

# Credibility theory features of **actuar**

Christophe Dutang  
ISFA, Université Claude Bernard Lyon 1

Vincent Goulet  
École d'actuariat, Université Laval

Xavier Milhaud  
ISFA, Université Claude Bernard Lyon 1

Mathieu Pigeon  
École d'actuariat, Université Laval

## 1 Introduction

Credibility models are actuarial tools to distribute premiums fairly among a heterogeneous group of policyholders (henceforth called *entities*). More generally, they can be seen as prediction methods applicable in any setting where repeated measures are made for subjects with different risk levels.

The credibility theory features of **actuar** consist of matrix `hachemeister` containing the famous data set of [Hachemeister \(1975\)](#) and function `cm` to fit hierarchical (including Bühlmann, Bühlmann-Straub) and regression credibility models. Furthermore, function `simul` can simulate portfolios of data satisfying the assumptions of the aforementioned credibility models; see the "simulation" vignette for details.

## 2 Hachemeister data set

The data set of [Hachemeister \(1975\)](#) consists of private passenger bodily injury insurance average claim amounts, and the corresponding number of claims, for five U.S. states over 12 quarters between July 1970 and June 1973. The data set is included in the package in the form of a matrix with 5 rows and 25 columns. The first column contains a state index, columns 2–13 contain the claim averages and columns 14–25 contain the claim numbers:

```

> data(hachemeister)
> hachemeister

```

	state	ratio.1	ratio.2	ratio.3	ratio.4	ratio.5	ratio.6	ratio.7
[1,]	1	1738	1642	1794	2051	2079	2234	2032
[2,]	2	1364	1408	1597	1444	1342	1675	1470
[3,]	3	1759	1685	1479	1763	1674	2103	1502
[4,]	4	1223	1146	1010	1257	1426	1532	1953
[5,]	5	1456	1499	1609	1741	1482	1572	1606

	ratio.8	ratio.9	ratio.10	ratio.11	ratio.12	weight.1	weight.2
[1,]	2035	2115	2262	2267	2517	7861	9251
[2,]	1448	1464	1831	1612	1471	1622	1742
[3,]	1622	1828	2155	2233	2059	1147	1357
[4,]	1123	1343	1243	1762	1306	407	396
[5,]	1735	1607	1573	1613	1690	2902	3172

	weight.3	weight.4	weight.5	weight.6	weight.7	weight.8	weight.9
[1,]	8706	8575	7917	8263	9456	8003	7365
[2,]	1523	1515	1622	1602	1964	1515	1527
[3,]	1329	1204	998	1077	1277	1218	896
[4,]	348	341	315	328	352	331	287
[5,]	3046	3068	2693	2910	3275	2697	2663

	weight.10	weight.11	weight.12
[1,]	7832	7849	9077
[2,]	1748	1654	1861
[3,]	1003	1108	1121
[4,]	384	321	342
[5,]	3017	3242	3425

### 3 Hierarchical credibility model

The linear model fitting function of R is named `lm`. Since credibility models are very close in many respects to linear models, and since the credibility model fitting function of **actuar** borrows much of its interface from `lm`, we named the credibility function `cm`.

Function `cm` acts as a unified interface for all credibility models supported by the package. Currently, these are the unidimensional models of [Bühlmann \(1969\)](#) and [Bühlmann and Straub \(1970\)](#), the hierarchical model of [Jewell \(1975\)](#) (of which the first two are special cases) and the regression model of [Hachemeister \(1975\)](#), optionally with the intercept at the barycenter of time ([Bühlmann and Gisler, 2005](#), Section 8.4). The modular design of `cm` makes it easy to add new models if desired.

This subsection concentrates on usage of `cm` for hierarchical models.

There are some variations in the formulas of the hierarchical model in the literature. We compute the credibility premiums as given in [Bühlmann and Jewell \(1987\)](#) or [Bühlmann and Gisler \(2005\)](#). We support three types of estimators of the between variance structure parameters: the unbiased estimators of [Bühlmann and Gisler \(2005\)](#) (the default), the slightly different version of

Ohlsson (2005) and the iterative pseudo-estimators as found in Goovaerts and Hoogstad (1987) or Goulet (1998).

