Credibility and Regression-Type Modeling

Vytaras Brazauskas* Harald Dornheim† Ponmalar Ratnam§

Revised: June 2013 (*Submitted*: June 2012)

Chapter Preview. This chapter introduces the reader to credibility and related regression modeling. The first section provides a brief overview of credibility theory, regression-type credibility, and discusses historical developments. The next section shows how some well-known credibility models can be embedded within the linear mixed model framework. Specific procedures on how such models can be used for prediction and standard ratemaking are given as well. Further, in Section 3, a step-by-step numerical example, based on the widely-studied Hachemeister's data, is developed to illustrate the methodology. All computations are done using the statistical software package "R". The fourth section identifies some practical issues with the standard methodology. In particular, its lack of robustness against various types of outliers is mentioned. Possible solutions that have been proposed in the statistical and actuarial literatures are discussed. Performance of the most effective proposals is illustrated on the Hachemeister's data set and compared to that of the standard methods. Suggestions for further reading are made in Section 5.

1 Introduction

1.1 Early Developments

Credibility theory is one of the oldest but still most common premium ratemaking techniques in insurance industry. The earliest works in credibility theory date back to the beginning of the 20th century, when Mowbray (1914) and Whitney (1918) laid the foundation for *limited fluctuation credibility theory*. It is a stability-oriented form of credibility, the main objective of which is to incorporate into the premium as much individual experience as possible while keeping the premium sufficiently stable. Despite numerous attempts, this approach never arrived at a unifying principle that covered all special cases and that opened new venues for generalization. Its range of applications is quite limited and thus it never became a full-fledged theory.

* Corresponding author: Vytaras Brazauskas, Ph.D., ASA, is a Professor in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA. *e-mail*: `vytaras@uwm.edu`

† Harald Dornheim, Ph.D., ASA, CERA, DAV, SAV, is an Actuarial Consultant in the Financial Services Department, KPMG Switzerland. *e-mail*: `bavaria_harry@gmx.de`

§ Ponmalar Ratnam, is a Ph.D. Candidate in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee. *e-mail*: `psratnam@uwm.edu`

Instead of solely focusing on the stability of the premium, the modern and more flexible approach to credibility theory concentrates on finding the most accurate estimate of an insured's pure risk premium. This is accomplished by striking a balance between the individual's risk experience and the average claim over all risk classes. While initial contributions to this area can be traced back to the 1920s (see Keffer, 1929), it is generally agreed that the systematic development of the field of *greatest accuracy credibility* started in the late 1960s with the seminal paper of Bühlmann (1967). A few years later, Bühlmann and Straub (1970) introduced a credibility model as a means to rate reinsurance treaties, which generalized previous results and became the cornerstone of greatest accuracy credibility theory. The model is one of the most frequently applied credibility models in insurance practice, and it enjoys some desirable optimality properties. For more historical facts and further discussion about credibility, see the classic textbook of Klugman, Panjer, Willmot (2012, Chapters 17–18).

1.2 Regression-Type Credibility

The first credibility model linked to regression was introduced by Hachemeister (1975) who employed it to model U.S. automobile bodily injury claims classified by state and with different inflation trends. Specifically, Hachemeister considered 12 periods, from the third quarter of 1970 to the second quarter of 1973, of claim data for bodily injury that are covered by a private passenger auto insurance. The response variable of interest to the actuary is the severity *average loss per claim*, denoted by y_{it}. It is followed over the periods $t = 1, \ldots, n_i$ for each state $i = 1, \ldots, m$. Average losses were reported for $n_1 = \cdots = n_m = 12$ periods and from $m = 5$ different states (see Appendix, Table A).

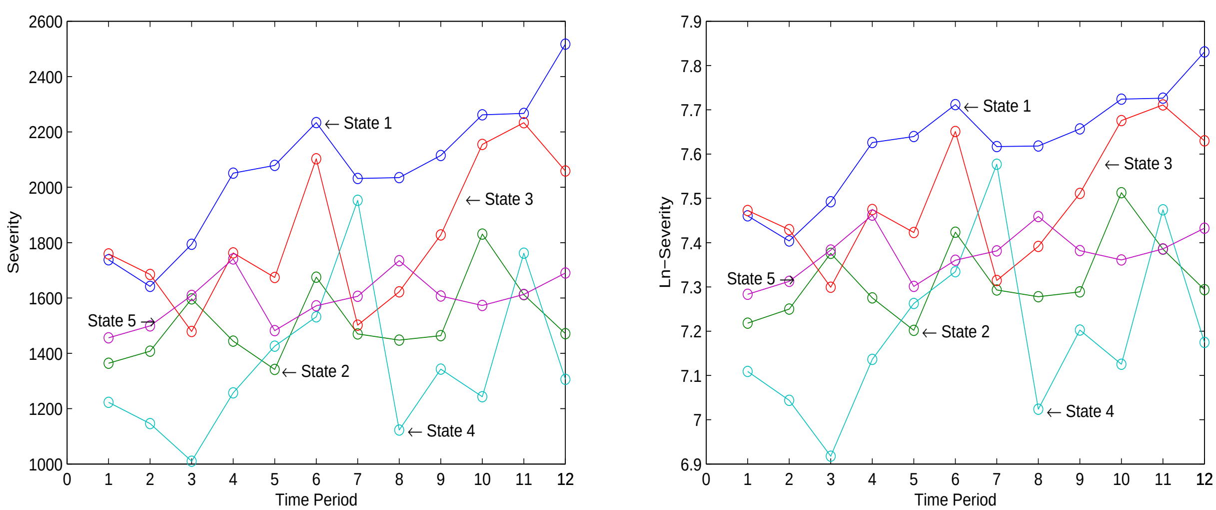

FIGURE 1: Multiple time series plot of the variable average loss per claim, y_{it}, and the logarithmic average loss per claim, $\ln(y_{it})$.

A multiple time series plot of the observed variable average loss per claim, y_{it}, and of the average loss per claim in logarithmic units, $\ln(y_{it})$, is provided in Figure 1 (log-claim modeling will be considered in Section 4). The plots indicate that states differ with respect to their within-state variability and severity. For instance, State 1 reports the highest average losses per claim, whereas State 4 seems to have larger variability compared to other states. For all five states we observe a small increase of severity over time. Since the response variable y_{it} grows over time t and varies from one state to another, this provides a hint about possible structure of *explanatory variables*. Therefore, Hachemeister originally suggested to use the linear trend model—a regression model—which can be viewed as a special case of the linear mixed models of **Chapter 8 (mixed)**:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\boldsymbol{u}_i + \varepsilon_{it},$$

where $\mathbf{x}'_{it} = \mathbf{z}'_{it} = (1, t)'$ are known designs for the *fixed effects* (parameter $\boldsymbol{\beta}$) and the subject-specific *random effects* (parameter $\boldsymbol{u}_i$), respectively, and ε_{it} denotes within-subject residuals.

Later in this chapter, more details and examples will be provided about the link between the linear mixed models and most popular credibility models (see Section 2.1). Also, as will be shown in Section 2.2, the general linear prediction problem for linear mixed models is closely related to credibility ratemaking. It turns out that *generalized least squares* and *best linear unbiased predictors* correspond to the well-known pricing formulas of credibility theory.

1.3 Recent Developments

Frees, Young, Luo (1999) have provided a longitudinal data (see **Chapter 7 (panel)**) analysis interpretation for the aforementioned and other additive credibility ratemaking procedures, which also remains valid in the framework of linear mixed models. The flexibility of linear mixed models for handling simultaneously within-risk variation and heterogeneity among risks makes them a powerful tool for credibility (see **Chapter 8 (mixed)** for details and generalizations of linear mixed models).

As is the case with many mathematical models, credibility models contain unknown *structural parameters* (or in the language of linear mixed models, fixed effects and variance components) that have to be estimated from the data. For statistical inference about fixed effects and variance components, likelihood-based methods such as (restricted) maximum likelihood estimators, (RE)ML, are commonly pursued. However, it is also known that while these methods offer most flexibility and full *efficiency* at the assumed model, they are extremely sensitive to small deviations from hypothesized normality of random components as well as to the occurrence of outliers. To obtain more reliable estimators for premium calculation and prediction of future claims, various *robust* methods have been successfully adapted to credibility theory in the actuarial literature (see, for example, Pitselis, 2002, 2008, 2012, Dornheim and Brazauskas, 2007, 2011b).

