HIERARCHICAL CREDIBILITY: ANALYSIS OF A RANDOM EFFECT LINEAR MODEL WITH NESTED CLASSIFICATION

by

Ragnar Norberg
University of Oslo

Abstract

A random coefficient regression model with multi-stage nested classification is considered. Linear Bayes estimators are obtained for random effects at all stages, and estimators of the structural parameters are proposed.

Key words: Random effects; Regression; Nested classification; Hierarchical credibility.
1. Introduction

In standard credibility analysis the unobservable characteristics of the individual risks are represented by i.i.d. (independent and identically distributed) random elements. This way of attack is appropriate when the portfolio can be regarded as a random sample from a population of risks. In many situations, however, the portfolio is composed of risk classes that presumably differ from one another with respect to basic risk conditions. In such cases more general models are called for that allow for stochastic dependence between risks belonging to one and the same class. As an example, consider workmen's compensation insurance written by firms to cover employee accidents: the entire portfolio can be divided into industries (heavy industry, transportation, agriculture, ...), each industry can be further divided into branches (heavy industry into smelting works, mining, engine works, ...), and so on until we end up with the individual firms.

In this example the risks are classified in a nested or hierarchical manner. Similar patterns arise in a number of lines of insurance, and they have become a separate issue of study under the heading "hierarchical credibility". The topic was launched by Jewell (1975), who - inspired by Taylor (1974) - considered a hierarchy with two stages. Taylor (1979) extended the analysis to hierarchies with a general number of stages. Hierarchical extensions of Hachemeister's (1974) regression model were treated by Sundt (1979, -1980), first in the case of two stages and then in the general multin-stage case. The two-stage regression case was analysed in a Bayesian setting in an early work by Lindley and Smith (1972).

To this small assembly of works on hierarchical credibility one could add an immense list of relevant references from general statistics; just like standard credibility essentially deals with random effect models with one-way classification, hierarchical credibility fits into the framework of random effect models with nested
classifications. What is peculiar to credibility, is that emphasis is laid on estimation of random effects and on representing the estimators as credibility weighted means. Moreover, credibility primarily has in view situations with unbalanced design and nonparametric families of distributions and, therefore, centers on distribution-free methods depending only on the first and second order moments of the observations. In general statistics random effects are by tradition studied under normality assumptions, which - incidentally - also leads to methods based on moments up to second order. Bayesian literature, broadly represented by Box and Tiao (1973), is concerned with estimation of hidden random effects; sampling theoretic literature, notably Scheffé (1959) and Swamy (1971), concentrates on inference about expected values and variance components (called structural parameters in the credibility terminology).

Those who should try to apply the existing theory on hierarchical credibility in practice, may face problems of two kinds. In the first place, there exists no explicit algorithm for calculation of credibility estimators from factual data in the general multi-stage regression case. In the second place, the problem of parameter estimation has received little attention. It is the intent of this paper to supplement the theory at these points. It is also hoped that the present treatment will open an easy way to fairly general results; the proofs are elementary and computationally oriented, the starting point being a model for the observations as they present themselves in practical applications.

The contents of the paper are organized in sections as follows. The hierarchical model is introduced in section 2. In section 3 well known results on linear Bayes estimation in the nonhierarchical regression model are reviewed and then extended to the hierarchical case. In section 4 estimators are proposed for the 1st and 2nd order moments appearing in the linear Bayes estimators, whereby empirical linear Bayes estimators are obtained.
We end this introduction by laying down a few notational conventions. The dimension of a vector or matrix may be indicated by a topscript; thus \( L^{q \times n} \) denotes a matrix \( L \) with \( q \) rows and \( n \) columns.

By \( \text{diag}(A_1^{n_1}, \ldots, A_m^{n_m}) \) is meant the \((\sum n_i) \times (\sum n_i)\) matrix with \( A_1, \ldots, A_m \) placed in consecutive order downwards the principal diagonal and zeros elsewhere. The symbols \( \text{Cov} \) and \( \text{Var} \) designate covariance and variance matrices, respectively. Thus, if \( x \) and \( y \) are random column vectors, we have \( \text{Cov}(x, y') = \text{E}(xy') - \text{ExEy}' \), and \( \text{Var} x = \text{Cov}(x, x') \). The conditional covariance of \( x \) and \( y \), given a random element \( \theta \), is denoted by \( \text{Cov}(x, y' | \theta) \). We remind of the relation

\[
\text{Cov}(x, y') = \text{Cov}[\text{E}(x | \theta), \text{E}(y' | \theta)] + \text{E} \text{Cov}(x, y' | \theta). \tag{1.1}
\]
2. A hierarchical regression model

Consider a system of quantities ranked in a hierarchical order as follows.