Consider an insurance portfolio where contracts are classified into cohorts. In our terminology, this is a two-level hierarchical classification structure. The observations are claim amounts  $S_{ijt}$ , where index  $i = 1, \dots, I$  identifies the cohort, index  $j = 1, \dots, J_i$  identifies the contract within the cohort and index  $t = 1, \dots, n_{ij}$  identifies the period (usually a year). To each data point corresponds a weight — or volume —  $w_{ijt}$ . Then, the best linear prediction for the next period outcome of a contract based on ratios  $X_{ijt} = S_{ijt}/w_{ijt}$  is

$$\begin{aligned}\hat{\pi}_{ij} &= z_{ij}X_{ijw} + (1 - z_{ij})\hat{\pi}_i \\ \hat{\pi}_i &= z_iX_{izw} + (1 - z_i)m\end{aligned}\tag{1}$$

with the credibility factors

$$\begin{aligned}z_{ij} &= \frac{w_{ij\Sigma}}{w_{ijk\Sigma} + s^2/a}, & w_{ij\Sigma} &= \sum_{t=1}^{n_{ij}} w_{ijt} \\ z_i &= \frac{z_{i\Sigma}}{z_{i\Sigma} + a/b}, & z_{i\Sigma} &= \sum_{j=1}^{J_i} z_{ij}\end{aligned}$$

and the weighted averages

$$\begin{aligned}X_{ijw} &= \sum_{t=1}^{n_{ij}} \frac{w_{ijt}}{w_{ij\Sigma}} X_{ijt} \\ X_{izw} &= \sum_{j=1}^{J_i} \frac{z_{ij}}{z_{i\Sigma}} X_{ijw}.\end{aligned}$$

The estimator of  $s^2$  is

$$\hat{s}^2 = \frac{1}{\sum_{i=1}^I \sum_{j=1}^{J_i} (n_{ij} - 1)} \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{t=1}^{n_{ij}} w_{ijt} (X_{ijt} - X_{ijw})^2.\tag{2}$$

The three types of estimators for parameters  $a$  and  $b$  are the following. First, let

$$\begin{aligned}A_i &= \sum_{j=1}^{J_i} w_{ij\Sigma} (X_{ijw} - X_{iww})^2 - (J_i - 1)s^2 & c_i &= w_{i\Sigma\Sigma} - \sum_{j=1}^{J_i} \frac{w_{ij\Sigma}^2}{w_{i\Sigma\Sigma}} \\ B &= \sum_{i=1}^I z_{i\Sigma} (X_{izw} - \bar{X}_{zzw})^2 - (I - 1)a & d &= z_{\Sigma\Sigma} - \sum_{i=1}^I \frac{z_{i\Sigma}^2}{z_{\Sigma\Sigma}},\end{aligned}$$

with

$$\bar{X}_{zzw} = \sum_{i=1}^I \frac{z_{i\Sigma}}{z_{\Sigma\Sigma}} X_{izw}.\tag{3}$$

(Hence,  $E[A_i] = c_i a$  and  $E[B] = db$ .) Then, the Bühlmann–Gisler estimators are

$$\hat{a} = \frac{1}{I} \sum_{i=1}^I \max\left(\frac{A_i}{c_i}, 0\right) \quad (4)$$

$$\hat{b} = \max\left(\frac{B}{d}, 0\right), \quad (5)$$

the Ohlsson estimators are

$$\hat{a}' = \frac{\sum_{i=1}^I A_i}{\sum_{i=1}^I c_i} \quad (6)$$

$$\hat{b}' = \frac{B}{d} \quad (7)$$

and the iterative (pseudo-)estimators are

$$\tilde{a} = \frac{1}{\sum_{i=1}^I (J_i - 1)} \sum_{i=1}^I \sum_{j=1}^{J_i} z_{ij} (X_{ijw} - X_{izw})^2 \quad (8)$$

$$\tilde{b} = \frac{1}{I - 1} \sum_{i=1}^I z_i (X_{izw} - X_{zzw})^2, \quad (9)$$

where

$$X_{zzw} = \sum_{i=1}^I \frac{z_i}{z_\Sigma} X_{izw}. \quad (10)$$

Note the difference between the two weighted averages (3) and (10). See [Belhadj et al. \(2009\)](#) for further discussion on this topic.

Finally, the estimator of the collective mean  $m$  is  $\hat{m} = X_{zzw}$ .

The credibility modeling function `cm` assumes that data is available in the format most practical applications would use, namely a rectangular array (matrix or data frame) with entity observations in the rows and with one or more classification index columns (numeric or character). One will recognize the output format of `simul` and its summary methods.

