Robust and Efficient Methods for Credibility When Claims Are Approximately Gamma-Distributed

Harald Dornheim* and Vytaras Brazauskas†

Abstract

As is well known in actuarial practice, excess claims (outliers) have a disturbing effect on the ratemaking process. To obtain better estimators of premiums, which are based on credibility theory, Künsch and Gisler and Reinhard suggested using robust methods. The estimators proposed by these authors are indeed resistant to outliers and serve as an excellent example of how useful robust models can be for insurance pricing. In this article we further refine these procedures by reducing the degree of heuristic arguments they involve. Specifically we develop a class of robust estimators for the credibility premium when claims are *approximately* gamma-distributed and thoroughly study their robustness-efficiency trade-offs in large and small samples. Under specific data-generating scenarios, this approach yields quantitative indices of estimators' strength and weakness, and it allows the actuary (who is typically equipped with information beyond the statistical model) to choose a procedure from a full menu of possibilities. Practical performance of our methods is illustrated under several simulated scenarios and by employing expert judgment.

1. Introduction and Preliminaries

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

Instead of solely focusing on the stability of the premium, the modern and more flexible approach to credibility theory concentrates on finding the most accurate estimate of an insured's pure risk premium. Initial contributions to this area can be traced back to the work of Keffer (1929) and Bailey (1945, 1950). However, it is generally agreed that the systematic development of the field of *greatest accuracy credibility* started in the late 1960s with fundamental contributions by Bühlmann (1967, 1969). Later Bühlmann and Straub (1970) introduced a credibility model as a means to rate reinsurance treaties, which generalized previous results and became the cornerstone of greatest accuracy credibility theory. The model is one of the most frequently applied credibility models in insurance practice, and it enjoys some desirable optimality properties. We will use the Bühlmann-Straub model as a reference for our robust model (defined in Section 2.1).

* Harald Dornheim is a Graduate Teaching Assistant, Department of Mathematical Sciences, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, WI 53201, dornheim@uwm.edu.

† Vytaras Brazauskas, PhD is Associate Professor, Department of Mathematical Sciences, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, WI 53201, vytaras@uwm.edu.

From a practical point of view, the Bühlmann-Straub model requires that unknown model parameters be estimated from the data. Methods of estimation proposed in the actuarial literature include purely parametric and standard non- and semiparametric procedures (see, e.g., Klugman, Panjer, and Willmot 2004, Section 16.5; Young 1997). However, as discussed by Norberg (1979) and Goulet (1998), the main practical problem with these approaches is that they are *nonrobust,* that is, they are sensitive to outliers and/or model misspecification. The problem of robustness in the Bayesian context, where one is concerned with the uncertainty in the prior distribution, is addressed by Young (1998). Another approach toward robustness, proposed by Künsch (1992) and Gisler and Reinhard (1993), is to use procedures that extract information from the majority of the data, then identify outliers (i.e., observations not in the majority) and treat them separately. Antecedents of this approach have been used by Swiss actuaries for the calculation of the pure risk premium in industrial fire insurance (see discussion by Gisler and Reinhard 1993).

The robust estimators proposed by Künsch (1992) and Gisler and Reinhard (1993) are indeed resistant to outliers and serve as an excellent example of how useful robust models can be for insurance pricing. In this article we further refine these procedures by reducing the degree of heuristic arguments they involve. Specifically we develop a class of robust estimators for the credibility premium when claims are *approximately* gamma-distributed and thoroughly study their robustness efficiency trade-offs in large and small samples. Under specific data-generating scenarios, this approach yields quantitative indices of estimators' strength and weakness, and it allows the actuary (who is typically equipped with information beyond the statistical model) to choose a procedure from a full menu of possibilities. For implementation of this approach in practice, expert judgment is also needed but at a different stage. That is, an expert would first have to decide on the level of exposure (of the whole portfolio) to extremes and then recommend a procedure that offers the necessary degree of protection while having high efficiency.

