Case Studies Using Credibility and Corrected Adaptively Truncated Likelihood Methods

by Harald Dornheim and Vytaras Brazauskas

ABSTRACT

Two recent papers by Dornheim and Brazauskas (2011a, 2011b) introduced a new likelihood-based approach for robust-efficient fitting of mixed linear models and showed that it possesses favorable large- and small-sample properties which yield more accurate premiums when extreme outcomes are present in the data. In particular, they studied regression-type credibility models that can be embedded within the framework of mixed linear models for which heavy-tailed insurance data are approximately log-location-scale distributed. The new methods were called *corrected adaptively truncated likelihood* methods (or CATL, for short). In this paper, we build upon that work and further explore how CATL methods can be used for pricing risks. We extend the area of application of standard credibility ratemaking to several well-studied examples from property and casualty insurance, health care, and real estate fields. The process of outlier identification, the ensuing model inference, and related issues are thoroughly investigated on the featured data sets. Throughout the case studies, performance of CATL methods is compared to that of other robust regression credibility procedures.

KEYWORDS

Adaptive robust-efficient estimation, mixed linear model, outlier detection, prediction, regression credibility ratemaking

1. Introduction

Credibility theory is one of the oldest but still most common premium ratemaking techniques in insurance industry, and it continues to attract the attention of practicing actuaries and academic researchers. Since the publication of the first regression-linked credibility model (Hachemeister, 1975), numerous extensions, improvements, and practical applications of this model have been proposed in the actuarial literature. Here is a short but representative list of recent papers on this topic:

- Frees, Young, and Luo (1999) gave a longitudinal data analysis interpretation of the standard (Bühlmann 1967; Bühlmann and Straub 1970) and other additive credibility ratemaking procedures. This interpretation also remains valid in the framework of mixed linear models. A few years later, the same authors presented several case studies involving such models (Frees, Young, and Luo 2001).
- Fellingham, Tolley, and Herzog (2005) analyzed a real data set provided by a major health insurance provider, which summarized claims experience for select health insurance coverages in Illinois and Wisconsin. Credibility methods, in conjunction with mixed linear models and Bayesian hierarchical models, were used to make next-year predictions of claims costs.
- Guszcza (2008) provided an excellent introduction to hierarchical models that can be viewed as an extension of traditional (linear) credibility models. Several hierarchical models were then used in a loss reserving exercise, to model loss development across multiple accident years.
- Klinker (2011) introduced generalized linear mixed models (GLMM) as a way of incorporating credibility in a generalized linear model setting. The application of a GLMM to a case study on ISO data revealed connections between this approach and the Bühlmann-Straub credibility.

As is the case with many mathematical models, credibility-type models contain unknown *structural parameters* (or, in the language of mixed linear models, *fixed effects* and *variance components*) that have to be estimated from the data. For statistical inference about fixed effects and variance components, traditional likelihood-based methods such as (restricted) maximum likelihood estimators, (RE)ML, are commonly pursued. However, it is also known that while these methods offer most flexibility and full efficiency at the assumed model, they are extremely sensitive to small deviations from the hypothesized normality of random components as well as to the occurrence of outliers. To obtain more reliable estimators for premium calculation and prediction of future claims, various robust methods have been successfully adapted to credibility theory in the actuarial literature (see, for example, Garrido and Pitselis 2000; Dornheim and Brazauskas 2007; Pitselis 2002, 2008).

At this point a short discussion about robust statistics—a well-established field in statistics—might aid the reader who is not familiar with this area. In a nutshell, robust statistics is concerned with *model mis-specification* (or using the actuarial terminology, *model risk*) and *data quality* (e.g., measurement errors, outliers, typos). The main focus is on parametric models, their fitting to the observed data, and identification of outliers. Fitting of the model, however, is accomplished by employing procedures that are designed to have limited sensitivity to changes in the underlying assumptions as well as to "unexpected" data points. The procedures that possess such properties are called *robust.* A very important tool for measuring performance of, and for providing guidance on how to construct robust estimators is the *influence function* (IF). The IF helps to quantify the estimator's robustness and efficiency, which usually are two competing criteria. Typically, robust and efficient estimators belong to one of three general classes of statistics: *L*-, *M*-, or *R*-statistics. (Here, *L* stands for *linear* in the "linear combinations of order statistics"; *M* stands for *maximum* in the "maximum likelihood type statistics"; *R* stands for *ranks* in the "statistics based on ranks".) It is not uncommon, however, to have estimators that can be reformulated within more than one of these classes. On the other hand, despite the existing overlap, each type of these statistics has its own appeal.

M-statistics, for example, are arguably the most amenable to generalization and often lead to theoretically optimal procedures. *R*-statistics can be recast in the context of hypothesis testing and enjoy a close relationship with the broad field of nonparametric statistics. *L*-statistics are fairly simple computationally and have a straightforward interpretation in terms of quantiles.

Further, in the initial stages of its development, research in the field of robust statistics was focused on quite simple parametric problems (e.g., estimation of the location parameter of a bell-shaped curve), but over the years it expanded into linear models, time series analysis and other more complex problems. Now, almost five decades later, the impact of robust statistics is felt in numerous applied and interdisciplinary areas, ranging from natural and social sciences to computer science, engineering, insurance, and finance. Robust procedures add the most value to the modeling process when the underlying model contains many unknown parameters and/or multiple assumptions, because the more complex the model, the higher the chance to misspecify it. Finally, robust statistics is parametric by its nature and thus differs from empirical nonparametric methods which have no (or very weak) parametric assumptions. The main advantages the parametric methods enjoy over the nonparametric ones are: (a) parametric models are easier to interpret, (b) they are parsimonious, i.e., have few parameters, and (c) they facilitate inference beyond the range of observed data. On the other hand, the attentive reader will recognize that some of these advantages are also risks (e.g., extrapolation beyond the data). Therefore, in this context, robust techniques are indispensable, since they focus on model risk management. For a comprehensive (and much deeper) introduction into the area of robust statistics, the reader should consult Maronna, Martin, and Yohai (2006). Note also that in the academic literature and actuarial practice there exist various methods that are labeled "robust" or "robust-efficient" but fall outside the scope of the present discussion. For example, quantile regression also has the robustness-efficiency quality, but the estimated variable/parameter is generally not the object of actuarial interest. Another popular approach uses a large loss adjustment before the model fitting, which is supposed to "robustify" the estimation process. This approach, however, has no theoretical justification and thus its statistical properties are unknown.

Let us get back to the robust methods at the intersection of credibility and mixed linear models. In this area, the most recent proposal—*corrected adaptively truncated likelihood* methods (CATL)—is designed for situations when heavy-tailed claims are approximately log-location-scale distributed (Dornheim and Brazauskas 2011a, 2011b). This new class of robust-efficient credibility estimators enjoys a number of desirable features.

- First, it provides full protection against within-risk outliers and observations that may have disturbing effects on the estimation process of the between-risk variability. (Note that the latter property cannot be guaranteed when one applies standard versions of *M*-estimators that solely robustify the estimation of the individual's claim experience.)
- Second, the estimators do not require expert judgment to find appropriate truncation points which are obtained adaptively from the data.
- Third, the CATL procedure automatically identifies and removes atypical data points without employing extensive graphical tools or including data-specific predictor variables into the model. This makes the modeling process easier and quicker. (Note that the emphasis on the automatic nature of CATL methods should not be viewed as the authors' endorsement of the fact that the decision-making process can now be delegated to computers. Understanding of the practical problem, its context, available data, and the economic consequences of the model-based decisions is the sole province of human intelligence . . . at least in the foreseeable future.)

In this paper, we further explore how the CATL approach can be used for pricing risks. We extend the area of application of standard credibility ratemaking to several well-studied examples from property and casualty insurance, health care, and real estate fields. We do not merely determine prices via CATL but rather walk the reader through the entire process of outlier identification and the associated statistical inference using examples. The featured data sets include exposure measures, heteroscedasticity, random and fixed effect covariates and outliers. Throughout the case studies performance of CATL methods is compared to that of other robust regression credibility procedures.

The rest of the paper is organized as follows. In Section 2, we present the CATL procedure for fitting heavy-tailed mixed linear models and the resulting formulas for robust-efficient credibility premiums. In Section 3, we analyze Hachemeister's bodily injury data using CATL and compare the findings with those of classical robust methods. For illustrative purposes, we fit a log-normal regression model (although other log-location-scale models could be considered as well) and show how exposure measures are incorporated to account for heteroscedasticity in the data. The impact of a single within-risk outlier on the computation of credibility estimates of future claims as well as their prediction intervals is discussed. The case study of Section 4 is related to health care data that contains Medicare costs for inpatient hospital charges classified by state. This data set is also contaminated by outlying risks and thus offers further insights into how robust procedures act on data. Moreover, we fit a more generic regression-type model that requires additional explanatory variables. Robust credibility estimates and prediction intervals for future Medicare costs are computed and compared to those of classical methods. In Section 5, we venture outside the actuary's comfort zone and demonstrate the usefulness of CATL procedures in the field of real estate. We analyze a widely studied data set of Green and Malpezzi (2003). Summary is provided in Section 6.

2. CATL credibility for heavy-tailed claims

In Section 2.1, we outline the corrected adaptively truncated likelihood procedure for fitting mixed linear models with heavy-tailed error components. Robust credibility ratemaking is briefly discussed in Section 2.2. More technical details on mixed linear models for heavy-tailed claims is presented in Appendix (Sections 7.2 and 7.3). For a complete description of these methods, see Dornheim (2009).

2.1. The CATL procedure

For robust-efficient fitting of the mixed linear model with normal random components, Dornheim (2009) and Dornheim and Brazauskas (2011a) developed adaptively truncated likelihood methods. Those methods were further generalized to log-location-scale models with symmetric or asymmetric errors and labeled *corrected adaptively truncated likelihood* methods, CATL (Dornheim 2009; Dornheim and Brazauskas 2011b). More specifically, utilizing the notation of Sections 7.2 and 7.3, the CATL estimators for location λ_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:

1. *Detection of within-risk outliers.*

   Consider the random sample

$$(\mathbf{x}_{i1}, \mathbf{z}_{i1}, log(y_{i1}), \upsilon_{i1}), \ldots, (\mathbf{x}_{i\tau_i}, \mathbf{z}_{i\tau_i}, log(y_{i\tau_i}), \upsilon_{i\tau_i}),\quad i = 1, \ldots, I.$$

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

$$(\mathbf{x}^*_{i1}, \mathbf{z}^*_{i1}, log(y^*_{i1}), \upsilon^*_{i1}), \ldots, (\mathbf{x}^*_{i\tau^*_i}, \mathbf{z}^*_{i\tau^*_i}, log(y^*_{i\tau^*_i}), \upsilon^*_{i\tau^*_i}),\quad i = 1, \ldots, I.$$

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

2. *Detection of between-risk outliers.*

   In the second step, the procedure searches the pre-cleaned sample (marked with *) and discards entire risks whose risk-specific profile expressed by the random effect significantly deviates from the overall portfolio profile. These risks are identified when their robustified Mahalanobis distance (this is a Wald-type statistic with robustly estimated input variables) exceeds some adaptive cutoff point. The process results in a "cleaned" sample of risks, denoted as

$$(\mathbf{x}_{i1}^{**}, \mathbf{z}_{i1}^{**}, log(y_{i1}^{**}), \upsilon_{i1}^{**}), \ldots, (\mathbf{x}_{i\tau_i^*}^{**}, \mathbf{z}_{i\tau_i^*}^{**}, log(y_{i\tau_i^*}^{**}), \upsilon_{i\tau_i^*}^{**}),\\ i = 1, \ldots, I^*.$$

Note that the number of remaining risks is I^* ($I^* \leq I$).

