$ . The idea fits perfectly into the framework of empirical Bayes theory, instituted by Robbins [37] [38], which was well developed at the time when the matter arose in the credibility context. The empirical linear Bayes procedure amounts to inserting statistical point estimators  $\mu^*, \lambda^*, \phi^*$  for the parameters involved in (1) and (17) to obtain an estimated LB estimator,

$$\bar{m}^* = z^* \hat{m} + (1 - z^*) \mu^*$$

(dropping now the index  $i$  of the given individual). The credibility literature has given much attention to the parameter estimation problem, which essentially is a matter of mean and variance component estimation in linear models. This is an established branch of statistical inference theory, well documented in textbooks and monographs, see e.g. [36] and [43].

Empirical Bayes theory works with certain criteria for assessing the performance of the estimators. An estimated Bayes estimator is called an *empirical Bayes estimator* if it converges in probability to the Bayes estimator as the amount of collateral data increases, and it is said to be *asymptotically optimal* (a.o.) if its MSE converges to the Bayes risk. Carrying these concepts over to LB estimation, Norberg [33] proved a.o. of the empirical LB estimator under the conditions that  $\mathbb{E}[\mu^* - \mu]^2 \rightarrow 0$  and that  $(\lambda^*, \phi^*) \rightarrow (\lambda, \phi)$  in probability (the latter condition is sufficient because  $z$  is a bounded function of  $(\lambda, \phi)$ ). Weaker conditions were obtained by Mashayeki [28]. We mention two more results obtained in the credibility literature that go beyond standard empirical Bayes theory: Neuhaus [31] considered the observational designs as i.i.d. replicates, and obtained asymptotic normality of the parameter estimators, hence possibilities of confidence estimation and testing of hypotheses. Hesselager [18] proved that the rate of convergence (to 0) of the MSE of the parameter estimators is inherited by the MSE of the empirical LB estimator.

**J. The Bayes point of view.** When no collateral data are available, the frequency theoretical empirical Bayes model does not apply. This appears to be the situation Mowbray had in mind (Paragraph C); his problem was to quote a premium for a unique, single standing risk or class of risks. The Bayesian approach to this problem is to place a prior  $G$  on the possible values of  $\theta$ , probabilities now representing subjective degrees of belief prior to any risk experience. This way the fixed effect  $\theta$  is turned into a random variable  $\Theta$  just as in the frequency theory set-up, only with a different interpretation, and the Bayes and linear Bayes analyses become the same as before.

**K. Hierarchical models.** The notion of hierarchies was introduced in credibility theory by Gerber and Jones [13], Taylor [44], and Jewell [24]. To explain the idea in its fully developed form, it is convenient to work with observations in “coordinate form” as is usual in statistical analysis of variance. With this device the B-S model in Paragraph F is cast as

$$X_{ij} = \mu + \vartheta_i + \varepsilon_{ij},$$

where  $\vartheta_i = m(\Theta_i) - \mu$  is the deviation of risk No.  $i$  from the overall mean risk level and  $\varepsilon_{ij} = X_{ij} - m(\Theta_i)$  is the erratic deviation of its year  $j$  result from the individual mean. The  $\vartheta_i$ ,  $i = 1, \dots, N$ , are i.i.d. with zero mean and variance  $\lambda$ , and the  $\varepsilon_{ij}$ ,  $j = 1, \dots, n_{ij}$ , are conditionally independent, given  $\Theta_i$ , and have zero mean and variances  $\text{Var } \varepsilon_{ij} = \phi/p_{ij}$ .

In the hierarchical extension of the model the data are of the form

$$X_{i_1 \dots i_s j} = \mu + \vartheta_{i_1} + \vartheta_{i_1 i_2} + \dots + \vartheta_{i_1 \dots i_s} + \varepsilon_{i_1 \dots i_s j},$$

$j = 1, \dots, n_{i_1 \dots i_s}$ ,  $i_s = 1, \dots, N_{i_1 \dots i_{s-1}}$ ,  $\dots$ ,  $i_r = 1, \dots, N_{i_1 \dots i_{r-1}}$ ,  $\dots$ ,  $i_1 = 1, \dots, N$ . The index  $i_1$  labels risk classes at first level (the coarsest classification), the index  $i_2$  labels risk (sub)classes at second level within a given first level class, and so on up to the index  $i_s$  which labels the individual risks (the finest classification) within a given class at level  $s - 1$ . The index  $j$  labels annual results for a given risk.