The article is organized as follows. In the remainder of this introduction we present the quantitative tools for investigation of the robustness and efficiency properties of estimators (Section 1.1) and describe the standard Bühlmann-Straub credibility model (Section 1.2). Further, in Section 2 our robust credibility model is developed. This section includes model description, introduction of estimators, and a comprehensive study of their large- and small-sample properties. Practical performance of our methods is illustrated in Section 3. Final discussion is provided in Section 4.

1.1 Robustness versus Efficiency

Simultaneous consideration of robustness and efficiency criteria is a relatively old idea in the statistical literature dating back to the late 1960s. In the actuarial literature, however, it has emerged recently but has already proven very successful in theory and in applications (see, e.g., Brazauskas 2003; Brazauskas and Serfling 2000, 2003; Marceau and Rioux 2001; Serfling 2002). Conceptually this approach is very similar to insurance contract: to get some (specific) amount of protection against damages that may be caused by disastrous events in the future one has to pay a prespecified premium right now. In statistical terms, the protection is understood as no or little effect on the bias and variance of the estimator by a fixed amount of contamination (outliers) in the data. At the same time, we have to accept the fact that such protection does not come free: that is, the estimators designed to resist outliers will sacrifice efficiency relative to performance when there is no contamination in the data. Such a sacrifice is the premium we have to pay.

To formalize these ideas and to quantify the ''robustness versus efficiency'' approach, we will use the following tools of robust statistics. (They are taken, either directly or with some modifications, from the references cited at the beginning of this section.) The most crucial tool among these is the *influence function;* all other quantitative measures stem from it.

- *Influence Function* (IF)

  For a parameter H(F) estimated by $H(\hat{F}_n)$, where $\hat{F}_n$ is an empirical distribution function estimating a distribution function F on the basis of a sample of size n, the associated influence function is defined as $\mathrm{IF}(x; \mathrm{H}(F)) = \partial/\partial s[\mathrm{H}((1 - s)F + s\,\delta_x)]_{s=0^+}$, where δ_x is the distribution placing all mass at the point x. The influence function approximates the contribution to the total estimation error that is made by an observation located at x. That is, for the sample observations $X_1, \ldots, X_n$,

$$\mathrm{H}(\hat{F}_n) - \mathrm{H}(F) \approx \frac{1}{n}\,[\mathrm{IF}(X_1; \mathrm{H}(F)) + \cdots + \mathrm{IF}(X_n; \mathrm{H}(F))].$$

Thus, the impact (or "influence") of observation X_i on the estimation error for a sample of size n is measured, approximately, by $n^{-1}\mathrm{IF}(X_i; \mathrm{H}(F))$.

- *Gross Error Sensitivity* (GES)
The gross error sensitivity (divided by n) measures the worst possible effect on the estimator due to contamination of the data. Since damaging contamination can be caused by low or high observations, we have to consider two versions—upper and lower—of the GES:

$$\mathrm{GES}^+ = \sup_{x\in\Omega^+} |\mathrm{IF}(x; \mathrm{H}(F))| \quad \text{and} \quad \mathrm{GES}^- = \sup_{x\in\Omega^-} |\mathrm{IF}(x; \mathrm{H}(F))|,$$

where Ω^+ (Ω^-) represents a set of upper (lower) x values. Then the most general definition of GES is $\mathrm{GES} = \max\{\mathrm{GES}^+, \mathrm{GES}^-\}$. Estimators with relatively low GES (GES > 0) are desired. Also, depending on the context, one may focus only on estimators with relatively low $\mathrm{GES}^+ > 0$ (or only relatively low $\mathrm{GES}^- > 0$).

- *Breakdown Point* (BP)
The breakdown point provides a guidance up to what distance from the model an approximation based on the IF can be used. In practice, the finite-sample BP is loosely characterized as the largest proportion of corrupted sample observations that the estimator can cope with. Similarly to GES, we have to allow the possibility that data corruption may occur because of low or high observations, and thus define two versions of BP:

> The *upper* (*lower*) breakdown point, denoted by BP^+ (BP^-), is the largest proportion of *upper* (*lower*) sample observations that may be taken to an *upper* (*lower*) limit without taking the estimator to an uninformative limit not depending on the parameter being estimated.

