Robust-Efficient Credibility Models with Heavy-Tailed Claims: A Mixed Linear Models Perspective

Harald Dornheim[1]
University of Wisconsin-Milwaukee

Vytaras Brazauskas[2]
University of Wisconsin-Milwaukee

Abstract

In actuarial practice, regression models serve as a popular statistical tool for analyzing insurance data and tariff ratemaking. In this paper, we consider classical credibility models that can be embedded within the framework of mixed linear models. For inference about fixed effects and variance components, likelihood-based methods such as (restricted) maximum likelihood estimators are commonly pursued. However, it is well-known that these standard and fully efficient estimators are extremely sensitive to small deviations from hypothesized normality of random components as well as to the occurrence of outliers. To obtain better estimators for premium calculation and prediction of future claims, various robust methods have been successfully adopted to credibility theory in the actuarial literature. The objective of this work is to develop robust and efficient methods for credibility when heavy-tailed claims are approximately log-location-scale distributed. To accomplish that, we first show how to express additive credibility models such as Bühlmann-Straub and Hachemeister as mixed linear models with symmetric or asymmetric errors. Then, we adjust *adaptively truncated likelihood* methods and compute highly robust credibility estimates for the ordinary but heavy-tailed claims part. Finally, we treat the identified excess claims separately and find robust-efficient credibility premiums. Practical performance of this approach is examined—via simulations—under several contaminating scenarios. A widely studied real-data set from workers' compensation insurance is used to illustrate functional capabilities of the new robust credibility estimators.

JEL Codes: C13, C14, C15, C16. *Subject Categories*: IM10, IM31, IM41, IM54. *Insurance Branch Categories*: IB 83.

Key words and phrases: Adaptive robust-efficient estimation; Asymmetric heavy-tailed residuals; Credibility ratemaking; Mixed linear model; Treatment of excess claims.

[1] Harald Dornheim, Ph.D., is Actuarial Consultant at PricewaterhouseCoopers, Birchstrasse 160, 8050 Zurich, Switzerland. E-mail: `harald.dornheim@ch.pwc.com`. This work was completed while the author was a Ph.D. student at the University of Wisconsin-Milwaukee.

[2] Corresponding author: Vytaras Brazauskas is Associate Professor, Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA. E-mail: `vytaras@uwm.edu`

1 Introduction

Regression credibility models have become one of the most frequently applied premium rating techniques in the North American insurance industry. The first credibility model linked to regression has been proposed by Hachemeister (1975) who employed it to model U.S. automobile bodily injury claims classified by state and with different inflation trends. In the following decade, a vast literature on diverse (non-)linear regression models that allow the actuary to explain risk characteristics in terms of covariates emerged (see, e.g., Sundt, 1979, 1980, De Vylder, 1985, Norberg, 1986). Later, the seminal work of McCullagh and Nelder (1989) on generalized linear models (GLMs) simplified the treatment of non-normal claim data that naturally arises in insurance, and gave a boost to further actuarial research (Haberman and Renshaw, 1996, Nelder and Verall, 1997).

To unify important additive credibility ratemaking procedures, Frees (1999) gave a longitudinal data analysis interpretation of the well-known credibility models of Bühlmann (1967), Bühlmann-Straub (1970) and Hachemeister (1975). Assuming that there is no time-dependency in the portfolio, similar interpretations remain valid in the framework of mixed linear models. The flexibility of mixed linear models for handling simultaneously within-risk variation and heterogeneity among risks make them a powerful tool for credibility. Further, the popularity of these models in practice lead to many well-developed estimation procedures for fixed effects and for variance components, which in the context of credibility theory are known as *structural parameters*. Furthermore, standard inferential methods for such models are likelihood-based (e.g., REML) and completely rely on the assumption of joint normality of random effects and error terms. While such techniques usually yield fully efficient estimators at the assumed model, extreme sensitivity of likelihood-based inference to violations of distributional assumptions and data contamination has been known for some time (see, e.g., Rocke, 1983). Therefore, the classical credibility models are examples of non-robust experience rating procedures.

To limit distorting effects of extremes on credibility weights and the ensuing ratemaking process, De Vylder (1976) and Gisler (1980) presented semi-linear credibility that applies models of Bühlmann and Straub (1967, 1970) to trimmed claim data. Further, pioneers that suggested to combine robust statistics with credibility theory are Kremer (1991) and Künsch (1992). Their main idea toward

1

robustification is to bound the influence of atypical observations on the individual claims experience by applying Huber's M-estimator (Huber, 1981), and then to use credibility estimators with robust means. This technique has been further refined by Gisler and Reinhard (1993), Garrido and Pitselis (2000), and Dornheim and Brazauskas (2007). Also, Pitselis (2008) extended this strategy to Hachemeister's regression credibility.

All these approaches are indeed resistant toward outliers. Yet it is generally agreed that M-estimators do not provide full robustness and high efficiency at the same time. In the present article, we address this issue and formulate a new robust-efficient methodology for credibility in the mixed linear model framework when claims are approximately log-location-scale distributed. As is well-known, actuaries frequently encounter skewed data both in life (Rosenberg *et al.*, 2007, Manning *et al.* 2005) and non-life (Klugman *et al.*, 2004) insurance. However, extreme-value statistics and the modeling of long-tailed distributions in regression analysis has only recently surfaced in the actuarial literature (Beirlant *et al.*, 2004, Sun *et al.*, 2008). Also, standard inferential methods for mixed linear models are based on the normality assumption, and they cannot be directly applied for skewed or heavy-tailed insurance data. To alleviate this problem, we follow the approach of Caroll and Ruppert (1988) and proceed as follows.

First, we take the logarithmic transformation of the observed claim amounts that follow some hypothesized log-location-scale model. In a second step, the transformed response variables are linked to explanatory variables where random effects are included in the linear predictor. Then, we generalize robust-efficient *adaptively truncated likelihood* (ATL) methods to mixed linear models with symmetric or asymmetric log-location-scale errors. As a result we obtain *corrected adaptively truncated likelihoods* (CATL). The modified three-step procedure inherits the properties of ATL methods, and provides truncated generalized least squares for location and variance components' estimators with hard-rejection weights adaptively computed from the long-tailed data (see Dornheim, 2009). More specifically, the corrected re-weighting mechanism automatically detects and removes outlying events within risks, then identifies and discards risks that do not fit into the overall portfolio structure, and finally it employs maximum likelihood procedures on the "clean" insurance data. Therefore, extremes that have significant distorting effects on the estimation of structural parameters and the computation of

credibility weights are completely dispelled from the data. In a fourth step, re-weighted estimates for location and scale are used to calculate robust credibility premiums for the ordinary but heavy-tailed claims part of the original data. Lastly, we employ robust regression to price separately excess claims and to find portfolio-unbiased robust credibility premiums. The new ratemaking procedure is called *corrected adaptively truncated likelihood credibility* (CATLC).

The newly designed class of robust credibility regression possesses several desirable features. First, the presented procedure automatically provides the highest possible degree of robustness while achieving high efficiency at the assumed long-tailed model when none of the claims are truncated. Second, this approach enables the actuary to robustify many credibility models that can be expressed as special cases of mixed linear models with ordinary claims following some log-location-scale family. Third, the robust credibility estimator provides protection against outlying risks which usually increase the between-risk variability. Forth, efficiency and detection of extremes can be controlled by the choice of the hypothetical model distribution without sacrificing robustness. Fifth, they can be easily embedded in available software.

The rest of the article is organized as follows. In Section 2, we present the formulation of mixed linear models and describe how some well-known credibility models can be embedded within this framework. In Section 3, the CATL procedure for fitting of heavy-tailed mixed linear models is designed. We also give a detailed introduction to defined rejection-rules and the entire detection process. Further, in Section 4 our robust credibility model is developed. This section includes model description, presentation of structural estimators, treatment of excess claims and final robust credibility estimators. Practical performances of our methods is examined via simulations in Section 5. In the subsequent section, a real-data example from workers' compensation insurance is analyzed. Summarizing comments are provided in Section 7.

2 Preliminaries

In this section, we first present the general mixed linear model and its classical, likelihood-based fitting methods. Then, in Section 2.2, we briefly describe how some popular (linear) credibility models are expressed as mixed linear models. The problem of prediction in mixed linear models and its application

to standard credibility ratemaking are discussed in Section 2.3.

2.1 The Mixed Linear Model

Let $\mathbf{Y}$ be an $m \times 1$ vector of total observations. Then, conditional on the random effect vectors $\{\boldsymbol{\alpha}_i \in \mathbb{R}^q, \, i = 1, \dots, I\}$, the response $\mathbf{Y}$ can naturally be grouped and decomposed into a set of independent τ_i-dimensional vectors $\mathbf{y}_1, \dots, \mathbf{y}_I$, such that $\sum_{i=1}^{I} \tau_i = m$. We consider the following mixed effects model:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, I, \tag{2.1}$$

where the $\tau_i \times p$ matrix $\mathbf{X}_i$ and $\tau_i \times q$ matrix $\mathbf{Z}_i$ are known designs for the fixed population parameter $\boldsymbol{\beta} \in \mathbb{R}^p$ and the subject-specific random effects $\boldsymbol{\alpha}_i \in \mathbb{R}^q$, respectively, and $\boldsymbol{\varepsilon}_i$ is a τ_i-dimensional vector of within-subject residuals.

Following the classical framework of the model (2.1), we assume that the random effects and error terms are: (a) both normally distributed, (b) both serially uncorrelated, and (c) independent of each other. More specifically, for $i = 1, \dots, I$,

$$\boldsymbol{\alpha}_i \sim N_q(\mathbf{0}, \mathbf{D}), \quad \boldsymbol{\varepsilon}_i \sim N_{\tau_i}(\mathbf{0}, \mathbf{R}_i), \quad \text{and} \quad Cov(\boldsymbol{\alpha}_i, \boldsymbol{\varepsilon}_i) = \mathbf{0},$$

where $\mathbf{D}$ is a $q \times q$ positive definite variance-covariance matrix of the form $\mathrm{diag}(\sigma_{\alpha_1}^2, \dots, \sigma_{\alpha_q}^2)$ and $\mathbf{R}_i = \sigma_\varepsilon^2 \mathbf{I}_{\tau_i \times \tau_i}$ represents the variance-covariance matrix of the residuals. Here $\mathbf{I}_{\tau_i \times \tau_i}$ denotes the $\tau_i \times \tau_i$-dimensional identity matrix. The assumption that random effects and error terms are uncorrelated is necessary for the proofs of statistical properties of robust-efficient adaptively truncated likelihood (ATL) methods (see Dornheim, 2009).

Hence, in view of the assumptions (a), (b), (c), we have the so-called *hierarchical* formulation of the mixed linear model (2.1), for which

$$\mathbf{y}_i \,|\, \boldsymbol{\alpha}_i \sim N_{\tau_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\alpha}_i, \mathbf{R}_i), \quad i = 1, \dots, I. \tag{2.2}$$

This formulation implies the *marginal* model, $\mathbf{y}_i \sim N_{\tau_i}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i(\boldsymbol{\theta}))$, with the covariance structure $\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i$, where $\boldsymbol{\theta} = (\sigma_{\alpha_1}^2, \dots, \sigma_{\alpha_q}^2, \sigma_\varepsilon^2)$ is a vector of variance components implicit in $\mathbf{V}_i$.

The regression parameter $\boldsymbol{\beta}$ common to all individuals i is estimated by the generalized least squares (GLS) estimator

$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = \left(\sum_{i=1}^{I} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{I} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i. \tag{2.3}$$

In the case of *known* variance components $\boldsymbol{\theta}$, this estimator is optimal and coincides with the maximum likelihood estimator of $\boldsymbol{\beta}$. Then, the random effects $\boldsymbol{\alpha}_i$ are determined by the best (with respect to the mean squared error criterion) linear unbiased predictor (BLUP)

$$\widehat{\boldsymbol{\alpha}}_{\text{BLUP},\,i}(\boldsymbol{\theta}) = \mathbf{D}\mathbf{Z}_i' \mathbf{V}_i^{-1}\big(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{\text{GLS}}(\boldsymbol{\theta})\big), \quad i = 1,\ldots,I. \tag{2.4}$$

In most practical situations, however, the parameter vector $\boldsymbol{\theta}$ is unknown and usually estimated by (asymptotically) fully efficient methods: maximum likelihood (ML) and restricted maximum likelihood (REML). The latter estimators are found using Henderson's mixed model equations; for details, see Searle *et al.* (1992, pp. 275–286). Once the estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\sigma}_{\alpha_1}^2, \ldots, \widehat{\sigma}_{\alpha_q}^2, \widehat{\sigma}_{\varepsilon}^2)$ are available, the variance-covariance matrices $\mathbf{V}_i(\boldsymbol{\theta})$ are estimated by $\mathbf{V}_i(\widehat{\boldsymbol{\theta}})$. That is, $\widehat{\mathbf{V}}_i = \mathbf{Z}_i \widehat{\mathbf{D}} \mathbf{Z}_i' + \widehat{\mathbf{R}}_i$, where $\widehat{\mathbf{D}} = \text{diag}(\widehat{\sigma}_{\alpha_1}^2, \ldots, \widehat{\sigma}_{\alpha_q}^2)$ and $\widehat{\mathbf{R}}_i = \widehat{\sigma}_{\varepsilon}^2 \mathbf{I}_{\tau_i \times \tau_i}$. The resulting estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i$, defined by (2.3) and (2.4) with $\mathbf{V}_i$, $\mathbf{D}$, $\mathbf{R}_i$ replaced by their corresponding REML estimators, are called *empirical* GLS estimator and *empirical* BLUP, respectively.

2.2 Credibility Theory Models in the Mixed Linear Model Framework

Here we demonstrate that some well-known additive credibility models can be interpreted as mixed linear 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. For a comprehensive overview of credibility theory the reader is referred to Goovaerts *et al.* (1987, 1990), Dannenburg *et al.* (1996), and Klugman *et al.* (2004). The following descriptions are taken, with some modifications, from Frees *et al.* (1999, 2004) and Bühlmann *et al.* (2005).