The latent variables are uncorrelated with zero means and variances  $\text{Var } \vartheta_{i_1 \dots i_r} = \lambda_r$  and  $\text{Var } \varepsilon_{i_1 \dots i_s j} = \phi/p_{i_1 \dots i_s j}$ . The variance component  $\lambda_r$  measures the variation between level  $r$  risk classes within a given level  $r - 1$  class.

The problem is to estimate the mean

$$m_{i_1 \dots i_r} = \mu + \vartheta_{i_1} + \dots + \vartheta_{i_1 \dots i_r}$$

for each class  $i_1 \dots i_r$ . The LB solution is a system of recursive relationships: Firstly, recursions upwards,

$$\bar{m}_{i_1 \dots i_r} = z_{i_1 \dots i_r} \hat{m}_{i_1 \dots i_r} + (1 - z_{i_1 \dots i_r}) \bar{m}_{i_1 \dots i_{r-1}}, \quad (25)$$

starting from  $\bar{m}_{i_1} = z_{i_1} \hat{m}_{i_1} + (1 - z_{i_1}) \mu$  at level 1. These are credibility formulas. Secondly, recursions downwards,

$$z_{i_1 \dots i_r} = \frac{\lambda_r \sum_{i_{r+1}=1}^{N_{i_1 \dots i_r}} z_{i_1 \dots i_r i_{r+1}}}{\lambda_r \sum_{i_{r+1}=1}^{N_{i_1 \dots i_r}} z_{i_1 \dots i_r i_{r+1}} + \lambda_{r+1}}, \quad (26)$$

$$\hat{m}_{i_1 \dots i_r} = \frac{\sum_{i_{r+1}=1}^{N_{i_1 \dots i_r}} z_{i_1 \dots i_r i_{r+1}} \hat{m}_{i_1 \dots i_r i_{r+1}}}{\sum_{i_{r+1}=1}^{N_{i_1 \dots i_r}} z_{i_1 \dots i_r i_{r+1}}}, \quad (27)$$

starting from (16) and (17) at level  $s$  (with  $i_1 \dots i_s$  added in the subscripts). There is also a set of recursive equations for the LB risks:

$$\bar{\rho}_{i_1 \dots i_r} = (1 - z_{i_1 \dots i_r}) \lambda_r + (1 - z_{i_1 \dots i_r})^2 \bar{\rho}_{i_1 \dots i_{r-1}}. \quad (28)$$

The formulas bear a resemblance to those in Paragraphs E, F, and G and are easy to interpret. They show how the estimator of any class mean  $m_{i_1 \dots i_r}$  depends on the parameters and on data more or less remote in the hierarchy. The recursion (25) was proved by Jewell [24] and extended to the regression case by Sundt [39] [40]. Displaying the complete data structure of the hierarchy, Norberg [34] established the recursions (26), (27), and (28) in a regression setting.

**L. Hilbert space methods.** For a fixed probability space the set  $\mathcal{L}^2$  of all square integrable random variables is a linear space and, when equipped with the inner product  $\langle X, Y \rangle = \mathbb{E}[XY]$ , it becomes a Hilbert space. The corresponding norm of an  $X$  in  $\mathcal{L}^2$  is  $\|X\| = \sqrt{\langle X, X \rangle}$ , and the distance between any  $X$  and  $Y$  in  $\mathcal{L}^2$  is  $\|X - Y\|$ . In this set-up the MSE (3) is the squared distance between the estimand and the estimator, and finding the best estimator in some family of estimators amounts to finding the minimum distance point to  $m$  in that family. If the family of admitted estimators is a closed linear subspace,  $\bar{\mathcal{M}} \subset \mathcal{L}^2$ , then a unique minimum distance point exists, and it is the random variable  $\bar{m} \in \bar{\mathcal{M}}$  such that

$$\langle m - \bar{m}, \check{m} \rangle = 0, \quad \forall \check{m} \in \bar{\mathcal{M}}. \quad (29)$$