3. *CATL estimators.*

   In the final step, the CATL procedure employs traditional likelihood-based methods, such as (restricted) maximum likelihood, on the cleaned sample and computes re-weighted parameter estimates $\hat{\beta}_{CATL}$ and $\hat{\theta}_{CATL} = (\hat{\sigma}^2_{\alpha_1}, \ldots, \hat{\sigma}^2_{\alpha_q}, \hat{\sigma}^2_{\varepsilon})$. Here, the subscript CATL emphasizes that the maximum likelihood type estimators are not computed on the original sample, i.e., the starting point of Step 1, but rather on the cleaned sample which is the end result of Step 2.

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

$$\hat{\lambda}_i = \mathbf{X}_i^{**}\hat{\beta}_{CATL} + \mathbf{Z}_i^{**}\hat{\alpha}_{rBLUP,i} + \hat{E}_{F_0}(\varepsilon_i), \quad i = 1, \ldots, I,$$

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

Remark 1 *(Within-risk outliers)*

To illustrate what the above-described procedure does in Step 1, let's take a look at the first graph of Section 3, where a multiple time series plot of Hachemeister's bodily injury data is given. Notice that in State 4, for example, the claim at time 7 deviates from the average trend and thus would be treated as a *within-risk* outlier. Procedures that do not remove such claims during the model-fitting process produce an inflated estimate of within-risk variability for State 4 (i.e., $\hat{\sigma}^2_{\varepsilon}$). □

Remark 2 *(Between-risk outliers)*

To better understand Step 2, let's examine the first graph of Section 4, where hospital claims over a period of six years are plotted for 54 states/regions across the United States. In the plot, we clearly see that *all* claims within each of the states 40, 48, and 54 are apart from the bulk of other states. They would be treated as *between-risk* outliers because each of these states affects the level of variability (heterogeneity) among the states. Procedures that treat these risks the same way as other risks produce significantly larger estimates of between-risk variability of random effects (i.e., $\hat{\sigma}^2_{\alpha}$). □

Remark 3 *(Mean correction factors)*

In view of the mixed linear model as described in Section 7.3, residuals that follow asymmetric log-location-scale distributions have no longer mean zero. Thus, the expectations $E(\log(\mathbf{y}_i))$ and $E(\log(\mathbf{y}_i|\alpha_i))$ differ from $\mathbf{X}_i\beta$ and λ_i, respectively. Therefore, to ensure that the estimators are targeting the right variables when estimating λ_i, we need to correct the equation towards the mean by adding $\hat{E}_{F_0}(\varepsilon_i)$. This also explains the word "corrected" in the acronym CATL. □

2.2. Robust credibility ratemaking

The re-weighted estimates for location, $\hat{\lambda}_i$, and structural parameters, $\hat{\theta}_{CATL} = (\hat{\sigma}^2_{\alpha_1}, \ldots, \hat{\sigma}^2_{\alpha_q}, \hat{\sigma}^2_{\varepsilon})$, are used to calculate robust credibility premiums for the ordinary but heavy-tailed claims part of the original data. The robust ordinary net premiums

$$\hat{\mu}_{it}^{ordinary} = \hat{\mu}_{it}^{ordinary}(\hat{\alpha}_{rBLUP,i}), \quad t = 1, \ldots, \tau_i + 1, i = 1, \ldots, I$$

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

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

$$\hat{O}_{it} = \begin{cases} -\hat{\mu}_{it}^{ordinary}, & for\ y_{it} < q_l. \\ (y_{it} - q_l) - \hat{\mu}_{it}^{ordinary}, & for\ q_l \le y_{it} < q_g. \\ (q_g - q_l) - \hat{\mu}_{it}^{ordinary}, & for\ y_{it} \ge q_g. \end{cases}$$

Further, let I_t denote the number of insureds in the portfolio at time t and let $T = \max_{1 \le i \le I} \tau_i$, the maximum horizon among all risks. For each period $t = 1, \ldots, T$, we find the mean cross-sectional overshot of excess claims $\hat{O}_{\cdot t} = I_t^{-1} \Sigma_{i=1}^{I_t} \hat{O}_{it}$, and fit robustly the random effects model

$$\hat{O}_{\cdot t} = \mathbf{o}_t \xi + \tilde{\varepsilon}_t, \quad t = 1, \ldots, T,$$

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

$$\mu_{it}^{extra} = \mathbf{o}_t \hat{\xi}.$$

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

$$\hat{\mu}_{i,\tau_i+1}^{CATL}(\hat{\alpha}_{rBLUP,i}) = \hat{\mu}_{i,\tau_i+1}^{ordinary}(\hat{\alpha}_{rBLUP,i}) + \hat{\mu}_{i,\tau_i+1}^{extra}, \quad i = 1, \ldots, I.$$

From the actuarial point of view, premiums assigned to the insured have to be positive. Therefore, we determine pure premiums by $max\{0, \hat{\mu}_{i,\tau_i+1}^{CATL}(\hat{\alpha}_{rBLUP,i})\}$.

3. Hachemeister's bodily injury data

The first credibility model linked to regression was proposed by Hachemeister in 1975. In the decades since the publication of his paper, the private passenger automobile data that he studied has been extensively analyzed by several authors in the actuarial literature. Goovaerts and Hoogstad (1987), Dannenburg, Kaas, and Goovaerts. (1996), Bühlmann and Gisler (1997), Frees, Young, and Luo (1999), Garrido and Pitselis (2000), and Pitselis (2002, 2008) used this data set to illustrate the effectiveness of diverse regression credibility ratemaking techniques.

3.1. Data characteristics

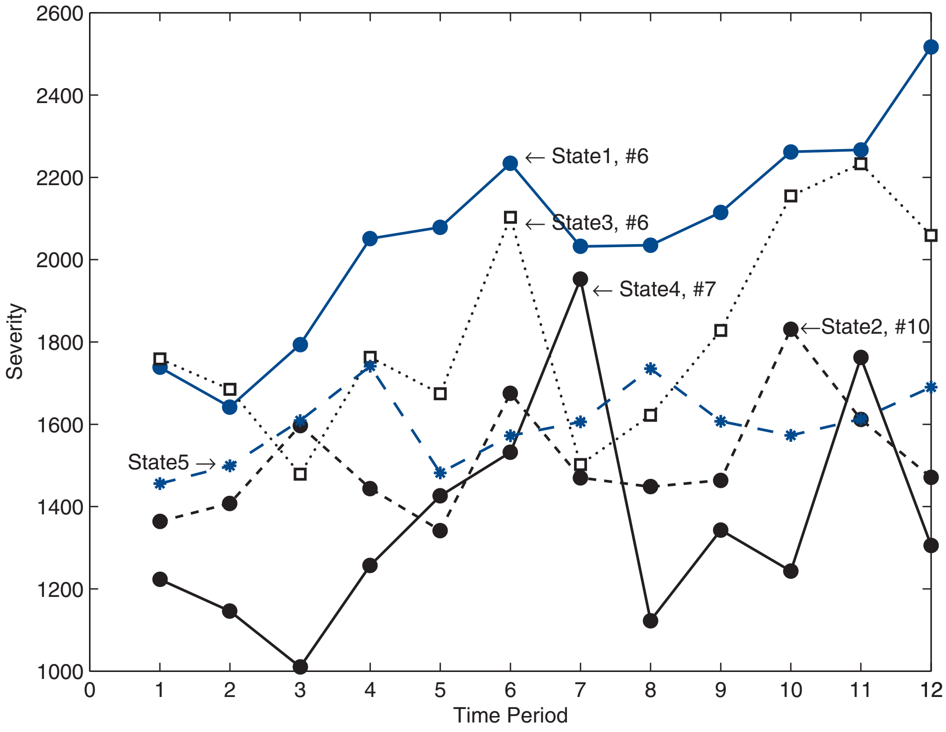
Hachemeister (1975) considered $\tau = 12$ periods, from the third quarter of 1970 to the second quarter of 1973, of claim data for bodily injury that are covered by a private passenger auto insurance. The response variable of interest to the actuary is the severity *average loss per claim,* denoted by y_{it}. It is followed over the periods $t = 1, \ldots, \tau_i = \tau$ for each state $i = 1, \ldots, I$. Average losses were reported from $I = 5$ different states (Appendix, Section 7.1).

A multiple time series plot of the observed variable average loss per claim, y_{it}, is provided in Figure 1. The plot indicates that states differ with respect to their within-state variability and severity. State 1 reports the highest average losses per claim, whereas State 4 seems to have larger variability compared to other states. For all five states we observe a small increase of severity over time. Since y_{it} varies over states and time (periods $t = 1, \ldots, 12$), these characteristics are important explanatory variables. Therefore, Hachemeister originally suggested using the linear trend model given by

$$y_{it} = \mathbf{x}_{it}\beta + \mathbf{z}_{it}\alpha_i + \varepsilon_{it}, \tag{1}$$

where $p = q = 2$ in (5) and $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t)$. This results in a random coefficients model of the form $\mathbf{y}_i = \mathbf{X}_i(\beta + \alpha_i) + \varepsilon_i$, with diagonal matrix $\mathbf{R}_i = \mathbf{Var}(y_i|\alpha_i) = Var(\varepsilon_i) = \sigma_\varepsilon^2\, diag(\upsilon_{i1}^{-1}, \ldots, \upsilon_{i\tau}^{-1})$, where $\upsilon_{it} > 0$ are some known (potential) volume measures. By assumption (b) in Section 7.2, we consider independent (unobservable) risk factors that have variance-covariance structure $D = diag(\sigma_{\alpha_1}^2, \sigma_{\alpha_2}^2)$. Notice that the quarterly observations #6 in State 1, #10 in State 2, #7 in State 4, and maybe #6 in State 3 seem to be apart from their state-specific inflation trends (see Figure 1).

Figure 1. Multiple time series plot of the variable average loss per claim, y_{it}



We use Hachemeister's regression model (1) as our reference model to facilitate comparison of CATL credibility with other estimators discussed by the authors mentioned above. We also consider a revised version of Hachemeister's model. When applying the linear trend model to bodily injury data, Hachemeister (1975) obtained unsatisfying model fits due to systematic underestimation of the regression line. Bühlmann and Gisler (1997) suggest taking the intercept of the regression line centered at the mean time (instead of the origin of the time axis). This ensures that the regression line stays between the individual and collective regression lines. Accordingly, we choose design matrices $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau})'$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{i\tau})'$, with $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t - G_i)'$, where $G_i = \upsilon_{i\cdot}^{-1} \sum_{t=1}^{\tau} t\,\upsilon_{it}$, is the center of gravity of the time range in risk i, and $\upsilon_{i\cdot} = \sum_{t=1}^{\tau} \upsilon_{it}$.