Stage 1: $\theta_{i_1}; i_1 = 1, \ldots, m$

Stage 2: $\theta_{i_1i_2}; i_2 = 1, \ldots, i_1; i_1 = 1, \ldots, m$

$\vdots$

Stage $r$: $\theta_{i_1 \ldots i_r}; i_r = 1, \ldots, m; i_1 \ldots i_{r-1}$

$\vdots$

Stage $s$: $\theta_{i_1 \ldots i_s}; i_s = 1, \ldots, m; i_1 \ldots i_{s-1}$

Observables: $x_{i_1 \ldots i_s}; i_s = 1, \ldots, m; i_1 \ldots i_{s-1}$

The hierarchical structure is visualized by Figure 1. (To be in perfect keeping with the notion of hierarchy, the figure ought to be turned upside down, the idea being that all quantities labeled by some relevant index $i_1 \ldots i_p$, with $p > r$, are governed by $\theta_{i_1}, \ldots, \theta_{i_1 \ldots i_r}$. The present solution is chosen for terminological reasons: it is convenient to speak of all quantities labeled by an $r$-tuple, e.g. $i_1 \ldots i_r$, as being at stage $r$. Accordingly, we shall rather think of $\theta_{i_1}, \ldots, \theta_{i_1 \ldots i_r}$ as "underlying" all quantities labeled $i_1 \ldots i_r \ldots i_p$.) One interpretation of the quantities is the following: $x_{i_1 \ldots i_s}$ and $\theta_{i_1 \ldots i_s}$ are, respectively, the observed claims experience and the hidden risk characteristics of a unit risk class $C_{i_1 \ldots i_s}$; this class is the $i_s$-th of $m_{i_1 \ldots i_{s-1}}$ classes constituting a hyper-class $C_{i_1 \ldots i_{s-1}}$ with
Figure 1. A hierarchy of classes at \( s \) stages
hidden risk characteristics $\theta_{i_1 \ldots i_{s-1}}$; in its turn this hyper-
class is the $i_{s-1}$-th of $m_{i_1 \ldots i_{s-2}}$ hyper-classes constituting a
"hyper-hyper-class" $C_{i_1 \ldots i_{s-2}}$ with hidden risk characteristics
$\theta_{i_1 \ldots i_{s-2}}$; and so on.

Formally, a class $C_{i_1 \ldots i_r}$ is identified with the quantities
that are specifically belonging to it, viz. all quantities at stages
$p > r$ labeled $i_1 \ldots i_r \ldots i_p$. In Figure 1 they are found in the
local hierarchy above and including $\theta_{i_1 \ldots i_r}$.

We denote by $\Theta_r$ the assembly of all latent quantities
$\theta_{i_1 \ldots i_p}$ at stages $p < r$; $r=1, \ldots, s$. It is convenient also to
introduce $\Theta_0 = \emptyset$.

All quantities introduced so far are envisaged as random
elements. We make the following model assumptions, the last three
of which add further contents to the notion of hierarchy.

(i) Each $x_{i_1 \ldots i_s}$ is a vector of dimension $n_{i_1 \ldots i_s}$, say. There
exist a vector-valued function $b_{q \times 1}$ and, for each
$(i_1, \ldots, i_s)$, a nonrandom, known matrix $Y_{i_1 \ldots i_s}^{n_{i_1 \ldots i_s}}$ such that

$$E(x_{i_1 \ldots i_s} | \Theta_s) = Y_{i_1 \ldots i_s} b_{i_1 \ldots i_s}$$

with

$$b_{i_1 \ldots i_s} = b(\theta_{i_1}, \theta_{i_1 i_2}, \ldots, \theta_{i_1 \ldots i_s}).$$  \hspace{1cm} (2.1)

All the random vectors $x_{i_1 \ldots i_s}$ and $b_{i_1 \ldots i_s}$ have finite
second order moments.

(ii) The stage 1 classes $C_{i_1}$ are independent, and their risk
characteristics $\theta_{i_1}$ are identically distributed; $i_1=1, \ldots, m$. 
(iii) Within each class \( C_{i_1 \ldots i_r} \) at stage \( r \leq s-1 \) assumption (ii) has the following analogue: conditionally, given \( \Theta_r \), the stage \( r+1 \) classes \( C_{i_1 \ldots i_r i_{r+1}} \) are independent, and their risk characteristics \( \Theta_{i_1 \ldots i_r} \) are identically distributed; \( i_{r+1} = 1, \ldots, m_i \ldots i_r \).

(iv) All the vectors \( (\Theta_{i_1}, \Theta_{i_1 i_2}, \ldots, \Theta_{i_1 \ldots i_s}) \) are identically distributed (implying that all hidden variables \( \Theta_{i_1 \ldots i_r} \) at a given stage \( r \) assume their values in the same space, which may be quite general).

We now organize the data class-wise as follows. For each class \( C_{i_1 \ldots i_r} \) let \( x_{i_1 \ldots i_r} \) and \( Y_{i_1 \ldots i_r} \) denote, respectively, the vector of all regressands and the matrix of all regressors belonging to that class. More specifically, we start from the already introduced elementary data \( x_{i_1 \ldots i_s} \) and \( Y_{i_1 \ldots i_s} \) related to classes at stage \( s \), and define the corresponding aggregate quantities at lower stages by the recursive relations

\[
x'_{i_1 \ldots i_r} = (x'_{i_1 \ldots i_{r-1}}, \ldots, x'_{i_1 \ldots i_{r-1}}, x'_{i_1 \ldots i_r m_{i_1} \ldots i_r})
\]

\[
Y'_{i_1 \ldots i_r} = (Y'_{i_1 \ldots i_{r-1}}, \ldots, Y'_{i_1 \ldots i_{r-1}}, Y'_{i_1 \ldots i_r m_{i_1} \ldots i_r})
\]

valid for all relevant \( (i_1, \ldots, i_r) \) and all \( r=1, \ldots, s-1 \). (In the following we shall feel free to drop such lengthy and self-evident specifications of domains of subscripts.) Clearly, \( x'_{i_1 \ldots i_r} \) and \( Y'_{i_1 \ldots i_r} \) are of order \( n_{i_1 \ldots i_r} x_1 \) and \( n_{i_1 \ldots i_r} x_q \), respectively, where the \( n_{i_1 \ldots i_r} \)'s are defined recursively as
\( n_{i_1 \ldots i_r} = \sum_{i_{r+1}} n_{i_1 \ldots i_r \cdot i_{r+1}} \).

For each \((i_1, \ldots, i_s)\) introduce
\[
b_{i_1 \ldots i_r} = \mathbb{E}(b_{i_1 \ldots i_s} | \theta_r) ; \quad r=1, \ldots, s. \tag{2.4}
\]

Now it is seen that each class \(C_{i_1 \ldots i_r}\) possesses a regression structure similar to that of the stage \(s\) classes as specified by assumption (i), namely
\[
\mathbb{E}(x_{i_1 \ldots i_r} | \theta_r) = y_{i_1 \ldots i_r} b_{i_1 \ldots i_r}, \tag{2.5}
\]
with elements defined by (2.2)-(2.4). Thus, in a hierarchy of order \(s\) there are embedded hierarchies of all orders \(r < s\).