In the remainder of the chapter, we first present the standard likelihood-based procedures for ratemaking, then provide a step-by-step numerical example, and conclude with a brief review of robust techniques and compare their performance to that of the standard methods. All computations are done using the statistical software package "R" and are based on Hachemeister's data.

2 Credibility and the LMM Framework

In this section, we start by briefly describing how some popular (linear) credibility models are expressed as linear mixed models, or LMM for short. The problem of prediction in linear mixed models and its application to standard credibility ratemaking are discussed in Section 2.2.

2.1 Credibility Models

Here we demonstrate that some well-known additive credibility models can be interpreted as linear mixed models which enjoy many desirable features. For instance, they allow the modeling of claims across risk classes and time as well as the incorporation of categorical and continuous explanatory characteristics for prediction of claims. The following descriptions are taken, with some modifications, from Bühlmann and Gisler (2005), Frees (2004), and Frees, Young, Luo (1999). The basic credibility models such as Bühlmann and Bühlmann-Straub can also be found in Klugman, Panjer, Willmot (2012, Chapter 18). The notation we use is similar to that of other chapters in this book (especially, **Chapter 8 (mixed)**), but may differ from the notation used elsewhere in the literature.

2.1.1 The Bühlmann Model

Let us consider a portfolio of different insureds or risks $i,\ i = 1, \dots, m$. For each risk i we have a vector of observations $\boldsymbol{y}_i = (y_{i1}, \dots, y_{in_i})'$, where y_{it} represents the *observed* claim amount (or loss ratio) of risk i at time $t,\ t = 1, \dots, n_i$, where n_i's are allowed to be unequal. Then, by choosing $p = q = 1$ and $\mathbf{X}_i = \mathbf{Z}_i = \mathbf{1}_{n_i}$ in equation **(17) of Chapter 8 (mixed)**, we arrive at

$$\boldsymbol{y}_i = \mathbf{1}_{n_i}\beta + \mathbf{1}_{n_i}u_i + \boldsymbol{\varepsilon}_i,$$

where $\beta = \mathrm{E}(y_{it}) = \mathrm{E}(\mathrm{E}(y_{it}|u_i))$ is the overall mean or *collective premium* charged for the whole portfolio, u_i denotes the *unobservable* risk parameter characterizing the subject-specific deviation from the collective premium β, and $\mathbf{1}_{n_i}$ represents the n_i-variate vector of ones. From the hierarchical formulation of linear mixed models (see **Section 2.1 of Chapter 8 (mixed)**), the risk premium $\mu_i = \mathrm{E}(y_{it}|u_i) = \beta + u_i$ is the *true premium* for an insured i if its risk parameter u_i were known. In addition, we obtain $\mathrm{G} = \mathrm{Var}(u_i) = \sigma_u^2$ and the variance-covariance matrices

$$\boldsymbol{\Sigma}_i = \mathbf{Var}(\boldsymbol{y}_i|u_i) = \mathbf{Var}(\boldsymbol{\varepsilon}_i) = \sigma_\varepsilon^2\,\mathbf{I}_{n_i \times n_i}.$$

Note that, in general, the *structural parameters* β, σ_u^2 and σ_ε^2 are unknown and must be estimated from the data. Also, viewing the Bühlmann model from this broader perspective provides insight about the explanatory variables for claims (or loss ratios) and possible generalizations.

Note 1 (*The Balanced Bühlmann Model*): When the number of observation periods is the same for all risks, i.e., $n_1 = \cdots = n_m$, the basic credibility model becomes the Balanced Bühlmann Model. □

2.1.2 The Bühlmann-Straub Model

The credibility model of Section 2.1.1 can be easily extended to the heteroscedastic model of Bühlmann and Straub (1970) by choosing the variance-covariance matrix as follows:

$$\boldsymbol{\Sigma}_i = \mathbf{Var}(\boldsymbol{y}_i|u_i) = \mathbf{Var}(\boldsymbol{\varepsilon}_i) = \sigma^2_\varepsilon \ \mathrm{diag}\left(\upsilon_{i1}^{-1}, \ldots, \upsilon_{in_i}^{-1}\right),$$

where $\upsilon_{it} > 0$ are known volume measures. These weights represent varying exposures toward risk for insured i over the period n_i. Practical examples of exposure weights include number of years at risk in motor insurance, sum insured in fire insurance, annual turnover in commercial liability insurance, among others (see Bühlmann and Gisler, 2005).

2.1.3 The Hachemeister Regression Model

Hachemeister's simple linear regression model is a generalization of the Bühlmann-Straub Model, which includes the time (as linear trend) in the covariates. To obtain the linear trend model, in equation **(17) of Chapter 8 (mixed)** we choose $p = q = 2$ and set $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})'$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})'$, where $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t)'$. This results in the random coefficients model of the form

$$\boldsymbol{y}_i = \mathbf{X}_i \ (\boldsymbol{\beta} + \boldsymbol{u}_i) + \boldsymbol{\varepsilon}_i,$$

with the diagonal matrix $\boldsymbol{\Sigma}_i$ defined as in Section 2.1.2. It is common to assume that (unobservable) risk factors u_1 and u_2 are independent with the variance-covariance matrix $\mathbf{G} = \mathrm{diag}(\sigma^2_{u_1}, \sigma^2_{u_2})$.

2.1.4 The Revised Hachemeister Regression Model

Application of the Hachemeister's model to bodily injury data (see Section 1.2) results in unsatisfying model fits which are due to systematic underestimation of the credibility regression line. To overcome this drawback, Bühlmann and Gisler (1997) suggested to take the intercept of the regression line at the "center of gravity" of the time variable, instead of the origin of the time axis. That is, choose design matrices $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})'$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})'$ with $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t - C_{i\bullet})'$, where

$$C_{i\bullet} = \upsilon_{i\bullet}^{-1} \sum_{t=1}^{n_i} t \ \upsilon_{it}$$

is the center of gravity of the time range in risk i, and $\upsilon_{i\bullet} = \sum_{t=1}^{n_i} \upsilon_{it}$. This modification ensures that the regression line stays between the individual and collective regression lines; and the model is called the *revised* Hachemeister regression model.

From a practical point of view, volumes are often equal enough across periods for a single risk to be considered constant in time, which yields similar centers of gravity between risks. Then, it is reasonable to use the center of gravity of the collective, which is defined by $C_{\bullet\bullet} = \upsilon_{\bullet\bullet}^{-1} \sum_{i=1}^{m} \sum_{t=1}^{n_i} t \ \upsilon_{it}$, where $\upsilon_{\bullet\bullet} = \sum_{i=1}^{m} \sum_{t=1}^{n_i} \upsilon_{it}$ (see Bühlmann and Gisler, 2005, Section 8.3).

2.2 Prediction and Ratemaking

In the linear mixed model defined by **(17) of Chapter 8 (mixed)**, let $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ and $\widehat{\boldsymbol{\theta}}$ be the likelihood-based estimates of the grand mean $\boldsymbol{\beta}$ and the variance component vector $\boldsymbol{\theta} = (\sigma^2_{u_1}, \ldots, \sigma^2_{u_q}, \sigma^2_{\varepsilon})$, respectively. Then, the minimum mean square error predictor of the random variable

$$W_i = \mathrm{E}(y_{i,n_i+1}|\boldsymbol{u}_i) = \mathbf{x}'_{i,n_i+1}\boldsymbol{\beta} + \mathbf{z}'_{i,n_i+1}\boldsymbol{u}_i$$

is given by the best linear unbiased predictor

$$\widehat{W}_{\text{BLUP},i} = \mathbf{x}'_{i,n_i+1}\ \widehat{\boldsymbol{\beta}}_{\text{GLS}} + \mathbf{z}'_{i,n_i+1}\ \widehat{\boldsymbol{u}}_{\text{BLUP},i}\ , \qquad i = 1, \ldots, m, \tag{2.1}$$

where $\mathbf{x}'_{i,n_i+1}$ and $\mathbf{z}'_{i,n_i+1}$ are known covariates of risk i in time period $n_i + 1$, and $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ and $\widehat{\boldsymbol{u}}_{\text{BLUP},i}$ are computed using equations **(26) and (29) of Chapter 8 (mixed)**, respectively. (See also the discussion following equation **(29)**.)