3.2. Model estimation, outlier detection and model inference

For Hachemeister's regression credibility model and its revised version (R), we use $ln(y_{it})$ as response variable in the framework of adaptively truncated likelihood credibility. Then, we fit (log-normal) models of the form

$$ln(y_{it}) = \mathbf{x}_{it}\beta + \mathbf{z}_{it}\alpha_i + \varepsilon_{it} = \lambda_{it} + \varepsilon_{it}, \qquad (2)$$

where $\varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon_{it}})$. First of all, we assume that error terms are serially uncorrelated and process variances are equal for all states, so that $\sigma^2_{\varepsilon_i} = \sigma^2_{\varepsilon}$. This yields the unweighted covariance matrix $\mathbf{R}_i = \mathbf{R} = \sigma^2_{\varepsilon}\,\mathbf{I}_{\tau\times\tau}$, $i = 1, \ldots, I,$ as special case of (1). Results of the fitted model using CATL based on Henderson's Mixed Model equation (H) are presented in Table 1. In real-data sets where the within-risk variability differs considerably from risk to risk, it is more appropriate to model unequal process variances when using CATL methods. As noted by Dornheim (2009), this assumption is rather significant for detection of atypical data points. Figure 1 indicates substantial differences in within-risk variability among the states. Hence, we fit model (2) twice, once when assuming equal (E) process variances and another time for

Table 1. Grand parameters, credibility adjusted estimates, and predictions of Hachemeister's bodily injury data (*Scenario 1*) based on *unweighted* regression

Estimation procedure	Grand parameters	Credibility Adjusted Estimates for State					Prediction (est. std. error) for State				
		1	2	3	4	5	1	2	3	4	5
REML H	1460 32	1685 60	1381 21	1545 42	1221 24	1470 20	2412 (110)	1651 (123)	2087 (193)	1533 (242)	1726 (81)
CATL HE	7.2874 0.0155	7.4379 0.0285	7.2372 0.0080	7.3710 0.0246	7.0773 0.0069	7.3136 0.0094	2461 (111)	1542 (75)	2188 (148)	1294 (65)	1695 (78)
CATL HU	7.2841 0.0164	7.4390 0.0283	7.2353 0.0083	7.3571 0.0211	7.0937 0.0138	7.2954 0.0107	2457 (111)	1542 (76)	2071 (193)	1451 (178)	1689 (59)
REML HR	1671 32	2051 46	1516 25	1817 38	1371 30	1601 23	2348 (121)	1680 (125)	2062 (194)	1569 (242)	1751 (85)
CATL HRE	7.3879 0.0156	7.6234 0.0268	7.2897 0.0089	7.5309 0.0236	7.1200 0.0093	7.3755 0.0096	2435 (112)	1550 (76)	2174 (149)	1311 (66)	1698 (78)
CATL HRU	7.3901 0.0169	7.6233 0.0270	7.2897 0.0090	7.4948 0.0199	7.1776 0.0178	7.3651 0.0108	2438 (112)	1552 (76)	2057 (194)	1482 (176)	1692 (59)

Note: For REML procedures, we report grand mean, $\hat{\beta}$, and corresponding credibility adjusted estimates, $\hat{\beta}_i$.
For CATL procedures, we provide grand location, $\hat{\lambda}$, and its subject-specific credibility adjusted locations, $\hat{\lambda}_i$.

unequal (U). Here is a summary of all the notations used in Tables 1–3:

- **REML H** and **REML HR:** REML fitting, based on Henderson's mixed model equations, of Hachemeister's model (**REML H**) and Hachemeister's-revised model (**REML HR**).
- **CATL HE** and **CATL HU:** CATL fitting, based on Henderson's mixed model equations, of Hachemeister's model using the assumption of equal process variances (**CATL HE**) and unequal process variance (**CATL HU**).
- **CATL HRE** and **CATL HRU:** CATL fitting, based on Henderson's mixed model equations, of Hachemeister's-revised model using the assumption of equal process variances (**CATL HRE**) and unequal process variance (**CATL HRU**).

Scenario 1 denotes the real-data set used by Hachemeister. We report fixed effects β (intercept and slope) and for each state credibility adjusted estimates and predicted average claim sizes (premiums for the next period, $t = 13$). Further, in Figure 2 we plot residuals from the fitted log-Hachemeister model (2) versus the potential exposure measure, *number of claims per period.* The megaphone-shaped picture reveals that the logarithmic transformation of average loss per claim did not remove heteroscedasticity of error components. Thus, additional weighting is required. In practice, it is common to observe situations, where risks with larger exposure measure typically exhibit lower variability (Kaas, Dannenburg, and Goovaerts 1997; Frees, Young, and Luo 2001). Here, it seems that *number of claims per period,* denoted by υ_{it}, significantly affects the within-risk variability. Indeed, the data set presented in the appendix supports this fact. In comparison to number of claims, State 4 reports high average losses per claim. This, in turn, yields to the increased within-state variability that is noticeable in Figure 1.

To obtain homoscedastic error terms, we fit models using υ_{it} as subject-specific weights. This model can be written as

$$ln(y_{it}) = \mathbf{x}_{it}\beta + \mathbf{z}_{it}\alpha_i + \varepsilon_{it}\upsilon_{it}^{1/2},$$

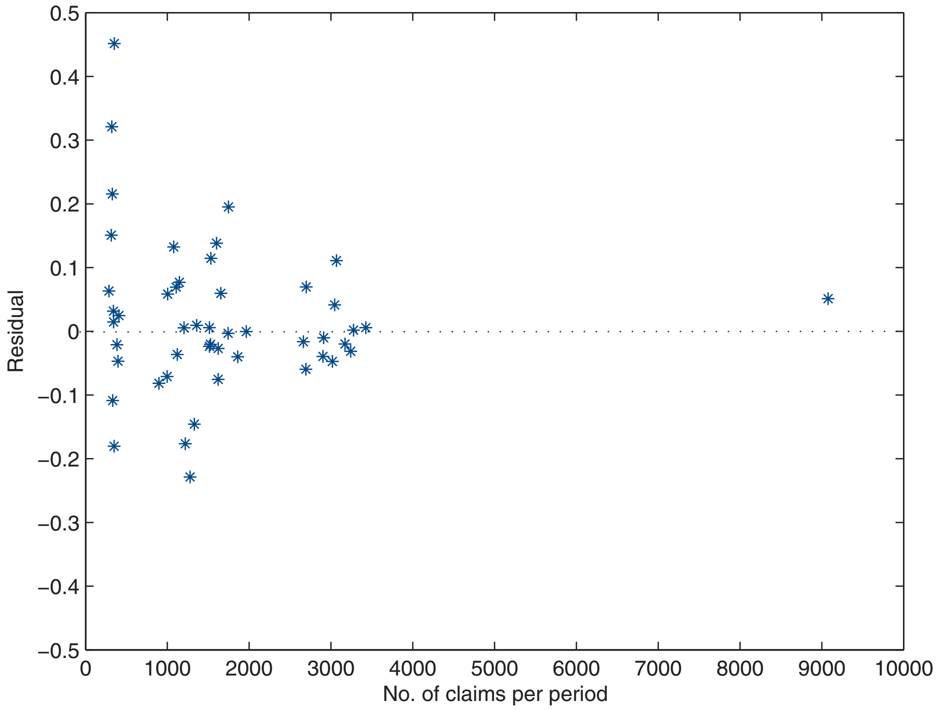
where $\{\varepsilon_{it}\}$ is an *i.i.d.* sequence of normally distributed noise terms. Similarly, the classical linear trend model as described by Hachemeister becomes

$$y_{it} = \mathbf{x}_{it}\beta + \mathbf{z}_{it}\alpha_i + \varepsilon_{it}\upsilon_{it}^{1/2}.$$

In view of the latter regression equation, Hachemeister's model can be considered as generalization of the weighted Bühlmann-Straub model.

Our findings for fitting of weighted regression models to Hachemeister's data are shown in Table 2, together with the results obtained from other prominent estimation procedures. The base model is the linear

Figure 2. Scatter plot of residuals versus number of claims per period



trend model estimated by Goovaerts and Hoogstad (1987). To investigate the influence of the pursued procedure for estimation in Hachemeister's regression model, Frees, Young, and Luo (1999) implemented several combinations of estimation methods (R = REML, M = Maximum likelihood), iteration methods (N = Newton-Raphson, F = Fisher-Scoring) and starting values (M = Rao's MIVQUE(0), S = Swamy's moment estimator) in SAS using the package PROC MIXED. Moreover, the authors distinguish between constraining the estimator of the variance-covariance matrix $\mathbf{D}$, $\hat{\mathbf{D}}$, to be positive definite (Y) or not (N). In order to limit distorting effects of within-risk outliers on the assessment of individual's claim experience, Pitselis (2002) applied MM- and GM-estimators to Hachemeister's regression model. We denote these approaches by MM-RC and GM-RC, respectively. As an extension of his joint work about robust estimation in the Bühlmann-Straub model (Garrido and Pitselis 2000), Pitselis (2008) also employs regression *M*-estimators (M-RC) that are based on Hampel's influence function approach (Hampel et al., 1986). Here is a summary of all the additional notations:

- **Base:** Linear trend model (Goovaerts and Hoogstad 1987).
- **RNMY** and **RNMN:** REML fitting; Newton-Raphson iteration method; Rao's MIVQUE(0) starting values; Estimator of variance-covariance matrix is constrained to be positive definite = Yes (**RNMY**) and = No (**RNMN**). (Frees, Young, and Luo 1999).
- **RFSY** and **RFSN:** REML fitting; Fisher-Scoring iteration method; Swamy's starting values; Estimator of variance-covariance matrix is constrained to be positive definite = Yes (**RFSY**) and = No (**RFSN**). (Frees, Young, and Luo 1999).
- **MNMY** and **MNMN:** Maximum likelihood fitting; Newton-Raphson iteration method; Rao's MIVQUE(0) starting values; Estimator of variance-covariance matrix is constrained to be positive definite = Yes (**MNMY**) and = No (**MNMN**). (Frees, Young, and Luo 1999).
- **MFSY** and **MFSN:** Maximum likelihood fitting; Fisher-Scoring iteration method; Swamy's starting values; Estimator of variance-covariance matrix is constrained to be positive definite = Yes (**MFSY**) and = No (**MFSN**). (Frees, Young, and Luo 1999).

