Study Note on "Greatest Accuracy" Credibility

We hope that the reader appreciates the similarity between what Frees refers to as "greatest accuracy" credibility and linear mixed models. The purpose of this Study Note is two-fold:

- To provide the traditional Bühlmann-Straub formulas commonly used in insurance. (In particular, this includes formulas for the variance components which are not emphasized in Frees.)
- To show by example the similarities and differences between the Bühlmann-Straub formulas and a linear random effects model.

Credibility procedures are in common use in insurance as a way to blend particular estimates based on highly relevant but very thin data with global estimates based on less relevant but less noisy data.

We will explain this with a numerical example similar to (but not identical to) the dental insurance example in Frees. The subject for which we estimate random effects in this case are employers, but they could be geographic territories, individual policyholders, etc., in other contexts.

Employer	Year	Count	Avg Cost
A	1	8	500
A	2	12	400
A	3	10	900
B	1	11	100
B	2	11	200
B	3	11	300
C	1	8	650
C	2	7	900
C	3	7	1450
D	1	50	350
D	2	55	400
D	3	60	500

The usual actuarial (Bühlmann-Straub) estimators[1] for expected value of the conditional variance $\hat{\sigma}^2$ (E(Var(y|α)) in Frees) and the variance of the conditional expectations $\hat{\tau}^2$ (Var(E(y|α)) in Frees) are somewhat different from what is given in Frees, namely:

$$\hat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{n-1}\sum_{j=1}^{n} w_{ij}(X_{ij} - X_i)^2$$

where X_{ij} represents the average cost for the ith employer in the jth year, X_i the employee-count weighted average cost for employer i across all year, w_{ij} the number of covered employees at the ith

[1] The formulas on this page and the following page are adapted from Bühlmann & Gisler,. *A Course in Credibility Theory and its Applications*, Springer, 2005, pp. 94-96

employer in the j[th] year, M the number of employers, and n the number of years. Note that this is an estimate of the variance for an individual, so that the model expects that the variance in per-employee average cost for an employer for a year is inversely proportional to the number of employees.

The estimate of $\hat{\tau}^2$ is the maximum of 0 (and if this is realized, it means there is likely no difference between the groups) and

$$\frac{1}{U}\left\{\left[\frac{1}{M-1}\sum_{i=1}^{M}\frac{w_i}{W}(X_i-\bar{X})^2\right]-\frac{\hat{\sigma}^2}{W}\right\}$$

where U is given by

$$U=\frac{1}{(M-1)}\sum_{i=1}^{M}\frac{w_i}{W}(1-\frac{w_i}{W})$$

and w_i is the sum of employee counts at employer i across all years, and W is the sum of all the w_i, and $\bar{X}$ is the average of the X_i weighted by the w_i.

Applying these formulas to the given data yields $\hat{\sigma}^2$= 598,919.5 and $\hat{\tau}^2$= 55,310.5. The ratio is K=10.83, which implies that the best linear unbiased prediction for a given employer's average cost in the following year is given by a weighted average of Z_i times the employer's historical average cost X_i and (1- Z_i) times the "complement of credibility" $\hat{\mu}_{new}$, where Z_i is given by $\frac{w_i}{w_i+K}$, and $\hat{\mu}_{new}$ is given by $\frac{\sum Z_i X_i}{\sum Z_i}$. In this case, $\hat{\mu}_{new}$ is 530.1, and a table of credibilities (Z_i) and best estimates is given by:

Employer	Employee-Years	Historical Avg Cost	Credibility	Complement	Best Prediction
A	30	593.3	73.5%	530.1	576.6
B	33	200.0	75.3%	530.1	281.6
C	22	984.1	67.0%	530.1	834.3
D	165	421.2	93.8%	530.1	427.9

Note that both the historical average cost and the best prediction columns have weighted averages of 462.2. One would have reason to be worried about a process that did not do this. It is important to note, however, that the complement of credibility (which can be interpreted as the expected average cost of a previously unknown employer) is 530.1. This is because employers are weighted fairly evenly, rather than by size, in determining this number, which makes sense, because the typical employer is small (even if most employees work for large employers). This is simply an example of the family paradox. (Ask a group of people what size family [them plus siblings] they are from. The average is likely to be much larger than the average family size [measured as number of children], because larger families are better represented in the sample!)