In the actuarial literature, $\widehat{W}_{\text{BLUP},i}$ is called a homogeneous estimator of W_i (Dannenburg, Kaas, Goovaerts, 1996) and it is used to predict the expected claim size $\mu_{i,n_i+1} = \mathrm{E}(y_{i,n_i+1}|\boldsymbol{u}_i)$ of risk i for time $n_i + 1$. This estimator is even optimal for non-normally distributed claims (Norberg, 1980).

Recall that the central objective of credibility is to price fairly heterogeneous risks based on the overall portfolio mean, M, and the risk's individual experience, M_i. This relation can be expressed by the general credibility pricing formula

$$P_i = \zeta_i\ M_i + (1 - \zeta_i)\ M = M + \zeta_i\ (M_i - M), \qquad i = 1, \ldots, m, \tag{2.2}$$

where P_i is the credibility premium of risk i, and $0 \le \zeta_i \le 1$ is known as the credibility factor. Note, a comparison of equation (2.1) with (2.2) implies that $\mathbf{x}'_{i,n_i+1}\ \widehat{\boldsymbol{\beta}}_{\text{GLS}}$ can be interpreted as estimate of M, and $\mathbf{z}'_{i,n_i+1}\ \widehat{\boldsymbol{u}}_{\text{BLUP},i}$ as predictor of the weighted, risk-specific deviation $\zeta_i\ (M_i - M)$. This relationship will be illustrated for the Bühlmann-Straub model and the revised Hachemeister regression model.

2.2.1 Example 1: The Bühlmann-Straub Model

In Section 2.1.2, we have seen that the Bühlmann-Straub Model can be formulated as random coefficients model of the form $\mathbf{E}(\mathbf{y}_i|u_i) = \mathbf{1}_{n_i}\beta + \mathbf{1}_{n_i}u_i$. Then, for future expected claims $\mu_i = \mathrm{E}(y_{i,n_i+1}|u_i)$ of risk i, Frees (2004) finds the best linear unbiased predictor $\widehat{\mu}_i = \widehat{\beta}_{\text{GLS}} + \widehat{u}_{\text{BLUP},i}$ with:

$$\widehat{\beta}_{\text{GLS}} = \bar{y}_\zeta \qquad \text{and} \qquad \widehat{u}_{\text{BLUP},i} = \zeta_i\ \left(\bar{y}_i - \widehat{\beta}_{\text{GLS}}\right) \tag{2.3}$$

where $\bar{y}_\zeta = (\sum_{i=1}^m \zeta_i)^{-1}\ \sum_{i=1}^m \zeta_i\ \bar{y}_i$, $\bar{y}_i = \upsilon_{i\bullet}^{-1} \sum_{t=1}^{n_i} \upsilon_{it}\ y_{it}$, and $\zeta_i = (1 + \sigma^2_\varepsilon/(\upsilon_{i\bullet}\sigma^2_u))^{-1}$. To compute formulas (2.3), one needs to estimate the structural parameters σ^2_u and σ^2_ε. The estimators $\widehat{\sigma}^2_u$ and $\widehat{\sigma}^2_\varepsilon$ are obtained from (RE)ML (i.e., as byproduct from Henderson's Mixed Model Equations) and coincide, when assuming normality, with the following nonparametric estimators:

$$\begin{aligned}
\widehat{\sigma}_\varepsilon^2 &= \frac{\sum_{i=1}^m \sum_{t=1}^{n_i} \upsilon_{it}(y_{it}-\bar{y}_i)^2}{\sum_{i=1}^m (n_i-1)} \qquad \text{and} \\
\widehat{\sigma}_u^2 &= \frac{\upsilon_{\bullet\bullet}}{\upsilon_{\bullet\bullet}^2 - \sum_{i=1}^m \upsilon_{i\bullet}^2} \left(\sum_{i=1}^m \upsilon_{i\bullet}(\bar{y}_i-\bar{y})^2 - \widehat{\sigma}_\varepsilon^2(m-1) \right),
\end{aligned}$$

where $\bar{y} = \upsilon_{\bullet\bullet}^{-1} \sum_{t=1}^m \upsilon_{i\bullet}\ \bar{y}_i$ (see also Klugman, Panjer, Willmot, 2012, Section 19.2).

2.2.2 Example 2: The Revised Hachemeister Regression Model

Here, we provide the necessary details for estimators in the revised Hachemeister regression model. For risk i one can estimate the expected claim amount $\mu_{i,n_i+1} = \mathrm{E}(y_{i,n_i+1}|\boldsymbol{u}_i)$ by the credibility estimator $\widehat{\mu}_{i,n_i+1} = (1, n_i+1)\ (\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} + \widehat{\boldsymbol{u}}_{\mathrm{BLUP},i}) = (1, n_i+1)\ \left((\mathbf{I}_{2\times2} - \boldsymbol{\zeta}_i)\ \widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} + \boldsymbol{\zeta}_i\ \boldsymbol{b}_i \right)$, with:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \left(\sum_{i=1}^m \boldsymbol{\zeta}_i \right)^{-1} \sum_{i=1}^m \boldsymbol{\zeta}_i \boldsymbol{b}_i \quad \text{and} \quad \widehat{\boldsymbol{u}}_{\mathrm{BLUP},i} = \boldsymbol{\zeta}_i \left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} \right),$$

where

$$\boldsymbol{b}_i = \mathbf{A}_i^{-1} \begin{bmatrix} \sum_{t=1}^{n_i} \upsilon_{it}\ y_{it} \\ \sum_{t=1}^{n_i} \upsilon_{it}\ y_{it}\ \ (t - C_{i\bullet}) \end{bmatrix}$$

is the estimated individual claim experience of risk i,

$$\boldsymbol{\zeta}_i = \mathrm{diag} \begin{bmatrix} \left(1 + \sigma_\varepsilon^2\ /\ (\sigma_{u_1}^2\ a_{i1})\right)^{-1} \\ \left(1 + \sigma_\varepsilon^2\ /\ \left(\sigma_{u_2}^2\ a_{i2}\right)\right)^{-1} \end{bmatrix}$$

is the credibility factor for risk i, and $\mathbf{A}_i = \mathrm{diag}(a_{i1}, a_{i2})$ with $a_{i1} = \upsilon_{i\bullet}$, $a_{i2} = \widetilde{\upsilon}_{i\bullet} = \sum_{t=1}^{n_i} \upsilon_{it}(t - C_{i\bullet})^2$, and $C_{i\bullet} = \upsilon_{i\bullet}^{-1} \sum_{t=1}^{n_i} t\ \upsilon_{it}$ is the center of gravity. We still have to estimate the process variance σ_ε^2 and variances of hypothetical means $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$. It is reasonable to estimate σ_ε^2 by the natural variance estimator $\widehat{\sigma}_\varepsilon^2 = m^{-1} \sum_{i=1}^m \widehat{\sigma}_{\varepsilon,i}^2$, where $\widehat{\sigma}_{\varepsilon,i}^2 = (n_i-2)^{-1} \sum_{t=1}^{n_i} \upsilon_{it}(y_{it} - \widehat{\mu}_{it})^2$ is a (conditionally) unbiased estimator of the within-risk variance $\sigma_{\varepsilon,i}^2$, and $\widehat{\mu}_{it}$ is the fitted value of the ith regression line in time t. The structural parameters $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$ are estimated by

$$\begin{aligned}
\widehat{\sigma}_{u_1}^2 &= c_1 \left[\frac{m}{m-1} \sum_{i=1}^m \frac{\upsilon_{i\bullet}}{\upsilon_{\bullet\bullet}}\ (b_{i,1} - \bar{b}_1)^2 - \frac{m\ \widehat{\sigma}_\varepsilon^2}{\upsilon_{\bullet\bullet}} \right] \quad \text{and} \\
\widehat{\sigma}_{u_2}^2 &= c_2 \left[\frac{m}{m-1} \sum_{i=1}^m \frac{\widetilde{\upsilon}_{i\bullet}}{\widetilde{\upsilon}_{\bullet\bullet}}\ (b_{i,2} - \bar{b}_2)^2 - \frac{m\ \widehat{\sigma}_\varepsilon^2}{\widetilde{\upsilon}_{\bullet\bullet}} \right],
\end{aligned}$$

where

$$c_1 = \frac{m-1}{m} \left\{ \sum_{i=1}^m \frac{\upsilon_{i\bullet}}{\upsilon_{\bullet\bullet}} \left(1 - \frac{\upsilon_{i\bullet}}{\upsilon_{\bullet\bullet}} \right) \right\}^{-1}, \quad \bar{b}_1 = \upsilon_{\bullet\bullet}^{-1} \sum_{i=1}^m \upsilon_{i\bullet}\ b_{i,1},$$

$$c_2 = \frac{m-1}{m} \left\{ \sum_{i=1}^m \frac{\widetilde{\upsilon}_{i\bullet}}{\widetilde{\upsilon}_{\bullet\bullet}} \left(1 - \frac{\widetilde{\upsilon}_{i\bullet}}{\widetilde{\upsilon}_{\bullet\bullet}} \right) \right\}^{-1}, \quad \text{and} \quad \bar{b}_2 = \widetilde{\upsilon}_{\bullet\bullet}^{-1} \sum_{i=1}^m \widetilde{\upsilon}_{i\bullet}\ b_{i,2}.$$