Table 2. Grand parameters, credibility adjusted estimates, and predictions of Hachemeister's bodily injury data (*Scenario 1*) based on *weighted* regression

Estimation procedure	Grand parameters	Credibility Adjusted Estimates for State					Prediction (est. std. error *) for State				
		1	2	3	4	5	1	2	3	4	5
Base[a]	1469 32	1696 57	1377 21	1540 41	1312 15	1421 26	2436	1650	2073	1507	1759
RNMY[a]	1461 32	1659 62	1396 18	1535 43	1199 25	1518 12	2468 (44)	1624 (96)	2093 (119)	1528 (204)	1680 (72)
RFSY[a]	1461 32	1659 62	1396 18	1535 43	1199 25	1518 12	2468 (44)	1624 (96)	2093 (119)	1527 (204)	1680 (72)
RNMN[a]	1501 28	1660 62	1434 13	1555 39	1394 5	1464 20	2464 (41)	1606 (76)	2067 (92)	1454 (156)	1720 (59)
RFSN[a]	1501 28	1660 62	1434 13	1555 39	1394 5	1464 20	2464 (41)	1606 (76)	2067 (92)	1453 (157)	1720 (59)
MNMY[a]	Initial estimate is not feasible.										
MFSY[a]	1461 32	1659 62	1396 18	1535 43	1197 26	1518 12	2468 (44)	1624 (95)	2093 (118)	1529 (201)	1680 (71)
MNMN[a]	Yes, but Hessian is not positive definite – leads to trying to take the inverse of a singular matrix.										
MFSN[a]	1502 28	1660 62	1435 13	1555 39	1397 5	1464 20	2463 (41)	1608 (75)	2065 (91)	1467 (153)	1720 (59)
M-RC[b] ($c = 1.5$)		1696 57	1377 21	1540 41	1312 15	1421 26	2437	1650	2073	1507	1759
M-RC[b] ($c = 1.345$)		1696 57	1377 21	1540 41	1312 15	1421 26	2437	1650	2073	1507	1759
GM-RC[c] ($k = 1$)		1680 57	1378 21	1544 42	1247 20	1452 22	2427	1648	2092	1505	1737
GM-RC[c] ($k = 2$)		1686 57	1377 21	1544 42	1247 20	1451 22	2427	1648	2092	1505	1737
MM-RC[c]		1680 57	1378 21	1544 42	1247 20	1452 22	2427	1648	2092	1505	1737
REML H	1492 30	1655 62	1414 16	1536 42	1354 12	1502 15	2465 (109)	1625 (122)	2077 (193)	1519 (248)	1695 (77)
CATL HE	7.2755 0.0174	7.3790 0.0335	7.2464 0.0084	7.3422 0.0228	7.1169 0.0114	7.2932 0.0110	2483 (64)	1575 (94)	2091 (193)	1455 (175)	1703 (57)
CATL HU	7.2819 0.0165	7.4348 0.0289	7.2349 0.0082	7.3530 0.0210	7.0901 0.0137	7.2967 0.0105	2471 (111)	1545 (74)	2065 (194)	1447 (174)	1691 (57)
REML HR	1675 34	2058 60	1517 22	1800 40	1399 32	1601 16	2451 (109)	1661 (123)	2065 (193)	1613 (242)	1706 (78)
CATL HRE	7.3856 0.0189	7.5955 0.0330	7.3026 0.0096	7.4872 0.0217	7.1200 0.0188	7.3755 0.0111	2484 (65)	1596 (95)	2086 (194)	1517 (172)	1713 (58)
CATL HRU	7.3867 0.0169	7.6216 0.0276	7.2894 0.0089	7.4859 0.0198	7.1710 0.0176	7.3658 0.0106	2450 (113)	1552 (74)	2049 (195)	1477 (172)	1693 (57)

Sources: [a] Frees, Young, and Luo (1999), [b] Pitselis (2008), [c] Pitselis (2002).
* Standard errors reported by Frees, Young, and Luo (1999) are computed.

- **M-RC:** Robust credibility based on *M*-estimators. (Pitselis 2008).
- **MM-RC** and **GM-RC:** Robust credibility based on multiple *M*-estimators (**MM-RC**) and generalized *M*-estimators (**GM-RC**). (Pitselis 2002).

The reliability of credibility estimates, $\hat{\mu}_{i,\tau+1}$, is of major concern to the insurer. In particular, down-biased predictions result in too-low portfolio premiums, which in turn lead to insufficient funding and threatening of the insurer's solvency in the long-run. To assess the quality of assigned credibility premiums we also report their standard errors in Table 2. These can be used to construct prediction intervals of the form BLUP (credibility estimate $\hat{\mu}_{i,\tau+1}$) plus and minus multiples of the standard error (Frees, Young, and Luo 1999). Here, we estimate the standard error of prediction, $\hat{\sigma}_{\hat{\mu}_{i,\tau+1}}$, from the data using the common nonparametric estimator

$$\hat{\sigma}_{\hat{\mu}_{i,\tau+1}} = \left[\widehat{MSE}(\hat{\mu}_i) - \widehat{bias}^2(\hat{\mu}_i)\right]^{1/2}$$

$$= \left[\upsilon_{i\cdot}^{-1}\sum_{t=1}^{\tau}\omega_{it}\upsilon_{it}(\hat{\mu}_{it}-y_{it})^2 - \left(\upsilon_{i\cdot}^{-1}\sum_{t=1}^{\tau}\omega_{it}\upsilon_{it}(\hat{\mu}_{it}-y_{it})\right)^2\right]^{1/2},$$

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

Discussion of Table 2

Let us start with Hachemeister's linear trend model. We observe that REML estimates based on Henderson's Mixed Model Equations, REML H, and the resulting predictions mirror findings of Frees, Young, and Luo (1999) who investigated the influence of several combinations of computation methods for classical credibility. When applying corrected adaptively truncated likelihood methods, CATL HE/U, for States 1, 3, and 5, we record credibility estimates (predictions) that are similar to those of standard approaches or robust regression techniques discussed by Pitselis (2002, 2008). For the second and fourth state, CATL HE/U determines slightly lower premiums. For instance, for State 4 CATL HU yields, 1447 whereas the base model by Goovaerts and Hoogstad (1987) finds 1507. This can be traced back to the removal of the suspicious observations #6 and #10 in State 2 and #7 in State 4. Robust methods employed by Pitselis (2002, 2008) hardly react to these potential outliers. This is mainly due to the subjectively chosen truncation points *c* and *k,* which may be inappropriate (too high). CATL HU also detects claim #4 in State 5 as an outlier and assigns negligible excess premiums (discounts) of −1.47 per risk. For the revised linear trend model (R) patterns are similar. □

Discussion of Figure 3

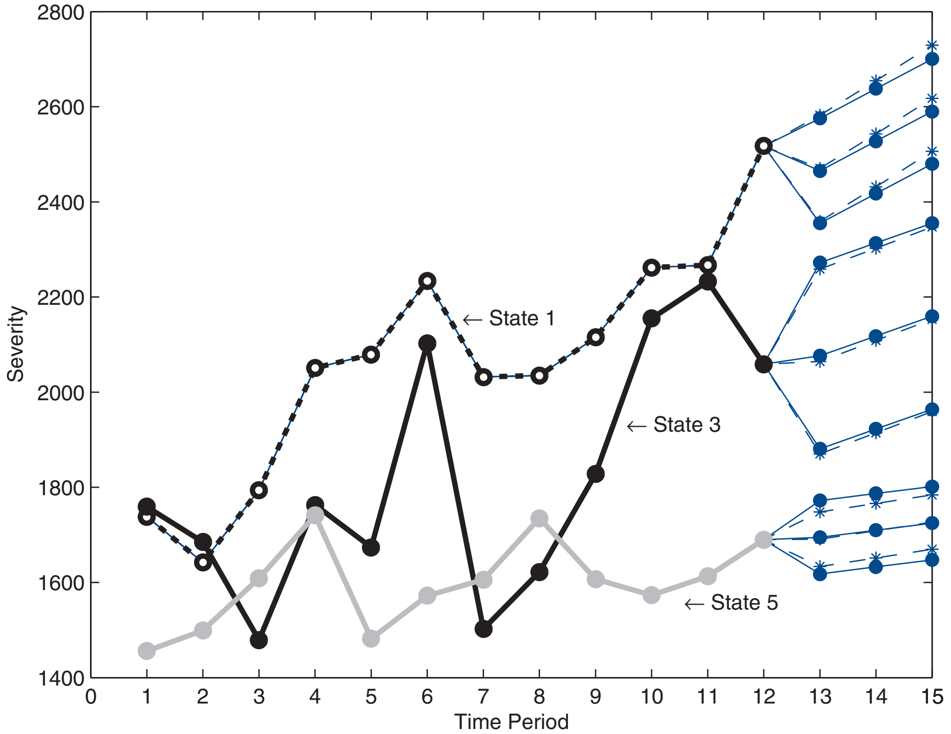
We display in Figure 3 the three-step predictions of the average loss per claim for States 1, 3, and 5 when using REML H and CATL HU. The thinner lines show predictions and one standard error of the prediction for each of the three states. These can be compared to Frees, Young, and Luo (1999, Figure 1). Note, the less variable States 1 and 5 have shorter prediction intervals compared to State 3. As expected, in *Scenario 1* predictions and intervals obtained from REML and CATL HU nearly coincide. □

To illustrate robustness of regression credibility methods that are based on M-RC, MM-RC, and GM-RC estimators for quantifying individual's risk experience, Pitselis (2002, 2008) replaces the last observation of the fifth state 1690 by 5000 (*Scenario 2*). We follow the same contamination strategy and summarize outcomes in Table 3. It turns out that the choice of the methodology has a major impact on the fitting of credibility models.

Discussion of Table 3

In the presence of a single outlier in *Scenario 2,* it is not surprising that standard statistical methods react dramatically. In the contaminated State 5, the REML H and base credibility estimate of Goovaerts and Hoogstad (1987) get highly attracted by the outlying

Figure 3. Selected severity predictions based on Hachemeister's model (weighted) and data. The thinner middle line marked by ● denotes one-, two-, and three-step predictions using REML H. The upper and lower lines complete the one standard error prediction interval. Predictions from the CATL HU procedure are marked by *



observation and increase from 1695 to 2542 and from 1759 to 2596, respectively. Further, it is well known that the occurrence of outliers distorts the estimation process of variance components and, hence, yields too-low credibility weights. As a consequence, credibility adjusted estimates are pulled toward the increased grand mean (most notably the slope component) which results in inflated credibility premiums and prediction intervals across all states. In particular, for REML H the estimated standard error in State 5 explodes from 77 to 829 due to enhanced within-risk variability that is caused by the single catastrophic event. Even though only the fifth contract is contaminated, the portfolio-unbiased regression M- and GM-credibility estimators with chosen tuning parameters $c = 1.345$ and $k = 2$, respectively, are significantly distorted. Indeed, these estimators provide protection in the assessment of the individual state experience. However, they reveal a substantial lack of robustness toward single large claims that have adverse effects on the between-risk variability as well. Consequently, the entire estimation process for states that have merely small claims (e.g., see State 2) also becomes distorted. These deficiencies have been removed by CATL estimators. We observe that CATL HU still provides reasonable credibility premiums for all contracts in the portfolio. Both the grand location, $\hat{\lambda} = (7.28, 0.016)$, the credibility adjusted estimates, $\hat{\lambda}_i$, the estimated standard errors, and predictions of expected claims remain stable. We shall emphasize that the latter changes only marginally from 1691 to 1689 in the contaminated State 5. The CATL HU identifies the single large claim and distributes uniformly an extraordinary premium of 4.33 to all contracts. □

In Figure 4 we visualize the impact of a single outlier on the premium ratemaking process. When using classical regression techniques such as REML, the

Table 3. Grand parameters, credibility adjusted estimates, and predictions of Hachemeister's contaminated bodily injury data based on *weighted* regression