The following moments are well defined under the assumptions (i)-(iv):
\[
\beta^{q \times 1} = \mathbb{E}(b_{i_1 \ldots i_r}) , \tag{2.6}
\]
\[
\Lambda^{q \times q}_r = \mathbb{E}(\text{Var}(b_{i_1 \ldots i_r} | \theta_{r-1})) , \tag{2.7}
\]
\[
\Phi^{n_{i_1 \ldots i_r} \times n_{i_1 \ldots i_r}} = \mathbb{E}(\text{Var}(x_{i_1 \ldots i_r} | \theta_r)) ; \quad \forall (i_1, \ldots, i_r); \quad r=1, \ldots, s. \tag{2.8}
\]

These quantities have a straightforward interpretation: \(\beta\) is the mean risk level in the portfolio, \(\Lambda_r\) measures the risk differentials between stage \(r\) classes in one and the same stage \(r-1\) class, and - in a sense - \(\Phi_{i_1 \ldots i_s}\) measures the variability in claims experience that is not explained by between class variations at stages \(1, \ldots, r\). The basic parameters are \(\beta\), the \(\Lambda_r\), and the \(\Phi_{i_1 \ldots i_s}\) at stage \(s\). In section 3 it will be shown that each \(\Phi_{i_1 \ldots i_r}\) at stage \(r<s\) is a function of \(\Lambda_{r+1}, \ldots, \Lambda_s\) and the \(\Phi_{i_1 \ldots i_r} \ldots i_s\).
3. Linear Bayes estimation by known parameters

Formulation of the problem. For each class $C_{i_1}...i_r$ we seek an estimator $\hat{b}_{i_1}...i_r$ of $b_{i_1}...i_r$ given by (2.4), the purpose being to minimize the expected weighted squared error or (overall) risk,

$$p_{i_1}...i_r (\hat{b}_{i_1}...i_r) = E\{b_{i_1}...i_r - \hat{b}_{i_1}...i_r\} \cdot W\{b_{i_1}...i_r - \hat{b}_{i_1}...i_r\},$$

(3.1)

where $W_{q\times q}$ is a nonrandom p.d.s. (positive definite symmetric) matrix. In view of the independence assumption (ii), only $X_{i_1}$ is relevant for the purpose. We confine ourselves to inhomogeneous linear estimators of the form

$$\hat{b}_{i_1}...i_r = g_0 + Gx_{i_1},$$

(3.2)

where $g_0^{q\times 1}$ and $G^{q\times n}$ are allowed to depend on $Y$ and the parameters of the distributions. The optimal estimator will be called the LB (linear Bayes) estimator, and the minimum value of the risk will be called the LB risk.

The case $s = 1$. The simple non-hierarchical random coefficient regression model was first studied in the context of credibility theory by Hachemeister (1975). It shall be treated in some detail since there exists no single reference that spells out the results in the form required here.

Thus, consider the model (i)-(iv) in the special case $s = 1$, whereby items (iii) and (iv) become void. The problem is to estimate $b_{i_1} = b(\theta_{i_1})$ (say) by means of $X_{i_1}$, the parameters (2.6)-(2.8) being assumed to be known. To save notation, drop all indices and consider the simple model relation

$$E(x^{n\times 1} | \theta) = Y^{n\times q} b^{q\times 1} (\theta)$$

(3.3)

and the parameters
\[ \beta_1 = \mathbb{E} b(\theta), \quad \Lambda = \text{Var} b(\theta), \quad \Phi = \text{Var}(x|\theta). \] (3.4)

**Lemma 3.1** (standard). Assume \( \Phi \) to be p.d.s. The LB estimator of \( b(\theta) \) is

\[
\hat{b} = \Lambda c + (I - \Lambda M) \beta, \tag{3.5}
\]

where \( c_{\text{q} \times 1} \) and \( M_{\text{q} \times \text{q}} \) are defined as

\[
c = Y'(YAY' + \Phi)^{-1} x, \tag{3.6a}
\]

\[
= (I - MA)Y'\Phi^{-1} x, \tag{3.6b}
\]

\[
M = Y'(YAY' + \Phi)^{-1} Y, \tag{3.7a}
\]

\[
= Y'\Phi^{-1} Y(AY'\Phi^{-1} Y + I)^{-1}. \tag{3.7b}
\]

Lemma 3.1 is a classical result in credibility and Bayesian regression. It is stated here in a form that is particularly well suited for our purposes. Notice that neither \( \Lambda \) nor \( Y \) need to be of full rank. In fact, formula (3.5) remains valid if \( \Lambda = 0 \) and even if \( n = 0 \) (no observations) if \( c \) and \( M \) are then interpreted as 0. In both cases \( \hat{b} \) reduces to \( \beta \), of course.

The so-called credibility matrix is defined as

\[
Z = \Lambda M \tag{3.8.a}
\]

\[
= \Lambda Y'(YAY' + \Phi)^{-1} Y \tag{3.8.b}
\]

\[
= I - (AY'\Phi^{-1} Y + I)^{-1}. \tag{3.8.c}
\]

In case \( n > q \) and \( Y \) has full rank \( q \), \( \hat{b} \) can be cast in the form of a credibility weighted mean,

\[
\tilde{b} = Z\hat{b} + (I-Z)\beta, \tag{3.9}
\]

with

\[
\hat{b} = (Y'\Phi^{-1} Y)^{-1} Y'\Phi^{-1} x. \tag{3.10}
\]
The identities in (3.6) and (3.7) turn out to be instrumental in the analysis of the general hierarchical model. They are also computationally convenient; if \( n < q \), use formulas (3.6a) and (3.7a), which involve inversion of an \( n \times n \) matrix; otherwise — provided that \( \Phi \) is easily inverted — use (3.6b) and (3.7b), which require inversion of a \( q \times q \) matrix. The merits of formula (3.9) is on the interpretative rather than the computational side, see e.g. Norberg (1980).

For the sake of completeness, and in the absence of a suitable reference, we prove the lemma. Then the proof of (3.9) will be an easy exercise.