3 Numerical Examples

In this section, we revisit Section 1.2 and model the Hachemeister's data set which, over the years, has been extensively analyzed by a number of authors in the actuarial literature. For example, Dannenburg, Kaas, Goovaerts (1996), Bühlmann and Gisler (1997), Frees, Young, Luo (1999), Pitselis (2008), and Dornheim and Brazauskas (2011b), among others, used this data set to illustrate the effectiveness of various regression-type credibility ratemaking techniques.

To get a feel for how things work, let us study a detailed example that shows how to fit and make predictions based on the Hachemeister model and the revised Hachemeister model. All computations are done using the statistical software package "R". Parts of the computer code are available in **actuar**, an R package for actuarial science, which is described in Dutang, Goulet, Pigeon (2008).

In order to fit the linear trend regression model of Section 2.1.3 to the Hachemeister's data set, we employ the following R-code:

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

The label `hachemeister` in the first line of the code reads the data set which is available in the **actuar** package. The last line produces predictions which are based on formula (2.1). The R-code yields the following credibility-adjusted parameter estimates and predictions for the five states:

State	*Parameter Estimates*		*Prediction*
i	$\widehat{\beta}_0 + \widehat{u}_{i,0}$	$\widehat{\beta}_1 + \widehat{u}_{i,1}$	$\widehat{\mu}_{i,12+1}$
1	1693.52	57.17	2436.75
2	1373.03	21.35	1650.53
3	1545.36	40.61	2073.30
4	1314.55	14.81	1507.07
5	1417.41	26.31	1759.40

In addition, the grand parameters are found by taking the average across all states; they are: $\widehat{\beta}_0 = 1468.77$ and $\widehat{\beta}_1 = 32.05$. Also, within-state variance is $\widehat{\sigma}^2_{\varepsilon} = 49,870,187$, and the other estimates of variance components are: $\widehat{\sigma}^2_{u_1} = 24,154.18$ and $\widehat{\sigma}^2_{u_2} = 301.81$. Note that the numerical values in the table above differ from the ones reported by Dutang, Goulet, Milhaud, Pigeon (2012), who used the same R-code but applied the reversed time variable, that is, `time = 12:1` instead of `time = 1:12`.

Fitting and predictions based on the revised Hachemeister model (see Sections 2.1.4 and 2.2.2) are much more involved. The complete program for running these tasks is presented in Appendix.

Comments explaining the formulas or a block of the program are listed between the signs `#` and `#`. To run the program, download the R-code to your computer and use the following command lines:

```
> source("HachemRevised.R")
> HachemRevised(5,12, hachemeister[,2:13], hachemeister[,14:25], 1:12,13)
```

Let us go through the main blocks of the program and review the numerical output of each block. The first group of commands (which is labeled "Program Initialization") defines the variables used in computations and sets their intial values, if necessary. The second group of commands, labeled "Center of Gravity & Recentered X", computes the collective center of gravity (the outcome is $C_{\bullet\bullet} = 6.4749$), and re-centers the design matrix X, which results in

$$\begin{bmatrix} 1 & -5.4749 \\ 1 & -4.4749 \\ 1 & -3.4749 \\ 1 & -2.4749 \\ 1 & -1.4749 \\ 1 & -0.4749 \\ 1 & 0.5251 \\ 1 & 1.5251 \\ 1 & 2.5251 \\ 1 & 3.5251 \\ 1 & 4.5251 \\ 1 & 5.5251 \end{bmatrix}$$

The third group of commands, labeled "Volume Measures & Other Constants", computes the constants required for further calculations. The results of this step are: $\tilde{\upsilon}_{\bullet\bullet} = 2,104,688$; $c_1 = 1.3202$; $c_2 = 1.3120$; $\bar{b}_1 = 1865.4040$; $\bar{b}_2 = 44.1535$. The fourth and fifth groups of commands, labeled "Variance Components" and "Parameters & Prediction", respectively, yield estimates of the structural parameters and deliver next-period predictions. The summarized results of these two program blocks are:

State	*Estimates of Structural Parameters*					*Prediction*	*Std. Error*
i	$\widehat{u}_{i,0}$	$\widehat{u}_{i,1}$	$\widehat{\beta}_0 + \widehat{u}_{i,0}$	$\widehat{\beta}_1 + \widehat{u}_{i,1}$	$\widehat{\sigma}^2_{\varepsilon,i}$	$\widehat{\mu}_{i,12+1}$	$\widehat{\sigma}_{\widehat{\mu}_{i,12+1}}$
1	385.71	27.06	2058.85	60.70	121,484,314	2847.94	110.13
2	−157.67	−12.60	1515.48	21.04	30,175,637	1789.01	123.15
3	127.71	6.58	1800.85	40.22	52,560,076	2323.70	195.49
4	−283.51	−2.40	1389.63	31.25	24,362,730	1795.83	242.22
5	−72.24	−18.63	1600.90	15.01	21,075,078	1796.00	76.39

Note that the grand parameters are found by taking the average across all states for $\widehat{\beta}_0 + \widehat{u}_{i,0}$, $\widehat{\beta}_1 + \widehat{u}_{i,1}$, and $\widehat{\sigma}^2_{\varepsilon,i}$; they are: $\widehat{\beta}_0 = 1673.14$, $\widehat{\beta}_1 = 33.64$, and $\widehat{\sigma}^2_{\varepsilon} = 49,931,567$. In addition, the estimates of variance components are: $\widehat{\sigma}^2_{u_1} = 93,021.43$ and $\widehat{\sigma}^2_{u_2} = 665.48$. The last line of the code screen-prints the results of the program.

Note 2 (*Potential Outliers*): A careful examination of Figure 1 reveals that the quarterly observations #6 in State 1, #10 in State 2, #7 in State 4, and maybe #6 in State 5 are somewhat apart from their state-specific inflation trends. This suggest the topic for the next section. That is, we would like to know how to identify outliers and, if they are present, what to do about them. $\square$

4 Theory versus Practice

The modeling approach of Section 3 is well-understood and widely used, but it is not very realistic in practice. In particular, insurance loss data are often highly-skewed, heavy-tailed and contain outliers. While complete treatment of these issues is beyond the scope of this chapter, in the following we will present some insights on how to modify and improve the standard methodology. First of all, in Section 4.1, we formulate heavy-tailed linear mixed models. Note that 'heavy-tailed' models are also known as 'long-tailed' or 'fat-tailed' (see **Chapter 9 (fat-tailed)**). Then, Section 4.2 introduces a three-step procedure for robust-efficient fitting of such models. Robust credibility ratemaking based on heavy-tailed linear mixed models (which are calibrated using the robust-efficient fitting procedure of Section 4.2) is described in Section 4.3. Finally, in Section 4.4, we revisit the earlier examples and illustrate performance of the robust methods using Hachemeister's data.