In geometric terms,  $\bar{m}$  is the projection of  $m$  onto  $\bar{\mathcal{M}}$ , the equations (29) are the normal equations stating that  $m - \bar{m}$  is orthogonal to  $\bar{\mathcal{M}}$ , and the Pythagoras relationship  $\|m\|^2 = \|\bar{m}\|^2 + \|m - \bar{m}\|^2$  gives

$$\bar{\rho} = \|m - \bar{m}\|^2 = \|m\|^2 - \|\bar{m}\|^2 .$$

The Hilbert space approach to linear estimation under the MSE criterion was taken by Gerber and Jones [13], De Vylder [9] and Taylor [45]. It adds insight into the structure of the problem and its solution, but can be dispensed with if  $\bar{\mathcal{M}}$  is of finite dimension since the problem then reduces to minimizing a finite-dimensional quadratic form. Hilbert space methods are usually needed in linear spaces  $\bar{\mathcal{M}}$  of infinite dimension. Paradoxically, maybe, for the biggest conceivable  $\bar{\mathcal{M}}$  consisting of all square integrable functions of the data, the best estimator (5) can be obtained from the purely probabilistic argument (4) without visible use of the normal equations. In the following situations the optimal estimator can only be obtained by solving the infinite-dimensional system of normal equations (29): The ‘‘semilinear credibility’’ problem [10], where  $X_1, \dots, X_n$  are conditionally i.i.d., given  $\Theta$ , and  $\bar{\mathcal{M}}$  consists of all estimators of the form  $\hat{m} = \sum_{i=1}^n f(X_i)$  with  $f(X_i)$  square integrable. The continuous time credibility problem, where the claims process  $X$  has been observed continually up to some time  $\tau$  and  $\bar{\mathcal{M}}$  consists of all estimators of the form  $\hat{m} = g_0 + \int_0^\tau g_t dX_t$  with constant coefficients  $g_t$ . In [19]  $X$  is of diffusion type and in [35] it is of bounded variation. The Hilbert space approach may simplify matters also in finite-dimensional problems of high complexity. An example is [7] on hierarchical credibility.

**M. Exact credibility.** The problem studied under this headline is: when is the linear Bayes estimator also Bayes? The issue is closely related to conjugate Bayes analysis (Paragraph D). In [21] [22] [23] Jewell showed that  $\bar{m} = \hat{m}$  if the likelihood is of exponential form with  $\bar{X}$  as canonical sufficient statistic and the prior is conjugate. Diaconis and Ylvisaker [11] completed the picture by proving that these conditions are also sufficient. Pertinently, since the spirit of credibility is very much non-parametric, Zehnwirth [49] pointed out that Bayes estimators are of credibility form in Ferguson’s non-parametric model (Paragraph D). These results require observations to be conditionally i.i.d., and thus already the B-S model falls outside their remit. In insurance applications, where non-parametric distributions and imbalanced designs are commonplace, the LB approach is justified by its practicability rather than its theoretical optimality properties. One may, however, show that LB estimators in a certain sense are restricted minimax under mild conditions.

**N. Linear sufficiency** A linear premium formula need not necessarily be linear in the claim amounts themselves. In the simple framework of Paragraphs D and E, the question is how to choose the sample statistic  $\hat{m}$ . Taylor [45] used results from parametric statistical inference theory to show that the best choice is the unbiased minimal sufficient estimator (when it exists). More in the vein of non-parametric credibility, and related to the choice of regressors problem in statistics, Neuhaus [32] gave a rule for discarding data that do not significantly help reducing the MSE. Related work was done by Witting [48] and Sundt [42] under the heading linear sufficiency.

**O. Recursive formulas.** The credibility premium is to be currently updated as claims data accrue. From a practical computational point of view it would be convenient if the new premium would be a simple function of the former premium and the new data. Key references on this topic are [13], [25], and - including regression and dynamical risk characteristics - [41].