Then, function `cm` works much the same as `lm`. It takes in argument: a formula of the form `~` terms describing the hierarchical interactions in a data set; the data set containing the variables referenced in the formula; the names of the columns where the ratios and the weights are to be found in the data set. The latter should contain at least two nodes in each level and more than one period of experience for at least one entity. Missing values are represented by NAs. There can be entities with no experience (complete lines of NAs).

In order to give an easily reproducible example, we group states 1 and 3 of the Hachemeister data set into one cohort and states 2, 4 and 5 into another. This shows that data does not have to be sorted by level. The fitted model using the iterative estimators is:

```
> X <- cbind(cohort = c(1, 2, 1, 2, 2), hachemeister)
> fit <- cm(~cohort + cohort:state, data = X, ratios = ratio.1:ratio.12,
+          weights = weight.1:weight.12, method = "iterative")
> fit
```

```
Call:
cm(formula = ~cohort + cohort:state, data = X, ratios = ratio.1:ratio.12,
    weights = weight.1:weight.12, method = "iterative")
```

Structure Parameters Estimators

Collective premium: 1746

Between cohort variance: 88981

Within cohort/Between state variance: 10952

Within state variance: 139120026

The function returns a fitted model object of class "cm" containing the estimators of the structure parameters. To compute the credibility premiums, one calls a method of predict for this class:

```
> predict(fit)
```

```
$cohort
[1] 1949 1543
```

```
$state
[1] 2048 1524 1875 1497 1585
```

One can also obtain a nicely formatted view of the most important results with a call to summary:

```
> summary(fit)
```

```
Call:
cm(formula = ~cohort + cohort:state, data = X, ratios = ratio.1:ratio.12,
    weights = weight.1:weight.12, method = "iterative")
```

Structure Parameters Estimators

Collective premium: 1746

Between cohort variance: 88981

Within cohort/Between state variance: 10952

Within state variance: 139120026

Detailed premiums

```
Level: cohort
      cohort Indiv. mean Weight Cred. factor Cred. premium
      1      1967      1.407 0.9196      1949
```

2	1528	1.596	0.9284	1543
---	------	-------	--------	------

Level: state

	cohort	state	Indiv. mean	Weight	Cred. factor	Cred. premium
1	1	2061		100155	0.8874	2048
2	2	1511		19895	0.6103	1524
1	3	1806		13735	0.5195	1875
2	4	1353		4152	0.2463	1497
2	5	1600		36110	0.7398	1585

The methods of predict and summary can both report for a subset of the levels by means of an argument levels. For example:

```
> summary(fit, levels = "cohort")
```

Call:

```
cm(formula = ~cohort + cohort:state, data = X, ratios = ratio.1:ratio.12,
    weights = weight.1:weight.12, method = "iterative")
```

Structure Parameters Estimators

Collective premium: 1746

Between cohort variance: 88981

Within cohort variance: 10952

Detailed premiums

Level: cohort

	cohort	Indiv. mean	Weight	Cred. factor	Cred. premium
1	1967		1.407	0.9196	1949
2	1528		1.596	0.9284	1543

```
> predict(fit, levels = "cohort")
```

\$cohort

[1] 1949 1543

The results above differ from those of [Goovaerts and Hoogstad \(1987\)](#) for the same example because the formulas for the credibility premiums are different.

## 4 Bühlmann and Bühlmann–Straub models

As mentioned above, the Bühlmann and Bühlmann–Straub models are simply one-level hierarchical models. In this case, the Bühlmann–Gisler and Ohlsson estimators of the between variance parameters are both identical to the usual [Bühlmann and Straub \(1970\)](#) estimator

$$\hat{a} = \frac{w_{\Sigma\Sigma}}{w_{\Sigma\Sigma}^2 - \sum_{i=1}^I w_{i\Sigma}^2} \left( \sum_{i=1}^I w_{i\Sigma} (X_{i\Sigma} - X_{\Sigma\Sigma})^2 - (I-1)\hat{s}^2 \right), \quad (11)$$

and the iterative estimator

$$\tilde{a} = \frac{1}{I-1} \sum_{i=1}^I z_i (X_{iw} - X_{zw})^2 \quad (12)$$

is better known as the Bichsel–Straub estimator.