4.1 Heavy-Tailed Linear Mixed Models

Suppose we are given a random sample $(\mathbf{x}_{i1}, \mathbf{z}_{i1}, y_{i1}, \upsilon_{i1}), \ldots, (\mathbf{x}_{in_i}, \mathbf{z}_{in_i}, y_{in_i}, \upsilon_{in_i})$, where $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ are known p- and q-dimensional row-vectors of explanatory variables and $\upsilon_{it} > 0$ some known volume measure. Assume the claims y_{it} follow a log-location-scale distribution with cdf of the form:

$$G(y_{it}) = F_0\left(\frac{\log(y_{it}) - \lambda_{it}}{\sigma_\varepsilon\ \upsilon_{it}^{-1/2}}\right), \qquad y_{it} > 0, \quad i = 1, \ldots, m, \quad t = 1, \ldots, n_i,$$

defined for $-\infty < \lambda_{it} < \infty$, $\sigma_\varepsilon > 0$, and where F_0 is the standard (i.e., $\lambda_{it} = 0, \sigma_\varepsilon = 1, \upsilon_{it} = 1$) cdf of the underlying location-scale family $F(\lambda_{it}, \sigma^2_\varepsilon/\upsilon_{it})$. Following regression analysis with location-scale models, we include the covariates $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ only through the location parameter λ_{it}. Then, the following linear mixed model may be formulated:

$$\log(\boldsymbol{y}_i) \ = \ \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}_i + \boldsymbol{\varepsilon}_i \ = \ \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, m,$$

where $\log(\boldsymbol{y}_i) = \Big(\log(y_{i1}), \ldots, \log(y_{in_i})\Big)'$ and $\boldsymbol{\lambda}_i$ is the n_i-dimensional vector of the within-subject locations λ_{it} that consist of the *population location* $\boldsymbol{\beta}$ and the subject-specific *location deviation* $\boldsymbol{u}_i$. While normality of $\boldsymbol{u}_i$ is still assumed, the error term $\boldsymbol{\varepsilon}_i$ now follows the n_i-dimensional multivariate cdf with location-scale distributions $F(0, \sigma^2_\varepsilon/\upsilon_{it})$ as margins. Examples of such marginal log-location-scale families F include lognormal, log-logistic, log-t, log-Cauchy, and Weibull, which after the logarithmic transformation become normal, logistic, t, Cauchy, and Gumbel (extreme-value), respectively. Special cases of the n_i-dimensional distributions $F_{n_i}(\boldsymbol{\lambda}_i, \boldsymbol{\Sigma}_i)$ are the well-known elliptical distributions such as multivariate normal and the heavy-tailed multivariate t.

4.2 Robust-Efficient Fitting

For robust-efficient fitting of the linear mixed model with normal random components, Dornheim (2009) and Dornheim and Brazauskas (2011a) developed adaptively truncated likelihood methods. Those methods were further generalized to log-location-scale models with symmetric or asymmetric errors and labeled *corrected adaptively truncated likelihood* methods, CATL (see Dornheim, 2009, and Dornheim and Brazauskas, 2011b). More specifically, the CATL estimators for location $\boldsymbol{\lambda}_i$ and variance components $\sigma^2_{u_1}, \ldots, \sigma^2_{u_q}, \sigma^2_{\varepsilon}$ can be found by the following three-step procedure:

1. *Detection of Within-Risk Outliers.*

   Consider the random sample

   $$\Big(\mathbf{x}_{i1}, \mathbf{z}_{i1}, \log(y_{i1}), \upsilon_{i1}\Big), \ \ldots, \Big(\mathbf{x}_{in_i}, \mathbf{z}_{in_i}, \log(y_{in_i}), \upsilon_{in_i}\Big), \qquad i = 1, \ldots, m.$$

   In the first step, the corrected re-weighting mechanism automatically detects and removes outlying events *within* risks whose standardized residuals computed from initial high breakdown-point estimators exceed some adaptive cut-off value. This threshold value is obtained by comparison of an empirical distribution with a theoretical one. Let us denote the resulting "pre-cleaned" random sample as

   $$\Big(\mathbf{x}^*_{i1}, \mathbf{z}^*_{i1}, \log(y^*_{i1}), \upsilon^*_{i1}\Big), \ \ldots, \Big(\mathbf{x}^*_{in^*_i}, \mathbf{z}^*_{in^*_i}, \log(y^*_{in^*_i}), \upsilon^*_{in^*_i}\Big), \qquad i = 1, \ldots, m.$$

   Note that for each risk i, the new sample size is n^*_i ($n^*_i \le n_i$).

2. *Detection of Between-Risk Outliers.*

   In the second step, the procedure searches the pre-cleaned sample (marked with *) and discards entire risks whose risk-specific profile expressed by the random effect significantly deviates from the overall portfolio profile. These risks are identified when their robustified Mahalanobis distance exceeds some adaptive cut-off point. The process results in

   $$\Big(\mathbf{x}^{**}_{i1}, \mathbf{z}^{**}_{i1}, \log(y^{**}_{i1}), \upsilon^{**}_{i1}\Big), \ \ldots, \Big(\mathbf{x}^{**}_{in^*_i}, \mathbf{z}^{**}_{in^*_i}, \log(y^{**}_{in^*_i}), \upsilon^{**}_{in^*_i}\Big), \qquad i = 1, \ldots, i^*,$$

   a "cleaned" sample of risks. Note that the number of remaining risks is i^* ($i^* \le m$).

3. *CATL estimators.*

   In the final step, the CATL procedure employs traditional likelihood-based methods, such as (restricted) maximum likelihood, on the cleaned sample and computes re-weighted parameter estimates $\widehat{\boldsymbol{\beta}}_{\text{CATL}}$ and $\widehat{\boldsymbol{\theta}}_{\text{CATL}} = (\widehat{\sigma}^2_{u_1}, \ldots, \widehat{\sigma}^2_{u_q}, \widehat{\sigma}^2_{\varepsilon})$. Here, the subscript CATL emphasizes that the maximum likelihood type estimators are not computed on the original sample, i.e., the starting point of Step 1, but rather on the cleaned sample which is the end result of Step 2.

Using the described procedure, we find the shifted *robust best linear unbiased predictor* for location:

$$\widehat{\boldsymbol{\lambda}}_i = \mathbf{X}_i^{**}\widehat{\boldsymbol{\beta}}_{\text{CATL}} + \mathbf{Z}_i^{**}\widehat{\boldsymbol{u}}_{\text{rBLUP},\,i} + \widehat{\mathbf{E}}_{F_0}(\boldsymbol{\varepsilon}_i), \qquad i = 1, \ldots, m,$$

where $\widehat{\boldsymbol{\beta}}_{\text{CATL}}$ and $\widehat{\boldsymbol{u}}_{\text{rBLUP},\,i}$ are standard likelihood-based estimators but computed on the clean sample from Step 2. Also, $\widehat{\mathbf{E}}_{F_0}(\boldsymbol{\varepsilon}_i)$ is the expectation vector of the n_i^*-variate cdf $F_{n_i^*}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_i)$. For symmetric error distributions we obtain the special case $\widehat{\mathbf{E}}_{F_0}(\boldsymbol{\varepsilon}_i) = 0$.

4.3 Robust Credibility Ratemaking

The re-weighted estimates for location, $\widehat{\boldsymbol{\lambda}}_i$, and structural parameters, $\widehat{\boldsymbol{\theta}}_{\text{CATL}} = (\widehat{\sigma}^2_{u_1}, \ldots, \widehat{\sigma}^2_{u_q}, \widehat{\sigma}^2_{\varepsilon})$, are used to calculate robust credibility premiums for the ordinary but heavy-tailed claims part of the original data. The robust ordinary net premiums

$$\widehat{\mu}_{it}^{\text{ordinary}} = \widehat{\mu}_{it}^{\text{ordinary}}(\widehat{\boldsymbol{u}}_{\text{rBLUP},i}), \quad t = 1, \ldots, n_i + 1, \quad i = 1, \ldots, m$$

are found by computing the empirical *limited expected value* (LEV) of the fitted log-location distribution of claims. The percentile levels of the lower bound q_l and the upper bound q_g used in LEV computations are usually chosen to be extreme, e.g., 0.1% for q_l and 99.9% for q_g.