Proof of Lemma 3.1. The LB estimator is (see e.g. Norberg, 1980)

\[
\tilde{b} = \mathbb{E} b(\theta) + \text{Cov}\{b(\theta), x'\}(\text{Var } x)^{-1}(x - \mathbb{E} x). \tag{3.11}
\]

In the presence of assumption (3.3) the moments appearing on the right of (3.11) can be expressed by the parameters in (3.4) as

\[
\mathbb{E} b(\theta) = \beta, \tag{3.12}
\]
\[
\text{Cov}\{b(\theta), x'\} = AY', \tag{3.13}
\]
\[
\text{Var } x = YAY' + \Phi, \tag{3.14}
\]
\[
\mathbb{E} x = Y\beta. \tag{3.15}
\]

The expressions (3.13) and (3.14) are easy consequences of (1.1). On inserting (3.12)-(3.15) in (3.11), we obtain

\[
\tilde{b} = \beta + AQ(x - Y\beta), \tag{3.16}
\]

where

\[
Q = Y'(YAY' + \Phi)^{-1}. \tag{3.17}
\]

The relations (3.16) and (3.17) are equivalent to (3.5), (3.6a), and (3.7a).

It remains to demonstrate the identities asserted in (3.6) and (3.7).
Postmultiplication by $YAY' + \Phi$ in (3.17) yields the equivalent relation

$$QYAY' + Q\Phi = Y'.$$

(3.18)

Further, postmultiplying by $\Phi^{-1}Y$ in (3.18) gives

$$QYAY'\Phi^{-1}Y + QY = Y'\Phi^{-1}Y,$$

from which we solve

$$QY = Y'\Phi^{-1}Y(\Delta Y'\Phi^{-1}Y + I)^{-1}.$$  

(3.19)

As $M = QY$ (compare (3.7a) and (3.17)), (3.19) is just the identity asserted in (3.7). Finally, substitute (3.19) back into (3.18) to obtain

$$Q = (I - MA)Y'\Phi^{-1},$$

which implies that (3.6a) and (3.6b) are identical.

To complete the proof, it must be established that $\Delta Y'\Phi^{-1}Y + I$ is always invertible when $\Phi$ is. This follows by use of the identity $|AB + I| = |BA + I|$, valid for any pair of matrices $A^{rxs}$ and $B^{sxr}$ (Zellner, 1971, p. 231). Putting $A = \Delta Y'$ and $B = \Phi^{-1}Y$, one gets

$$|\Delta Y'\Phi^{-1}Y + I| = |\Phi^{-1}YAY' + I|$$

$$= |\Phi^{-1}||YAY' + \Phi|.$$  

(3.20)

Since $\Phi$ is p.d.s., so is both $\Phi^{-1}$ and $YAY' + \Phi$, hence both determinants in (3.20) are strictly positive. It can be concluded that $\Delta Y'\Phi^{-1}Y + I$ has a non-zero determinant and, therefore, is invertible. \[\Box\]

The hierarchical case, $s > 1$. We now turn to the genuine hierarchical model and set out by demonstrating some useful recursive relations. On inserting (2.2) in the definition of the dispersion
matrices in (2.8) and recalling the independence assumption (iii), it is seen that

$$\Phi_{i_1 \ldots i_r} = \text{diag}[E \text{Var}(x_{i_1 \ldots i_r} | \theta_{i_{r+1}}) | \theta_{r+1} = 1, \ldots, m_i \ldots i_r].$$ (3.21)

Applying (1.1) to the conditional variances in (3.21), noticing that $\theta_r = \theta_{r+1}$, and then using (2.5), (2.7), and (2.8), yields

$$E \text{Var}(x_{i_1 \ldots i_r} | \theta_r) = E[\text{Var}(E(x_{i_1 \ldots i_{r+1}} | \theta_r) | \theta_{r+1})]$$

$$= E \text{Var}(Y_{i_1 \ldots i_{r+1}} | \theta_r)$$

$$+ E \text{Var}(x_{i_1 \ldots i_r} | \theta_r)$$

$$= Y_{i_1 \ldots i_{r+1}} \Lambda_{r+1} Y_{i_1 \ldots i_{r+1}} + \Phi_{i_1 \ldots i_{r+1}}.$$ (3.22)

Combination of (3.21) and (3.22) gives the recursive formula

$$\Phi_{i_1 \ldots i_r} = \text{diag}(Y_{i_1 \ldots i_r} \Lambda_{r+1} Y_{i_1 \ldots i_r} + \Phi_{i_1 \ldots i_{r+1}})$$

$$= (I-M_{i_1 \ldots i_r} \Lambda_{r+1})^{-1} Y_{i_1 \ldots i_r}$$

(3.23)

For each $i_1 \ldots i_r$ define, by virtue of (3.6) and (3.7),

$$c_{i_1 \ldots i_r} = Y_{i_1 \ldots i_r} \Phi_{i_1 \ldots i_r}^{-1} x_{i_1 \ldots i_r}.$$ (3.24a)

$$M_{i_1 \ldots i_r} = Y_{i_1 \ldots i_r} \Phi_{i_1 \ldots i_r}^{-1} Y_{i_1 \ldots i_r}.$$ (3.25a)

(The positive definiteness of each $\Phi_{i_1 \ldots i_r}$, which is necessary for the identities in (3.24) and (3.25) to hold true, is a trivial consequence of (3.23) and the positive definiteness of each $\Phi_{i_1 \ldots i_s}$ at
at stage s.) By use of (2.2), (2.3), (3.23), and (3.24a), it follows that

\[ Y'_i \ldots i_r \Phi^{-1} i_1 \ldots i_r x'_i \ldots i_r = \sum_{i_{r+1}} (Y'_i \ldots i_r i_{r+1} + \Phi i_1 \ldots i_r i_{r+1})^{-1} x'_i \ldots i_r i_{r+1} \]

\[ = \sum_{i_{r+1}} c_{i_1} \ldots i_r i_{r+1} \]

whence, by substitution in (3.24b),

\[ c_{i_1} \ldots i_r = (I-M_{i_1} \ldots i_r) i_{r+1} c_{i_1} \ldots i_r i_{r+1} \]