To fit the Bühlmann model using `cm`, one simply does not specify any weights:

```
> cm(~state, hachemeister, ratios = ratio.1:ratio.12)
```

Call:

```
cm(formula = ~state, data = hachemeister, ratios = ratio.1:ratio.12)
```

Structure Parameters Estimators

Collective premium: 1671

Between state variance: 72310

Within state variance: 46040

In comparison, the results for the Bühlmann–Straub model using the Bichsel–Straub estimator are:

```
> cm(~state, hachemeister, ratios = ratio.1:ratio.12,  
+    weights = weight.1:weight.12)
```

Call:

```
cm(formula = ~state, data = hachemeister, ratios = ratio.1:ratio.12,  
    weights = weight.1:weight.12)
```

Structure Parameters Estimators

Collective premium: 1684

Between state variance: 89639

Within state variance: 139120026

## 5 Regression model of Hachemeister

The regression model of [Hachemeister \(1975\)](#) is a generalization of the Bühlmann–Straub model. If data shows a systematic trend, the latter model will typically under- or over-estimate the true premium of an entity. The idea of [Hachemeister](#) was to fit to the data a regression model where the parameters are a credibility weighted average of an entity’s regression parameters and the group’s parameters.

In order to use `cm` to fit a credibility regression model to a data set, one simply has to supply as additional arguments `regformula` and `regdata`. The

first one is a formula of the form  $\sim$  terms describing the regression model and the second is a data frame of regressors. That is, arguments `regformula` and `regdata` are in every respect equivalent to arguments `formula` and `data` of `lm`, with the minor difference that `regformula` does not need to have a left hand side (and is ignored if present). For example, fitting the model

$$X_{it} = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1, \dots, 12$$

to the original data set of [Hachemeister \(1975\)](#) is done with

```
> fit <- cm(~state, hachemeister,
+         regformula = ~ time, regdata = data.frame(time = 1:12),
+         ratios = ratio.1:ratio.12, weights = weight.1:weight.12)
> fit

Call:
cm(formula = ~state, data = hachemeister, ratios = ratio.1:ratio.12,
    weights = weight.1:weight.12, regformula = ~time, regdata = data.frame(time = 1:12))
```

Structure Parameters Estimators

```
Collective premium: 1469 32.05

Between state variance: 24154 2700.0
                        2700 301.8
Within state variance: 49870187
```

Computing the credibility premiums requires to give the “future” values of the regressors as in `predict.lm`:

```
> predict(fit, newdata = data.frame(time = 13))

[1] 2437 1651 2073 1507 1759
```

It is well known that the basic regression model has a major drawback: there is no guarantee that the credibility regression line will lie between the collective and individual ones. This may lead to grossly inadequate premiums, as [Figure 1](#) shows.

The solution proposed by [Bühlmann and Gisler \(1997\)](#) is simply to position the intercept at the barycenter of time instead of at time origin (see also [Bühlmann and Gisler, 2005](#), Section 8.4). In mathematical terms, this essentially amounts to using an orthogonal design matrix. By setting the argument `adj.intercept` to `TRUE` in the call, `cm` will automatically fit the credibility regression model with the intercept at the barycenter of time. The resulting regression coefficients have little meaning, but the predictions are sensible:

```
> fit2 <- cm(~state, hachemeister,
+         regformula = ~ time, regdata = data.frame(time = 1:12),
+         adj.intercept = TRUE,
+         ratios = ratio.1:ratio.12, weights = weight.1:weight.12)
> summary(fit2, newdata = data.frame(time = 13))
```



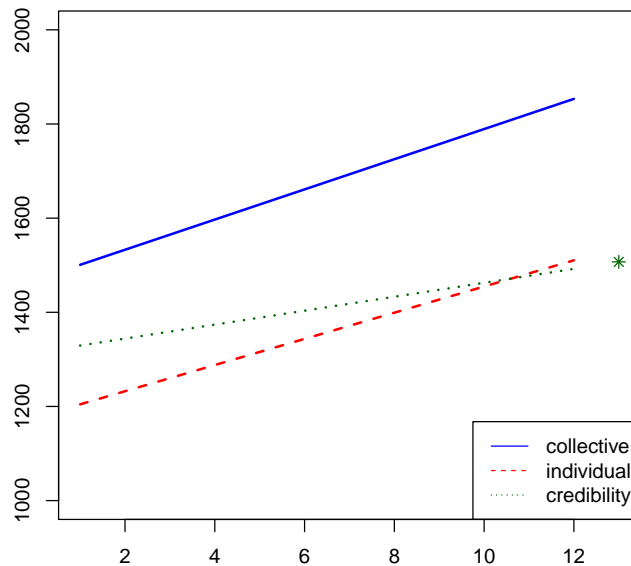


Figure 1: Collective, individual and credibility regression lines for State 4 of the Hachemeister data set. The point indicates the credibility premium.

```
Call:
cm(formula = ~state, data = hachemeister, ratios = ratio.1:ratio.12,
  weights = weight.1:weight.12, regformula = ~time, regdata = data.frame(time = 1:12),
  adj.intercept = TRUE)
```