Then, robust regression is employed to price separately identified excess claims. The risk-specific excess claim amount of insured i at time t is defined by

$$\widehat{O}_{it} = \begin{cases} -\widehat{\mu}_{it}^{\text{ordinary}}, & \text{for } y_{it} < q_l. \\ (y_{it} - q_l) - \widehat{\mu}_{it}^{\text{ordinary}}, & \text{for } q_l \leq y_{it} < q_g. \\ (q_g - q_l) - \widehat{\mu}_{it}^{\text{ordinary}}, & \text{for } y_{it} \geq q_g. \end{cases}$$

Further, let m_t denote the number of insureds in the portfolio at time t and let $N = \max_{1 \leq i \leq m} n_i$, the maximum horizon among all risks. For each period $t = 1, \ldots, N$, we find the mean cross-sectional overshot of excess claims $\widehat{O}_{\bullet t} = m_t^{-1} \sum_{i=1}^{m_t} \widehat{O}_{it}$, and fit robustly the random effects model

$$\widehat{O}_{\bullet t} = \mathbf{o}_t \boldsymbol{\xi} + \tilde{\varepsilon}_t, \qquad t = 1, \ldots, N,$$

where $\mathbf{o}_t$ is the row-vector of covariates for the hypothetical mean of overshots $\boldsymbol{\xi}$. Here we choose $\mathbf{o}_t = 1$, and let $\widehat{\boldsymbol{\xi}}$ denote a robust estimate of $\boldsymbol{\xi}$. Then, the premium for extraordinary claims, which is common to all risks i, is given by

$$\mu_{it}^{\text{extra}} = \mathbf{o}_t \widehat{\boldsymbol{\xi}}.$$

Finally, the portfolio-unbiased robust regression credibility estimator is defined by

$$\widehat{\mu}^{\text{CATL}}_{i,n_i+1}(\widehat{\boldsymbol{u}}_{\text{rBLUP},i}) = \widehat{\mu}^{\text{ordinary}}_{i,n_i+1}(\widehat{\boldsymbol{u}}_{\text{rBLUP},i}) + \mu^{\text{extra}}_{i,n_i+1}, \qquad i = 1, \ldots, m.$$

From the actuarial point of view, premiums assigned to the insured have to be positive. Therefore, we determine pure premiums by $\max\left\{0, \widehat{\mu}^{\text{CATL}}_{i,n_i+1}(\widehat{\boldsymbol{u}}_{\text{rBLUP},i})\right\}$.

4.4 Numerical Examples Revisited

For Hachemeister's regression credibility model and its revised version, we use $\log(y_{it})$ as response variable and fit it using the CATL method. In Table 1, we report loss predictions for individual states, which were computed using CATL and compared to those obtained by other authors. Specifically, the BASE method, which is the linear trend model used by Goovaerts and Hoogstad (1987), and the M-RC, MM-RC, GM-RC procedures which were studied by Pitselis (2002, 2008, 2012).

As discussed by Kaas, Dannenburg, Goovaerts (1997) and Frees, Young, Luo (2001), in practice it is fairly common to observe situations where risks with larger exposure measure exhibit lower variability. As one can infer from Table A (see Appendix), the *number of claims per period*, denoted by υ_{it}, significantly affects the within-risk variability. Also, State 4 reports high average losses per claim, which in turn yields to increased within-state variability (see Figure 1). Thus, to obtain homoscedastic error terms, we fit models using υ_{it} as subject-specific weights. This model can be written as

$$\log(y_{it}) = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_{it}\boldsymbol{u}_i + \varepsilon_{it}\ \upsilon_{it}^{1/2},$$

where ε_{it} is a sequence of independent normally distributed noise terms.

To assess the quality of credibility predictions, $\widehat{\mu}_{i,12+1}$, we also report their standard errors. These can be used to construct prediction intervals of the form BLUP (credibility estimate $\widehat{\mu}_{i,12+1}$) plus and minus multiples of the standard error. We estimate the standard error of prediction, $\widehat{\sigma}_{\widehat{\mu}_{i,12+1}}$, from the data using the common nonparametric estimator

$$\widehat{\sigma}_{\widehat{\mu}_{i,12+1}} = \left[\widehat{\text{MSE}}(\widehat{\mu}_i) - \widehat{\text{bias}}^2(\widehat{\mu}_i)\right]^{1/2} = \left[\upsilon_{i\bullet}^{-1}\sum_{t=1}^{12}\omega_{it}\upsilon_{it}(y_{it}-\widehat{\mu}_{it})^2 - \left(\upsilon_{i\bullet}^{-1}\sum_{t=1}^{12}\omega_{it}\upsilon_{it}(y_{it}-\widehat{\mu}_{it})\right)^2\right]^{1/2},$$

where $\widehat{\mu}_{it}$ denotes the credibility estimate obtained from the pursued regression method, $\upsilon_{i\bullet}$ is the total number of claims in state i, and ω_{it} is the hard-rejection weight for the observed average loss per claim y_{it} when employing the CATL procedure. For non-robust REML where no data points are truncated we put $\omega_{it} = 1$ as special case.

Table 1: Individual state predictions for Hachemeister's bodily injury data based on various model-fitting procedures (if available, estimated standard errors are provided in parentheses).

Fitting Procedure	*Prediction for State* 1	2	3	4	5
BASE	2436	1650	2073	1507	1759
M-RC ($c = 1.5$)	2437	1650	2073	1507	1759
GM-RC ($k = 1$)	2427	1648	2092	1505	1737
MM-RC	2427	1648	2092	1505	1737
REML	2465 (109)	1625 (122)	2077 (193)	1519 (248)	1695 (77)
CATL	2471 (111)	1545 (74)	2065 (194)	1447 (174)	1691 (57)
REML*	2451 (109)	1661 (123)	2065 (193)	1613 (242)	1706 (78)
CATL*	2450 (113)	1552 (74)	2049 (195)	1477 (172)	1693 (57)

* Based on the revised Hachemeister model (see Section 2.2.2).

Several conclusion emerge from Table 1. First, we note that the REML and REML* estimates (predictions) are based on Henderson's Mixed Model Equations and thus they slightly differ from those of Section 3. Second, for States 1, 3, and 5, all techniques, standard and robust, yield similar predictions. For the second and fourth states, however, CATL produces slightly lower predictions. For instance, for State 4 it results in 1447 whereas the base model finds 1507. This can be traced back to the truncation of the suspicious observations #6 and #10 in State 2 and #7 in State 4. Third, CATL also identifies claim #4 in State 5 as outlier and, as a result, assigns a small discount of -1.47 to each risk. For the revised model (i.e., for REML* and CATL*), prediction patterns are similar.

Table 2: Individual state predictions for *contaminated* Hachemeister's data based on various model-fitting procedures (if available, estimated standard errors are provided in parentheses).

Fitting Procedure	*Prediction for State* 1	2	3	4	5
BASE	2501	1826	2181	1994	2596
M-RC ($c = 1.5$)	2755	1979	2396	1841	2121
GM-RC ($k = 1$)	2645	1868	2311	1723	1964
MM-RC	2649	1870	2315	1724	1943
REML	2517 (119)	1852 (150)	2206 (204)	1987 (255)	2542 (829)
CATL	2477 (111)	1550 (74)	2071 (194)	1452 (174)	1689 (60)
REML*	2455 (110)	1949 (166)	2229 (204)	2141 (275)	2629 (818)
CATL*	2459 (112)	1559 (74)	2057 (195)	1484 (172)	1694 (60)

* Based on the revised Hachemeister model (see Section 2.2.2).

To illustrate robustness of regression credibility methods that are based on M-RC, MM-RC, and GM-RC estimators for quantifying individual's risk experience, Pitselis (2002, 2008) replaces the last observation of the fifth state, 1690, by 5000. We follow the same contamination strategy and summarize our findings in Table 2. As one can see from Table 2, the choice of the model-fitting methodology has major impact on predictions. Indeed, in the presence of a single outlier, we find that robust procedures provide stability and reasonable adjustment to predictions whereas standard methods overreact. For example, in the contaminated State 5 the REML and BASE credibility estimates get inflated by the outlying observation and increase from 1695 to 2542 and from 1759 to 2596, respectively. Further, since outliers usually distort the estimation process of variance components and thus yield too low credibility weights, predictions across all states increase significantly. Note also a dramatic explosion of standard errors (e.g., the standard error of REML in State 5 jumps from 77 to 829), which is due to increased within-risk variability that was caused by the contaminating observation. Furthermore, not all robust credibility predictions react equally to data contamination. The CATL based predictions change only slightly when compared to the non-contaminated data case, but those of M-RC, GM-RC, and MM-RC shift upwards by 10%-20%. For instance, predictions for State 5 change from: 1691 to 1689 (for CATL), 1759 to 2121 (for M-RC), 1737 to 1964 (for GM-RC), 1737 to 1943 (for MM-RC). This can be explained by the fact that the latter group of procedures does not provide protection against large claims that influence the between-risk variability. Note also that in all but contaminated State 5 the CATL predictions slightly increase while the corresponding standard errors remain unchanged. On the other hand, State 5 prediction is practically unchanged, but the CATL method "penalizes" the state by increasing its standard error (i.e., the standard error has changed from 57 to 60). An increase in standard error implies a credibility reduction for State 5.