Estimation procedure	Grand parameters	Credibility Adjusted Estimates for State					Prediction (est. std. error) for State				
		1	2	3	4	5	1	2	3	4	5
Base		1526 75	1423 31	1427 58	1409 45	1296 100	2501	1826	2181	1994	2596
M-RC[b] ($c = 1.5$)		1650 85	1316 51	1499 69	1256 45	1380 57	2755	1979	2396	1841	2121
M-RC[b] ($c = 1.345$)		1638 86	1304 52	1487 70	1308 41	1365 58	2756	1980	2397	1841	2119
GM-RC[c] ($k = 1$)		1571 83	1266 46	1434 67	1138 45	1339 48	2645	1868	2311	1723	1964
GM-RC[c] ($k = 2$)		1573 82	1266 46	1435 67	1140 45	1334 49	2640	1867	2307	1719	1978
MM-RC[c]		1568 83	1264 47	1432 68	1133 45	1315 48	2649	1870	2315	1724	1943
REML H	1377 65	1547 75	1305 42	1391 63	1315 52	1327 93	2517 (119)	1852 (150)	2206 (204)	1987 (255)	2542 (829)
CATL HE	7.2539 0.0123	7.4061 0.0296	7.2371 0.0098	7.3377 0.0169	7.1090 0.0116	7.2924 0.0109	2424 (75)	1589 (95)	2006 (199)	1447 (175)	1700 (60)
CATL HU	7.2825 0.0163	7.4345 0.0290	7.2354 0.0081	7.3530 0.021	7.0909 0.0135	7.2986 0.0100	2477 (111)	1550 (74)	2071 (194)	1452 (174)	1689 (60)
REML HR	1812 72	2032 65	1630 50	1809 63	1700 67	1887 12	2455 (110)	1949 (166)	2229 (204)	2141 (275)	2629 (818)
CATL HRE	7.3326 0.0123	7.5981 0.0256	7.3025 0.0110	7.4837 0.0144	7.1800 0.0128	7.3643 0.0113	2360 (94)	1597 (96)	1967 (203)	1450 (174)	1700 (61)
CATL HRU	7.3865 0.0168	7.6216 0.0277	7.2894 0.0088	7.4859 0.0198	7.1709 0.0175	7.3645 0.0101	2459 (112)	1559 (74)	2057 (195)	1484 (172)	1694 (60)

Source: [b] Pitselis (2008), [c] Pitselis (2002).

contaminated contract elevates the grand mean of all states, most of all the forecasted credibility premium and its prediction interval in State 5. Clearly, CATL provides robustness for adequate inference in Hachemeister's regression model.

4. Medicare data

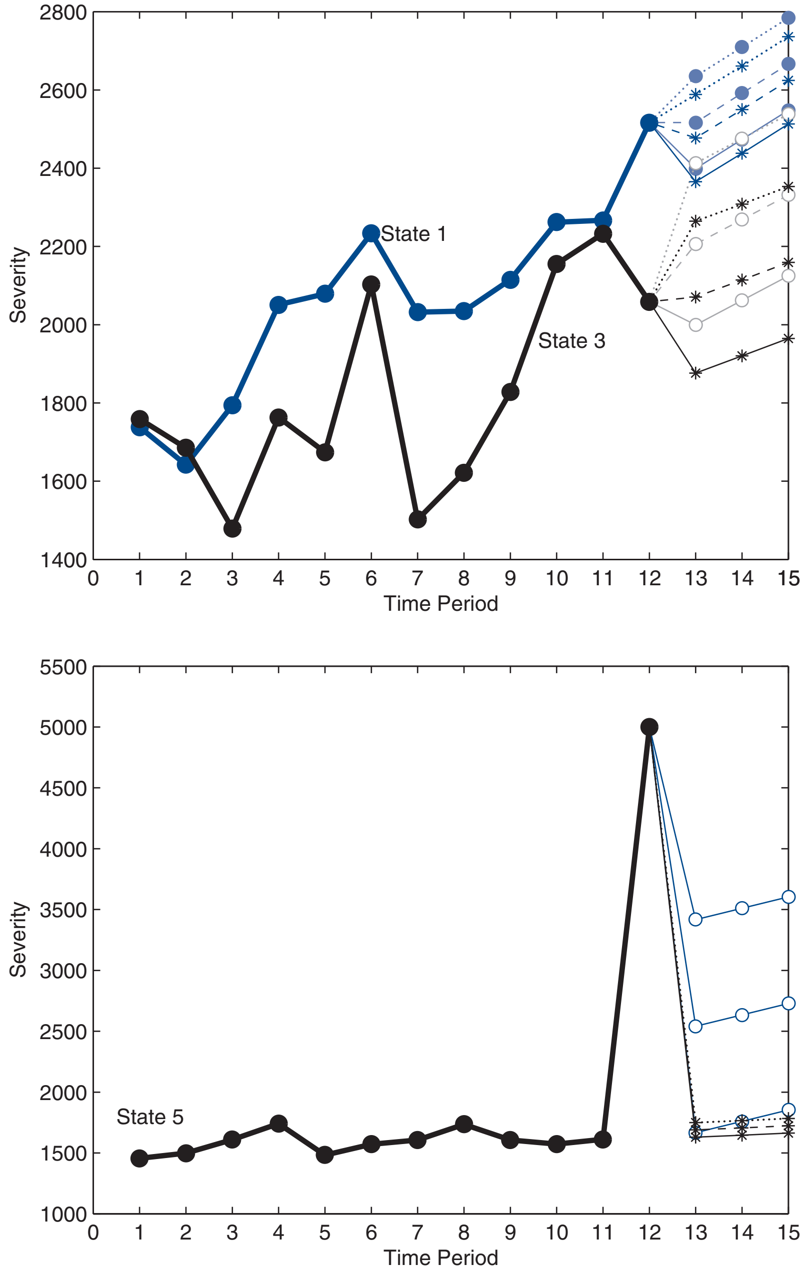
In this section, we apply CATL procedures to a health care data set that contains Medicare costs for inpatient hospital charges. This real-data set has been examined by Frees, Young, and Luo (2001) to demonstrate that general mixed linear models can be used to produce credibility estimates of future claims.

4.1. Data characteristics

The underlying health care data set for Figure 5 has been published by the Health Care Financing Administration and reports inpatient hospital charges over $\tau = 6$ years, from 1990 to 1995. These charges have been covered by the Medicare program in $I = 54$ states across the United States, including the District of Columbia, Virgin Islands, Puerto Rico, and an unspecified category, in addition to the 50 states.

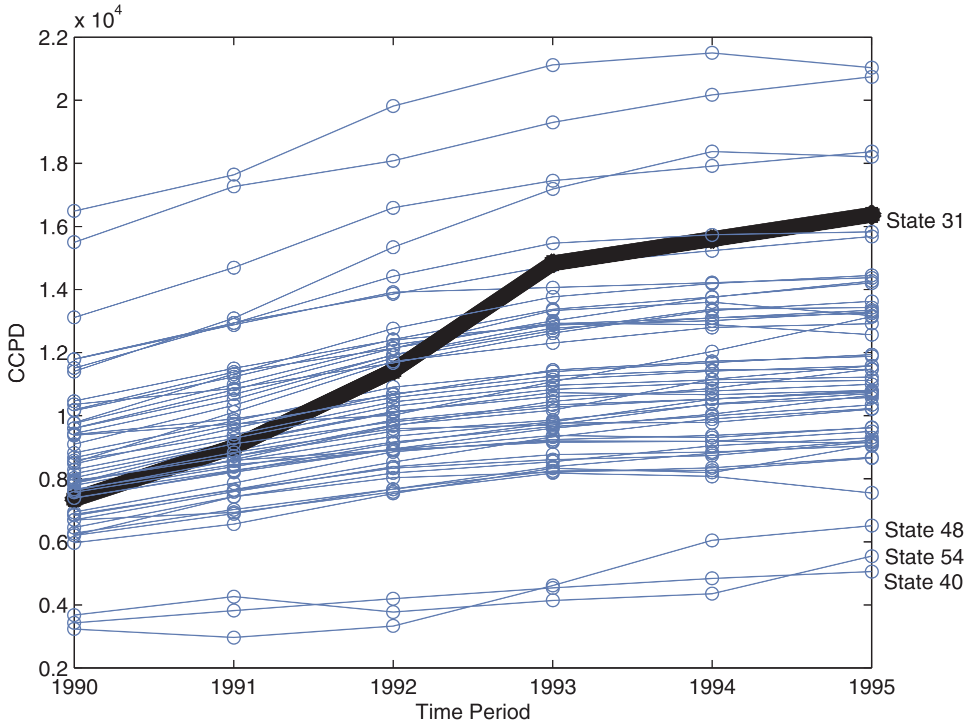
The Medicare program reimburses hospital claims on per-stay basis, hence, the dependent variable is *covered claims per discharge,* denoted by CCPD. The multiple time series plot in Figure 5 indicates that there is rather substantial variability among the states and some variability over the time. Therefore, it is natural to explain the response at least by the regressors state and time. Frees, Young, and Luo (2001) investigated several possible explanatory variables through graphical tools and found that the component *average hospital stay per discharge in days,* denoted by AVE_DAYS, is a statistically significant predictor.

Figure 4. Selected severity predictions based on Hachemeister's model (weighted) and data. The thinner middle line marked by ○ denotes one-, two-, and three-step predictions using REML H. The upper and lower lines complete the one standard error prediction interval. Predictions from CATL HU are marked by *



More interestingly, the state of New Jersey (State 31) reveals a notably greater increase of CCPD over the years 1990 to 1993. Frees, Young, and Luo (2001) identify this *outlying* growth of medical costs and take it explicitly into account in their model by incorporating a state-specific correction term for the time slope in New Jersey. Further analysis (e.g., employing scatter plots) shows that the second observation of the 54th state is explained by an unusually high hospital utilization AVE_DAYS. Thus, this data point that has large distorting effects on classical estimation procedures, such as REML, has been manually removed by

Figure 5. Multiple time series plot of covered claims per discharge (denoted by CCPD) over the years 1990 to 1995



the authors (see Frees, Young, and Luo 2001, Table 1, Figure 4). In Figure 5 we observe that two other outlying risks are State 40 and 48. These states report relatively low CCPD over time and, thus, have a distorting effect on the estimation of between-state variability.