2.2.1 The Basic Credibility Model of Bühlmann

Let us consider a portfolio of different insureds (or risks) i, $i = 1, \ldots, I$. For each risk i we have a vector of observations $\mathbf{y}_i = (y_{i1}, \ldots, y_{i\tau_i})'$, where y_{it} represents the *observed* claim amount (or loss ratio) of risk i at time t, $t = 1, \ldots, \tau_i$. Note that Bühlmann (1967) allows for unequal horizons τ_i. Then, when choosing $p = q = 1$ and $\mathbf{X}_i = \mathbf{Z}_i = \mathbf{1}_{\tau_i}$, equation (2.1) yields

$$\mathbf{y}_i = \mathbf{1}_{\tau_i}\beta + \mathbf{1}_{\tau_i}\alpha_i + \boldsymbol{\varepsilon}_i,$$

where $\beta = \mathrm{E}(y_{it}) = \mathrm{E}(\mathrm{E}(y_{it}|\alpha_i))$ is the overall mean or collective premium charged for the whole portfolio, α_i denotes the *unobservable* risk parameter characterizing the subject-specific deviation from the collective premium β, and $\mathbf{1}_{\tau_i}$ represents the τ_i-variate vector of ones. From the hierarchical formulation of mixed linear models, given by (2.2), the risk premium $\mu_i = \mathrm{E}(y_{it}|\alpha_i) = \beta + \alpha_i$ is the *true* premium for an insured i if its risk parameter (random effect) α_i were known. We also obtain the variance-covariance matrices $\mathbf{R}_i = \mathbf{Var}(\boldsymbol{y}_i|\alpha_i) = \mathbf{Var}(\boldsymbol{\varepsilon}_i) = \sigma_\varepsilon^2 \mathbf{I}_{\tau_i \times \tau_i}$, and $\mathrm{D} = \mathrm{Var}(\alpha_i) = \sigma_\alpha^2$. In credibility theory the parameters β, σ_α^2, and σ_ε^2 are called structural parameters that are generally unknown and must be estimated from the data.

Remark 1: *The Balanced Bühlmann Model*

For equal number of observations, i.e., $\tau_i \equiv \tau$, $i = 1, \ldots, I$, the basic credibility model becomes the Balanced Bühlmann Model. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

2.2.2 The Bühlmann-Straub Model

Let us continue with the same setup as in Section 2.2.1. Then, the previous credibility model can easily be extended to the heteroscedastic model of Bühlmann-Straub (1970) if we choose the variance-covariance matrix $\mathbf{R}_i = \mathbf{Var}(\boldsymbol{y}_i|\alpha_i) = \mathbf{Var}(\boldsymbol{\varepsilon}_i) = \sigma_\varepsilon^2 \, \mathrm{diag}\left(v_{i1}^{-1}, \ldots, v_{i\tau_i}^{-1}\right)$, where $v_{it} > 0$ are known volume measures. These weights represent varying exposures toward risk for insured i over the period τ_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 (Bühlmann *et al.*, 2005).

2.2.3 The Hachemeister Regression Model

Hachemeister's simple linear regression model is a generalization of the Bühlmann-Straub Model. This model includes the time (as linear trend) in the covariates and has originally been developed by Hachemeister (1975) to investigate bodily injury data. To obtain Hachemeister's linear trend model, in (2.1) we choose $p = q = 2$ and set $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i})'$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{i\tau_i})'$, where $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t)'$. This results in the random coefficients model of the form $\mathbf{y}_i = \mathbf{X}_i \, (\boldsymbol{\beta} + \boldsymbol{\alpha}_i) + \boldsymbol{\varepsilon}_i$, with diagonal matrix $\mathbf{R}_i$ as in Section 2.2.2. By assumption (b) in Section 2.1, we consider independent (unobservable) risk factors that have variance-covariance structure $\mathbf{D} = \mathrm{diag}(\sigma_{\alpha_1}^2, \sigma_{\alpha_2}^2)$.

2.2.4 The Hachemeister Regression Model (Revisited)

Though Hachemeister (1975) developed a very promising regression approach to credibility, he obtained unsatisfying model fits when applying his linear trend model to bodily injury data. This is due to underestimation of the credibility regression line. To overcome this drawback, Bühlmann *et al.* (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). Therefore, we choose design matrices $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau_i})'$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{i\tau_i})'$ with $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t - G_{i\bullet})'$, where $G_{i\bullet} = v_{i\bullet}^{-1} \sum_{t=1}^{\tau_i} t \, v_{it}$ is the center of gravity of the time range in risk i, and $v_{i\bullet} = \sum_{t=1}^{\tau_i} v_{it}$. 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 $G_{\bullet\bullet} = v_{\bullet\bullet}^{-1} \sum_{i=1}^{I} \sum_{t=1}^{\tau_i} t \, v_{it}$, where $v_{\bullet\bullet} = \sum_{i=1}^{I} \sum_{t=1}^{\tau_i} v_{it}$ (see Bühlmann *et al.*, 2005, Section 8.3). $\qquad\square$

2.3 Prediction and Standard Credibility Ratemaking

In this section, we describe the general linear prediction problem for mixed linear models and its relationship to credibility ratemaking. It turns out that generalized least squares and best linear unbiased predictors correspond to the classical pricing formulas of credibility theory.

In the mixed linear model, defined by (2.1), let $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ and $\widehat{\boldsymbol{\theta}}$ be the likelihood estimates of the grand mean $\boldsymbol{\beta}$ and the variance component vector $\boldsymbol{\theta}$, respectively. Then, Norberg (1980) has shown that the

minimum mean square error predictor of the random variable $W_i = \mathrm{E}(y_{i,\tau_i+1}|\boldsymbol{\alpha}_i) = \mathbf{x}_{i,\tau_i+1}\boldsymbol{\beta} + \mathbf{z}_{i,\tau_i+1}\boldsymbol{\alpha}_i$ is given by the best linear unbiased predictor

$$W_{\mathrm{BLUP},i} = \mathbf{x}_{i,\tau_i+1}\,\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} + \mathbf{z}_{i,\tau_i+1}\,\widehat{\boldsymbol{\alpha}}_{\mathrm{BLUP},i}\,,\quad i = 1,\ldots,I, \tag{2.5}$$

where $\mathbf{x}'_{i,\tau_i+1} \in \mathbb{R}^p$ and $\mathbf{z}'_{i,\tau_i+1} \in \mathbb{R}^q$ are known covariates of risk i in time period τ_i+1. In the actuarial literature, $W_{\mathrm{BLUP},i}$ is called homogeneous estimator of W_i (Dannenburg *et al.*, 1996) and it is used to predict the expected claim size $\mu_{i,\tau_i+1} = \mathrm{E}(y_{i\tau_i+1}|\boldsymbol{\alpha}_i)$ of risk i in time $\tau_i + 1$. This estimator is even optimal for non-normally distributed claims (Norberg, 1980).

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

$$P_i = \zeta_i\,m + (1 - \zeta_i)\,M = M + \zeta_i\,(m - M),\quad i = 1,\ldots,I, \tag{2.6}$$

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 (2.5) with (2.6) implies that $\mathbf{x}_{i,\tau_i+1}\,\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ can be interpreted as estimate of M, and $\mathbf{z}_{i,\tau_i+1}\,\widehat{\boldsymbol{\alpha}}_{\mathrm{BLUP},\,i}$ as predictor of the weighted, risk-specific deviation $\zeta_i\,(m - M)$. This relationship will be exemplified for the Bühlmann-Straub Model. Also, we follow Bühlmann *et al.* (2005, Section 8.4) and present credibility estimators for Hachemeister's regression model (revisited) where the center of gravity was included in the design. For more examples (e.g., Hachemeister's linear trend model or the basic credibility model of Bühlmann) the reader is referred to Frees (2001 or 2004, Section 4.7).

2.3.1 Example 1: The Bühlmann-Straub Model

In Section 2.2.2, we have seen that the Bühlmann-Straub Model can be formulated as random coefficients model of the form $\mathbf{E}(\mathbf{y}_i|\alpha_i) = \mathbf{1}_{\tau_i}\beta + \mathbf{1}_{\tau_i}\alpha_i$. Then, for future expected claims $\mu_i = \mathrm{E}(y_{i,\tau_i+1}|\alpha_i)$ of risk i, Frees (1999, 2004) finds the best linear unbiased predictor $\widehat{\mu}_i = \widehat{\beta}_{\mathrm{GLS}} + \widehat{\alpha}_{\mathrm{BLUP},i}$ with:

$$\widehat{\beta}_{\mathrm{GLS}} = \bar{y}_\zeta \quad\text{and}\quad \widehat{\alpha}_{\mathrm{BLUP},i} = \zeta_i\left(\bar{y}_i - \widehat{\beta}_{\mathrm{GLS}}\right), \tag{2.7}$$

where $\bar{y}_\zeta = \left(\sum_{i=1}^I \zeta_i\right)^{-1} \sum_{i=1}^I \zeta_i\,\bar{y}_i\,,\quad \bar{y}_i = v_{i\bullet}^{-1}\sum_{t=1}^{\tau_i} v_{it}\,y_{it},\ \text{and}\ \zeta_i = (1 + \sigma_\varepsilon^2/(v_{i\bullet}\sigma_\alpha^2))^{-1}$. To apply standard credibility formulas, given by (2.7), one needs to estimate the structural parameters

σ_α^2 and σ_ε^2. The estimators $\widehat{\sigma}_\alpha^2$ and $\widehat{\sigma}_\varepsilon^2$ 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:

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{I} \sum_{t=1}^{\tau_i} v_{it}(y_{it} - \bar{y}_i)^2}{\sum_{i=1}^{I}(\tau_i - 1)} \quad \text{and} \quad \widehat{\sigma}_\alpha^2 = \frac{v_{\bullet\bullet}}{v_{\bullet\bullet}^2 - \sum_{i=1}^{I} v_{i\bullet}^2} \left(\sum_{i=1}^{I} v_{i\bullet}(\bar{y}_i - \bar{y})^2 - \widehat{\sigma}_\varepsilon^2(I - 1) \right),$$

where $\bar{y} = v_{\bullet\bullet}^{-1} \sum_{t=1}^{I} v_{i\bullet} \, \bar{y}_i$. For further discussion, see Klugman, Panjer, and Willmot (2004, Section 16.5) and Goulet (1998).

2.3.2 Example 2: The Hachemeister Regression Model (Revisited)

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

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

where

$$\boldsymbol{b}_i = \mathbf{A}_i^{-1} \begin{bmatrix} \sum_{t=1}^{\tau_i} v_{it} \, y_{it} \\ \sum_{t=1}^{\tau_i} v_{it} \, y_{it} \, (t - G_{i\bullet}) \end{bmatrix} \quad \text{is the estimated individual claim experience of risk } i,$$

$$\boldsymbol{\zeta}_i = \mathrm{diag} \begin{bmatrix} \left(1 + \sigma_\varepsilon^2 \, / \, (\sigma_{\alpha_1}^2 \, a_{i1})\right)^{-1} \\ \left(1 + \sigma_\varepsilon^2 \, / \, (\sigma_{\alpha_2}^2 \, a_{i2})\right)^{-1} \end{bmatrix} \quad \text{is the credibility factor for risk } i,$$

and $\mathbf{A}_i = \mathrm{diag}(a_{i1}, a_{i2})$ with $a_{i1} = v_{i\bullet}$, $a_{i2} = \widetilde{v}_{i\bullet} = \sum_{t=1}^{\tau_i} v_{it}(t - G_{i\bullet})^2$, and $G_{i\bullet} = v_{i\bullet}^{-1} \sum_{t=1}^{\tau_i} t \, v_{it}$ is the center of gravity. We still have to estimate the process variance σ_ε^2 and variances of hypothetical means $\sigma_{\alpha_1}^2$ and $\sigma_{\alpha_2}^2$. Following Bühlmann *et al.* (Section 8.4, 2005), it is reasonable to estimate σ_ε^2 by the natural variance estimator $\widehat{\sigma}_\varepsilon^2 = I^{-1} \sum_{i=1}^{I} \widehat{\sigma}_{\varepsilon,i}^2$, where $\widehat{\sigma}_{\varepsilon,i}^2 = (\tau_i - 2)^{-1} \sum_{t=1}^{\tau_i} v_{it}(y_{it} - \widehat{\mu}_{it})^2$ 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 $\sigma_{\alpha_1}^2$ and $\sigma_{\alpha_2}^2$ are estimated by

$$\widehat{\sigma}_{\alpha_1}^2 = c_1 \left[\frac{1}{I - 1} \sum_{i=1}^{I} \frac{v_{i\bullet}}{v_{\bullet\bullet}} \, (b_{i,1} - \bar{b}_1)^2 - \frac{I \, \widehat{\sigma}_\varepsilon^2}{v_{\bullet\bullet}} \right] \quad \text{and} \quad \widehat{\sigma}_{\alpha_2}^2 = c_2 \left[\frac{1}{I - 1} \sum_{i=1}^{I} \frac{\widetilde{v}_{i\bullet}}{\widetilde{v}_{\bullet\bullet}} \, (b_{i,2} - \bar{b}_2)^2 - \frac{I \, \widehat{\sigma}_\varepsilon^2}{\widetilde{v}_{\bullet\bullet}} \right],$$

where $c_1 = \frac{I-1}{I} \left\{ \sum_{i=1}^{I} \frac{v_{i\bullet}}{v_{\bullet\bullet}} \left(1 - \frac{v_{i\bullet}}{v_{\bullet\bullet}}\right) \right\}^{-1}$, $\bar{b}_1 = v_{\bullet\bullet}^{-1} \sum_{i=1}^{I} v_{i\bullet} \, b_{i,1}$, $c_2 = \frac{I-1}{I} \left\{ \sum_{i=1}^{I} \frac{\widetilde{v}_{i\bullet}}{\widetilde{v}_{\bullet\bullet}} \left(1 - \frac{\widetilde{v}_{i\bullet}}{\widetilde{v}_{\bullet\bullet}}\right) \right\}^{-1}$, and $\bar{b}_2 = \widetilde{v}_{\bullet\bullet}^{-1} \sum_{i=1}^{I} \widetilde{v}_{i\bullet} \, b_{i,2}$.

3 Robust Fitting of Heavy-Tailed Mixed Linear Models

In this section, we present our robust approach for fitting of heavy-tailed mixed linear models. First, we consider heavy-tailed regression models in the framework of mixed linear models. Thereafter, in Section 3.2 we modify ATL procedures for robust-efficient estimation when residuals are from symmetric or asymmetric heavy-tailed log-location-scale families.

3.1 Heavy-Tailed Mixed Linear Models

For standard mixed linear models it is assumed that residuals are normally distributed. Here, we discuss a more general case where residuals are from log-location-scale families. This class of models is commonly pursued for failure time analysis or modeling of heavy-tailed data.