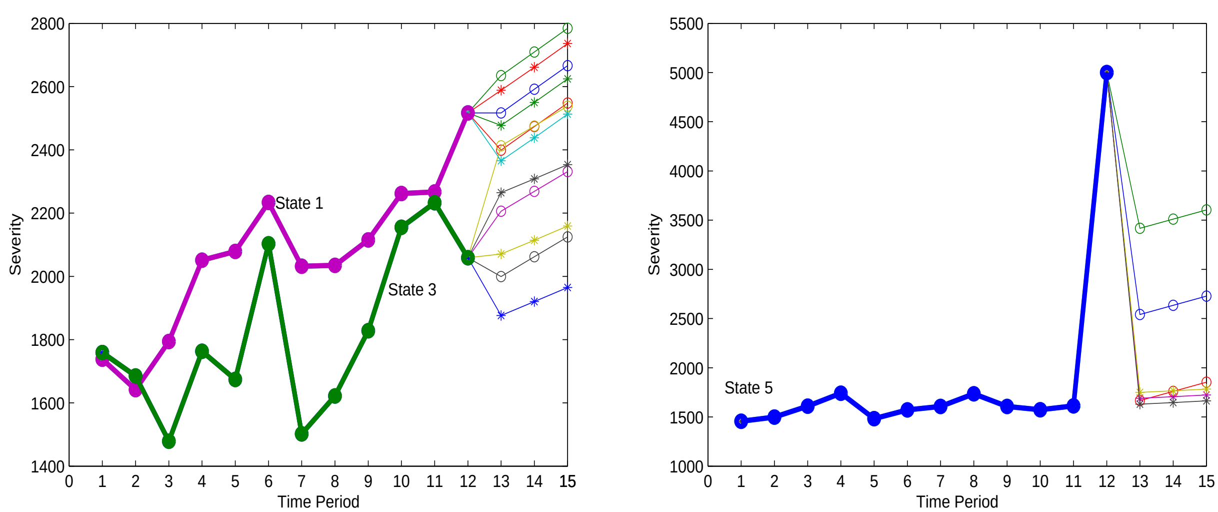

FIGURE 2: Selected point and interval predictions for *contaminated* Hachemeister's data. The thin lines connecting 'o' denote one-, two-, and three-step predictions using REML. The corresponding CATL predictions are marked by '$*$'.

As the last point of this discussion, Figure 2 illustrates the impact of data contamination on the ratemaking process. The left panel plots claim severity over time for two non-contaminated states, State 1 and State 3, with the last three time periods representing the point and interval predictions. The right panel plots corresponding results for the contaminated State 5. In both cases the REML based inference leads to elevated point predictions and wider intervals (which are constructed by adding and subtracting one standard error to the point prediction). As one would expect, the REML and CATL predictions are most disparate in State 5.

5 Further Reading

The regression-type credibility models of this chapter have been extended and generalized in several directions. To get a broader view of this topic, we encourage the reader to consult other papers and textbooks. For example, to learn how to model correlated claims data, we recommend reading Frees, Young, Luo (1999, 2001) and Frees (2004). Further, to gain a deeper understanding of and appreciation for robust credibility techniques, the reader should review Garrido and Pitselis (2000), Pitselis (2002, 2004, 2008, 2012), and Dornheim and Brazauskas (2007, 2011a,b). If the reader is not familiar with the philosophy and methods of robust statistics, then the book by Maronna, Martin, Yohai (2006) will provide a gentle introduction into the subject. Finally, an introduction to and developments of hierarchical credibility modeling can be found in Sundt (1979, 1980), Norberg (1986), Bühlmann and Jewell (1987), and Belhadj, Goulet, Ouellet (2009).

References

[1] Belhadj, H., Goulet, V., and Ouellet, T. (2009). On parameter estimation in hierarchical credibility. *ASTIN Bulletin*, **39**(2), 495–514.

[2] Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, **4**, 199–207.

[3] Bühlmann, H., and Gisler, A. (1997). Credibility in the regression case revisited. *ASTIN Bulletin*, **27**, 83–98.

[4] Bühlmann, H., and Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer, New York.

[5] Bülmann, H., and Jewell, W.S. (1987). Hierarchical credibility revisited. *Bulletin of the Swiss Association of Actuaries*, **87**, 35–54.

[6] Bühlmann, H., and Straub, E. (1970). Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **70**, 111–133.

[7] Dannenburg, D. R., Kaas, R., and Goovaerts, M. J. (1996). *Practical Actuarial Credibility Models*. Amsterdam: Institute of Actuarial Science and Economics, University of Amsterdam.

[8] Dornheim, H. (2009). *Robust-Efficient Fitting of Mixed Linear Models: Theory, Simulations, Actuarial Extensions, and Examples*. Ph.D. dissertation, University of Wisconsin-Milwaukee, Milwaukee.

[9] Dornheim, H., and Brazauskas, V. (2007). Robust-efficient methods for credibility when claims are approximately gamma-distributed. *North American Actuarial Journal*, **11**(3), 138–158.

[10] Dornheim, H., and Brazauskas, V. (2011a). Robust-efficient fitting of mixed linear models: Methodology and theory. *Journal of Statistical Planning and Inference*, **141**(4), 1422–1435.

[11] Dornheim, H., and Brazauskas, V. (2011b). Robust-efficient credibility models with heavy-tailed claims: A mixed linear models perspective. *Insurance: Mathematics and Economics*, **48**(1), 72–84.

[12] Dutang, C., Goulet, V., and Pigeon, M. (2008). **actuar**: An R package for actuarial science. *Journal of Statistical Software*, **25**(7), 1–37.

[13] Dutang, C., Goulet, V., Milhaud, X., and Pigeon, M. (2012). Credibility theory features of **actuar**. `http://cran.r-project.org/web/packages/actuar/index.html`

[14] Frees, E. W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge University Press, Cambridge.

[15] Frees, E. W., Young, V. R., and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, **24**, 229–247.

[16] Frees, E. W., Young, V. R., and Luo, Y. (2001). Case studies using panel data models. *North American Actuarial Journal*, **5**(4), 24–42. Supplemental material is available at: `http://research3.bus.wisc.edu/course/view.php?id=129`

[17] Garrido, J., and Pitselis, G. (2000). On robust estimation in Bühlmann-Straub's credibility model. *Journal of Statistical Research*, **34**(2), 113–132.

[18] Goovaerts, A. S., and Hoogstad, W. (1987). *Credibility Theory, Surveys of Actuarial Studies*. National-Nederlanden N.V., Rotterdam.

[19] Hachemeister, C. A. (1975). Credibility for regression models with applications to trend. In *Credibility: Theory and Applications*, ed. P. M. Kahn, Academic Press, New York, pp. 129–163.

[20] Kaas, R., Dannenburg, D., and Goovaerts, M. (1997). Exact credibility for weighted observations. *ASTIN Bulletin*, **27**, 287–295.

[21] Keffer, R. (1929). An experience rating formula. *Transactions of the Actuarial Society of America*, **30**, 130–139.

[22] Klugman, S.A., Panjer, H.H., and Willmot, G.E. (2012). *Loss Models: From Data to Decisions*, 3rd edition. Wiley, New York.

[23] Maronna, R.A., Martin, D.R., and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*. Wiley, New York.

[24] Mowbray, A. H. (1914). How extensive a payrol exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society*, **I**, 25–30.

[25] Norberg, R. (1980). Empirical Bayes credibility. *Scandinavian Actuarial Journal*, **1980**, 177–194.

[26] Norberg, R. (1986). Hierarchical credibility: Analysis of a random effect linear model with nested classification. *Scandinavian Actuarial Journal*, **1986**, 204–222.