The general definition of BP then is $\mathrm{BP} = \min\{\mathrm{BP}^+, \mathrm{BP}^-\}$. For asymptotic comparisons, the limit (as the sample size tends to ∞) of BP will be used. Interestingly, as we will see later, the asymptotic BP is a function of the underlying IF. Estimators with relatively high BP ($0 < \mathrm{BP} \le 0.50$) are desired. Again, depending on the context, estimators with only good $0 < \mathrm{BP}^+ < 1$ (or $0 < \mathrm{BP}^- < 1$) properties may be preferred.

- *Relative Efficiency* (RE)
Suppose that, for estimation of a parameter $\mathrm{H}(F)$, there exist two competing statistical procedures U_n and V_n, where $n \ge 1$ is the sample size. Then, to compare their performances at the model F, we use the ratio of their *mean squared errors* (MSEs):

$$\mathrm{RE}(V_n, U_n) = \frac{\mathrm{MSE}(U_n)}{\mathrm{MSE}(V_n)} = \frac{[\mathbf{E}(U_n) - \mathrm{H}(F)]^2 + \mathbf{Var}(U_n)}{[\mathbf{E}(V_n) - \mathrm{H}(F)]^2 + \mathbf{Var}(V_n)},$$

which is called the relative efficiency of procedure U_n relative to procedure V_n. For large-sample comparisons, the asymptotic relative efficiency may be defined as the limit of $\mathrm{RE}(V_n, U_n)$ as $n \to \infty$, or, equivalently, since all estimators considered in this article are asymptotically unbiased, as

$$\mathrm{ARE}(V_n, U_n) = \frac{\text{asymptotic } \mathbf{Var}(U_n)}{\text{asymptotic } \mathbf{Var}(V_n)}.$$

As will be seen in Section 2.3.3, the asymptotic variance of an estimator is functionally related to the IF. Also, for our choice of distribution function F (i.e., for a gamma distribution), the standard credibility estimator has optimum asymptotic variance; thus competing (robust) estimators will have $0 < \text{ARE} < 1$ at F.

1.2 Standard Credibility Model

Let us consider a portfolio of different insureds or risks i, $i = 1, \ldots, I$, where each risk i is characterized by an *unobservable* risk parameter θ_i. For $i = 1, \ldots, I$, we have a vector of observations $(X_{i1}, \ldots, X_{i\tau_i})$, where X_{it} represents the *observed* claim amount (or loss ratio) of risk i during time period t, $t = 1, \ldots, \tau_i$. The mathematical assumptions of the Bühlmann-Straub model are the following:

- The hidden risk parameters $\theta_1, \ldots, \theta_I$ are independent and identically distributed
- For $t = 1, \ldots, \tau_i$, the random variables X_{it}, given θ_i, are conditionally independent
- The random variables $X_{i1}, \ldots, X_{i\tau_i}$ have finite variances
- For $i = 1, \ldots, I$ and $t = 1, \ldots, \tau_i$: $\mathbf{E}(X_{it}|\theta_i) = \mu(\theta_i)$ and $\mathbf{Var}(X_{it}/\theta_i) = v(\theta_i)/w_{it}$, where w_{it} are known volume measures.

Additionally, to simplify theoretical derivations of Section 2, we will slightly modify the last assumption by considering the case that volume measures of risk i do not change over time:

- For $i = 1, \ldots, I$: $w_{i1} = \cdots = w_{i\tau_i} \equiv w_i$.

Of course, this somewhat reduces the generality of the model. From a practical point of view, however, it is not too restrictive of an assumption because, as one referee aptly commented, ''volumes are often equal enough across periods for a single risk to be considered constant in time. Indeed, volumes vary usually way more between than within risks.''