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

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

defined for $-\infty < \lambda_{it} < \infty$, $\sigma_\varepsilon > 0$, and where F_0 is the standard (i.e., $\lambda_{it} = 0, \sigma_\varepsilon = 1, v_{it} = 1$) cdf of the underlying location-scale family $F(\lambda_{it}, \sigma_\varepsilon^2 / v_{it})$. 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 mixed linear model, given by (2.1), becomes

$$\log(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i = \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i \quad i = 1, \ldots, I, \tag{3.1}$$

where $\log(\mathbf{y}_i) = (\log(y_{i1}), \ldots, \log(y_{i\tau_i}))'$ and $\boldsymbol{\lambda}_i$ is the τ_i-dimensional vector of the within-subject locations λ_{it} that consist of the *population location* $\boldsymbol{\beta} \in \mathbb{R}^p$ and the subject-specific *location deviation* $\boldsymbol{\alpha}_i \in \mathbb{R}^q$. While assumptions (b) and (c) in Section 2.1 remain valid for random components, in the mixed linear model given by (3.1) we replace (a) by assumption (a'), for which

$$\boldsymbol{\alpha}_i \sim N_q(\mathbf{0}, \mathbf{D}) \quad \text{and} \quad \boldsymbol{\varepsilon}_i \sim F_{\tau_i}(\mathbf{0}, \mathbf{R}_i), \quad i = 1, \ldots, I,$$

where $F_{\tau_i}(\mathbf{0}, \mathbf{R}_i)$ is the τ_i-dimensional multivariate cdf with location-scale distributions $F(0, \sigma_\varepsilon^2 / v_{it})$ as margins, and $\mathbf{D} = \mathrm{diag}(\sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_q}^2)$ and $\mathbf{R}_i = \sigma_\varepsilon^2 \, \mathrm{diag}\left(v_{i1}^{-1}, \ldots, v_{i\tau_i}^{-1}\right)$ are positive-definite variance-

covariance matrices. Hence, from (a'), (b), (c), we obtain the hierarchical formulation of the heavy-tailed mixed linear model that is given by

$$\log(\mathbf{y}_i) \,|\, \boldsymbol{\alpha}_i \; \sim \; F_{\tau_i}\big(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\alpha}_i, \, \mathbf{R}_i\big), \quad i = 1, \ldots, I. \tag{3.2}$$

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, Student's t, Cauchy, and Gumbel (extreme-value), respectively. Special cases of the τ_i-dimensional distributions $F_{\tau_i}\big(\boldsymbol{\lambda}_i, \mathbf{R}_i\big)$ are the well-known elliptical distributions such as multivariate normal (see Section 2.1) and the heavy-tailed multivariate Student's t with ν degrees of freedom.

3.2 ATL estimation: Log-location-scale Families

3.2.1 (RE)ML for Asymmetric Errors

The objective of traditional fitting procedures in model (2.1) is first to find the likelihood-based estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ and the predictors $\widehat{\boldsymbol{\alpha}}_{\mathrm{BLUP},\,i}$, and then to predict the random variable $\mathrm{E}(\mathbf{y}_i|\boldsymbol{\alpha}_i)$ with the best linear unbiased predictor. In view of the mixed linear model defined by (3.1), residuals that follow *skewed* log-location-scale distributions have no longer mean zero. Thus, the expectations $\mathrm{E}(\log(\mathbf{y}_i))$ and $\mathbf{E}(\log(\mathbf{y}_i)|\boldsymbol{\alpha}_i)$ differ from $\mathbf{X}_i\boldsymbol{\beta}$ and the location $\boldsymbol{\lambda}_i$, respectively. Therefore, when fitting model (2.1) with (RE)ML we use *shifted* linear predictors of location $\boldsymbol{\lambda}_i$:

$$\widehat{\boldsymbol{\lambda}}_i = \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} + \mathbf{Z}_i\widehat{\boldsymbol{\alpha}}_{\mathrm{BLUP},\,i} + \widehat{\mathbf{E}}_{F_{\tau_i}}(\boldsymbol{\varepsilon}_i), \tag{3.3}$$

where $\widehat{\mathbf{E}}_{F_{\tau_i}}(\boldsymbol{\varepsilon}_i)$ is the expectation vector of the τ_i-variate cdf $F_{\tau_i}(\mathbf{0}, \widehat{\mathbf{R}}_i)$. As we will see, this correction requires that the first two moments of the asymmetric log-location-scale error model F_{τ_i} exist.

3.2.2 Corrected ATL Procedures for Asymmetric Errors

For robust-efficient fitting of the mixed linear model with normal random components, Dornheim (2009) developed adaptively truncated likelihood (ATL) methods. We adopt his procedure to more general log-location-scale error models defined by (3.2). Specifically, the corrected ATL estimators for location $\boldsymbol{\lambda}_i$ and variance components $\sigma^2_{\alpha_1}, \ldots, \sigma^2_{\alpha_q}, \sigma^2_{\varepsilon}$ can be found by the following three-step

procedure:

STEP 1: Detection of Within-Risk Outliers

Consider the random sample $(\mathbf{x}_{i1}, \mathbf{z}_{i1}, \log(y_{i1}), v_{i1}), \ldots, (\mathbf{x}_{i\tau_i}, \mathbf{z}_{i\tau_i}, \log(y_{i\tau_i}), v_{i\tau_i})$. From (3.2) we continue with the linear relationship

$$\log(y_{it}) = (\mathbf{x}_{it}, \mathbf{z}_{it})(\boldsymbol{\beta}', \boldsymbol{\alpha}_i')' + \varepsilon_{it}, \quad t = 1, \ldots, \tau_i, \tag{3.4}$$

where $(\mathbf{x}_{it}, \mathbf{z}_{it})' \in \mathbb{R}^{p+q}$ is the stacked predictor and $(\boldsymbol{\beta}', \boldsymbol{\alpha}_i')'$ is the corresponding $(p+q)$-dimensional regression parameter vector. For risk-specific location $(\boldsymbol{\beta}', \boldsymbol{\alpha}_i')'$ and scale σ_{ε_i} we compute pairs of initial highly robust estimators $(\mathbf{T}_{\tau_i}, S_{\tau_i})$. If $S_{\tau_i} > 0$, we evaluate the standardized residuals

$$r_{it} = \left[\log(y_{it}) - (\mathbf{x}_{it}, \mathbf{z}_{it})\mathbf{T}_{\tau_i}\right] v_{it}^{1/2} / S_{\tau_i}, \quad t = 1, \ldots, \tau_i, \tag{3.5}$$

where $v_{it} > 0$ is the known exposure of risk i in period t. We find the pooled order statistics $r_{(j)}$ of r_{it}, where $i = 1, \ldots, I$; $t = 1, \ldots, \tau_i$, and $j = 1, \ldots, m = \sum_{i=1}^{I} \tau_i$. Then, the portion of gross-errors in the data can be measured by

$$\begin{aligned} d_g &= \max_{j > j_0^g} \left[F_0(r_{(j)}) - \frac{(j-1)}{m} \right]^+, \quad \text{and} \\ d_l &= \max_{j < j_0^l} \left[\frac{(j-1)}{m} - F_0(r_{(j)}) \right]^+, \end{aligned}$$

where the notation $[\cdot]^+$ is the positive part, $j_0^g = \max\{j : r_{(j)} < \eta^g\}$, $j_0^l = \min\{j : r_{(j)} > \eta^l\}$, $\widehat{F}_r(u) = m^{-1} \sum_{j=1}^m \mathbb{1}\{r_j \leq u\}$ is the empirical distribution function of standardized residuals, and $\eta^g = F_0^{-1}(0.975)$ and $\eta^l = F_0^{-1}(0.025)$. The adaptive cutoff values are given by

$$\begin{aligned} c_g &= \min\left\{u : \widehat{F}_r(u) \geq 1 - d_g\right\} = r_{(j_g)}, \quad \text{with } j_g = m - \lfloor md_g \rfloor, \ j_g > j_0^g, \ c_g > \eta^g, \quad \text{and} \\ c_l &= \max\left\{u : \widehat{F}_r(u) \leq d_l\right\} = r_{(j_l)}, \quad \text{with } j_l = \lfloor md_l \rfloor, \ j_l < j_0^l, \ c_l < \eta^l. \end{aligned}$$

Finally, we define the hard-rejection weights $w_{it} = 1$, for $c_l < r_{it} < c_g$, and $= 0$, for $r_{it} \leq c_l$ or $r_{it} \geq c_g$. We detect those $\lfloor md_g \rfloor$ and $\lfloor md_l \rfloor$ observations as outlying that have largest and smallest residuals, respectively. Let $j_m = m - \lfloor md_g \rfloor - \lfloor md_l \rfloor$. Then, the "pre-cleaned" random sample is of the form $\mathbf{S}^* = \{(\mathbf{X}_i^*, \mathbf{Z}_i^*, \log(\mathbf{y}_i^*), \boldsymbol{v}_i^*), \mathbf{X}_i^* \in \mathbb{R}^{t_{\tau_i} \times p}, \mathbf{Z}_i^* \in \mathbb{R}^{t_{\tau_i} \times q}, \log(\mathbf{y}_i^*) \in \mathbb{R}^{t_{\tau_i}}, \boldsymbol{v}_i^* \in \mathbb{R}^{t_{\tau_i}}, i = 1, \ldots, I\}$.

It is obtained from the original sample $\mathbf{S} = \{(\mathbf{X}_i, \mathbf{Z}_i, \log(\mathbf{y}_i), \boldsymbol{v}_i), \mathbf{X}_i \in \mathbb{R}^{t \times p}, \mathbf{Z}_i \in \mathbb{R}^{t \times q}, \log(\mathbf{y}_i) \in \mathbb{R}^t, \boldsymbol{v}_i \in \mathbb{R}^t, i = 1, \ldots, I\}$ by truncation (deletion) of those rows t for which $w_{it} = 0$.

STEP 2: Detection of Between-Risk Outliers

We are given the pre-cleaned random sample $\mathbf{S}^*$ and suppose $\widehat{\boldsymbol{\beta}}_0$ and $\widehat{\boldsymbol{\alpha}}_i$ are initial high breakdown estimates of $(\boldsymbol{\beta}', \boldsymbol{\alpha}_i')'$. For estimates $\{(\widehat{\alpha}_{1j}, \ldots, \widehat{\alpha}_{Ij}), j = 1, \ldots, q\}$ of the independent and identically distributed random effects $\boldsymbol{\alpha}_i \sim N_q(\mathbf{0}, \mathbf{D})$, we determine highly robust multivariate S-estimates of scale $\widehat{\sigma}_{0\alpha_j}$. Then, it is natural to put $\widehat{\mathbf{D}}_0 = \operatorname{diag}(\widehat{\sigma}_{0\alpha_1}^2, \ldots, \widehat{\sigma}_{0\alpha_q}^2)$ and $\widehat{\mathbf{R}}_{0i} = \widehat{\sigma}_{0\varepsilon}^2 \operatorname{diag}(v_{i1}^{-1}, \ldots, v_{it_{\tau_i}}^{-1}) f_{c_m}$ such that $\widehat{\mathbf{V}}_{0i}(\widehat{\boldsymbol{\theta}}_0) = \mathbf{Z}_i^* \widehat{\mathbf{D}}_0 \mathbf{Z}_i^{*\prime} + \widehat{\mathbf{R}}_{0i} \in \mathbb{R}^{t_{\tau_i} \times t_{\tau_i}}$, where $\widehat{\boldsymbol{\theta}}_0 = (\widehat{\sigma}_{0\alpha_1}^2, \ldots, \widehat{\sigma}_{0\alpha_q}^2, \widehat{\sigma}_{0\varepsilon}^2)$ and f_{c_m} is a Fisher consistency correction factor as defined in (3.6). Once we have found robustified Mahalanobis distances

$$\widehat{d}_i := \left\{ \left(\log(\mathbf{y}_i^*) - \mathbf{X}_i^* \widehat{\boldsymbol{\beta}}_0 \right)' \left[\widehat{\mathbf{V}}_{0i}(\widehat{\boldsymbol{\theta}}_0) \right]^{-1} \left(\log(\mathbf{y}_i^*) - \mathbf{X}_i^* \widehat{\boldsymbol{\beta}}_0 \right) \right\}^{1/2},$$

where $\widehat{d}_i^2$ is approximately $\chi_{t_{\tau_i}}^2$-distributed with (possibly) unequal degrees of freedom t_{τ_i}, $i = 1, \ldots, I$, we apply an accurate normal approximation for the cumulative distribution function of the chi-square distribution, and establish standardized robust Mahalanobis distances through the transformation

$$\widehat{d}_i^* = \left\{ \left(\frac{\widehat{d}_i^2}{t_{\tau_i}} \right)^{1/6} - \frac{1}{2} \left(\frac{\widehat{d}_i^2}{t_{\tau_i}} \right)^{1/3} + \frac{1}{3} \left(\frac{\widehat{d}_i^2}{t_{\tau_i}} \right)^{1/2} - \left(\frac{5}{6} - \frac{1}{9(t_{\tau_i})} - \frac{7}{648(t_{\tau_i})^2} + \frac{25}{2187(t_{\tau_i})^3} \right) \right\}$$
$$\times \left\{ \frac{1}{18(t_{\tau_i})} + \frac{1}{162(t_{\tau_i})^2} - \frac{37}{11664(t_{\tau_i})^3} \right\}^{-1/2},$$

such that approximately $\widehat{d}_i^* \sim N(0,1)$. Since $\widehat{d}_i^*$ is strictly increasing in $\widehat{d}_i^2$, outlying individuals i that have typically large $\widehat{d}_i^2$ will inherit a large positive $\widehat{d}_i^*$. Let $\widehat{F}_{\widehat{d}_i^*}(u) = I^{-1} \sum_{i=1}^I \mathbb{1}\left\{ \widehat{d}_i^* \le u \right\}$ be the empirical distribution of standardized robust Mahalanobis distances. Then, the portion of outliers in the data can be measured by

$$\gamma_I = \max_{i > i_0} \left[\Phi\left(\widehat{d}_{(i)}^* \right) - \frac{i-1}{I} \right]^+,$$

where $i_0 = \max\left\{ i : \widehat{d}_{(i)}^* < \xi \right\}$, $\widehat{d}_{(i)}^*$ is the i-th order statistic of $\widehat{d}_i^*$, and ξ denotes a large upper quantile of the standard normal distribution, e.g. $\xi = \Phi^{-1}(0.975)$. The adaptive cutoff value is found by $c_I = \widehat{F}_{\widehat{d}_i^*}^{-1}(1 - \gamma_I) = \widehat{d}_{(i_I)}^*$, with $i_I = I - \lfloor \gamma_I I \rfloor$, $i_I > i_0$, and $c_I > \xi$. We introduce the hard-rejection weight function $\omega_i\left(\widehat{d}_i^*; c_I \right) = \mathbb{1}\left\{ \widehat{d}_i^*/c_I < 1 \right\}$ and propose to eliminate those subjects i for which $\omega_i\left(\widehat{d}_i^*; c_I \right) = 0$. Removal of $\lfloor \gamma_I I \rfloor$ subjects from the pre-cleaned sample $\mathbf{S}^* =$

$\{(\mathbf{X}_1^*, \mathbf{Z}_1^*, \log(\mathbf{y}_1^*), \boldsymbol{v}_1^*), \dots, (\mathbf{X}_I^*, \mathbf{Z}_I^*, \log(\mathbf{y}_I^*), \boldsymbol{v}_I^*)\}$, results in $\mathbf{S}_{i_I}^c = \{(\mathbf{X}_1^c, \mathbf{Z}_1^c, \log(\mathbf{y}_1^c), \boldsymbol{v}_1^c), \dots, (\mathbf{X}_{i_I}^c, \mathbf{Z}_{i_I}^c,$
$\log(\mathbf{y}_{i_I}^c), \boldsymbol{v}_{i_I}^c)\}$, a "cleaned" random sample of length i_I.