Similarly, (3.25) gives

\[ Y'_i \ldots i_r \Phi^{-1} i_1 \ldots i_r Y'_i \ldots i_r = \sum_{i_{r+1}} M_{i_1} \ldots i_r i_{r+1} \]

and

\[ M_{i_1} \ldots i_r = \sum_{i_{r+1}} M_{i_1} \ldots i_r i_{r+1} (\Lambda_{i_1} i_{r+1} M_{i_1} \ldots i_r i_{r+1} + I)^{-1} \]

We are now in a position to construct LB estimators of all random effects \( b_{i_1} \ldots i_r \) given by (2.4). Consider first the problem of estimating effects at stage 1. Application of Lemma 3.1 to the regression (2.5) with \( r = 1 \), gives:

**Lemma 3.2.** The LB estimator of \( b_{i_1} \) is

\[ \tilde{b}_{i_1} = \Lambda_{i_1} c_{i_1} + (I-\Lambda_{i_1} M_{i_1}) \beta \]

\[ = \sum_{i_{1+1}} \Lambda_{i_1} i_{1+1} c_{i_1} \ldots i_{1+1} \]

where \( c_{i_1} \) and \( M_{i_1} \) are defined by (3.24) and (3.25).
To construct the LB estimator \( \tilde{b}_{i_1 \ldots i_r} \) of a general \( b_{i_1 \ldots i_r} \), we call on a beautiful result proved in the univariate case by Jewell (1975) and generalized to the regression case by Sundt (1980).

Lemma 3.3 (Jewell/Sundt). LB estimators at neighbouring stages are related by the formula

\[
\tilde{b}_{i_1 \ldots i_r} = A_r c_{i_1 \ldots i_r} + (I - A_r M_{i_1 \ldots i_r}) \tilde{b}_{i_1 \ldots i_{r-1}}. 
\]  

(3.31)

By comparison of formulas (3.30) and (3.31), it is seen that \( \tilde{b}_{i_1 \ldots i_r} \) is obtained by formally treating \( b_{i_1 \ldots i_r} \) as a stage 1 effect in the local hierarchy of \( c_{i_1 \ldots i_{r-1}} \), with \( \tilde{b}_{i_1 \ldots i_{r-1}} \) in the place of \( \beta \).

Sundt proves (3.31) by verifying that the expression on the right hand side satisfies the normal equations determining \( \tilde{b}_{i_1 \ldots i_r} \).

We shall give an alternative, direct argument, which may present an interest of its own.

Proof of Lemma 3.3. Since LB estimators depend only on the 1st and 2nd order moments of the distributions, it suffices to prove the result for one particular choice of distributions with the required moment structure. We pick the following. Suppose that all the latent quantities \( \theta_{i_1 \ldots i_r} \) are mutually independent, each \( \theta_{i_1 \ldots i_r} \) having a \( q \)-variate normal distribution with mean \( \theta_i \) and variance \( A_r \), and that, conditional on \( \theta_s \), each \( x_{i_{1s}} \) has a \( n_{i_{1s}} \)-variate normal distribution with mean \( Y_{i_1 \ldots i_s} b_{i_1 \ldots i_s} \) and variance \( \phi_{i_1 \ldots i_s} \), where \( b_{i_1 \ldots i_s} = \beta + \theta_{i_1} + \cdots + \theta_{i_s} \) (and, consequently, \( b_{i_1 \ldots i_r} = \beta + \theta_{i_1} + \cdots + \theta_{i_r} \)).

In this model the joint distribution of the \( b_{i_1 \ldots i_r} \)'s and the \( x_{i_1 \ldots i_s} \)'s is multivariate normal, hence the conditional mean
of \( b_{i_1 \ldots i_r} \), given the \( x_{i_1 \ldots i_s} \)'s, is a linear function (Anderson, 1958, p.29). Now it is well known that this conditional mean is the unrestricted Bayes estimator of \( b_{i_1 \ldots i_r} \), and - being of the form (3.2) - it is also LB. Therefore, in the present case

\[
\tilde{b}_{i_1 \ldots i_r} = E(\tilde{b}_{i_1 \ldots i_r} | x_{i_1})
\]

(3.32a)

\[
= E[E(\tilde{b}_{i_1 \ldots i_r} | \theta_{r-1}, x_{i_1}) | x_{i_1}]
\]

(3.32b)

the last equation being a consequence of assumption (iii). Under the present assumptions the inner expectation in (3.32b) assumes the form

\[
E(\tilde{b}_{i_1 \ldots i_r} | \theta_{r-1}, x_{i_1 \ldots i_r}) = \Lambda_r c_{i_1 \ldots i_r} + (I-\Lambda_r m_{i_1 \ldots i_r}) b_{i_1 \ldots i_{r-1}}
\]

(3.33)

This assertion follows by noting that, conditional on \( \theta_{r-1} \), the local hierarchy within \( C_{i_1 \ldots i_{r-1}} \) has the same distributional structure as the global hierarchy, with \( b_{i_1 \ldots i_{r-1}} \), \( \Lambda_r \), \( \ldots \), \( \Lambda_s \), \( \Phi_{i_1 \ldots i_s} \) playing the roles of \( \beta \), \( \Lambda_1 \), \( \ldots \), \( \Lambda_s \), \( \Phi_{i_1 \ldots i_s} \). It follows that the conditional mean on the left of (3.33) is - formally - the LB estimator of \( b_{i_1 \ldots i_r} \) in the conditional model, which, according to Lemma 3.2, is the expression on the right of (3.33). Now, upon substituting (3.33) in (3.32b) and then using (3.32a), this time at stage \( r-1 \), we arrive at (3.31). \( \Box \)

The results above are now summarized as a complete algorithm for computation of LB estimators.