#### Structure Parameters Estimators

Collective premium: -1675 117.1

Between state variance: 93783 0  
0 8046

Within state variance: 49870187

#### Detailed premiums

```
Level: state
state Individ. coef. Credibility matrix Adj. coef. Cred. premium
1      -2062.46    0.9947 0.0000      -2060.41    2457
      216.97     0.0000 0.9413        211.10
```

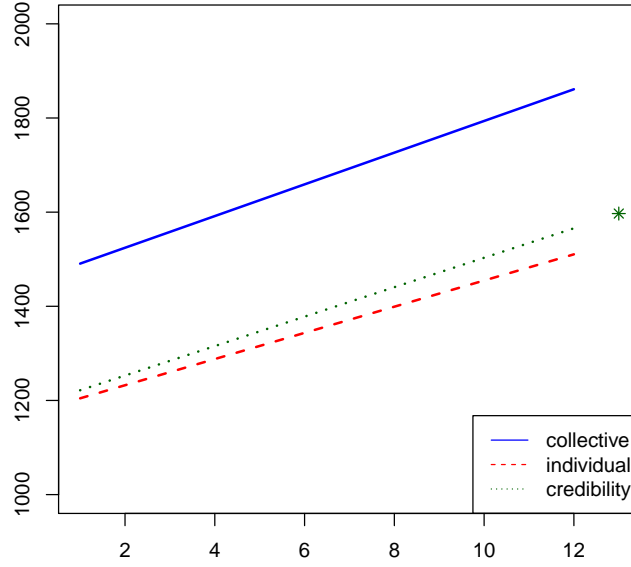


Figure 2: Collective, individual and credibility regression lines for State 4 of the Hachemeister data set when the intercept is positioned at the barycenter of time. The point indicates the credibility premium.

2	-1509.28	0.9740	0.0000	-1513.59	1651
	59.60	0.0000	0.7630	73.23	
3	-1813.41	0.9627	0.0000	-1808.25	2071
	150.60	0.0000	0.6885	140.16	
4	-1356.75	0.8865	0.0000	-1392.88	1597
	96.70	0.0000	0.4080	108.77	
5	-1598.79	0.9855	0.0000	-1599.89	1698
	41.29	0.0000	0.8559	52.22	

Figure 2 shows the beneficial effect of the intercept adjustment on the premium of State 4.

## References

H. Belhadj, V. Goulet, and T. Ouellet. On parameter estimation in hierarchical credibility. *ASTIN Bulletin*, 39(2), 2009.

- H. Bühlmann. Experience rating and credibility. *ASTIN Bulletin*, 5:157–165, 1969.
- H. Bühlmann and A. Gisler. Credibility in the regression case revisited. *ASTIN Bulletin*, 27:83–98, 1997.
- H. Bühlmann and A. Gisler. *A course in credibility theory and its applications*. Springer, 2005. ISBN 3-5402575-3-5.
- H. Bühlmann and W. S. Jewell. Hierarchical credibility revisited. *Bulletin of the Swiss Association of Actuaries*, 87:35–54, 1987.
- H. Bühlmann and E. Straub. Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries*, 70:111–133, 1970.
- M. J. Goovaerts and W. J. Hoogstad. *Credibility theory*. Number 4 in Surveys of actuarial studies. Nationale-Nederlanden N.V., Netherlands, 1987.
- V. Goulet. Principles and application of credibility theory. *Journal of Actuarial Practice*, 6:5–62, 1998.
- C. A. Hachemeister. Credibility for regression models with application to trend. In *Credibility, theory and applications*, Proceedings of the Berkeley Actuarial Research Conference on Credibility, pages 129–163, New York, 1975. Academic Press.
- W. S. Jewell. The use of collateral data in credibility theory: a hierarchical model. *Giornale dell'Istituto Italiano degli Attuari*, 38:1–16, 1975.
- E. Ohlsson. Simplified estimation of structure parameters in hierarchical credibility. Presented at the Zurich ASTIN Colloquium, 2005. URL <http://www.actuaries.org/ASTIN/Colloquia/Zurich/Ohlsson.pdf>.