STEP 3: Corrected Adaptively Truncated Likelihood Estimators

Based on the adaptively truncated random sample $\mathbf{S}_{i_I}^c$, we apply fully efficient likelihood methods such as (restricted) maximum likelihood, and compute re-weighted parameter estimates $\widehat{\boldsymbol{\beta}}_{\text{ATL}}$ and $\widehat{\boldsymbol{\theta}}_{\text{ATL}} = (\widehat{\sigma}_{\alpha_1}^2, \dots, \widehat{\sigma}_{\alpha_q}^2, \widehat{\sigma}_{\varepsilon}^2)$. Then, the maximum likelihood estimator $\widehat{\boldsymbol{\tau}}_{\text{ATL}} = (\widehat{\boldsymbol{\beta}}_{\text{ATL}}', \widehat{\boldsymbol{\theta}}_{\text{ATL}}')'$ is defined by the non-linear constrained minimization problem

$$\widehat{\boldsymbol{\tau}}_{\text{ATL}} = \arg\min_{\boldsymbol{\tau} \in \Omega} L(\boldsymbol{\tau}),$$

where

$$L(\boldsymbol{\tau}) = \sum_{i=1}^{I} \frac{1}{2} \omega_i\left(\widehat{d}_i^*; c_I\right) \left\{ (\log(\mathbf{y}_i^*) - \mathbf{X}_i^* \boldsymbol{\beta})' \mathbf{V}_i^{*-1} (\log(\mathbf{y}_i^*) - \mathbf{X}_i^* \boldsymbol{\beta}) + \frac{\kappa}{2} \log|\mathbf{V}_i^*| \right\}, \qquad \boldsymbol{\tau} = (\boldsymbol{\beta}', \boldsymbol{\theta}')' \in \Omega,$$

is the negative adaptively corrected log-likelihood and $\kappa = \int_{-c_I}^{c_I} u^2 d\Phi(u) \big/ [2(\Phi(c_I) - \Phi(0))]^2$ is the empirical (data-dependent) Fisher consistency correction factor. Suppose the parameter $\boldsymbol{\tau}$ is identifiable and $\widehat{\boldsymbol{\tau}}_{\text{ATL}} \in \Omega = \Omega_{\boldsymbol{\beta}} \times \Omega_{\boldsymbol{\theta}}$, where $\Omega_{\boldsymbol{\beta}} = \{\beta_k, k = 1, \dots, p\}$ and $\Omega_{\boldsymbol{\theta}} = \{\theta_l > 0, l = 1, \dots, q+1\}$. Then, $\widehat{\boldsymbol{\tau}}_{\text{ATL}}$ is a solution of the unbiased estimating equations $\partial L(\boldsymbol{\tau})/\partial \boldsymbol{\tau} = 0$. We define the *robust-efficient generalized least squares estimator* by

$$\widehat{\boldsymbol{\beta}}_{\text{rGLS}}(\widehat{\boldsymbol{\theta}}_{\text{ATL}}) = \left(\sum_{i=1}^{i_I} \mathbf{X}_i^{c'} [\widehat{\mathbf{V}}_i^c]^{-1} \mathbf{X}_i^c \right)^{-1} \sum_{i=1}^{i_I} \mathbf{X}_i^{c'} [\widehat{\mathbf{V}}_i^c]^{-1} \mathbf{y}_i^c,$$

where $\widehat{\mathbf{V}}_i^c(\widehat{\boldsymbol{\theta}}_{\text{ATL}}) = \mathbf{Z}_i^c \widehat{\mathbf{D}} \mathbf{Z}_i^{c'} + \widehat{\mathbf{R}}_i$, with $\widehat{\mathbf{D}} = \text{diag}(\widehat{\sigma}_{\alpha_1}^2, \dots, \widehat{\sigma}_{\alpha_q}^2)$ and $\widehat{\mathbf{R}}_i = \widehat{\sigma}_{\varepsilon}^2 \text{diag}(v_{i1}^{-1}, \dots, v_{it_{\tau_i}}^{-1})$. The *robust best linear unbiased predictor* is calculated from

$$\widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i}(\widehat{\boldsymbol{\theta}}_{\text{ATL}}) = \widehat{\mathbf{D}} \mathbf{Z}_i^{c'} [\widehat{\mathbf{V}}_i^c]^{-1} (\mathbf{y}_i^c - \mathbf{X}_i^c \widehat{\boldsymbol{\beta}}_{\text{rGLS}}(\widehat{\boldsymbol{\theta}}_{\text{ATL}})), \quad i = 1, \dots, I.$$

Following comments in the previous section, from (3.3) we obtain the *robust best linear unbiased predictor for location* by

$$\widehat{\boldsymbol{\lambda}}_i = \mathbf{X}_i^c \widehat{\boldsymbol{\beta}}_{\text{rGLS}} + \mathbf{Z}_i^c \widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i} + \widehat{\mathbf{E}}_{F_0}(\boldsymbol{\varepsilon}_i), \quad i = 1, \dots, I,$$

where $\widehat{\mathbf{E}}_{F_0}(\boldsymbol{\varepsilon}_i)$ is the expectation vector of the t_{τ_i}-variate cdf $F_{t_{\tau_i}}(\mathbf{0}, \widehat{\mathbf{R}}_i)$. This completes our three-step *corrected adaptively truncated likelihood* (CATL) procedure.

Remark 2: *Initial High-Breakdown Estimators for Asymmetric Errors*

For the fixed effects model $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \widetilde{\boldsymbol{\varepsilon}}$ we assume that approximately $\widetilde{\boldsymbol{\varepsilon}} \sim N_{\tau_i}(\mathbf{0}, \mathbf{V}_i(\boldsymbol{\theta}))$, and compute multivariate S-estimates $\widehat{\boldsymbol{\beta}}_0$. Then, for each risk we fit multiple regression models of the form $\log(\mathbf{y}_i) - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_0 = \mathbf{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i, \ i = 1, \dots, I$, where the noise term $\boldsymbol{\varepsilon}_i$ follows some (asymmetric) long-tailed distribution F_{τ_i} as described in Section 3.1. We employ highly robust but *corrected S-*estimators and find estimates $\widehat{\boldsymbol{\alpha}}_i$ for location $\boldsymbol{\alpha}_i$. This class of estimators has been used by Marazzi *et al.* (2004) and inherits robustness and efficiency properties from usual S-estimators while producing Fisher-consistent estimates of $(\boldsymbol{\alpha}_i, \sigma_{\varepsilon_i})$ when the standard location-scale model is F_0. Note that, when the distribution of the error term $\boldsymbol{\varepsilon}_i$ belongs to a class of exponential families including the log–gamma distribution, one might also consider initial high–breakdown estimators as proposed by Bianco *et al.* (2005). Lastly, we obtain the initial robust estimators $\mathbf{T}_{\tau_i} = (\widehat{\boldsymbol{\beta}}_0', \widehat{\boldsymbol{\alpha}}_i')'$ and $S_{\tau_i} = \widehat{\sigma}_{0\varepsilon}$. More details about initial robust estimation in the mixed linear model and theoretical and computational aspects of ATL estimators are available in Dornheim (2009). □

Remark 3: *Important Special Cases*

We calibrate corrected S-estimators for maximum asymptotic breakdown point and consistency for normal, Cauchy, Student's t-, Gumbel and logistic distribution. Results of calibration constants l_0 for location, s_0 for scale, and k_0 for full robustness using Tukey's smooth biweight function (Marazzi *et al.*, 2004) are summarized in Table 1. □

Table 1. The consistency factors l_0, s_0 and k_0 of corrected S-estimators
for selected location-scale families.

Factor	Standard Location-Scale Family						
	Cauchy	t_3	t_8	t_{20}	Normal	Gumbel	Logistic
l_0	0.000	0.000	0.000	0.000	0.000	-0.135	0.755
s_0	1.000	1.000	1.000	1.000	1.000	1.000	1.000
k_0	2.436	1.791	1.633	1.581	1.548	1.718	1.745

Remark 4: *Fisher Consistency Factor*

We assume that the hypothesized error model F_0 has finite second moments (c_g and c_l may be unbounded when no outliers are present). Then, for the hard-rejection rule we obtain Fisher-correction factors f_{c_m} that are given by

$$f_{c_m} = \int_{c_l}^{c_g} u^2 dF_0(u) - \Big[\int_{c_l}^{c_g} u dF_0(u) \Big]^2. \tag{3.6}$$

$\square$

Remark 5: *Types of Outliers: What can be Detected?*

The designed procedure provides protection at the observational level and at the risk level. The unusual claims that might be removed can be large, small, or medium (so-called *leverage* points). Their magnitude, however, is not the reason why they might be treated as outliers. It is our choice of the model assumptions that makes them look as "misfits". To make these ideas more transparent, let us consider a simple example. Suppose a set of claims, with a significant proportion of large claims, was actually generated by t_5 (Student's t with 5 degrees of freedom) but we assumed that the underlying model is t_{10}, a thinner-tailed distribution. Then, our procedure would recognize that the given data set and the assumed model are not in tune, would identify and remove outliers, and apply a MLE-type method on the clean data. After that, when we turn to pricing, the identified outliers would be recycled with their "excesses" being redistributed across all risk premiums (see Section 4.1). On the other hand, if we assumed a thicker-tailed distribution (e.g., t_2), the procedure would react the same way, except that now those outlier-excesses might be negative. Finally, if made the right distributional assumption, i.e., t_5, then the proposed procedure would treat such observations as representative claims and no removals would occur. (That is where the full-efficiency property becomes so crucial.) In the latter case, the pricing formula in Section 4 would have no "extra premium", i.e., $\mu_t^{\text{extra}} = 0$.

It should also be noted that by pre-cleaning the data from observational outliers first, we make risks in some sense homogeneous (up to unknown risk characteristics), and that allows us to identify risk-outliers in Step 2. In theory, it is even possible to remove all risks because of within-risk outliers, when in fact they are quite "usual" if properly compared with their own risk characteristics. In order to avoid excessive truncation at the risk level, we follow the ideas of Huggins and Staudte (1994) and Welsh and Richardson (1997). $\square$

4   Robust Credibility Model

In this section we develop our robust-efficient regression credibility model. First, we present description of the model. Then, in Section 4.2 we introduce a class of robust-efficient adaptively truncated credibility estimators for risk premiums that cover the majority of claims. Thereafter, we discuss the treatment of truncated large claims and present the final robust credibility pricing formula in Section 4.3.

4.1   Model Description

Let us continue with the setup of Section 2.3. Following Gisler and Reinhard (1993) we divide the true risk premium μ_{it} into two parts —a risk premium for the *ordinary* claims, $\mu_{it}^{\text{ordinary}}$, and a risk premium for the *extraordinary* claims, μ_{it}^{extra}— and estimate each component separately. The extraordinary premium represents the expected claims load generated mainly by extraordinary events (e.g., big fires or hurricanes), whose occurrence is rare but usually leads to outlier observations of the affected loss ratios. Thus, it is reasonable to assume that

$$\mu_{it}^{\text{extra}} = \mu_t^{\text{extra}}, \qquad \text{for } i = 1, \ldots, I; \ t = 1, \ldots, \tau_i.$$

In this setting the ordinary premium $\mu_{it}^{\text{ordinary}}$ is estimated using the robust CATL procedure, which automatically identifies ordinary claims. Then, the robust credibility estimator is given by

$$\widehat{\mu}_{it} = \widehat{\mu}_{it}^{\text{ordinary}} + \mu_{it}^{\text{extra}},$$

such that $\mu_t = \mathrm{E}(y_{it}) = \mathrm{E}(\mathrm{E}(y_{it}|\boldsymbol{\alpha}_i)) = \mathrm{E}(\widehat{\mu}_{it})$. Specific estimators are presented in subsequent sections.

4.2   Estimation of Structural Parameters

We are given claim data of the form $(\mathbf{X}_i, \mathbf{Z}_i, \mathbf{y}_i, \boldsymbol{v}_i)$, $i = 1, \ldots, I$, where $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{i\tau_i})$ are known volume measures of the observed claim vector $\mathbf{y}_i$. Then, we set forth with the following distributional assumptions:

The log-transformed observation vectors $log(\mathbf{y}_i)$ are distributed according to some multivariate

location-scale model $F_{\tau_i}(\boldsymbol{\lambda}_i, \mathbf{R}_i)$, defined by (3.1), where $\mathbf{R}_i = \sigma_{\varepsilon_i}^2 \, diag(v_{i1}^{-1}, \ldots, v_{i\tau_i}^{-1})$. The known designs $\mathbf{X}_i \in \mathbb{R}^{p \times \tau_i}$ and $\mathbf{Z}_i \in \mathbb{R}^{q \times \tau_i}$ of explanatory variables are linked to the subject -specific location vector $\boldsymbol{\lambda}_i$ by the linear relationship $\boldsymbol{\lambda}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\alpha}_i$.

This statement which assumes that all (logarithmic transformed) claims are ordinary, is necessary for the choice of the right hypothetical model distribution F_0 and, hence, for defined rejection rules in *STEP 1* of the CATL procedure.

Note that the choice of the designs $\mathbf{X}_i$ and $\mathbf{Z}_i$ already determines the standard credibility model pursued by the actuary. This has been illustrated in Section 2.2. The objective of CATL procedures is first to clean the given sample $(\mathbf{X}_i, \mathbf{Z}_i, \log(\mathbf{y}_i), \boldsymbol{v}_i)$, $i = 1, \ldots, I$, then to estimate robustly the structural parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta} = (\sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_q}^2, \sigma_\varepsilon^2)$, and, finally, to predict the risk-specific location vectors $\boldsymbol{\lambda}_i$. The corresponding CATL-estimators $\widehat{\boldsymbol{\beta}}_{\text{rGLS}}$ and $\widehat{\boldsymbol{\theta}}$, and *robust credibility predictors* $\widehat{\boldsymbol{\lambda}}_i$ are presented in Section 3.2. The chosen credibility model specifies entirely all covariates. Therefore, there is no need to take into consideration the occurrence of outliers in the designs $\mathbf{X}_i$ and $\mathbf{Z}_i$. Hence, we find robust ordinary net premiums $\widehat{\mu}_{it}^{\text{ordinary}} = \widehat{\mu}_{it}^{\text{ordinary}}(\widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i})$, $i = 1, \ldots, I$, $t = 1, \ldots, \tau_i + 1$, by computation of the empirical *limited expected value* (LEV) that is given by

$$
\begin{aligned}
\widehat{\mu}_{it}^{\text{ordinary}} &= \widehat{\text{LEV}}_{F_0}\left[\log(Y_{it}); q_l(p_l), q_g(p_g), \widehat{\lambda}_{it}, \widehat{\sigma}_{\varepsilon_i}^2/v_{it}\right] \qquad (4.1) \\
&= \int_{q_l}^{q_g} (y_{it} - q_l) \, dF_0\left(\log(y_{it}); \widehat{\lambda}_{it}, \widehat{\sigma}_{\varepsilon_i}^2/v_{it}\right) + (q_g - q_l)(1 - p_g), \quad y_{it} > 0,
\end{aligned}
$$

where $\widehat{\lambda}_{it} = \mathbf{x}_{it}\widehat{\boldsymbol{\beta}}_{\text{rGLS}} + \mathbf{z}_{it}\widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i} + \widehat{\text{E}}_{F_0}(\varepsilon_{it})$, and $q_l = \exp\left\{F_0^{-1}\left(p_l; \widehat{\lambda}_{it}, \widehat{\sigma}_{\varepsilon_i}^2/v_{it}\right)\right\}$ and $q_g = \exp\left\{F_0^{-1}\left(p_g; \widehat{\lambda}_{it}, \widehat{\sigma}_{\varepsilon_i}^2/v_{it}\right)\right\}$ are some lower and upper quantiles, respectively, given the limiting probabilities $0 \leq p_l < p_g \leq 1$. It is common to choose $p_l \leq 0.001$ and $p_g \geq 0.999$. In insurance, for the loss Y_{it} the quantiles q_l and q_g can be interpreted as *deductible* and *upper limit*, respectively (Klugman *et al.*, 2004).