4.2. Model fitting, outlier detection, prediction

For CATL we fit an error components model of the form

$$ln(CCPD_{it}) = \beta_1 + \beta_2 YEAR_t + \beta_3 AVE\_DAYS_{it} + \alpha_{i1} + \alpha_{i2} YEAR_t + \varepsilon_{it}, \qquad (3)$$

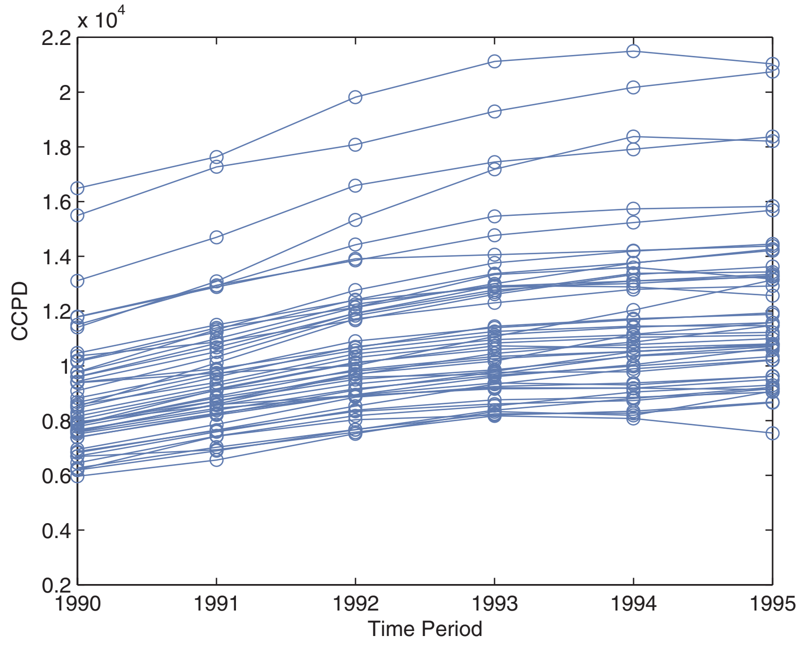
where $\beta = (\beta_1, \beta_2, \beta_3)'$ is the grand location, $\alpha_i = (\alpha_{i1}, \alpha_{i2})'$ are the unobservable random effects, and $\mathbf{x}_{it} = (1, t, AVE\_DAYS_{it})$ and $\mathbf{z}_{it} = (1, t)$ are the corresponding explanatory variables for $i = 1, \ldots, I = 54$ states followed over the periods $t = 1, \ldots, \tau = 6$. Frees, Young, and Luo (2001) found that no weighting by some exposure measure is required to accommodate potential heteroscedasticity in the data. Therefore, we model the serially uncorrelated error terms by the unweighted variance-covariance matrix $\mathbf{R}_i = \sigma_\varepsilon^2 \mathbf{I}_{\tau\times\tau}$, where $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ for all states. This model can be viewed as an extension of Hachemeister's linear trend model in Section 3. It is similar to the preferred mixed linear model in Frees, Young, and Luo (2001), denoted by Model 6. Note, however, that their model includes an additional special interaction variable to represent the atypical large time slope in State 31 and is given by

$$CCPD_{it} = \beta_1 + YEAR_t(\beta_2 + \beta_4 1\{State = 31\}) + \beta_3 AVE\_DAYS_{it} + \alpha_{i1} + \alpha_{i2} YEAR_t + \varepsilon_{it}, \qquad (4)$$

where $\beta_2 + \beta_4 1\ \{State = 31\} + \alpha_{i2}$ can be interpreted as the New Jersey specific total contribution of the regressor YEAR.

In our analysis, the CATL procedure detects and rejects all suspicious data points that have been identified by laborious graphical tools in Frees, Young, and Luo (2001) and, therefore, supersedes the modeling

Figure 6. Multiple time series plot of covered claims per discharge (denoted CCPD) over the years 1990 to 1995 after data *cleaning*



of New Jersey-specific covariates. Results of the detection process using CATL with equal process variances are visualized in Figure 6. We see that State 31 (New Jersey) having an extraordinary high time slope, States 40, 48 and 54 having unusual low intercepts, and the second observation of the 54th state (though not visible in Figure 6) have been entirely removed.

In Table 4 we provide results of fitted regression models based on Equation (4) for Model 6 in Frees, Young, and Luo (2001), and on Equation (3) for the CATL *approaches 1* and *2*. Here, *approach 1* denotes the Medicare data set containing all 54 states, whereas the outlying States 31 and 54 have been deleted for *approach 2*. Note that our REML estimates do not incorporate the New Jersey-specific time slope.

Table 4. Comparisons of diverse fitted regression credibility models

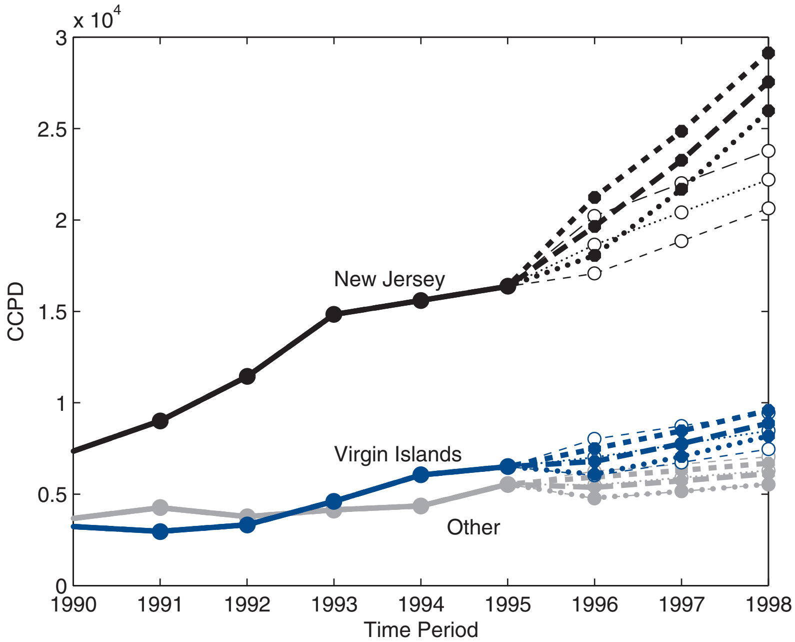
	Parameter Estimates of Model Variables			
Model	Intercept (β_1)	Year (β_2)	Year (State = 31) (β_4)	AVE_DAYS (β_3)
Model 6*	4,827	753.1	1,540.81	348.3
REML 1	7,932.6	675.3		21.9
REML 2	5,447.9	744.5		301.8
CATL 1	8.491	0.081		0.061
CATL 2	8.504	0.081		0.059

Source: * Frees, Young, and Luo (2001) using REML.

Discussion of Table 4

When the entire data set is considered (*approach 1*) we observe that the standard method REML 1 gets attracted by the outlying data points and produces estimates of fixed effects that differ significantly from those of Model 6. For instance, β_3 is 348.3 for Model 6 versus 21.9 for REML 1. Also, REML 1 finds the time slope $\beta_2 = 675.3$ for all states. However, the mean slope of the state of New Jersey should be much steeper ($\beta_2 + \beta_4 = 753.1 + 1540.81$) as indicated by Model 6 and Figure 7. Once the States 31 and 54 have been eliminated (REML 2), these standard estimates are comparable to those in Model 6 but still do not account for the extraordinary increase of hospital costs in State 31. As expected, the robust CATL procedure provides protection against catastrophic events and we do not observe any significant differences between the estimates obtained when employing CATL 1 or CATL 2. Also, Figure 7 shows that

Figure 7. The years 1990–1995 represent actual CCPD for selected states New Jersey (State 31), Virgin Islands (State 48), and "Other" (State 54). For 1996–1998, the middle line, marked by ○, denotes the one-, two-, and three-step predictions using REML. The upper and lower lines complete the two standard deviations interval. Predictions and intervals obtained from CATL are marked by ●



the robust procedures allow for the greater rate of inflation of Medicare costs in New Jersey. □

For the computation of one-, two-, and three-step predictions over the years 1996–1998, we assume that the most recent subject-specific average hospital utilizations per discharge (AVE_DAYS_{it}) remain constant, i.e., $AVE\_DAYS_{i6} = AVE\_DAYS_{i7}$. Then, in Table 5 we give the grand mean $\hat{\beta}$ and selected credibility adjusted estimates $\hat{\beta}_i$ when using REML 1. Likewise, the grand location and credibility-adjusted locations are summarized for CATL 1. The resulting credibility estimates for the expected $CCPD_{it}$ as well as the estimated standard errors for construction of prediction intervals, are registered over the forecasted periods $t = 7, 8, 9$, that represent the years 1996 to 1998.

Table 5. Grand parameters, credibility adjusted estimates, and predictions of Medicare data for selected states New Jersey (State 31), Virgin Islands (State 48) and "Other" (State 54)

Estimation procedure	Grand parameters	Credibility Adj. Est. for State			3-step Prediction (est. std. error) for State		
		31	48	54	31	48	54
REML 1	7,932.6	5,954.0	1,679.3	2,592.7	18,644 (788.3)	7,017 (498.3)	5,509 (372.9)
	675.3	1,781.6	721.8	391.5	20,426	7,739	5,900
	21.9	21.9	21.9	21.9	22,207	8,461	6,292
CATL 1	8.4904	8.1043	7.1184	7.6646	19,648 (792.2)	6,790 (350.2)	5,365 (285.7)
	0.0814	0.1685	0.1327	0.0647	23,269	7,766	5,729
	0.0606	0.0606	0.0606	0.0606	27,555	8,880	6,118

Note: For REML fitting, eq. (4.2), we report grand mean, $\hat{\beta}$, and corresponding credibility adjusted estimates, $\hat{\beta}_i$.
For CATL fitting, eq. (4.1), we provide grand location, $\hat{\beta}$, and its subject-specific credibility adjusted locations, $\hat{\beta}_i$.

Discussion of Figure 7

Most of the findings that are outlined in Table 5 for the selected states New Jersey, Virgin Islands, and "Other" can be visualized in Figure 7. We see that credibility estimates for Virgin Islands and "Other" are similar. However, we observe that the prediction intervals obtained from CATL are somewhat shorter. For instance, this is due to the removal of the second data point in the 54th state. Let us focus on predictions in the state of New Jersey. As indicated by results recorded in Table 4, it becomes apparent that REML significantly underestimates the inflation trend of Medicare costs. In contrast, CATL produces suitable forecasts for CCPDs. We shall also emphasize that intervals based on CATL are shorter. □

5. Housing prices in U.S. metropolitan areas

In the case study of this section, we analyze annual housing prices in U.S. metropolitan statistical areas (MSAs). In view of the previous two sections, where Section 3 falls directly in traditional property/casualty field and Section 4 follows the health insurance practice, this type of data is not what actuaries concentrate on currently in their work. With this choice of data we clearly venture outside the actuary's comfort zone, but hope to provide additional insights, highlight the versatility of these advanced techniques, and to broaden their area of application.

In this section, we employ graphical tools and statistical criteria to

- view data,
- find atypical data points, and
- evaluate the reliability of the newly designed robust approach as diagnostic tool.

We also compare fitted parameters of robust CATL with classical REML. The data set we consider is taken from Frees (2004) and had been studied originally by Green and Malpezzi (2003).

We investigate the influence of the demand-side variables *annual percentage growth of per capita income* (PERYPC) and *annual percentage growth of population* (PERPOP) on the response NARSP. The response variable NARSP represents the MSA's average sale price in logarithmic units and is based on transactions reported through the Multiple Listing Service, National Association of Realtors. The balanced data set contains $I = 36$ MSAs with $n = 9$ annual observations each, over the years 1986 to 1994. We fit an error components model of the form

$$NARSP_{it} = \mu + \beta_1 PERYPC_{it} + \beta_2 PERPOP_{it} + \beta_3 YEAR_i + \alpha_i + \varepsilon_{it},$$

where $i = 1,\ldots, I$, $t = 1,\ldots, n_i = n$, μ is the population grand mean, α_i is the metropolitan-specific random intercept and $\beta = (\beta_1, \beta_2, \beta_3)'$ is the fixed effects of time-dependent explanatory variables $PERYPC_{it}$, $PERPOP_{it}$, and $YEAR_t$, respectively. We assume that the error terms are serially uncorrelated, so that $\mathbf{R}_i = \sigma^2_{\varepsilon_i}\mathbf{I}_{n\times n}$.