**Theorem 3.4.** LB estimators are found for all random effects \( b_{i_1 \ldots i_r} \) by first calculating all the quantities \( c_{i_1 \ldots i_r} \) and \( m_{i_1 \ldots i_r} \) by the recursive relations (3.27) and (3.29), starting from stage \( s \), and then calculating all the \( \tilde{b}_{i_1 \ldots i_r} \) by the recursive relation (3.31), starting from (3.30) at stage 1.
It is noteworthy that the LB estimators do not depend on the weighting $W$.

Comments. Sundt's (1980) setting is different from the present one; in terms of our definitions, he focuses attention at one stage class $C_1 \ldots C_s$ and works with statistics $x_1', \ldots, x_s'$, where $x_s = x_1' \ldots x_s'$ and each $x_r' : r=1, \ldots, s-1$; consists of the stage $r+1$ statistics $x_1' \ldots x_r' h_{r+1}$; $h_{r+1} \neq i_{r+1}$. The starting point of our model is the assembly of all observables $x_1' \ldots x_s'$ in unit risk classes, and Theorem 3.4 explains how to actually calculate LB estimators from these data.

The proof given here of Lemma 3.3 is elementary and constructive. The technique can be carried over to a number of complex LB estimation problems that otherwise would require either heavy algebra or refined Hilbert space methods (de Vylder, 1976) for their solution. It serves, inter alia, to justify the switch from conditional to unconditional moments in Taylor's proof of Lemma 3.3 for the univariate case (Taylor, 1979, Theorem 4, item 1).

If $Y_{i_1} \ldots i_r$ is of rank $q$, then (3.31) can be rewritten in the form of a credibility weighted mean,

$$\bar{b}_{i_1} \ldots i_r = Z_{i_1} \ldots i_r \bar{b}_{i_1} \ldots i_r + (I-Z_{i_1} \ldots i_r) \bar{b}_{i_1} \ldots i_{r-1},$$

(3.34)

where

$$\bar{b}_{i_1} \ldots i_r = (Y_{i_1} \ldots i_r \phi^{-1}_{i_1} \ldots i_r Y_{i_1} \ldots i_r)^{-1} Y_{i_1} \ldots i_r \phi^{-1}_{i_1} \ldots i_r \bar{b}_{i_1} \ldots i_r \phi^{-1}_{i_1} \ldots i_r$$

(3.35)

and

$$Z_{i_1} \ldots i_r = A_r M_{i_1} \ldots i_r.$$  

(3.36)

The demonstration of these formulas is straightforward by virtue of (3.8)-(3.10). From a computational point of view (3.31) is more convenient than (3.34). Besides, the case where $Y_{i_1} \ldots i_r$ is not of full rank $q$ is of great practical relevance: insurance companies are constantly facing the problem of fixing premiums for new written


- 17 -
business. If a new risk or class of risks can be classified with other risks, possibly in a hierarchical system as visualized in Figure 1, then its premium can be determined in a rational manner within the framework of a random effect model, the point being that the model specifies a relationship between the risks. An insurance company working with a fixed effect model, as advocated by Gerber (1982), would be at a loss when confronted with the problem of assessing a risk with no claims experience of its own. For a further discussion of these matters, see Norberg (1985).

**Linear Bayes risks.** The LB risk for the problem of estimating $b_{i_1 \ldots i_r}$ is easily shown to be (Norberg, 1980)

$$\tilde{\rho}_{i_1 \ldots i_r} = \text{tr}(W \Delta_{i_1 \ldots i_r}) ,$$  \hspace{1cm} (3.37)

where $\Delta_{i_1 \ldots i_r}$ is the LB risk matrix given by

$$\Delta_{i_1 \ldots i_r} = \text{Var} b_{i_1 \ldots i_r} - \text{Cov}(b_{i_1 \ldots i_r}, x') (\text{Var} x_{i_1})^{-1} \text{Cov}(x_{i_1}, b_{i_1 \ldots i_r}).$$  \hspace{1cm} (3.38)

Notice that the LB risks depend on the weighting matrix $W$, whereas the LB risk matrices - just like the LB estimators - do not.

The following theorem states that the LB risks in (3.37) can be calculated by a recursive procedure once the matrices $M_{i_1 \ldots i_r}$ have been found by use of the recursion (3.29).

**Theorem 3.5.** The LB risk matrices in (3.38) can be calculated by the recursive relation

$$\Delta_{i_1 \ldots i_r} = (I - Z_{i_1 \ldots i_r}) \{ \Delta_{i_1 \ldots i_r - 1} (I - Z_{i_1 \ldots i_r})' + \Lambda_{i_r} \},$$  \hspace{1cm} (3.39)

starting from stage 1 with

$$\Delta_{i_1} = (I - Z_{i_1}) \Lambda_{i_1} ,$$  \hspace{1cm} (3.40)

the matrices $Z_{i_1 \ldots i_r}$ are given by (3.29) and (3.36).
Proof: To demonstrate (3.40), substitute (3.13) and (3.14) with $b_1, x_1, Y_1, \Lambda_1, \Phi_1$ in the roles of $b(\theta), x, Y, \Lambda, \Phi$ into (3.38) for $r=1$. This gives

\[ \Delta_i = \Lambda_i - \Lambda_i Y_i' \left( Y_i \Lambda_i Y_i' + \Phi_i \right)^{-1} Y_i \Lambda_i, \]

which, by (3.25a) and (3.36), is just formula (3.40).