[27] Pitselis, G. (2002). Application of GM and MM estimators to regression credibility. *Proceedings of the 2nd Conference in Actuarial Science and Finance*, Samos, Greece.

[28] Pitselis, G. (2004). A seemingly unrelated regression model in a credibility framework. *Insurance: Mathematics and Economics*, **34**, 37–54.

[29] Pitselis, G. (2008). Robust regression credibility: The influence function approach. *Insurance: Mathematics and Economics*, **42**, 288–300.

[30] Pitselis, G. (2012). A review on robust estimators applied to regression credibility. *Journal of Computational and Applied Mathematics*, **239**, 231–249.

[31] Sundt, B. (1979). A hierarchical regression credibility model. *Scandinavian Actuarial Journal*, **1979**, 107–114.

[32] Sundt, B. (1980). A multi-level hierarchical credibility regression model. *Scandinavian Actuarial Journal*, **1980**, 25–32.

[33] Whitney, A. W. (1918). The theory of experience rating. *Proceedings of the Casualty Actuarial Society*, **IV**, 275–293.

Appendix

Table A. Hachemeister's bodily injury data set comprising average loss per claim, y_{it}, and the corresponding number of claims per period, υ_{it}.

Period	*Average loss per claim in State*					*Number of claims per period in State*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
1	1738	1364	1759	1223	1456	7861	1622	1147	407	2902
2	1642	1408	1685	1146	1499	9251	1742	1357	396	3172
3	1794	1597	1479	1010	1609	8706	1523	1329	348	3046
4	2051	1444	1763	1257	1741	8575	1515	1204	341	3068
5	2079	1342	1674	1426	1482	7917	1622	998	315	2693
6	2234	1675	2103	1532	1572	8263	1602	1077	328	2910
7	2032	1470	1502	1953	1606	9456	1964	1277	352	3275
8	2035	1448	1622	1123	1735	8003	1515	1218	331	2697
9	2115	1464	1828	1343	1607	7365	1527	896	287	2663
10	2262	1831	2155	1243	1573	7832	1748	1003	384	3017
11	2267	1612	2233	1762	1613	7849	1654	1108	321	3242
12	2517	1471	2059	1306	1690	9077	1861	1121	342	3425

Source: Hachemeister (1975), Figure 3.

The R-code for the revised Hachemeister's model (for details, see Sections 2.1.4, 2.2.2, and 3):

```
HachemRevised <- function(number_of_states, timeline_number, Y,W,t, time_to_predict) {

##########  PROGRAM INITIALIZATION  #####

I = number_of_states       # number of rows in each matrix #
N = timeline_number         # number of columns in each matrix #
k = time_to_predict        # prediction horizon #
X = matrix(nrow=12,ncol=2); X[,1] = rep(1,12); X[,2] = t
ivector = seq(from=1,to=I,by=1)
tvector = seq(from=1,to=N,by=1)
WYear = vector(length=N); WTilda = 0; WTotal = 0
WS = vector(length=I); WSTotal = 0
c1 = 0; c2 = 0
b1 = vector(length=I); b2 = vector(length=I)
b1bar = 0; b2bar = 0
b = matrix(nrow=2,ncol=5)
mu = matrix(nrow=5,ncol=12)
XCent = matrix(nrow=12,ncol=2)       # X centered #
SigmaTemp = matrix(nrow=5,ncol=12)
SigmaTemp1 = matrix(nrow=5,ncol=12)
SigmaTemp2 = vector(length=I)
SigmaTemp3 = vector(length=I)
StdError = vector(length=I); SigmaSq = 0
Tau1Temp1 = vector(length=I); Tau2Temp1 = vector(length=I)
Tau1Sq = 0; Tau2Sq = 0; Tau1Sq1 = 0; Tau2Sq1 = 0
k0 = 0; k1 = 0
A11 = vector(length=I); A22 = vector(length=I)
cred = matrix(nrow=2,ncol=2)
sumcred = vector(length=2); sumcredwt = vector(length=2)
halpha = matrix(nrow=2,ncol=5); hbeta = matrix(nrow=2,ncol=5)
paramet = matrix(nrow=2,ncol=5)
cred_prem = matrix(nrow=1,ncol=5)

##########  CENTER OF GRAVITY & RECENTERED X  #####

for (t in tvector) {
WYear[t] = sum(W[,t])
WTilda = WTilda + (WYear[t] * t)
WTotal = WTotal + WYear[t] }       # collective center of gravity W..  #
XCent = as.matrix(cbind(rep(1,12),X[,2] - WTilda/WTotal))       # X at center of gravity #

##########  VOLUME MEASURES & OTHER CONSTANTS  #####

for (i in ivector) {
WS[i] = sum(XCent[,2]^2 * W[i,])
WSTotal = WSTotal + WS[i] }
for(i in ivector) {
c1 = c1 + (sum(W[i,])/WTotal) * (1 - sum(W[i,])/WTotal)
c2 = c2 + (WS[i]/WSTotal) * (1 - WS[i]/WSTotal)
b1[i] = sum(Y[i,] * t(W[i,]))/sum(W[i,])
b2[i] = sum(t(W[i,]) * (Y[i,] - b1[i]) * XCent[,2])/WS[i]
b[1,i] = b1[i]; b[2,i] = b2[i]
b1bar = b1bar + (b1[i] * sum(W[i,])/WTotal)
b2bar = b2bar + (b2[i] * WS[i]/WSTotal)
```

```
HachemData = data.frame(Y=Y[i,],X=XCent[,2])
Reg = lm(Y~X, data=HachemData)       # Individual Regression #
mu[i,] = Reg$fitted.values }
c1 = 1/c1 * (I-1)/I; c2 = 1/c2 * (I-1)/I

##########  VARIANCE COMPONENTS  #####

SigmaTemp = Y-mu; SigmaTemp1=(Y-mu)^2
for(i in ivector) {
SigmaTemp2[i] = sum(SigmaTemp1[i,] * t(W[i,]))
SigmaTemp3[i] = sum(SigmaTemp[i,] * t(W[i,]))/sum(W[i,])
StdError[i] = sqrt((SigmaTemp2[i]/sum(W[i,])) - (SigmaTemp3[i])^2) }
SigmaTemp2 = SigmaTemp2/(N-2); SigmaSq = sum(SigmaTemp2)/I
Tau1Temp1 = (b1-b1bar)^2; Tau2Temp1 = (b2-b2bar)^2
for(i in ivector) {
Tau1Sq1 = Tau1Sq1 + sum(W[i,]) * Tau1Temp1[i]
Tau2Sq1 = Tau2Sq1 + WS[i] * Tau2Temp1[i] }
Tau1Sq = (I*c1/WTotal) * (Tau1Sq1/(I-1) - SigmaSq)
Tau2Sq = (I*c2/WSTotal) * (Tau2Sq1/(I-1) - SigmaSq)
k0 = SigmaSq/Tau1Sq; k1 = SigmaSq/Tau2Sq

##########  PARAMETERS & PREDICTION  #####

for (i in ivector) {
A11[i] = sum(W[i,])/(sum(W[i,])+k0)
A22[i] = WS[i]/(WS[i]+k1)
cred = as.matrix(rbind(c(A11[i],0),c(0,A22[i])))
sumcredwt = sumcredwt + (cred% * %b[,i])
sumcred[1] = sumcred[1] + A11[i]; sumcred[2] = sumcred[2] + A22[i] }
hbeta = sumcredwt/sumcred
for (i in ivector) {
cred = as.matrix(rbind(c(A11[i],0),c(0,A22[i])))
halpha[,i] = cred% * %(b[,i] - hbeta); paramet[,i] = halpha[,i] + hbeta }
cred_prem = t(as.matrix(c(1, k)))% * %paramet

##########  RESULTS  #####

result1 = list(collective_center_of_gravity=WTilda/WTotal, X_at_center_of_gravity=XCent)
result2 = list(vtilda_double_dot=WSTotal, c1=c1, c2=c2, b1bar=b1bar, b2bar=b2bar)
result3 = list(alpha=halpha, beta_plus_alpha=paramet, sigmasq=SigmaTemp2)
result4 = list(result1, Volume_Measure_and_Other_Components=result2, Structural_Parameters=result3,
    Prediction=cred_prem, Std._Error=StdError); return(result4) }
```