Remark 6: *Important Special Cases*

a. The net premium principle defined by (4.1) can always be evaluated regardless of the existence of the expectation of the underlying heavy-tailed model distribution. Well-known examples where no moments exist are log-Student's t-distributions with $\nu \geq 1$ degrees of freedom.

b. If the first moment exists, from (4.1) we obtain the usual net premium

$$
\begin{aligned}
\widehat{\mu}_{it}^{\text{ordinary}} &= \widehat{\mathrm{E}}_{F_0}\Big[\log(Y_{it}); \widehat{\lambda}_{it}, \widehat{\sigma}_{\varepsilon_i}^2/v_{it}\Big] = \widehat{\mathrm{LEV}}_{F_0}\Big[\log(Y_{it}); q_l(p_l) = 0, q_g(p_g) \to \infty, \widehat{\lambda}_{it}, \widehat{\sigma}_{\varepsilon_i}^2/v_{it}\Big] \\
&= \int_0^\infty y_{it}\, dF_0\Big(\log(y_{it}); \widehat{\lambda}_{it}, \widehat{\sigma}_{\varepsilon_i}^2/v_{it}\Big), \qquad y_{it} > 0,
\end{aligned}
$$

with $q_l(p_l) = 0$ as $p_l \to 0$ and $q_g(p_g) \to \infty$ as $p_g \to 1$. $\qquad\qquad\qquad\square$

4.3 Robust Credibility Ratemaking

4.3.1 Treatment of Excess Claims

Let us define extraordinary premiums for identified excess claims. From Section 4.2, we are given the ordinary premiums $\widehat{\mu}_{it}^{\text{ordinary}}$ and claims y_{it}. The risk-specific excess claim amount of insured i at time t is defined by $\widehat{O}_{it} = (y_{it} - q_l) - \widehat{\mu}_{it}^{\text{ordinary}}$, for $y_{it} > q_l$, $= (q_g - q_l) - \widehat{\mu}_{it}^{\text{ordinary}}$, for $y_{it} > q_g$, and $= -\widehat{\mu}_{it}^{\text{ordinary}}$, for $y_{it} < q_l$. Further, let I_t denote the number of insureds in the portfolio at the time of t and $\tau_i^* = \max_{1 \le i \le I} \tau_i$, the maximum horizon among all risks. For each period $t = 1, \ldots, \tau_i^*$, we find the mean cross-sectional overshot of excess claims $\widehat{O}_{\bullet t} = I_t^{-1} \sum_{i=1}^{I_t} \widehat{O}_{it}$, and fit robustly the random effects model

$$
\widehat{O}_{\bullet t} = \mathbf{o}_t \boldsymbol{\xi} + \varepsilon_t^*, \qquad t = 1, \ldots, \tau_i^*,
$$

where $\mathbf{o}_t$ is the n-dimensional row-vector of covariates for the hypothetical mean of overshots $\boldsymbol{\xi} \in \mathbb{R}^n$. Let $\widehat{\boldsymbol{\xi}}$ be the robust estimate of $\boldsymbol{\xi}$. Common robust estimators pursued are LMS or S-estimators that are tuned for normality. Then, the extraordinary premium common to all risks i is given by

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

In practice, the data collected does not provide much information about the distribution of outliers. Hence, we recommend to select the intercept model or the linear trend model with covariates $\mathbf{o}_t = 1$ and $\mathbf{o}_t = (1, t)$, respectively. Latter allows to model possible time-dependency for occurrence of excess claims. We do not assume any time-dependency of excess claims and, therefore, choose $\mathbf{o}_t = 1$.

4.3.2 Robust Credibility Estimator

Given the credibility estimates of expected ordinary premiums, $\widehat{\mu}_{i,\tau_i+1}^{\text{ordinary}}(\widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i})$, and excess claims, $\mu_{i,\tau_i+1}^{\text{extra}}$, we find the portfolio-unbiased and robust regression credibility estimator defined by

$$\widehat{\mu}_{i,\tau_i+1}^{\text{CATLC}}(\widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i}) = \widehat{\mu}_{i,\tau_i+1}^{\text{ordinary}}(\widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i}) + \mu_{i,\tau_i+1}^{\text{extra}}, \qquad i = 1, \ldots, I,$$

where CATLC denotes the robust-efficient *corrected adaptively truncated likelihood credibility* estimator. From the actuarial point of view, premiums assigned to the insured have to be positive. Therefore, we suggest to determine pure premiums by $\max\left\{0, \widehat{\mu}_{i,\tau_i+1}^{\text{CATLC}}(\widehat{\boldsymbol{\alpha}}_{\text{rBLUP},i})\right\}$.

5 Simulations and Practical Issues

In this section, we illustrate—via simulations—how the proposed new robust credibility estimators work in practice and how they compare with classical ratemaking procedures in credibility. We introduce a contaminating model that allows us to explore short- and long-run performances of our methods under several data generating scenarios. The two types of contamination we consider are:

- contamination of error distribution (ε-distribution),

- contamination of random effects ($\boldsymbol{\alpha}$-distribution).

The study we perform aims to illustrate two things. First, in the short-run case (where we simulate a single portfolio as actuaries and statisticians would face in practice), the reader can see how similar are the classical and robust estimates of credibility premiums, credibility weights, structural parameters and overall portfolio premium when there is no data contamination. On the other hand, when contamination is present, large disparities occur between the corresponding classical and robust estimates. Second, to establish a pattern and to "prove" that this was not an accidental success story of the robust procedure, we repeat the comparison many times ($R = 1000$) and monitor which methodology performs better. Below are the specific settings of the study.

5.1 Study Design

5.1.1 The Central Models

In the framework of credibility, we fit mixed linear models, defined by (3.1), with intercept, normal random effects and error terms that follow some heavy-tailed log-location-scale model F. We continue with the following two artificial examples.

Example 1: Bühlmann-Straub with Weibull-distributed Claims

From results in Sections 2.2.2 and 3.1, for the Bühlmann-Straub model we are given the regression equation $\log(y_{it}) = \beta + \alpha_i + \varepsilon_{it}$ with risk-specific location $\lambda_{it} = \lambda_i = \beta + \alpha_i$, $i = 1, \ldots, I$; $t = 1, \ldots, \tau_i$. The structural parameters of the hypothesized central model are $\beta = 0$, $\sigma_\alpha^2 = 1.0$ and

$$(W1) \quad \sigma_{\varepsilon_i} = 1.0 \quad \text{and} \quad (W2) \quad \sigma_{\varepsilon_i} \sim U(0.5, 1.5),$$

where $U(u_1, u_2)$ denotes the uniform distribution on the interval (u_1, u_2). This allows us to model risks that are from the same log-location-scale family but have different behavior in their tails. In view of the hierarchical formulation (3.2), the logarithmic claims $\log(y_{it})$ are supposed to be log-Weibull (i.e., Gumbel) distributed. That is, we assume the marginal log-location-scale error model with asymmetric density function

$$f_0(r; \lambda_{it}, \sigma_{\varepsilon_i}^2/v_{it}) = \sigma_{\varepsilon_i}^{-1} v_{it}^{1/2} \exp\left[\left(\frac{r - \lambda_{it}}{\sigma_{\varepsilon_i} v_{it}^{-1/2}}\right) - \exp\left(\frac{r - \lambda_{it}}{\sigma_{\varepsilon_i} v_{it}^{-1/2}}\right)\right], \quad -\infty < r < \infty.$$

Example 2: Hachemeister (revisited) with Log-t-distributed Claims

As a second example, we choose the revised Hachemeister regression model where log-claims follow Student's t-distribution with (known) $\nu \geq 3$ degrees of freedom. Then from Sections 2.3.2 and 3.1, we obtain the regression model $\log(y_{it}) = (1, t - G_{i\bullet})(\boldsymbol{\beta} + \boldsymbol{\alpha}_i) + \varepsilon_{it}$, where $\lambda_{it} = (1, t - G_{i\bullet})(\boldsymbol{\beta} + \boldsymbol{\alpha}_i)$, $i = 1, \ldots, I$; $t = 1, \ldots, \tau_i$, and $G_{i\bullet} = v_{i\bullet}^{-1} \sum_{t=1}^{\tau_i} t \, v_{it}$ is the center of gravity. We assume that the grand location λ increases by 5% per period and risk-specific locations λ_i deviate on average by $\sqrt{0.001}$ or $\sim 3.16\%$. Thus, we set the true fixed effect $\boldsymbol{\beta} = (0, 0.05)'$ and the variance components $\sigma_{\alpha_1}^2 = 1.0$ and $\sigma_{\alpha_2}^2 = 0.001$. Similar to previous example we model

$$(L1) \quad \sigma_{\varepsilon_i} = 1.0 \quad \text{and} \quad (L2) \quad \sigma_{\varepsilon_i} \sim U(0.5, 1.5).$$

Then, the log-location-scale t-model is given by the symmetric density function

$$f_0(r; \nu, \lambda_{it}, \sigma_{\varepsilon_i}^2/v_{it}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \left[1 + \left(\frac{[(r-\lambda_{it})/(\sigma_{\varepsilon_i}v_{it}^{-1/2})]^2}{\nu}\right)\right]^{-(\nu+1)/2}, \quad -\infty < r < \infty,$$

This model includes the special case where claims are lognormal as $\nu \to \infty$.

5.1.2 The Portfolio Structure

We generate an unbalanced portfolio of $I = 45$ risks. It contains 15 risks with the observation period $\tau_1 = \cdots = \tau_{15} = 8$ years of experience, 15 risks with $\tau_{16} = \cdots = \tau_{30} = 10$ years of experience, and 15 risks with $\tau_{31} = \cdots = \tau_{45} = 12$ years of experience. For simplicity we assume that given a risk i, the volume measures stay constant over the horizon $\tau_i + 1$; that is, $v_{it} \equiv v_i$, for $1 \leq t \leq \tau_i + 1$. Then, in each group of 15 risks, there are five risks with volumes $v_{i_1} = \cdots = v_{i_5} = 1$ (small volumes), five risks with volumes $v_{i_6} = \cdots = v_{i_{10}} = 2$ (medium volume), and five risks with volumes $v_{i_{11}} = \cdots = v_{i_{15}} = 3$ (large volumes), where $1 \leq i_1 < i_2 < \cdots < i_{15} \leq 45$.

5.1.3 The Contaminating Model

Let δ denote the proportion of outliers occurring in random components. Then, to study the performance of our methods via simulations, we need a model that would allow us to generate random effects from approximate normal distribution, and error terms from approximate (log-)Weibull and (log)-t-distribution, respectively. This can be accomplished by employing subsequent contaminating models.

(C_0) Normal random effects and Weibull/Student's t-distributed errors (central model with $\delta = 0$).

(C_1) Weibull/Student's t-distributed errors but first component α_{i1} (intercept) of random effects vector $\boldsymbol{\alpha}_i$ is generated from $N(3, 1/9)$ with probability δ.

(C_2) Normal random effects with vertical outliers, where errors with probability δ are generated from the uniform distribution on the interval (a_{it}, b_{it}), denoted by $U(a_{it}, b_{it})$, with the probability density function given by $f_U(r) = 1/(b_{it} - a_{it})$, for $a_{it} < r < b_{it}$, and $= 0$, elsewhere. Parameters

a_{it} and b_{it} are selected as follows:

$$a_{it} = 7 \, \text{LEV}_{F_0}\left[\log(y_{it}); q_l(p_l), q_u(p_u), \lambda_{it}, \sigma^2_{\varepsilon_i}/v_{it}\right] \quad \text{and}$$

$$b_{it} = 13 \, \text{LEV}_{F_0}\left[\log(y_{it}); q_l(p_l), q_u(p_u), \lambda_{it}, \sigma^2_{\varepsilon_i}/v_{it}\right].$$

The lower parameter a_{it} represents a threshold that can be exceeded by the assumed model with probability $F_0^{-1}\left(\log(\text{LEV}_{F_0}[\log(y_{it}); q_l, q_u, \lambda_{it}, \sigma^2_{\varepsilon_i}/v_{it}])\right)$. For instance, 0.0009 (for $\lambda_{it} = 0$ and $\sigma_{\varepsilon_i} = v_{it} = 1$), or 0.0237 (for $\lambda_{it} = 0$, $\sigma_{\varepsilon_i} = 2$, and $v_{it} = 1$), or 0.0309 (for $\lambda_{it} = 0$, $\sigma_{\varepsilon_i} = 3$, and $v_{it} = 1$) when log-Weibull is the central model. The upper parameter b_{it} ensures that the expected value of extraordinary claims is 10 times the ordinary premium $\mu_{it} = \text{LEV}_{F_0}\left[\log(y_{it}); q_l, q_u, \lambda_{it}, \sigma^2_{\varepsilon_i}/v_{it}\right]$, which is reasonable. Note, for log-t claims in Example 2 we choose $p_l = 0.001$ and $p_u = 0.999$ whereas the expected value is used for log-Weibull distributed claims in Example 1.

Each risk is generated according to a δ-contamination model of type (C_1). However, at the same time we allow for contamination of type (C_2) with probability δ. The proportions of contamination are: $\delta = 0.0$ (*Scenario 1*), $\delta = 0.01$ (*Scenario 2*), $\delta = 0.05$ (*Scenario 3*), and $\delta = 0.10$ (*Scenario 4*).

The choice of $U(a_{it}, b_{it})$ is simple and reflects what one would encounter in practice. Insurance portfolios typically generate claims, most of which are relatively small and few are very large. Hence, the chosen uniform distribution ensures that a small fraction of atypical claims consistently appear in data sets of our study. But these data points are not necessarily always the largest and blend in with the heavy-tailed F_0-distributed observations. This setup allows that within-risk observations produced by the central model can exceed those by the contaminating distribution. This means $U(a_{it}, b_{it})$ data are likely to stem from the hypothesized model F_0 which makes it impossible to arrive at conclusions about the specific contaminating model. Similar arguments justify the usage of the normal model $N(3, 1/9)$ as contaminating distribution for producing outlying risks in the portfolio.