The risk premium $\mu(\theta_i)$ is the *true* premium for an insured i if its risk parameter θ_i were known. Since in practice θ_i is mostly unknown, the true premium is estimated by the credibility estimator $\hat{\mu}(\theta_i)$. It follows from the Bühlmann-Straub model that the best (in the MSE sense) *linear Bayesian* credibility estimator is given by

$$\hat{\mu}(\theta_i) = (1 - \alpha_i)\mu + \alpha_i \bar{X}_i, \tag{1.1}$$

where $\mu = \mathbf{E}(\mu(\theta_i)) = \mathbf{E}(X_{it})$ is the overall mean or collective premium charged for the whole portfolio, $\bar{X}_i = w_{i\bullet}^{-1} \sum_{t=1}^{\tau_i} w_{it} X_{it} \equiv \tau_i^{-1} \sum_{t=1}^{\tau_i} X_{it}$ is the weighted mean of the individual experience of risk i, $\alpha_i = (1 + v/(w_{i\bullet}\sigma^2))^{-1} \equiv (1 + v/(\tau_i w_i \sigma^2))^{-1}$ is the credibility factor of risk i, and $w_{i\bullet} = \sum_{t=1}^{\tau_i} w_{it} \equiv \tau_i w_i$ is the total volume of risk i. Also, $v = \mathbf{E}(v(\theta_i))$ and $\sigma^2 = \mathbf{Var}(\mu(\theta_i))$. Here parameters μ, v, and σ^2 are called the *structural parameters* that are generally unknown and must be estimated from the data. (Estimators for these parameters are presented in Section 3.1.)

2. ROBUST CREDIBILITY MODEL

In this section we develop our robust credibility model. First, in Section 2.1 we present motivation for, and description of, the model. Then in Section 2.2 we introduce a class of robust-efficient estimators for summarizing the individual experience of an insured. Large-sample properties of these estimators are studied in Section 2.3. Finally, since in typical insurance portfolios the number of observation periods for a risk is not large, the estimators have to be corrected for (approximate) unbiasedness in small samples. In Section 2.4 we derive the necessary small-sample adjustments for the estimators, thus making the robust model more realistic.

2.1 Model Description

Let us continue with the same setup as in Section 1.2. The main idea of the robust credibility approach is to divide the true risk premium $\mu(\theta_i)$ in the Bühlmann-Straub model into two parts—a risk premium for the *ordinary* claims, $\mu_{\text{ordinary}}(\theta_i)$, and a risk premium for the *extraordinary* claims, $\mu_{\text{extra}}(\theta_i)$—and to 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. Further, as Gisler and Reinhard (1993) argue, "the bulk of the data contains very little information with respect to 'outlier-events'" and therefore "all risks in the portfolio can be considered as equally exposed to outlier events." Thus, it is reasonable to assume that

$$\mu_{\text{extra}}(\theta_i) \approx \mu_{\text{extra}}, \qquad \text{for } i = 1, \ldots, I.$$

In this setting the ordinary premium $\mu_{\text{ordinary}}(\theta_i)$ is estimated using a robust procedure (say, T_i), which automatically identifies what the ordinary observation (say, T_{it}) is, and estimation of the extraordinary premium μ_{extra} is based on the overshot of the excess claims T^*_{it}, $t = 1, \ldots, \tau_i$, $i = 1, \ldots, I$. Then, similar to formula (1.1), the robust credibility estimator is given by

$$\tilde{\mu}(\theta_i) = \hat{\mu}_{\text{ordinary}}(\theta_i) + \mu_{\text{extra}} = [(1 - \beta_i)\mu_{\text{robust}} + \beta_i T_i] + \mu_{\text{extra}}, \tag{2.1}$$

where the robust structural parameters $\mu_{\text{robust}} = \mathbf{E}(T_i)$, $\mu_{\text{extra}} = \mathbf{E}(T^*_{it})$, $v_{\text{robust}} = \mathbf{E}(w_{it}\mathbf{Var}(T_{it}|\theta_i)) \equiv \mathbf{E}(w_i\mathbf{Var}(T_{it}|\theta_i))$, and $\sigma^2_{\text{robust}} = \mathbf{Var}(\mathbf{E}(T_i|\theta_i))$, and the robust credibility factor $\beta_i = (1 + v_{\text{robust}}/(w_{i\cdot}\sigma^2_{\text{robust}}))^{-1} \equiv (1 + v_{\text{robust}}/(\tau_i w_i \sigma^2_{\text{robust}}))^{-1}$ must be estimated from the data. (Estimators for these parameters and definitions of the ordinary observation T_{it} and the overshot of the excess claims T^*_{it} are presented in Section 3.1.)