5.1. Data characteristics

In this subsection, we characterize the housing sales price data using graphical and common statistical tools. Table 6 summarizes the response variable NARSP, by MSA. We see that MSAs vary in their average and median NARSPs. In particular, average sales prices recorded for metropolitan areas #3, #5, #6 and #19 both start and close at comparatively high price levels. We also observe that the within-MSA variability (Std. Dev.) indicates substantial differences among MSAs. For instance, we record an extremely high variability of 0.24 for MSA #29 versus nearly negligible variations of as low as 0.03 for MSA #32. These MSAs are considered to be outlying with respect to their unusually high/low within-MSA variability. These patterns are highlighted when considering the total variation of yearly percentage-change in NARSP. (Note that the *total variation* of yearly percentage-change in NARSP is computed as the sum of the eight absolute percentage-changes for each NARSP observed over a 9-year period.)

Table 6. Summary statistics of average housing sales prices (NARSP), by MSA. Each MSA has *n* = 9 observations. The label * (#) marks outlying MSAs identified by using CATL with *identically* (*non-identically*) distributed error terms. For selected criteria the top *4* and the bottom *4* ranks are given in parentheses

Summary Statistics of average house sale prices (NARSP), by MSA						
MSA	Mean	Median	Minimum	Maximum	Std. Dev.	Total Variation
1	4.41	4.39	4.22	4.61	0.132	0.087
2	4.43	4.43	4.36	4.52	0.055 (33)	0.047 (34)
3#*	5.23 (2)	5.29 (2)	4.91	5.39	0.174	0.127 (4)
*4**	4.73	4.82	4.42	4.93	0.200(3)	0.131 (3)
5#*	5.10 (4)	5.18 (3)	4.77	5.23	0.170	0.106
6#*	5.44 (1)	5.54 (1)	5.11	5.56	0.178 (4)	0.097
7	4.53	4.49	4.40	4.76	0.114	0.103
8	4.50	4.49	4.40	4.64	0.090	0.054
9	4.41	4.42	4.28	4.51	0.078	0.051
10	4.24	4.27	4.11	4.33	0.076	0.057
*11**	4.73	4.76	4.46	4.97	0.209 (2)	0.119
12	4.31	4.31	4.08	4.51	0.150	0.101
13	4.57	4.57	4.30	4.75	0.155	0.139 (2)
14	4.17 (34)	4.22	3.92	4.34	0.155	0.103
15	4.49	4.49	4.36	4.62	0.087	0.059
16	4.32	4.31	4.18	4.47	0.092	0.067
17	4.61	4.67	4.29	4.72	0.152	0.099
18	4.27	4.35	3.96	4.42	0.172	0.116
19#*	5.17(3)	5.16 (4)	5.08	5.21	0.043 (35)	0.039 (36)
20	4.37	4.38	4.22	4.45	0.077	0.052
21	4.34	4.37	4.16	4.44	0.091	0.086
22	4.23	4.22 (33)	4.03	4.44	0.159	0.098
23	4.37	4.38	4.15	4.57	0.149	0.097
24	4.39	4.39	4.20	4.59	0.149	0.089
25	4.39	4.40	4.18	4.55	0.133	0.085
26	4.16 (35)	4.14 (35)	4.03	4.30	0.111	0.067
*27**	4.09 (36)	4.12 (36)	3.97	4.20	0.084	0.097
*28**	4.20(33)	4.18 (34)	4.14	4.31	0.054 (34)	0.052
*29**	4.40	4.38	4.14	4.76	0.236 (1)	0.142 (1)
30	4.67	4.69	4.41	4.78	0.123	0.087
31	4.38	4.36	4.26	4.47	0.072	0.050 (33)
*32**	4.52	4.54	4.48	4.55	0.026 (36)	0.044 (35)
33	4.27	4.26	4.12	4.39	0.096	0.098
34	4.23	4.24	4.15	4.36	0.077	0.075
35	4.31	4.24	4.22	4.59	0.125	0.094
36	4.45	4.44	4.25	4.69	0.165	0.100
Summary Statistics Portfolio						
Total	4.48	4.41	3.92	5.56	0.332	

Table 7. Summary statistics of average housing sales prices (NARSP), by Year. Each year summarizes I = 36 MSAs. The total summarizes (9 × 36) 324 observations

Year	Mean	Median	Minimum	Maximum	Std. Dev.
	4.31	4.24	3.92	5.11	0.28
	4.36	4.26	3.98	5.21	0.29
	4.40	4.32	4.03	5.36	0.33
	4.46	4.36	3.98	5.56	0.36
	4.50	4.39	3.97	5.56	0.35
	4.54	4.45	4.04	5.55	0.34
	4.57	4.50	4.12	5.54	0.32
	4.60	4.52	4.17	5.52	0.30
	4.62	4.56	4.20	5.55	0.29
Total	4.48	4.41	3.92	5.56	0.33

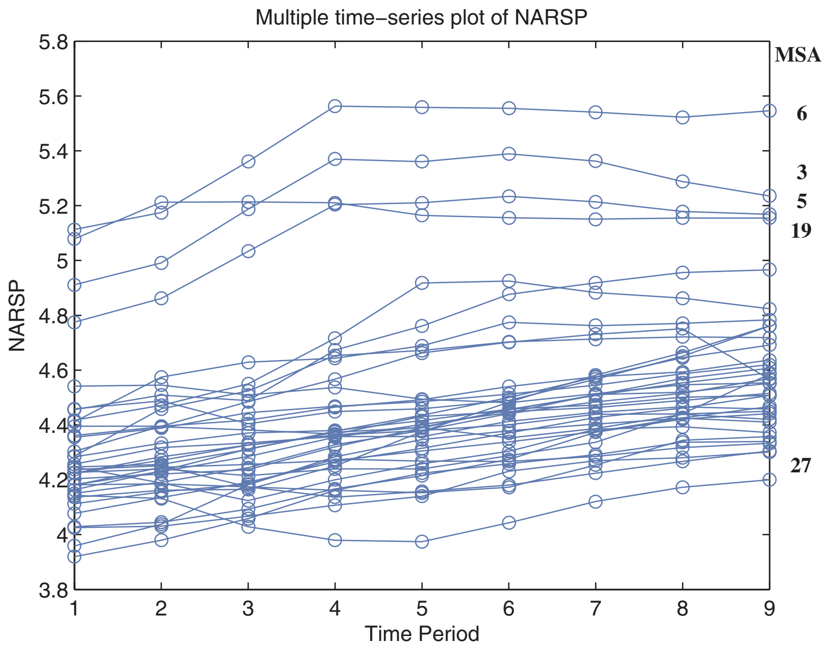
Table 7 provides summary statistics for NARSP, by year. We observe that almost all statistical quantities increase over time except for the overall variation (Std. Dev.) in the data. These do not hint at any significant time-dependency.

Figure 8 is a multiple time series plot of NARSP and provides a visualization of the findings in Tables 6 and 7. We see that not only are overall NARSP increasing in time but also that average house sale prices increase for each MSA. Indeed, as Figure 8 indicates, there is substantial variability among MSAs, not just a simple time trend. This plot also reveals that MSAs #3, #5, #6, #19, and #27 stay permanently away from the majority of data, which is seen from unusually high (low) subject-specific intercepts. Finally, this preliminary analysis will help us to understand how few atypical MSAs can have disturbing effects on classical estimation of variance components. It turns out that the robust CATL procedure can provide major improvements in model fit over the non-robust REML.

5.2. Model fitting and outlier detection

We find CATL estimates under the following assumptions for residuals: identically distributed, i.e., $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, and non-identically distributed, i.e., $\varepsilon_{it} \sim N(0, \sigma_{\varepsilon_i}^2)$. As discussed in a previous section, this assumption is rather significant for detection of atypical data points. In the current example, the assumption of equal error variances allows us to identify such

Figure 8. Multiple time series plot of NARSP over n = 9 years, 1986–1994. The line segments connect metropolitan statistical areas (MSAs).



MSAs as outlying when their individual within-subject variation is comparatively high or low. By contrast, when assuming unequal error variances, the CATL procedures only detect those MSAs for which random effects appear to be extreme.

Discussion of Table 8

CATL estimates for the grand mean $\hat{\mu}$ are lower than those from REML, which was expected. This effect can be traced back to the identification and elimination of extreme MSAs having rather high NARSP over time (Table 6). As a result, we record substantially reduced estimates for variance components σ_α^2 and σ_ε^2. In particular, when removing MSAs #3, #4, #5, #6, #11, #19, #27, #28, #29, #32 from the data, the between-MSA variability $\hat{\sigma}_\alpha^2$ drops dramatically from 0.097 (REML) to as low as 0.018 (CATL, identical). This is accompanied by a reduction of the overall within-subject variability $\hat{\sigma}_\varepsilon^2$ from 0.005 (REML) to 0.003 (CATL, identical). That is, when assuming identically distributed error terms, the CATL methods eliminate MSAs where either individual random intercepts are extreme or their within-subject variability differs clearly from the most typical one. Note that the CATL procedure reliably removes all extraordinary MSAs that have been detected using laborious and time-consuming graphical tools. For CATL methods with non-identical error variances the latter type of extremes is ignored, hence fewer MSAs are truncated from the data, and thus, the estimated variance component $\hat{\sigma}_\alpha^2$ is slightly increased to 0.026. Now, only the MSAs #3, #5, #6, #19 that have atypical high intercept are marked as extraordinary. We see that the estimates for fixed effects are of the same magnitude. □

Table 8. Fitted values for the housing sales price data

Estimation Procedure	Fixed Effects				Variance Components	
	$\hat{\mu}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_\alpha^2$
REML	4.34	–0.01	–0.00	0.04	0.005	0.097
CATL (identical)	4.22	–0.01	0.00	0.04	0.003	0.018
CATL (non-identical)	4.27	–0.01	–0.00	0.04	0.004	0.026

6. Summary

In this paper, we have illustrated how corrected adaptively truncated likelihood methods, CATL, can be used for robust-efficient fitting of general regression-type models with heavy-tailed data. We have seen that this procedure is a flexible and effective risk-pricing tool. This has been demonstrated through three well-studied examples from the fields of property and casualty insurance, health care, and real estate. In particular, we have shown the entire process of data analysis, model selection, outlier identification, and the associated statistical inference.

Further, a number of observations can be made based on our analysis. In particular, for Hachemeister's bodily injury data, we notice that CATL methods (a) allow to mitigate heteroscedasticity through explicit incorporation of weighting and/or use of logarithmic transformation; (b) accommodate within-risk variability through modeling of subject-specific process variance; (c) provide high robustness against outliers occurring both within and between risks through adaptive detection rules that automatically identify and reject excess claims in samples of small size; (d) provide robust credibility premiums; (e) demonstrate efficiency and reliability of detection rules, therefore making the use of graphical tools (for identification of outliers) and expert judgment (for the choice of truncation points) non-essential; and (f) compete well against established robust methods. Analysis of the Medicare and the house-pricing data sets just reinforces our findings.