5.2 Simulation Results: Credibility Premium Calculation

In this section we present results of the Monte-Carlo study that has been conducted to investigate finite-sample properties of newly introduced robust credibility estimators. Here, we test short- and long-run performances of new truncation methods for credibility with classical pricing models of Bühlmann-

Straub (1970) and Hachemeister in the revised case (see Bühlmann *et al.*, 1997).

The primary interest of this study is on the properties such as unbiasedness (as measure of robustness) and relative efficiency (RE). The latter quantitative tool is commonly defined by $E(\boldsymbol{\mu}) = \sum_{k=1}^{R} \left(||\widehat{\boldsymbol{\mu}}_k^{\text{std}} - \boldsymbol{\mu}||^2 \Big/ ||\widehat{\boldsymbol{\mu}}_k^{\text{CATLC}} - \boldsymbol{\mu}||^2 \right)$, where $||\cdot||$ is the L_2-norm, and $\widehat{\boldsymbol{\mu}}_k^{\text{std}}$ and $\widehat{\boldsymbol{\mu}}_k^{\text{CATLC}}$ denote the k-th credibility estimates of the true ordinary premium vector $\boldsymbol{\mu} = (\mu_{1,\tau_1+1}, \ldots, \mu_{I,\tau_I+1})$ when employing standard and CATL credibility estimators, respectively. The number of portfolios we simulate is denoted by R. In our study we choose $R = 1$ (short-run) and $R = 1,000$ (long-run) replications. We also report mean squared errors defined by $\text{MSE}(\widehat{\boldsymbol{\mu}}; \boldsymbol{\mu}) = R^{-1} \sum_{k=1}^{R} I^{-1} ||\widehat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}||^2$.

To gain some insight about robustness of competing strategies for estimation of true premiums we quantify the standardized bias of the credibility estimators by $bias^*(\widehat{\boldsymbol{\mu}}) = R^{-1} \sum_{k=1}^{R} bias^*(\widehat{\boldsymbol{\mu}}_k)$, where $bias^*(\widehat{\boldsymbol{\mu}}_k) = I^{-1} \sum_{i=1}^{I} (\widehat{\mu}_{i,\tau_i+1,k} - \mu_{i,\tau_i+1,k}) \Big/ \sum_{i=1}^{I} \mu_{i,\tau_i+1,k}$. Further, since portfolio-unbiasedness is an indispensable property for the insurer we also focus on the coverage of future claims, $Y_{i,\tau_i+1,k}$, that are predicted by credibility estimators, $\widehat{\mu}_{i,\tau_i+1,k}$. Hence, it is natural to consider the criterion $C(\mathbf{Y}; \widehat{\boldsymbol{\mu}}) = R^{-1} \sum_{k=1}^{R} \left[\sum_{i=1}^{I} \widehat{\mu}_{i,\tau_i+1,k} \Big/ \sum_{i=1}^{I} Y_{i,\tau_i+1,k} \right]$, where $C(\mathbf{Y}; \widehat{\boldsymbol{\mu}}) \geq 1$ indicates sufficient funding for future reported claims in the portfolio.

5.2.1 Short-run Performance

To get a first impression about the performance of newly introduced CATLC estimators, we generate one portfolio of 45 risks under Scenario 1 and 3 each. In Tables 2 and 3 we provide estimates of structural parameters, credibility premiums, credibility weights, and performance measures when Bühlmann-Straub was applied to Weibull-distributed claims (Example 1) and Hachemeister (revisited) to lognormal-distributed claims (Example 2). Let us start with Example 1.

DISCUSSION OF TABLE 2: In the clean *Scenario 1*, the classical Bühlmann-Straub model (BS) outperforms the robust CATL procedure. The MSE of CATLC is 64% larger than that of BS which was expected. The total estimated premiums $\widehat{P}$ are similar, although BS ($\widehat{P} = 41.95$) yields a somewhat more accurate estimate of the true portfolio premium ($P = 44.89$) than CATLC ($\widehat{P} = 38.31$). Further, notice that the ordinary premium $\widehat{P}_{\text{ordinary}}$ obtained from CATLC (41.95) coincides with the total Bühlmann-Straub premium. As anticipated, the extraordinary premium $\widehat{\mu}_{\text{extra}}$ charged per risk is

Table 2. Selected quantities of interest in the estimation process, based on the Bühlmann -Straub (BS) and CATLC approaches for (approximately) Weibull-distributed claims *(W2)*, under *Scenario 1* and *3*.

Model	Scenario 1					
BS	Credibility Premiums (true ordinary)			Credibility Weights		
	$\widehat{\mu}_1 = 1.28\ (0.68)$	$\widehat{\mu}_6 = 3.66\ (4.29)$	$\widehat{\mu}_{11} = 1.24\ (1.72)$	$\widehat{\zeta}_1 = 0.61$	$\widehat{\zeta}_6 = 0.76$	$\widehat{\zeta}_{11} = 0.83$
	$\widehat{\mu}_{16} = 1.16\ (2.23)$	$\widehat{\mu}_{21} = 0.79\ (0.83)$	$\widehat{\mu}_{26} = 1.91\ (1.61)$	$\widehat{\zeta}_{16} = 0.66$	$\widehat{\zeta}_{21} = 0.80$	$\widehat{\zeta}_{26} = 0.85$
	$\widehat{\mu}_{31} = 0.77\ (0.39)$	$\widehat{\mu}_{36} = 1.11\ (1.12)$	$\widehat{\mu}_{41} = 0.94\ (0.89)$	$\widehat{\zeta}_{31} = 0.70$	$\widehat{\zeta}_{36} = 0.83$	$\widehat{\zeta}_{41} = 0.88$
	Structural Parameters					
	$\widehat{\mu} = 0.99,\quad \widehat{\sigma}_\varepsilon^2 = 2.63,\quad \widehat{\sigma}_\alpha^2 = 0.77,\quad \widehat{P} = 41.95\quad (P = 44.89)$					
	Performance Measures					
	$\mathrm{MSE}(\widehat{\boldsymbol{\mu}}) = 0.11,\quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.00,\quad \mathrm{C}(\mathbf{Y};\widehat{\boldsymbol{\mu}}) = 1.15$					
CATLC	Credibility Premiums (ordinary estimated)			Credibility Weights		
	$\widehat{\mu}_1 = 0.15\ (0.23)$	$\widehat{\mu}_6 = 4.01\ (4.09)$	$\widehat{\mu}_{11} = 1.19\ (1.27)$	$\widehat{\zeta}_1 = 0.68$	$\widehat{\zeta}_6 = 0.87$	$\widehat{\zeta}_{11} = 0.91$
	$\widehat{\mu}_{16} = 0.76\ (0.83)$	$\widehat{\mu}_{21} = 0.61\ (0.69)$	$\widehat{\mu}_{26} = 2.04\ (2.12)$	$\widehat{\zeta}_{16} = 0.79$	$\widehat{\zeta}_{21} = 0.88$	$\widehat{\zeta}_{26} = 0.93$
	$\widehat{\mu}_{31} = 0.84\ (0.92)$	$\widehat{\mu}_{36} = 1.33\ (1.41)$	$\widehat{\mu}_{41} = 0.63\ (0.71)$	$\widehat{\zeta}_{31} = 0.83$	$\widehat{\zeta}_{36} = 0.91$	$\widehat{\zeta}_{41} = 0.94$
	Structural Parameters					
	$\widehat{\mu}^{\mathrm{extra}} = -0.08,\quad \widehat{\lambda} = -0.80,\quad \widehat{\sigma}_\varepsilon^2 = 0.94,\quad \widehat{\sigma}_\alpha^2 = 1.00,\quad \widehat{P} = 38.31\quad (\widehat{P}_{\mathrm{ordinary}} = 41.95)$					
	Performance Measures					
	$\mathrm{MSE}(\widehat{\boldsymbol{\mu}}) = 0.18,\quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.00,\quad \mathrm{C}(\mathbf{Y};\widehat{\boldsymbol{\mu}}) = 0.99$					

Model	Scenario 3					
BS	Credibility Premiums (true ordinary)			Credibility Weights		
	$\widehat{\mu}_1 = 2.46\ (0.93)$	$\widehat{\mu}_6 = 2.00\ (0.88)$	$\widehat{\mu}_{11} = 1.79\ (0.99)$	$\widehat{\zeta}_1 = 0.38$	$\widehat{\zeta}_6 = 0.55$	$\widehat{\zeta}_{11} = 0.65$
	$\widehat{\mu}_{16} = 7.24\ (1.52)$	$\widehat{\mu}_{21} = 1.71\ (1.02)$	$\widehat{\mu}_{26} = 1.70\ (1.26)$	$\widehat{\zeta}_{16} = 0.43$	$\widehat{\zeta}_{21} = 0.60$	$\widehat{\zeta}_{26} = 0.69$
	$\widehat{\mu}_{31} = 2.33\ (1.42)$	$\widehat{\mu}_{36} = 1.75\ (1.12)$	$\widehat{\mu}_{41} = 3.14\ (1.99)$	$\widehat{\zeta}_{31} = 0.48$	$\widehat{\zeta}_{36} = 0.65$	$\widehat{\zeta}_{41} = 0.73$
	Structural Parameters					
	$\widehat{\mu} = 3.09,\quad \widehat{\sigma}_\varepsilon^2 = 522.43,\quad \widehat{\sigma}_\alpha^2 = 42.94,\quad \widehat{P} = 139.08\quad (P = 65.70 + 32.85 = 98.55)$					
	Performance Measures					
	$\mathrm{MSE}(\widehat{\boldsymbol{\mu}}) = 12.58,\quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.02,\quad \mathrm{C}(\mathbf{Y};\widehat{\boldsymbol{\mu}}) = 1.35$					
CATLC	Credibility Premiums (estimated ordinary)			Credibility Weights		
	$\widehat{\mu}_1 = 1.41\ (0.37)$	$\widehat{\mu}_6 = 1.81\ (0.78)$	$\widehat{\mu}_{11} = 2.22\ (1.18)$	$\widehat{\zeta}_1 = 0.70$	$\widehat{\zeta}_6 = 0.82$	$\widehat{\zeta}_{11} = 0.89$
	$\widehat{\mu}_{16} = 1.95\ (0.92)$	$\widehat{\mu}_{21} = 2.00\ (0.97)$	$\widehat{\mu}_{26} = 2.28\ (1.25)$	$\widehat{\zeta}_{16} = 0.70$	$\widehat{\zeta}_{21} = 0.87$	$\widehat{\zeta}_{26} = 0.91$
	$\widehat{\mu}_{31} = 2.23\ (1.19)$	$\widehat{\mu}_{36} = 2.25\ (1.21)$	$\widehat{\mu}_{41} = 3.99\ (2.96)$	$\widehat{\zeta}_{31} = 0.80$	$\widehat{\zeta}_{36} = 0.89$	$\widehat{\zeta}_{41} = 0.92$
	Structural Parameters					
	$\widehat{\mu}^{\mathrm{extra}} = 1.03,\quad \widehat{\lambda} = -0.35,\quad \widehat{\sigma}_\varepsilon^2 = 0.95,\quad \widehat{\sigma}_\alpha^2 = 0.97,\quad \widehat{P} = 105.88\quad (\widehat{P}_{\mathrm{ordinary}} = 59.53)$					
	Performance Measures					
	$\mathrm{MSE}(\widehat{\boldsymbol{\mu}}) = 2.08,\quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.01,\quad \mathrm{C}(\mathbf{Y};\widehat{\boldsymbol{\mu}}) = 1.03$					

small (-0.08). *Scenario 3* represents a 5% contamination of the portfolio, therefore it is not surprising that the robust regression credibility estimators perform significantly better than non-robust BS. For instance, the MSE of BS is now 6 times greater than that of CATLC. Also, observe that the total premium gained from CATLC ($\widehat{P} = 105.88$) is much closer to the theoretical $P = 98.55$ than the one obtained from BS ($\widehat{P} = 139.08$). While both approaches are portfolio-unbiased, now the standard model of BS provides clearly higher coverage of future claims than needed. The main reason for the difference in the accuracy of estimation lies in the estimation procedure of structural parameters σ_ε^2 and σ_α^2. While the robust estimators of the variance components remain stable in the presence of outliers, the non-robust estimator of the process variance σ_ε^2 and between risk variability σ_α^2 get highly inflated (i.e., σ_ε^2 increases from 2.63 to 522.43 and σ_α^2 from 0.77 to 42.94 using BS), distorts estimates of credibility weights, and consequently the whole procedure performs poorly (i.e., has high MSE and coverages). In sum, the BS model gives less credibility to individual experience, thus risk premiums are too much pulled toward the overall mean which has been attracted and distorted by few outlying observations. $\square$

DISCUSSION OF TABLE 3: For the Hachemeister's revisited (HR) model, we unexpectedly find that the CATLC yields optimal estimation results under both —clean and contaminated— scenarios. Indeed, under *Scenario 1* the MSE of HR is 46% larger than that of CATLC. This better performance of CATLC is also accompanied by a more accurate estimate of the total premium $\widehat{P} = 121.49$ (true portfolio premium $P = 132.44$) compared to $\widehat{P} = 111.41$ for the HR. When there is a 5% contamination in the data, the superiority of CATLC becomes even more evident. Now, the MSE of HR is about 469% larger than that of CATLC due to distorted estimates of structural parameters. The total premium collected by HR is 57%, and the coverage of future claims is 83% higher than the required one. Notice again that CATLC's total premiums ($\widehat{P} = 399.69 = 334.19 + 45 \times 1.46$) are close to the true premium ($P = 427.31$) while still providing sufficient funds for future losses. $\square$

Table 3. Selected quantities of interest in the estimation process, based on the Hachemeister revisited (HR) and CATLC approaches for (approximately) Lognormal-distributed Claims *(L2)*, under *Scenario 1* and *3*.