2.2 Estimators

To select a class of robust estimators and study its properties, we have to make certain distributional assumptions about the claim data X_{it} (given θ_i). Using the motivation provided in examples of Gisler and Reinhard (1993), we assume the following:

For $i = 1, \ldots, I$, conditional variables $X_{i1}|\theta_i, \ldots, X_{i\tau_i}|\theta_i$ are distributed according to the gamma distribution with parameters $w_i\gamma$ and θ_i/w_i, denoted GAMMA $(w_i\gamma, \theta_i/w_i)$ and having the pdf

$$f(x|\theta_i) = \frac{(w_i/\theta_i)^{w_i\gamma}}{\Gamma(w_i\gamma)} x^{w_i\gamma - 1} e^{-xw_i/\theta_i}, \qquad \text{for } x > 0, \tag{2.2}$$

where the shape parameter $\gamma > 0$ and volume measures w_i are known.

Note that the above statement is equivalent to assuming that *all* claims are ordinary, which is necessary for the introduction and development of robust procedures. Once those robust estimators are properly defined, they will remain valid under the less stringent assumption of "the bulk of claims are ordinary" because of the way they are designed. Also, the justification for using the scale parameter θ_i as the risk parameter rather than the shape parameter γ is that, from one risk period to another, the probability of having "small" or "large" claims is the same; what differs is how "small" and "large" claims will be distributed between "good" and "bad" risks.

As discussed in Section 2.1, the role of a robust procedure is to estimate the ordinary premium $\mu_{\text{ordinary}}(\theta_i)$, that is, the mean of ordinary claims, which, in view of (2.2), is equal to $\mathbf{E}(T_i|\theta_i) = \mathbf{E}(\bar{X}_i|\theta_i) = \mathbf{E}(X_{it}|\theta_i) = \gamma\theta_i$. Thus, practically, our aim is to design a robust estimator for the scale parameter θ_i (because γ is known). Following the general guidelines of Huber (1981, Chapter 5) and, in part, some specific suggestions of Künsch (1992) and Gisler and Reinhard (1993), we consider a class of *M*-estimators for scale parameters and choose estimator T_i as a solution of the following equation:

$$\sum_{t=1}^{\tau_i} \psi\left(\frac{X_{it}}{T_i}; c_{\tau_i}\right) = 0, \tag{2.3}$$

where

$$\psi(t; c_{\tau_i}) = \begin{cases} \max\{-a_i, \min\{t - c_{\tau_i}, b_i\}\}, & \text{for } t > 0; \\ 0, & \text{for } t \le 0. \end{cases}$$

Here constants $0 < a_i < 1$ and $b_i > -a_i$ control the level of *lower* and *upper* trimming, respectively, and $c_{\tau_i} \equiv c_{\tau_i}(w_i\gamma, a_i, b_i)$ denotes the so-called *Fisher consistency factor* (see Hampel et al. 1986, p. 83). The case $a_i = 1$ implies that there is no trimming of lowest sample observations and is quite typical for insurance applications, where lower outliers are of no concern. (In this article we will present theoretical results for arbitrary a_i, but quantitative illustrations will be given only for $a_i = 1$.) To find an asymptotic approximation of c_{τ_i}, denoted $c_\infty \equiv c_\infty(w_i\gamma, a_i, b_i)$, one works with the corresponding integral equation:

$$\int \psi\left(\frac{x}{\mathbf{E}(T_i|\theta_i)}; c_\infty\right) f(x|\theta_i)\, dx = 0. \tag{2.4}$$