**P. A view to related work outside actuarial science.** Credibility Theory in actuarial science and Linear Bayes in statistics are non-identical twins: imperfect communication between the two communities caused parallel studies and discoveries and also some rediscoveries. Several works cited here are from statistics, and they are easily identified by inspection of the list of references. At the time when the greatest accuracy theory gained momentum Linear Bayes theory was an issue also in statistics [17]. Later notable contributions are [43], [47], and [3] on random coefficient regression models, [14] and [15] on standard LB theory, [30] on continuous time linear filtering, and [36] on parameter estimation and hierarchical models.

Linear estimation and prediction is a major project also in engineering, control theory, and operations research. Kalman's [26] linear filtering theory covers many early results in Credibility and Linear Bayes and, in one respect, goes far beyond as the latent variables are seen as dynamical objects.

Even when seen in this bigger perspective of its environments, credibility theory counts as a prominent scientific area with claim to a number of significant discoveries and with a wealth of special models arising from applications in practical insurance.

## References

- [1] Bailey, A.L. (1945): A generalized theory of credibility. *Proceedings of the Casualty Actuarial Society* **32**, 13-20.
- [2] Bailey, A.L. (1950): Credibility procedures, La Place's generalization of Bayes' rule, and the combination of collateral knowledge with observed data. *Proceedings of the Casualty Actuarial Society* **37**, 7-23. Discussion in **37**, 94-115.
- [3] Bunke, H. and Gladitz, J. (1974): Empirical linear Bayes decision rules for a sequence of linear models with different regressor matrices. *Mathematische Operationsforschung und Statistik* **5**, 235-244.
- [4] Bühlmann, H. (1967): Experience rating and credibility. *ASTIN Bulletin* **4**, 199-207.
- [5] Bühlmann, H. (1969): Experience rating and credibility. *ASTIN Bulletin* **5**, 157-165.
- [6] Bühlmann, H. and Straub, E. (1970): Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **70**, 111-133.
- [7] Bühlmann, H. and Jewell, W. (1987): Hierarchical credibility revisited. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 35-54.
- [8] De Groot, M. (1970): *Optimal Statistical Decisions*, McGraw-Hill.
- [9] De Vylder, F. (1976a): Geometrical credibility. *Scandinavian Actuarial Journal* **1976**, 121-149.
- [10] De Vylder, F. (1976b): Optimal semilinear credibility. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **76**, 27-40.
- [11] Diaconis, P. and Ylvisaker, D. (1979): Conjugate priors for exponential families. *Annals of Statistics* **7**, 269-281.
- [12] Ferguson, T.S. (1972): A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209-230.
- [13] Gerber, H.U. and Jones, D.A. (1975): Credibility formulas of the updating type. *Transaction of the Society of Actuaries* **27**, 31-46, and in

*Credibility: Theory and applications* (ed. P.M. Kahn), 89-105, Academic Press.

- [14] Goldstein, M. (1975a): Approximate Bayes solution to some nonparametric problems. *Annals of Statistics* **3**, 512-517.
- [15] Goldstein, M. (1975b): Bayesian nonparametric estimates. *Annals of Statistics*. **3**, 736-740.
- [16] Hachemeister, C. (1975): Credibility for regression models with application to trend. In *Credibility: Theory and applications* (ed. P.M. Kahn), 129-163. Academic Press.
- [17] Hartigan, J.A. (1969): Linear Bayesian methods *J. Royal Statist. Soc., B.* **31**, 446-454.
- [18] Hesselager, O. (1992): Rates of risk convergence of empirical linear Bayes estimators. *Scandinavian Actuarial Journal* **1992**, 88-94.
- [19] Hiss, K. (1991): Lineare Filtration und Kreditabilitätstheorie. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **1991**, 85-103.
- [20] Jewell, W.S. (1973): Multidimensional credibility. *Actuarial Research Clearing House* **4**.
- [21] Jewell, W.S. (1974a): Credible means are exact Bayesian for simple exponential families. *ASTIN Bulletin* **8**, 77-90.
- [22] Jewell, W.S. (1974b): Exact multidimensional credibility. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **74**, 193-214.
- [23] Jewell, W.S. (1974c): Regularity conditions for exact credibility. *ASTIN Bulletin* **8**, 336-341.
- [24] Jewell, W.S. (1975): The use of collateral data in credibility theory: A hierarchical model. *Giornale dell' Istituto degli Attuari Italiani* **38**, 1-16.
- [25] Jewell, W.S. (1976): Two classes of covariance matrices giving simple linear forecasts. *Scandinavian Actuarial Journal* **1976**, 15-29.