To prove (3.39), apply the same trick as in the proof of Lemma 3.3. In the normal model specified there the LB risk matrix coincides with the unrestricted Bayes risk matrix, that is,

\[ \Delta_i = \mathbb{E} \text{Var}(b_i | x_i). \tag{3.41} \]

Applying principle (1.1) to the conditional variance and arguing as in the proof of Lemma 3.3, rewrite (3.41) as

\[
\Delta_i = \mathbb{E} \text{Var}(b_i | \theta_{r-1}, x_i) | x_i \\
+ \mathbb{E} \left[ \text{Var}(b_i | \theta_{r-1}, x_i) | x_i \right] \\
= \mathbb{E} \text{Var}(b_i | \theta_{r-1}, x_i ... i_r) | x_i \\
+ \mathbb{E} \mathbb{E} \left[ \text{Var}(b_i | \theta_{r-1}, x_i ... i_r) | \theta_{r-1} \right]. \tag{3.42}
\]

Now substitute (3.33) in the first term in (3.42) and observe that the inner expectation in the second term is - formally - the Bayes risk for the problem of estimating $b_i ... i_r$ as a stage effect in the local hierarchy of $C_i, ... i_r$, conditional on $\theta_{r-1}$, so that in principle (3.40) applies. It follows that

\[
\Delta_i = \mathbb{E} \text{Var} \left[ (I- \Lambda^-) b_i ... i_r-1 | x_i \right] + (I- \Lambda^-) \Lambda_r, \]

which by (3.36) and (3.41) is the same as (3.39). \qed
4. Estimation of parameters; empirical linear Bayes estimators

For the purpose of parameter estimation only classes with observations are relevant, and so all \( n_{i_1\ldots i_s} \) 's are now taken to be greater than 0.

Estimation of the second order moments in (2.7) and (2.8) is in general not feasible unless further structure is imposed on the matrices \( \Phi_{i_1\ldots i_s} \). Therefore, in the present section it will be assumed - as is standard in regression theory - that there exist a positive function \( v \) and nonrandom, known p.d.s. matrices \( P_{i_1\ldots i_s} \) such that

\[
\text{Var}(x_{i_1\ldots i_s} | \theta_s) = v(\theta_{i_1}, \ldots, \theta_{i_1\ldots i_s})P_{i_1\ldots i_s}^{-1}
\]

Then

\[
\Phi_{i_1\ldots i_s} = \Phi P_{i_1\ldots i_s}^{-1}
\]

where

\[
\Phi = \text{Ev}(\theta_{i_1}, \ldots, \theta_{i_1\ldots i_s}),
\]

and the relevant set of parameters becomes

\[
\beta, \Lambda_1, \ldots, \Lambda_s, \phi.
\]

(4.1)

The starting point for constructing estimators are the expressions

\[
E x_{i_1\ldots i_s} = Y_{i_1\ldots i_s} \beta,
\]

(4.2)

\[
E(x_{i_1\ldots i_s} x_{j_1\ldots j_s}') = Y_{i_1\ldots i_s}(\beta\beta' + \sum_{r=1}^s \delta_{i_1\ldots i_r, j_1\ldots j_r} \Lambda_r)Y_{j_1\ldots j_s}' + \delta_{i_1\ldots i_s, j_1\ldots j_s} \Phi P_{i_1\ldots i_s}^{-1}
\]

(4.3)

where \( \delta_{i_1\ldots i_r, j_1\ldots j_r} \) equals 1 if \( (i_1, \ldots, i_r) = (j_1, \ldots, j_r) \) and 0 otherwise. Relation (4.3) is obtained by repeated use of principle (1.1).
A class of unbiased estimators of the parameters in (4.1) is constructed as follows. For each \( i_1, \ldots, i_s \), denote by \( r_{i_1 \ldots i_s} \) the rank of the regressor matrix \( Y_{i_1 \ldots i_s} \), and let \( L(Y_{i_1 \ldots i_s}) \) be the \( r_{i_1 \ldots i_s} \)-dimensional linear space spanned by the columns of \( Y_{i_1 \ldots i_s} \). Let \( A_{i_1 \ldots i_s} \) and \( B_{i_1 \ldots i_s} \) be matrices of order \( n_{i_1 \ldots i_s} \times q \) and \( n_{i_1 \ldots i_s} \times (n_{i_1 \ldots i_s} - r_{i_1 \ldots i_s}) \), respectively, such that the columns of \( A_{i_1 \ldots i_s} \) span \( L(Y_{i_1 \ldots i_s}) \) and the columns of \( B_{i_1 \ldots i_s} \) is a basis of the orthocomplement of \( L(Y_{i_1 \ldots i_s}) \).

By (4.2), a straightforward estimator of \( \beta \) is

\[
\beta^* = \left( \sum_{i_1, \ldots, i_s} A'_{i_1 \ldots i_s} Y_{i_1 \ldots i_s} \right)^{-1} \left( \sum_{i_1, \ldots, i_s} A'_{i_1 \ldots i_s} x_{i_1 \ldots i_s} \right). \tag{4.4}
\]

For each \( i_1, \ldots, i_s \), introduce the statistic

\[
w_{i_1 \ldots i_s} = x'_{i_1 \ldots i_s} B_{i_1 \ldots i_s} \right)^{-1} x_{i_1 \ldots i_s}.
\]

By (4.3) and fact that the columns of \( Y_{i_1 \ldots i_s} \) are orthogonal to those of \( B_{i_1 \ldots i_s} \),

\[
E w_{i_1 \ldots i_s} = \phi \left( \sum_{i_1, \ldots, i_s} \text{tr}(B_{i_1 \ldots i_s} B'_{i_1 \ldots i_s} P^{-1}) \right).
\]

Therefore, an unbiased estimator of \( \phi \) is

\[
\phi^* = \left( \sum_{i_1, \ldots, i_s} \frac{\text{tr}(B_{i_1 \ldots i_s} B'_{i_1 \ldots i_s} P^{-1})}{w_{i_1 \ldots i_s}} \right)^{-1} \left( \sum_{i_1, \ldots, i_s} \frac{w_{i_1 \ldots i_s}}{w_{i_1 \ldots i_s}} \right). \tag{4.6}
\]

with \( w_{i_1 \ldots i_s} \) defined by (4.5).
Estimators of $\Lambda_1, \ldots, \Lambda_s$ are based on the quantities

$$G_r = \{(\sum_{i_1, \ldots, i_s} (\sum_{j_{r+1}, \ldots, j_s} A_{i_1}^{j_1} \cdots i_r^{j_r} \cdots j_s^{j_{r+1}} Y_{i_1}^{j_1} \cdots i_r^{j_r} \cdots j_s^{j_{r+1}} i_s)^{-1} \}