Model	Scenario 1					
HR	Credibility Premiums (true ordinary)			Credibility Weights		
	$\widehat{\mu}_1 = 2.07 \ (4.30)$	$\widehat{\mu}_6 = 0.68 \ (0.32)$	$\widehat{\mu}_{11} = 0.90 \ (0.61)$	$\widehat{\zeta}_1 = (0.65, 0.03)$	$\widehat{\zeta}_6 = (0.79, 0.03)$	$\widehat{\zeta}_{11} = (0.85, 0.03)$
	$\widehat{\mu}_{16} = 0.84 \ (1.70)$	$\widehat{\mu}_{21} = 2.10 \ (2.90)$	$\widehat{\mu}_{26} = 1.92 \ (1.82)$	$\widehat{\zeta}_{16} = (0.70, 0.02)$	$\widehat{\zeta}_{21} = (0.82, 0.03)$	$\widehat{\zeta}_{26} = (0.87, 0.04)$
	$\widehat{\mu}_{31} = 1.86 \ (5.70)$	$\widehat{\mu}_{36} = 3.21 \ (5.75)$	$\widehat{\mu}_{41} = 1.36 \ (1.00)$	$\widehat{\zeta}_{31} = (0.73, 0.01)$	$\widehat{\zeta}_{36} = (0.85, 0.02)$	$\widehat{\zeta}_{41} = (0.89, 0.05)$
	Structural Parameters					
	$\widehat{\boldsymbol{\mu}} = (2.13, 0.06), \quad \widehat{\sigma}^2_\varepsilon = 12.60, \quad \widehat{\boldsymbol{\sigma}}^2_{\boldsymbol{\alpha}} = (8.63, 0.01), \quad \widehat{P} = 111.41 \quad (P = 132.44)$					
	Performance Measures					
	$\text{MSE}(\widehat{\boldsymbol{\mu}}) = 2.12, \quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.00, \quad \text{C}(\mathbf{Y}; \widehat{\boldsymbol{\mu}}) = 0.98$					
CATLC	Credibility Premiums (ordinary estimated)			Credibility Weights		
	$\widehat{\mu}_1 = 2.73 \ (2.50)$	$\widehat{\mu}_6 = 0.64 \ (0.40)$	$\widehat{\mu}_{11} = 1.06 \ (0.82)$	$\widehat{\zeta}_1 = (0.89, 0.14)$	$\widehat{\zeta}_6 = (0.94, 0.25)$	$\widehat{\zeta}_{11} = (0.96, 0.33)$
	$\widehat{\mu}_{16} = 0.84 \ (0.61)$	$\widehat{\mu}_{21} = 1.99 \ (1.75)$	$\widehat{\mu}_{26} = 2.01 \ (1.77)$	$\widehat{\zeta}_{16} = (0.91, 0.25)$	$\widehat{\zeta}_{21} = (0.96, 0.40)$	$\widehat{\zeta}_{26} = (0.97, 0.49)$
	$\widehat{\mu}_{31} = 2.76 \ (2.52)$	$\widehat{\mu}_{36} = 6.23 \ (5.99)$	$\widehat{\mu}_{41} = 1.78 \ (1.54)$	$\widehat{\zeta}_{31} = (0.93, 0.36)$	$\widehat{\zeta}_{36} = (0.96, 0.53)$	$\widehat{\zeta}_{41} = (0.97, 0.63)$
	Structural Parameters					
	$\widehat{\mu}^{\text{extra}} = 0.24, \quad \widehat{\boldsymbol{\lambda}} = (-0.18, 0.04), \quad \widehat{\sigma}^2_\varepsilon = 0.83, \quad \widehat{\boldsymbol{\sigma}}^2_{\boldsymbol{\alpha}} = (1.38, 0.005), \quad \widehat{P} = 121.49 \quad (\widehat{P}_{\text{ordinary}} = 110.70)$					
	Performance Measures					
	$\text{MSE}(\widehat{\boldsymbol{\mu}}) = 1.45, \quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.00, \quad \text{C}(\mathbf{Y}; \widehat{\boldsymbol{\mu}}) = 1.06$					

Model	Scenario 3					
HR	Credibility Premiums (true ordinary)			Credibility Weights		
	$\widehat{\mu}_1 = 6.25 \ (1.81)$	$\widehat{\mu}_6 = 3.43 \ (1.29)$	$\widehat{\mu}_{11} = 2.06 \ (0.58)$	$\widehat{\zeta}_1 = (0.39, 0.36)$	$\widehat{\zeta}_6 = (0.57, 0.53)$	$\widehat{\zeta}_{11} = (0.67, 0.63)$
	$\widehat{\mu}_{16} = 56.21 \ (50.45)$	$\widehat{\mu}_{21} = 5.12 \ (1.79)$	$\widehat{\mu}_{26} = 8.30 \ (2.80)$	$\widehat{\zeta}_{16} = (0.45, 0.52)$	$\widehat{\zeta}_{21} = (0.62, 0.69)$	$\widehat{\zeta}_{26} = (0.71, 0.77)$
	$\widehat{\mu}_{31} = 5.31 \ (0.66)$	$\widehat{\mu}_{36} = 2.51 \ (1.31)$	$\widehat{\mu}_{41} = 13.10 \ (3.77)$	$\widehat{\zeta}_{31} = (0.50, 0.66)$	$\widehat{\zeta}_{36} = (0.67, 0.79)$	$\widehat{\zeta}_{41} = (0.75, 0.85)$
	Structural Parameters					
	$\widehat{\boldsymbol{\mu}} = (7.82, 1.45), \quad \widehat{\sigma}^2_\varepsilon = 1,060.30, \quad \widehat{\boldsymbol{\sigma}}^2_{\boldsymbol{\alpha}} = (288.05, 46.34), \quad \widehat{P} = 673.61 \quad (P = 284.87 + 142.43 = 427.31)$					
	Performance Measures					
	$\text{MSE}(\widehat{\boldsymbol{\mu}}) = 1,219.35, \quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.03, \quad \text{C}(\mathbf{Y}; \widehat{\boldsymbol{\mu}}) = 1.83$					
CATLC	Credibility Premiums (estimated ordinary)			Credibility Weights		
	$\widehat{\mu}_1 = 2.08 \ (0.63)$	$\widehat{\mu}_6 = 2.86 \ (1.40)$	$\widehat{\mu}_{11} = 2.16 \ (0.71)$	$\widehat{\zeta}_1 = (0.85, 0.16)$	$\widehat{\zeta}_6 = (0.93, 0.31)$	$\widehat{\zeta}_{11} = (0.95, 0.37)$
	$\widehat{\mu}_{16} = 6.26 \ (4.81)$	$\widehat{\mu}_{21} = 2.49 \ (1.04)$	$\widehat{\mu}_{26} = 4.22 \ (2.76)$	$\widehat{\zeta}_{16} = (0.86, 0.19)$	$\widehat{\zeta}_{21} = (0.93, 0.41)$	$\widehat{\zeta}_{26} = (0.95, 0.57)$
	$\widehat{\mu}_{31} = 3.47 \ (2.02)$	$\widehat{\mu}_{36} = 3.13 \ (1.68)$	$\widehat{\mu}_{41} = 5.06 \ (3.61)$	$\widehat{\zeta}_{31} = (0.90, 0.43)$	$\widehat{\zeta}_{36} = (0.95, 0.59)$	$\widehat{\zeta}_{41} = (0.96, 0.67)$
	Structural Parameters					
	$\widehat{\mu}^{\text{extra}} = 1.46, \quad \widehat{\boldsymbol{\lambda}} = (0.29, 0.04), \quad \widehat{\sigma}^2_\varepsilon = 1.03, \quad \widehat{\boldsymbol{\sigma}}^2_{\boldsymbol{\alpha}} = (1.45, 0.01), \quad \widehat{P} = 399.69 \quad (\widehat{P}_{\text{ordinary}} = 334.19)$					
	Performance Measures					
	$\text{MSE}(\widehat{\boldsymbol{\mu}}) = 214.39, \quad bias^*(\widehat{\boldsymbol{\mu}}) = 0.01, \quad \text{C}(\mathbf{Y}; \widehat{\boldsymbol{\mu}}) = 1.09$					

5.2.2 Long-run Performance

To establish a pattern showing that new credibility methods perform better than classical in the long-run, we generate 1,000 portfolios according to the above described specifications. Let us start with Example 1.

Table 4. Example 1: Performance of CATLC versus Bühlmann-Straub (BS) for (approximately) Weibull-distributed claims *(W1)* and *(W2)*, under *Scenario 1, 2, 3* and *4.*

		Measure							
		MSE $(\widehat{\boldsymbol{\mu}})$		E$(\widehat{\boldsymbol{\mu}})$	winning ratio	$bias^*(\widehat{\boldsymbol{\mu}})$		C$(\mathbf{Y};\widehat{\boldsymbol{\mu}})$		failed
Model	Scenario	BS	CATLC		BS : CATLC	BS	CATLC	BS	CATLC	BS
Weibull *(W1)*	1	0.40	0.66	0.64	916 : 84	0.00	0.00	1.04	1.06	0
	2	1.02	1.16	1.15	357 : 643	0.00	0.00	1.07	1.11	0
	3	10.5	5.86	2.63	385 : 615	0.01	0.01	1.16	1.21	0
	4	41.45	29.85	3.51	356 : 644	0.02	0.02	1.24	1.32	0
Weibull *(W2)*	1	0.51	1.04	0.57	920 : 80	0.00	0.00	1.04	1.05	0
	2	1.45	1.88	1.11	634 : 366	0.00	0.00	1.07	1.05	0
	3	11.80	7.78	3.44	409 : 591	0.01	0.01	1.17	1.19	0
	4	66.50	31.70	4.17	372 : 628	0.02	0.02	1.28	1.37	0

DISCUSSION OF TABLE 4: As expected, under *Scenario 1* and *(W1)* where $\sigma_\varepsilon^2 = 1$, standard credibility estimators are optimal most of the time (not necessarily always). In detail, BS outperforms CATL procedures 91.6% of the time with E$(\widehat{\boldsymbol{\mu}}) = 0.64$. Both methods are portfolio-unbiased and ensure coverage of future claims. Under contamination, however, robust procedures take advantage. For example, in *Scenario 4* CATLC "beats" BS 64.4% of the time with E$(\widehat{\boldsymbol{\mu}})$ increasing to 3.51. For *(W2)* these findings are just reinforced. In the last column, we report the number of failed simulations when performing the classical approach. The robust approach never failed. □

Next, for the other example, where Hachemeister's revisited model is fitted to log-*t*-distributed claims, the simulation study is summarized in a similar fashion. The summary is presented in Tables 5 and 6.

DISCUSSION OF TABLE 5: Across all scenarios the CATL procedure produces best results and never fails. Even in the clean scenario, Hachemeister (HR) is not a top performer anymore, yielding the leader's position to the robust CATLC. Indeed, in *Scenario 1* we observe the following scores

of "winning ratios" between HR and CATLC: 500 : 500 (lognormal), 89 : 911 (log-t_5), and 12 : 988 (log-t_3). Under contamination these ratios remain stable or even change slightly in favor of HR for the following reason. When claims are log-t distributed, then log-transformed claims are t-distributed, hence, symmetric. Thus, robustified REML methods that are tuned for symmetric t-distribution gain efficiency. And, for high levels of contamination by large outliers, the resulting approximate log-t-distribution becomes almost symmetric, which in turn favors standard regression credibility methods that are based on classical REML. Still, robust regression credibility estimators are more efficient and produce on average better predictors for future claims while keeping portfolio-unbiasedness. Findings for subject-specific process variances $\sigma^2_{\varepsilon_i}$ *(L2)* are presented in Table 6. □

Table 5. Example 2: Performance of CATLC versus Hachemeister revisited (HR) for (approximately) Log-t-distributed Claims with selected degrees of freedom, under *(L1)* and *Scenario 1, 2, 3* and *4*.

| | | Measure | | | | | | | | | |
| | | \multicolumn{2}{c}{MSE ($\widehat{\boldsymbol{\mu}}$)} | E($\widehat{\boldsymbol{\mu}}$) | winning ratio | \multicolumn{2}{c}{$bias^*(\widehat{\boldsymbol{\mu}})$} | \multicolumn{2}{c}{C($\mathbf{Y};\widehat{\boldsymbol{\mu}}$)} | failed |
Model	Scenario	HR	CATLC		HR : CATLC	HR	CATLC	HR	CATLC	HR
Lognormal	1	5.07	4.31	1.24	500 : 500	0.00	0.00	1.01	1.01	19
	2	12.08	9.53	2.04	389 : 611	0.00	0.00	1.04	0.99	38
	3	79.08	51.88	3.69	309 : 691	0.01	0.01	1.14	0.99	114
	4	267.69	281.98.	3.53	366 : 634	0.02	0.01	1.19	1.05	115
Log-t_5	1	38.95	12.05	3.83	89 : 911	0.00	0.00	1.14	0.96	113
	2	130.34	21.14	5.36	91 : 909	0.00	0.00	1.15	0.97	126
	3	679.43	85.46	12.99	190 : 810	0.01	0.00	1.23	0.96	147
Log-t_3	1	162.81	50.09	4.34	12 : 988	0.00	-0.01	1.31	0.99	268
	2	475.30	92.21	7.17	32 : 968	0.00	-0.01	1.32	0.96	264
	3	1,897.32	224.17	9.91	109 : 891	0.00	-0.02	1.35	0.94	290

Table 6. Example 2: Performance of CATLC versus Hachemeister revisited (HR) for (approximately) Log-t-distributed Claims with selected degrees of freedom, under *(L2)* and *Scenario 1, 2, 3* and *4*.

Model	Scenario	MSE ($\widehat{\boldsymbol{\mu}}$) HR	MSE ($\widehat{\boldsymbol{\mu}}$) CATLC	E($\widehat{\boldsymbol{\mu}}$)	winning ratio HR : CATLC	$bias^*(\widehat{\boldsymbol{\mu}})$ HR	$bias^*(\widehat{\boldsymbol{\mu}})$ CATLC	C($\mathbf{Y};\widehat{\boldsymbol{\mu}}$) HR	C($\mathbf{Y};\widehat{\boldsymbol{\mu}}$) CATLC	failed HR
Lognormal	1	6.24	7.59	1.16	625 : 375	0.00	0.00	1.01	1.02	69
	2	19.57	22.63	1.52	484 : 516	0.00	0.00	1.07	1.01	77
	3	104.46	92.85	2.66	421 : 579	0.01	0.01	1.13	1.02	86
	4	270.59	380.56	2.34	472 : 528	0.01	0.01	1.16	1.01	118
Log-t_5	1	407.56	184.23	3.56	202 : 798	0.00	0.01	1.26	0.95	249
	2	4,894.27	330.34	5.72	200 : 800	0.00	0.01	1.67	0.96	312
	3	4,963.46	655.61	7.34	331 : 669	0.01	0.01	1.56	0.99	331
Log-t_3	1	6,872.34	1,781.65	4.76	81 : 919	0.01	-0.02	2.43	0.97	411
	2	1.64×10^4	3,091.37	4.92	132 : 868	0.00	-0.01	2.31	0.91	431
	3	5.82×10^4	6,299.45	5.85	300 : 700	0.00	-0.01	1.99	0.90	427

6 Real-Data Example: Workers' Compensation Data

In this section, we analyze a standard example from workers' compensation insurance. This data set contains losses due to permanent partial disability and has already been studied by Klugman (1992), Frees *et al.* (2001) and Antonio *et al.* (2007).

6.1 Data Characteristics

As is common in non-life insurance, Frees *et al.* (2001) consider standard mixed linear models, where the response is the logarithmic transformation of the pure premium PP, defined as the *Loss* per *Payroll*. This transformation helps to remove the right-skewness in the original claim data (see Antonio *et al.*, 2007) and makes standard statistical software, which is developed to handle normally distributed residuals, easily applicable to long-tailed insurance data. Unfortunately, premiums obtained from this approach are difficult to interpret and do not yield credibility premiums for the individual risk classes. Thus, we present final credibility premiums for the original loss data.