Furthermore, for the readers who are interested in implementing CATL methods, the matlab computer code is available from the authors upon request. Also, we shall emphasize again that one should be careful about the automatic nature of "data cleaning" and model fitting exhibited by CATL methods (see comments c and e above). This aspect of the methodology is an important improvement, as it makes the process of statistical modeling easier and quicker. However, it does *not* imply that the actuary may now ignore such crucial issues as understanding of the practical problem, its context, available data, and the economic consequences of the model-based decisions.

Finally, the case studies in this paper were focused and did illustrate various statistical aspects of the CATL approach. What is equally important and interesting is to compare this methodology with some well-established and actuarially sound techniques. For instance, one method that is often used in practice is to set the overall portfolio rate level and then use predictive modeling techniques to allocate the rates between classes. This will be a topic for future research project.

Acknowledgments

The authors gratefully acknowledge the support provided by a grant from the Casualty Actuarial Society. We also are very appreciative of useful comments, suggestions, and valuable insights provided by two anonymous referees that helped to significantly improve the paper.

References

Bühlmann, H., "Experience Rating and Credibility," *ASTIN Bulletin* 4, 1967, pp. 199–207.

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

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

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

Dornheim, H., "Robust-Efficient Fitting of Mixed Linear Models: Theory, Simulations, Actuarial Extensions, and Examples," Ph.D. thesis, 2009, Department of Mathematical Sciences, University of Wisconsin–Milwaukee.

Dornheim, H., and V. Brazauskas, "Robust-Efficient Methods for Credibility when Claims are Approximately Gamma-Distributed," *North American Actuarial Journal* 11(3), 2007, pp. 138–158.

Dornheim, H., and V. Brazauskas, "Robust-Efficient Fitting of Mixed Linear Models: Methodology and Theory," *Journal of Statistical Planning and Inference* 141(4), 2011a, pp. 1422–1435.

Dornheim, H., and V. Brazauskas, "Robust-Efficient Credibility Models with Heavy-Tailed Claims: A Mixed Linear Models Perspective," *Insurance: Mathematics and Economics,* 48(1), 2011b, pp. 72–84.

Fellingham, G.W., H.D. Tolley, and T.N. Herzog, "Comparing Credibility Estimates of Health Insurance Claims Costs," *North American Actuarial Journal,* 9(1), pp. 1–12.

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

Frees, E.W., V.R. Young, and Y. Luo, "A Longitudinal Data Analysis Interpretation of Credibility Models," *Insurance: Mathematics and Economics* 24, 1999, pp. 229–247.

Frees, E.W., V.R. Young, and Y. Luo, "Case Studies Using Panel Data Models," *North American Actuarial Journal* 5(4), 2001, pp. 24–42. Supplemental material is available at: http://research3.bus.wisc.edu/course/view.php?id=129

Garrido, J., and G. Pitselis, "On Robust Estimation in Bühlmann-Straub's Credibility Model," *Journal of Statistical Research* 34(2), 2000, pp. 113–132.

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

Green, R.K., and S. Malpezzi, *A Primer on U.S. Housing Markets and Policy,* Washington, D.C.: Urban Institute Press, 2003.

Guszcza, J., "Hierarchical Growth Curve Models for Loss Reserving," Casualty Actuarial Society *E-Forum,* Fall 2008, pp. 146–173.

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

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, *Robust Statistics: The Approach Based on Influence Functions.,* New York: Wiley, 1986.

Kaas, R., D. Dannenburg, and M. Goovaerts, "Exact Credibility for Weighted Observations," *ASTIN Bulletin* 27, 1997, pp. 287–295.

Klinker, F., "Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility Into a Generalized Linear Model Setting," Casualty Actuarial Society *E-Forum,* Winter 2011 (Volume 2), pp. 1–25.

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

Pitselis, G., "Application of GM and MM Estimators to Regression Credibility," *Proceedings of the 2nd Conference in Actuarial Science and Finance,* Samos, Greece, 2002.

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

Appendix

A.1. Data

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

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

Source: Hachemeister (1975), Figure 3.

A.2. Mixed linear models and credibility

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

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\alpha_i + \varepsilon_i, \quad i = 1, \ldots, I, \tag{5}$$

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 $\beta \in \mathbb{R}^p$ and the subject-specific random effects $\alpha_i \in \mathbb{R}^q$, respectively, and ε_i is a τ_i-dimensional vector of within-subject residuals.

Following the classical framework of the model (5), 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 $i = 1, \ldots, I,$

$$\alpha_i \sim N_q(\mathbf{0}, \mathbf{D}), \ \varepsilon_i \sim N_{\tau_i}(\mathbf{0}, \mathbf{R}_i), \ and \ Cov(\alpha_i, \varepsilon_i) = \mathbf{0},$$

where $\mathbf{D}$ is a $q \times q$ positive definite variance-covariance matrix of the form $diag(\sigma^2_{\alpha_1}, \ldots, \sigma^2_{\alpha_q})$ and $\mathbf{R}_i = \sigma^2_\varepsilon \mathbf{I}_{\tau_i \times \tau_i}$ 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. Hence, in view of the assumptions (a), (b), (c), we have the so-called *hierarchical* formulation of the mixed linear model (5), for which

$$\mathbf{y}_i | \alpha_i \sim N_{\tau_i}(\mathbf{X}_i\beta + \mathbf{Z}_i\alpha_i, \mathbf{R}_i), \quad i = 1, \ldots, I.$$

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

The regression parameter β common to all individuals i is estimated by the generalized least squares (GLS) estimator

$$\hat{\beta}_{GLS} = \left(\sum_{i=1}^{I} \mathbf{X}_{i'}\mathbf{V}_i^{-1}\mathbf{X}_i \right)^{-1} \sum_{i=1}^{I} \mathbf{X}_{i'}\mathbf{V}_i^{-1}\mathbf{y}_i. \tag{6}$$

In the case of *known* variance components θ, this estimator is optimal and coincides with the maximum likelihood estimator of β. Then, the random effects α_i are determined by the best (with respect to the mean squared error criterion) linear unbiased predictor (BLUP)

$$\hat{\alpha}_{BLUP,i}(\theta) = \mathbf{D}\mathbf{Z}_i'\mathbf{V}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_{GLS}(\theta)\right), \quad i = 1, \ldots, I. \tag{7}$$

In practice, however, the parameter vector θ is unknown and usually estimated by (asymptotically) fully efficient methods: maximum likelihood (ML) and restricted maximum likelihood (REML). Once the estimates $\hat{\theta} = (\hat{\sigma}^2_{\alpha_1}, \ldots, \hat{\sigma}^2_{\alpha_q}, \hat{\sigma}^2_\varepsilon)$ are available, the variance-covariance matrices $\mathbf{V}_i(\theta)$ are estimated by $\mathbf{V}_i(\hat{\theta})$. That is, $\hat{\mathbf{V}}_i = \mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_{i'} + \hat{\mathbf{R}}_i$, where $\hat{\mathbf{D}} = diag(\hat{\sigma}^2_{\alpha_1}, \ldots, \hat{\sigma}^2_{\alpha_q})$ and $\hat{\mathbf{R}}_i = \hat{\sigma}^2_\varepsilon \mathbf{I}_{\tau_i \times \tau_i}$.

It turns out that GLS and BLUP correspond to the classical pricing formulas of credibility theory. Indeed, the minimum mean square error predictor of the random variable $W_i = E(y_{i,\tau_i+1} | \alpha_i) = \mathbf{x}_{i,\tau_i+1}\beta + \mathbf{z}_{i,\tau_i+1}\alpha_i$ is given by the best linear unbiased predictor

$$W_{BLUP,i} = \mathbf{x}_{i,\tau_i+1}\hat{\beta}_{GLS} + \mathbf{z}_{i,\tau_i+1}\hat{\alpha}_{BLUP,i}, \quad i = 1, \ldots, I, \tag{8}$$

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 $\tau_i + 1$. In the actuarial literature, $W_{BLUP,i}$ is used to predict the expected claim size $\mu_{i,\tau_i+1} = E(y_{i\tau_i+1} | \alpha_i)$ of risk i in time $\tau_i + 1$. As is well-known, 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

$$P_i = \zeta_i m + (1 - \zeta_i) M = M + \zeta_i (m - M), \quad i = 1, \ldots, I, \tag{9}$$

where P_i is the credibility premium of risk i, and $0 \le \zeta_i \le 1$ is known as the credibility factor. A comparison of equation (8) with (9) implies that $\mathbf{x}_{i,\tau_i+1}\hat{\beta}_{GLS}$ can be interpreted as estimate of M, and $\mathbf{z}_{i,\tau_i+1}\hat{\alpha}_{BLUP,i}$ as predictor of the weighted, risk-specific deviation $\zeta_i(m - M)$.

A.3. Heavy-tailed mixed linear models

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

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

defined for $-\infty < \lambda_{it} < \infty$, $\sigma_\varepsilon > 0$, and where F_0 is the standard (i.e., $\lambda_{it} = 0$, $\sigma_\varepsilon = 1$, $\upsilon_{it} = 1$) cdf of the underlying location-scale family $F(\lambda_{it}, \sigma^2_\varepsilon/\upsilon_{it})$. Following regression analysis with location-scale models, we include the covariates $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ only through the location parameter λ_{it}. Then, the mixed linear model, given by (5), becomes

$$log(\mathbf{y}_i) = \mathbf{X}_i\beta + \mathbf{Z}_i\alpha_i + \varepsilon_i = \lambda_i + \varepsilon_i \quad i = 1, \ldots, I, \quad (10)$$

where $log(\mathbf{y}_i) = (log(y_{i1}), \ldots, log(y_{i\tau_i}))'$ and λ_i is the τ_i-dimensional vector of the within-subject locations λ_{it} that consist of the *population location* $\beta \in \mathbb{R}^p$ and the subject-specific *location deviation* $\alpha_i \in \mathbb{R}^q$. While assumptions (*b*) and (*c*) in Section 7.2 remain valid for random components, in the mixed linear model given by (10) we replace (*a*) by assumption (*a*′), for which

$$\alpha_i \sim N_q(\mathbf{0}, \mathbf{D}) \quad \textit{and} \quad \varepsilon_i \sim F_{\tau_i}(\mathbf{0}, \mathbf{R}_i), \quad i = 1, \ldots, I,$$

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

$$log(\mathbf{y}_i)|\alpha_i \sim F_{\tau_i}(\mathbf{X}_i\beta + \mathbf{Z}_i\alpha_i, \mathbf{R}_i), \quad i = 1, \ldots, I.$$

Examples of such marginal log-location-scale families F include log-normal, 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 $F_{\tau_i}(\lambda_i, \mathbf{R}_i)$ are the well-known elliptical distributions such as multivariate normal (see Section 7.2) and the heavy-tailed multivariate Student's *t* with ν degrees of freedom.