The estimator T_i, given by equation (2.3), is very similar to the estimators of Künsch (1992, equation (2.7)), and Gisler and Reinhard (1993, equation (33)), yet it is different from both, mostly in the degree of how much heuristics is involved in its definition. For example, the first paper uses the function $\psi(t; 1)$, instead of $\psi(t; c_{\tau_i})$, and recommends to (always) select constant b_i between 1 and 2 by providing intuitive, rather than mathematical (theoretical or empirical), arguments for it. The second paper defines $\psi(t, w_{it}) = w_{it} \min\{t - 1, c/\sqrt{w_{it}}\}$, where constant c is either $\sqrt{w_{\bullet\bullet}/(I\tau)}$ or $\sqrt{\text{median}_{t=1,\ldots,\tau, i=1,\ldots,I}(w_{it})}$ and $\tau \equiv \tau_1 = \cdots = \tau_I$. Here the upper trimming level is a function of the risk volume, which presents challenges in studying estimators' asymptotic behavior. Our approach incorporates volume w_i into the estimator T_i through $c_{\tau_i}(w_i\gamma, a_i, b_i)$. We make recommendations about the choice of b_i after a full picture of robustness efficiency trade-offs is available. Also, and very importantly, it is relatively easy to study asymptotics of our estimator.

Finally, we note that the choice of $a_i = 1$ and $b_i \to \infty$ will imply $c_{\tau_i}(w_i\gamma, a_i = 1, b_i \to \infty) \to 1$ which, in turn, will lead to $T_i \to \bar{X}_i$, the standard estimator of individual experience. Thus, similar to the procedures introduced by the above authors, the proposed class of estimators also includes the standard estimator as a limiting case.

2.3 Large-Sample Properties

Since estimator T_i belongs to the class of *M*-estimators, its large-sample properties (e.g., consistency, asymptotic normality) and large-sample measures (e.g., breakdown point, gross-error sensitivity, asymptotic relative efficiency) can be directly derived from general theory for *M*-estimators, which is available in Huber (1981) or Hampel et al. (1986), for example. Thus, in Sections 2.3.1–2.3.3 we will closely follow these references.

2.3.1 Fisher Consistency Factors

To find asymptotic Fisher consistency factors c_∞, we have to solve equation (2.4). First, by noting that procedure T_i is designed to estimate the average experience of risk i, that is, $\mathbf{E}(T_i|\theta_i) = \mathbf{E}(\bar{X}_i|\theta_i) = \gamma\theta_i$, we introduce the substitution of variables $z = x/(\gamma\theta_i)$, and hence rewrite equation (2.4) as

$$\int \psi(z; c_\infty) f(\gamma\theta_i z|\theta_i)\, dz = 0. \tag{2.5}$$

Next, it is easy to show that $\psi(z; c_\infty) = \psi_1(z)\mathbf{1}\{-b_i < c_\infty \le a_i\} + \psi_2(z)\mathbf{1}\{c_\infty > a_i\}$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function, $\psi_1(z) = z - c_\infty$, for $0 \le z < c_\infty + b_i$, and $\psi_1(z) = b_i$, for $z \ge c_\infty + b_i$, and $\psi_2(z) = -a_i$, for $0 \le z < c_\infty - a_i$, $\psi_2(z) = z - c_\infty$, for $c_\infty - a_i \le z < c_\infty + b_i$, and $\psi_2(z) = b_i$, for $z \ge c_\infty + b_i$. Taking all this together, straightforward integration of equation (2.5) yields that factors c_∞ can be found by solving (numerically) the following equation:

$$\mathbf{1}\{c_\infty > -b_i\}[b_i + \Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma + 1) - (c_\infty + b_i)\Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma)]$$
$$+ \mathbf{1}\{c_\infty > a_i\}[(c_\infty - a_i)\Gamma^*(w_i\gamma(c_\infty - a_i); w_i\gamma) - \Gamma^*(w_i\gamma(c_\infty - a_i); w_i\gamma + 1)] = 0, \quad (2.6)$$