It's also important to note that the credibility Z_i is computed based on the number of employees (number of exposure units, or in some sense the number of expected claims), in distinction to Frees, where is computed based on the number of claims. It is better in both theory and practice to use the number of expected claims. In theory, because that agrees with linear random effects models. In practice, because one needs to recognize the superior experience of a policyholder who never has a claim, whereas giving them zero credibility treats them like an average policyholder!

Now, let us compare this with using the lmer function in the R package lme4[2]:

```
model1<-lmer(avgcost ~ 1 + (1 | employer),datadental,weights = insdcount)
```

This is how one specifies a random intercepts model in lmer.

Let's look at the basic output:

```
summary(model1)
Linear mixed model fit by REML ['lmerMod']
Formula: avgcost ~ 1 + (1 | employer)
   Data: datadental
Weights: insdcount

REML criterion at convergence: 156.7

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-0.8034 -0.6887 -0.2569  0.3013  1.9649 

Random effects:
 Groups   Name        Variance Std.Dev.
 employer (Intercept)  80631   284.0   
 Residual             612538   782.6   
Number of obs: 12, groups:  employer, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)    534.8      156.1   3.426
```

The estimate of the residual variance is similar to that from the Bühlmann-Straub formulas. The REML procedure and the Bühlmann-Straub procedure use different estimators, though they are estimating the same model. The estimate of the fixed effect for the intercept corresponds to the estimate of $\hat{\mu}_{new}$ in the Bühlmann-Straub setup, and is within 1%. Again, they do not agree precisely because estimators used are different.

Note that the model object does not explicitly provide the random effect estimates that give best linear unbiased predictions, but these are available via the ranef function:

```
ranef(model1)
```

[2] Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker, "Fitting Linear Mixed-Effects Models Using lme4", *Journal of Statistical Software*, 67(1), 2015, 1-48. doi:10.18637/jss.v067.i01.

```
$employer
  (Intercept)
A    46.72135
B  -272.13364
C   333.98227
D  -108.56998
```

Adding these to the intercept of 534.8 yields lmer's BLUPs for each employer.

Note that you could easily reproduce the output of ranef with just the information in the model summary. This is because the BLUPs depend only on the variance components and on the fixed effect intercept, and they depend on them in precisely the same way the BLUPs depend on $\hat{\sigma}^2$, $\hat{\tau}^2$, and $\hat{\mu}_{new}$. The estimates of the variance components are the only difference between the processes. (The fixed-effects intercept is different from $\hat{\mu}_{new}$ only because each is calibrated to make the weighted average best predictions equal the weighted average historical costs.) In fact, we can reproduce the previous table using lmer instead of Bühlmann-Straub:

Employer	Employee-Years	Historical Avg Cost	Credibility	Complement	Best Prediction
A	30	593.3	79.8%	534.8	581.5
B	33	200.0	81.3%	534.8	262.7
C	22	984.1	74.3%	534.8	868.8
D	165	421.2	95.6%	534.8	426.2

Also note that only the ratio K of variance components and the complement of credibility matter, at the end of the day. Everything else is simple arithmetic. Insurance professionals sometimes choose a value of K based on long experience across multiple datasets and use it across multiple similar problems without reestimating it. The underlying motivation is that variance estimates are noisy, and it's useful to include as much information as possible in arriving at K, even beyond the dataset under consideration. On occasion, insurance professionals also occasionally modify the complement of credibility to take into account knowledge that is not contained in the data. These can be reasonable thing to do, but must be appropriately documented as they may have impacts on the estimates that the audience for the results of the analysis might not expect. In particular, any change to the complement will cause the historical average cost and the average best prediction to differ and should be carefully noted.