- [26] Kalman, R. (1960): A new approach to linear filtering and prediction problems. *Transactions ASME, Journal of Basic Engineering* **83D**, 35-45.
- [27] Longley-Cook, L.H. (1962): An introduction to credibility theory. *Proceedings of the Casualty Actuarial Society* **49** , 194-221.
- [28] Mashayeki, M (2002): On asymptotic optimality in empirical Bayes credibility. *Insurance: Mathematics & Economics* **31**, 285-295.
- [29] Mowbray, A.H. (1914): How extensive a payroll exposure is necessary to give a dependable pure premium. *Proceedings of the Casualty Actuarial Society* **1**, 24-30.
- [30] Näther, W. (1984): Bayes estimation of the trend parameter in random fields. *Statistics* **15**, 553-558.
- [31] Neuhaus, W. (1984): Inference about parameters in empirical linear Bayes estimation problems. *Scandinavian Actuarial Journal* **1984**, 131-142.
- [32] Neuhaus, W. (1985): Choice of statistics in linear Bayes estimation. *Scandinavian Actuarial Journal* **1985**, 1-26.
- [33] Norberg, R. (1980): Empirical Bayes credibility. *Scandinavian Actuarial Journal* **1980**, 177-194.
- [34] Norberg, R. (1986): Hierarchical credibility: analysis of a random effect linear model with nested classification. *Scandinavian Actuarial Journal* **1986**, 204-222.
- [35] Norberg, R. (1992): Linear estimation and credibility in continuous time. *ASTIN Bulletin* **22**, 149-165.
- [36] Rao, C.R. and Kleffe, J. (1988): *Estimation of Variance Components and Applications*. North-Holland.
- [37] Robbins, H. (1955): An empirical Bayes approach to statistics. *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1*, 157-163. University of California Press.
- [38] Robbins, H. (1964): The empirical Bayes approach to statistical problems. *Annals of Mathematical Statistics* **35**, 1-20.

- [39] Sundt, B. (1979): A hierarchical credibility regression model. *Scandinavian Actuarial Journal* **1979**, 107-114.
- [40] Sundt, B. (1980): A multi-level hierarchical credibility regression model. *Scandinavian Actuarial Journal* **1980**, 25-32.
- [41] Sundt, B. (1981): Recursive credibility estimation. *Scandinavian Actuarial Journal* **1981**, 3-21.
- [42] Sundt, B. (1991): Linearly sufficient fun with credibility. *Insurance: Mathematics & Economics* **10**, 69-74.
- [43] Swamy, P.A.V.B. (1971): Statistical inference in random coefficient regression models. *Lecture Notes in Operations Research and Mathematical Systems* 55. Springer-Verlag.
- [44] Taylor, G.C. (1974): Experience rating with credibility adjustment of the manual premium. *ASTIN Bulletin* **7**, 323-336.
- [45] Taylor, G.C. (1977): Abstract credibility. *Scandinavian Actuarial Journal* **1977**, 149-168.
- [46] Whitney, A.W. (1918): The theory of experience rating. *Proceedings of the Casualty Actuarial Society* **4**, 274-292.
- [47] Wind, S. (1973): An empirical Bayes approach to multiple linear regression. *Annals of Statistics* **1**, 93-103.
- [48] Witting, T (1987): The linear Markov property in credibility theory. *ASTIN Bulletin* **17**, 71-84.
- [49] Zehnwirth, B (1977): The mean credibility formula is a Bayes rule. *Scandinavian Actuarial Journal* **1977**, 212-216.