The workers' compensation insurance data set comprises $I = 100$ occupation (risk) classes that have positive (non-zero) loss ratios over $\tau = 7$ years (Frees *et al.* 1999, 2001). The variable *Loss* is

the amount paid (on a yearly basis) and *Payroll* is a time-dependent exposure (volume) measure. The latter explanatory variable has a major impact on the within risk variability and is used as weights to account for heteroscedasticity. Note that the multiple time series plots of the response PP and LnPP, respectively, do not indicate any time-dependency (see Figure 1).

Figure 1. On the left, pure premiums (PP) plotted over $\tau = 7$ years. The line segments connect occupation classes. On the right, logarithmic pure premiums (LnPP) over $\tau = 7$ years.
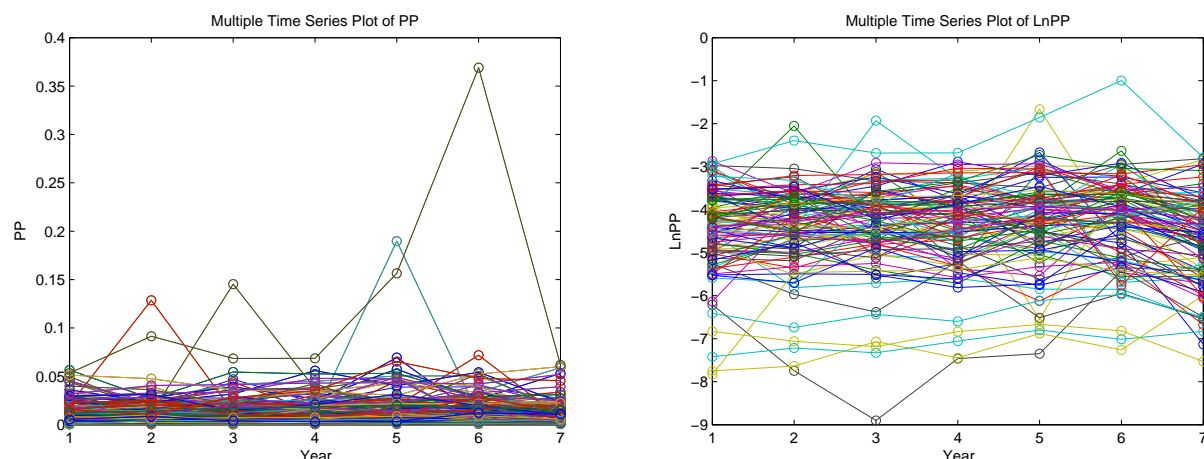
6.2 Model Fitting and Outlier Detection

Among the models with PP as dependent variable, Frees *et al.* (2007) found that the simple linear mixed model given by

$$\mathrm{PP}_{it} = \beta + \alpha_i + \varepsilon_{it}(\mathrm{Payroll}_{it})^{1/2}, \qquad i = 1, \ldots, I, \ \ t = 1, \ldots, \tau,$$

produces best fits. It shall be emphasized that this model, for which the weight *Payroll* captures heteroscedasticity among occupation classes, is the classical credibility model proposed by Bühlmann and Straub (1970). The structural parameters are obtained from REML with the error terms assumed to be serially uncorrelated, that is, $R_{it} = \sigma_\varepsilon^2 / \mathrm{Payroll}_{it}$. For the *corrected adaptively truncated likelihood credibility* (CATLC) model, we fit the log-linear model

$$\log(\mathrm{PP}_{it}) = \beta + \alpha_i + \varepsilon_{it}(\mathrm{Payroll}_{it})^{1/2}, \qquad i = 1, \ldots, I, \ \ t = 1, \ldots, \tau, \tag{6.1}$$

with location $\lambda_{it} = \beta + \alpha_i$ and normally distributed residuals $\varepsilon_{it}(\mathrm{Payroll}_{it})^{1/2} \sim N(0, \sigma_\varepsilon^2)$.

Figure 2. Logarithmic pure premiums (PP) plotted over $\tau = 7$ years based on the *cleaned* data set. The line segments connect occupation classes $\tau = 7$ years.
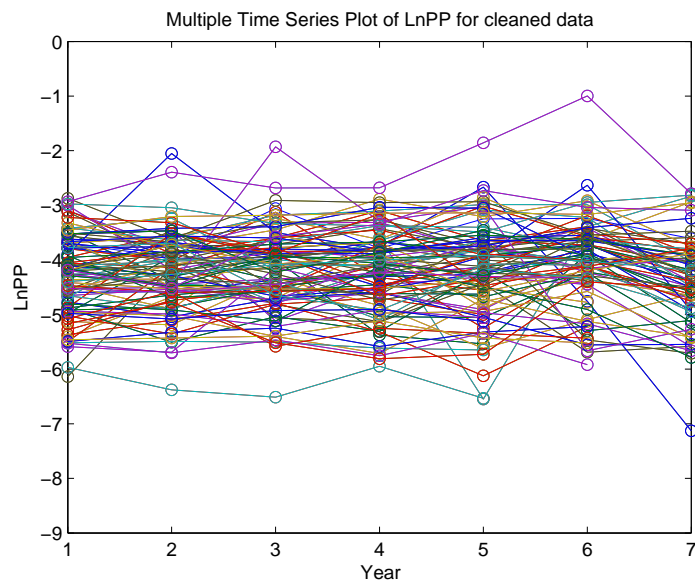
When the standard credibility model of Bühlmann-Straub (1970) is fitted with the REML procedure, the estimated portfolio mean is $\widehat{\beta}_{\mathrm{GLS}} = 0.016$. The robust CATL procedure detects 51 within-risk outliers while 94 out of 100 occupation classes are treated as ordinary. Fitting model (6.1), we find that the robust grand location $\widehat{\beta}_{\mathrm{rGLS}} = -4.261$ is slightly higher than the non-robust fit of -4.46 in Frees *et al*. (2001). This is mainly due to the removal of risks that report comparatively low claim sizes to the insurer over the entire observation period. What is important to understand here is that small claims may threaten, in the long run, the solvency of the insurance company. Indeed, if there is too many (or too few) of them, that can distort the estimation of variance components which in turn may yield too low total classical credibility premiums collected from the insured. And this threatens the solvency of the insurer in the long run. Further, assuming that the expected within-risk variability is $R_{\tau_i+1} \approx I^{-1} \sum_{i=1}^{I} \sum_{t=1}^{\tau_i} \tau_i^{-1} \left(\widehat{\sigma}_\varepsilon^2 / \mathrm{Payroll}_{it} \right) = 0.1906$, for each risk, we obtain the total ordinary credibility premium of 1.862 and the total extraordinary premium of -0.110. Hence, the resulting robust portfolio premium collected from all insureds is 1.752, whereas the corresponding standard credibility premium obtained from the Bühlmann-Straub model is 1.609. We notice that the aver-

age total portfolio premium is 1.848. This indicates that CATLC produces more accurate and less portfolio-biased credibility premiums which is of main interest to the insurer.

7 Summary

In this article, we have proposed a three-step procedure for robust-efficient fitting of regression credibility models when claims are heavy-tailed log-location-scale distributed. The designed procedure, which we call corrected adaptively truncated likelihood credibility (CATLC), provides high robustness against outliers occurring both within and between risks. Through adaptive detection rules, excess claims are automatically identified and rejected. Then, classical but corrected likelihood methods are employed on cleaned insurance data to find ordinary credibility premiums. Estimation techniques from robust regression are used to price excess claims and, finally, to calculate robust and portfolio-unbiased credibility premiums. Practical performance of the newly designed class of robust regression credibility has been investigated under several simulated scenarios and in a real-data example from workers' compensation insurance. We have also performed additional case studies using CATLC for pricing of risks in insurance and finance and arrived at similar conclusions. These studies will be presented in a parallel paper.

In summary, the CATLC procedure:

- automatically finds the most robust and efficient estimator available and requires no expert judgment for the choice of truncation points;

- provides accurate estimates of true premiums while ensuring portfolio unbiasedness;

- provides protection toward outliers that influence the within and/or the between risk variability, which in turn have distorting effects on credibility weights;

- provides credibility premiums for reported claims data which are easy to interpret;

- can be applied to any credibility model that can be expressed as mixed linear model where log-transformed claims are location-scale distributed.

A natural continuation in this line of research would be to develop CATLC procedures for credibility models with serially correlated error terms and various dependency structures between risks. We did not pursue such generalizations here because of the mathematical/technical challenges they present. For *non-robust* estimation in these more general models, however, some interesting proposals are already available in the actuarial literature (see Cossette and Luong, 2003, and Pitselis, 2004a,b).

References

[1] Antonio, K., and Beirlant, J., 2007. Actuarial Statistics with Generalized Linear Mixed Models. *Insurance: Mathematics and Economics*, **40**(1), 58–76.

[2] Bianco, A. M., Ben, M. G., and Yohai, V. J., 2005. Robust estimation for linear regression with asymmetric errors. *The Canadian Journal of Statistics*, **33**(4), 511–528.

[3] Bühlmann, H., 1967. Experience Rating and Credibility. *ASTIN Bulletin*, **4**, 199–207.

[4] Bühlmann, H., and Gisler, A., 1997. Credibility in the Regression Case Revisited. *ASTIN Bulletin*, **27**, 83–98.

[5] Bühlmann, H., and Gisler, A., 2005. A Course in Credibility Theory and its Applications. New York: Springer.

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

[7] Caroll, R. J., and Ruppert, D., 1988. Transformation and Weighting in Regression. New York: Chapman & Hall Ltd.

[8] Cossette, H., and Luong, A., 2003. Generalized Least Squares Estimators for Covariance Parameters for Credibility Regression Models with Moving Average Errors. *Insurance: Mathematics and Economics*, **32**, 281–293.

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

[10] De Vylder, F., 1976. Optimal Semilinear Credibility. *Bulletin of Swiss Association of Actuaries*, 25–33.

[11] De Vylder, F., 1985. Nonlinear Regression in Credibility Theory. *Insurance: Mathematics and Economics*, **4**, 163–172.

[12] 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.

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

[14] Frees, E. W., and Young, V. R., 1999. A Longitudinal Data Analysis Interpretation of Credibility Models. *Insurance: Mathematics and Economics*, **24**, 229–247.

[15] Frees, E. W., Young, V. R., and Luo, Y., 2001. Case Studies Using Panel Data Models. *North American Actuarial Journal*, **5**(4), 24–42.

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

[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] Gisler, A., 1980. Optimum Trimming of Data in the Credibility Model. *Bulletin of Swiss Association of Actuaries*, 313–325.

[19] Gisler, A., and Reinhard, P., 1993. Robust Credibility. *ASTIN Bulletin*, **23**(1), 117–143.

[20] Goovaerts, A. S., and Hoogstad, W., 1987. Credibility Theory, Surveys of Actuarial Studies. Rotterdam: National-Nederlanden N.V..

[21] Goovaerts, A. S., Kaas, R., van Heerwaarden, A. E., and Bauwelinckx, T., 1990. Effective Actuarial Methods. Insurance Series **3**, Amsterdam: North-Holland.

[22] Goulet, V., 1998. Principles and Application of Credibility Theory. *Journal of Actuarial Practice*, **6**(1-2), 5–62.

[23] Haberman, S., and Renshaw, A. E., 1996. Generalized Linear Models and Actuarial Science. *The Statistician: Journal of the Institute of Statisticians*, **45**, 407–436.

[24] Hachemeister, C. A., 1975. Credibility for Regression Models with Applications to Trend. In *Credibility: Theory and Applications*, ed. P. M. Kahn, New York: Academic Press, pp. 129–163.

[25] Huber, P. J., 1981. Robust Statistics. New York: Wiley.

[26] Huggins, R. M., and Staudte, R. G., 1994. Variance Components Models for Cell Dependent Populations. *Journal of the American Statistical Association*, **89**, 19–29.

[27] Klugman, S., 1992. Bayesian Statistics in Actuarial Science with Emphasis on Credibility. Boston: Kluwer.

[28] Klugman, S., Panjer, H., and Willmot, G., 2004. Loss Models: From Data to Decision. 2nd edition, New York: Wiley.

[29] Kremer, E., 1991. Large Claims in Credibility. *Blätter Deutsche Gesellschaft für Versicherungswirtschaft*, **XX**, 123–150.

[30] Künsch, H. R., 1992. Robust Methods for Credibility. *ASTIN Bulletin*, **22**(1), 33–49.

[31] Manning, W. G., Basu, A., and Mullahy, J., 2005. Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data. *Journal of Health Economics*, **24**, 465–488.

[32] Marazzi, A., and Yohai, V. J., 2004. Adaptively truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference*, **122**, 271–291.

[33] McCullagh, P., and J.A. Nelder, 1989. Generalized Linear Models. 2nd edition, Chapman and Hall: London.

[34] Nelder, J. A., and Verall, R.J., 1997. Credibility Theory and Generalized Linear Models. *ASTIN Bulletin*, **27**, 71–82.

[35] Norberg, R., 1980. Empirical Bayes Credibility. *Scandinavian Actuarial Journal*, 177–194.

[36] Norberg, R., 1986. Hierarchical credibility: Analysis of a Random Effect Linear Model with Nested Classification. *Scandinavian Actuarial Journal*, 204–222.

[37] Rocke, D. M., 1983. Robust Statistical Analysis of Interlaboratory Studies. *Biometrika*, 70, 421–431.

[38] Pitselis, G., 2004a. A Seemingly Unrelated Regression Model in a Credibility Framework. *Insurance: Mathematics and Economics*, **34**, 37–54.

[39] Pitselis, G., 2004b. Credibility Models with Cross-Section Effect and with both Cross-Section and Time Effects. *Blätter der DGVFM*, **26**(4), 643–663.

[40] Pitselis, G., 2008. Robust Regression Credibility: The Influence Function Approach. *Insurance: Mathematics and Economics*, **42**, 288–300.

[41] Rosenberg, M. A., Frees, E. W., Sun, J., Johnson, P. H., and Robinson, J. M., 2007. Predictive Modeling with Longitudinal Data: A Case Study with Wisconsin Nursing Homes. *North American Actuarial Journal*, **11**(3), 54–69.

[42] Searle, S. R., Casella, G., and McCulloch, C. E., 1992. Variance Components. New York: Wiley.

[43] Sun, J., Frees, E. W., and Rosenberg, M. A., 2008. Heavy-Tailed Longitudinal Data Modeling Using Copulas. *Insurance: Mathematics and Economics*, **42**(2), 817–830.

[44] Sundt, B., 1979. A Hierarchical Regression Credibility Model. *Scandinavian Actuarial Journal*, 107–114.

[45] Sundt, B., 1980. A Multi-level Hierarchical Credibility Regression Model. *Scandinavian Actuarial Journal*, 25–32.

[46] Welsh, A. H., and Richardson, A. M., 1997. Approaches to the Robust Estimation of Mixed Models. In *Handbook of Statistics*, Vol. 15, eds. G. Maddala and C. R. Rao, Amsterdam: Elsevier Science, pp. 343–384.