where the sum $\sum_{(r)}$ extends over all $i_1, \ldots, i_s$ for which the indicated inversions are valid. It is obvious how to interpret the summations in the cases $r=0$ and $r=s$; for $r=s$ the sum over $j_{r+1}, \ldots, j_s$ includes only $i_1, \ldots, i_s$. It is seen that $G_r$ involves all products of observations in different stage $r+1$ classes within the same stage $r$ class. Thus, all products $x_{i_1}^{j_1} \cdots x_{i_s}^{j_s}$ are utilized in the estimation. Using (4.3), it is easily checked that

$$EG_r = \begin{cases} \beta \beta' & ; r=0; \\ \beta \beta' + \Lambda_1 + \cdots + \Lambda_r & ; r=1, \ldots, s-1; \\ \beta \beta' + \Lambda_1 + \cdots + \Lambda_s + \phi H & ; r=s; \end{cases}$$

with

$$H = \phi \{(\sum_{i_1, \ldots, i_s} (A_{i_1}^{i_1} \cdots i_s^{i_s})^{-1} A_{i_1}^{i_1} \cdots i_s^{i_s} P_{i_1}^{-1} \cdots i_s^{i_s} A_{i_1}^{i_1} \cdots i_s^{i_s})^{-1} \}

It follows that a set of symmetric unbiased estimators of the $\Lambda_r$ is given by

$$\Lambda_r^* = \frac{1}{2} \{(G_r + G_r'^{-1}) - G_r'^{-1} - G_r'^{-1} - \delta_{r,s}^* (H+H'^{-1})\}; \quad r=1, \ldots, s;$$

where the $G_r$ and $H$ are defined by (4.7) and (4.8), respectively. Note that the matrices in (4.9) need not be positive definite. Thus the old problem of negative estimates of variance components may arise.
The estimator in (4.4) is well defined if and only if the regressor matrix of all the observation units, \( Y = (Y_1', \ldots, Y_m')' \), is of full rank \( q \). The estimator in (4.6) is well defined if at least one \( n_{1'} \ldots i_{1'} = r_{1'} \ldots i_{1'} \) is greater than 0. Both these requirements are very weak. The estimator \( A_r^* \) in (4.9) is well defined if \( G_r \) in (4.7) is, which requires that there exists at least one stage \( r \) class \( C_{i_1} \ldots i_r \) with a stage \( r+1 \) subclass \( C_{i_1} \ldots i_r i_{r+1} \) such that the matrix obtained by deleting the block \( Y_{i_1} \ldots i_r i_{r+1} \) from \( Y_{i_1} \ldots i_r \) is of rank \( q \). If any one of these requirements is not fulfilled, the corresponding parameter is not identifiable from the observations.

The question of how to choose the matrices \( A_{i_1} \ldots i_s \) and \( B_{i_1} \ldots i_s \) shall not be discussed in any detail here. If the design is balanced, it is, of course, optimal to take all \((A_{i_1} \ldots i_s, B_{i_1} \ldots i_s)\) equal (Norberg, 1977), but this case is of little interest in insurance and other non-experimental fields. In general no uniformly optimal choice can be made. A "natural" choice, paying regard to the amounts of information contained in the individual risk classes, is \( A_{i_1} \ldots i_s = Y_{i_1} \ldots i_s P_{i_1} \ldots i_s \) and, if \( r_{i_1} \ldots i_s = q_{i_1} \ldots i_s \), \( B_{i_1} \ldots i_s \)

\[
P_{i_1} \ldots i_s Y_{i_1} \ldots i_s (Y_{i_1} \ldots i_s P_{i_1} \ldots i_s Y_{i_1} \ldots i_s)^{-1} Y_{i_1} \ldots i_s P_{i_1} \ldots i_s.
\]

Then the expression in (4.4) becomes

\[
(Y_{i_1} \ldots i_s P_{i_1} \ldots i_s Y_{i_1} \ldots i_s)^{-1} \sum_{i_1, \ldots, i_s} Y_{i_1} \ldots i_s P_{i_1} \ldots i_s
\]

which would be the best linear unbiased estimator if \( A_1 = \cdots = A_s = 0 \), and (4.5) reduces to the ordinary sample estimate of
\[ v(\theta_{i_1 \ldots i_S}, \ldots, \theta_{i_1 \ldots i_S}) \text{ multiplied by the degrees of freedom,} \]

\[ n_{i_1 \ldots i_S} - q_{i_1 \ldots i_S}. \text{ If the second order moments were known, then} \]

the best linear unbiased estimator of \( \beta \) would be

\[ \{Y'(\text{Var } x)^{-1} Y\}^{-1} Y'(\text{Var } x)^{-1} x, \quad (4.10) \]

where \( x = (x'_1, \ldots, x'_m)' \) and \( Y = (Y'_1, \ldots, Y'_m)' \). By independence
between classes at stage 1, the expression in (4.10) is equal to

\[ \left( \sum_{i_1 = 1}^{m} Y'_{i_1} \Phi_{i_1}^{-1} Y_{i_1} \right)^{-1} \sum_{i_1 = 1}^{m} Y'_{i_1} \Phi_{i_1}^{-1} x_{i_1}, \]

or, by inserting (3.26) and (3.28),

\[ \left( \sum_{i_1 = 1}^{m} M_{i_1} \right)^{-1} \sum_{i_1} c_{i_1}, \quad (4.11) \]

which can be calculated recursively by (3.27) and (3.29). A genuine
estimator \( \beta^* \) is obtained by replacing all second order moments in
(4.11) by their estimators.

The estimators defined by (4.4), (4.6), and (4.9) are consistent as \( m \) - the number of stage 1 classes - increases, provided
that the ranks \( r_{i_1 \ldots i_S} \) and the precisions \( p_{i_1 \ldots i_S} \) do not tend
to too small values, roughly speaking. This result is, however,
mainly of theoretical interest to insurance people because in their
applications there will usually be a limited number of classes at
stage 1.

Finally, empirical LB estimators of random effects are ob-
tained upon replacing the parameters occurring in the formulas in
section 3 by their estimators in (4.4), (4.6), and (4.9).
References