where Γ^* denotes the incomplete gamma function, defined as $\Gamma^*(y; m) = (1/\Gamma(m)) \int_0^y x^{m-1}e^{-x}\,dx$, for $y > 0$, and $\Gamma^*(y; m) = 0$, for $y \leq 0$. Using standard techniques of calculus, it can be shown that the left-hand side of (2.6) is 0, for $c_\infty \leq -b_i$, has a jump (of size b_i) at $c_\infty = -b_i$, and is a continuous and monotonically decreasing function of c_∞ on the interval $(-b_i, \infty)$, which approaches $-a_i$ (< 0) as c_∞ tends to ∞. This implies that, on the interval $(-b_i, \infty)$, equation (2.6) has *no solution,* if $b_i < 0$, and a *unique solution,* if $b_i > 0$. (The case $b_i \to 0$ yields $c_\infty \to 0$.) Hence, from now on we will focus on the most interesting case of $b_i > 0$.

REMARK 1 *Important Special Cases*

For $a_i = 1$ and $b_i > 0$, similar arguments to those above imply that equation (2.6) has a unique solution on the interval (0, 1). Here $c_\infty \to 1$ for $b_i \to \infty$, which corresponds to $T_i \to \bar{X}_i$. □

Table 1 presents numerical solutions of equation (2.6) for various choices of model parameters.

2.3.2 Robustness Measures

As discussed in Section 1.1, the key tool for studying robustness and efficiency properties of an estimator is its influence function. We do not have to compute it by definition because some general results are already known. That is, the influence function of an M-estimator of scale parameter (in our case, $\mathbf{E}(T_i|\theta_i)$) is given by

$$\mathrm{IF}(x; \mathbf{E}(T_i|\theta_i)) = \frac{\psi(x/\mathbf{E}(T_i|\theta_i); c_\infty)\mathbf{E}(T_i|\theta_i)}{\int \psi'(y/\mathbf{E}(T_i|\theta_i); c_\infty)(y/\mathbf{E}(T_i|\theta_i))f(y|\theta_i)\,dy}$$

(see Huber 1981, Section 5.2). Thus, our main task here is to evaluate the integral in the denominator. As we did for the derivations of Section 2.3.1, we express function ψ in terms of ψ_1 and ψ_2, then

Table 1

Values of c_∞, for $a_i = 1$ and Selected $w_i\gamma > 0$ and $b_i > 0$

	$w_i\gamma$									
b_i	1	2	3	4	5	6	7	8	9	10
0.1	0.383	0.551	0.644	0.704	0.746	0.777	0.801	0.821	0.837	0.851
0.2	0.507	0.673	0.757	0.808	0.842	0.867	0.886	0.901	0.912	0.922
0.3	0.589	0.748	0.823	0.867	0.895	0.915	0.929	0.940	0.949	0.956
0.4	0.650	0.801	0.867	0.904	0.928	0.944	0.955	0.963	0.970	0.975
0.5	0.698	0.839	0.899	0.930	0.950	0.962	0.971	0.977	0.982	0.985
0.6	0.738	0.869	0.922	0.949	0.964	0.974	0.981	0.986	0.989	0.992
0.7	0.770	0.893	0.939	0.962	0.975	0.983	0.988	0.991	0.994	0.995
0.8	0.798	0.912	0.953	0.972	0.982	0.988	0.992	0.995	0.996	0.997
0.9	0.821	0.927	0.963	0.979	0.987	0.992	0.995	0.997	0.998	0.999
1.0	0.841	0.939	0.971	0.984	0.991	0.995	0.997	0.998	0.999	0.999
1.1	0.859	0.949	0.977	0.988	0.994	0.997	0.998	0.999	0.999	1.000
1.2	0.874	0.958	0.982	0.991	0.996	0.998	0.999	0.999	1.000	1.000
1.3	0.888	0.965	0.986	0.994	0.997	0.999	0.999	1.000	1.000	1.000
1.4	0.900	0.971	0.989	0.995	0.998	0.999	1.000	1.000	1.000	1.000
1.5	0.910	0.975	0.991	0.997	0.999	0.999	1.000	1.000	1.000	1.000
1.6	0.920	0.979	0.993	0.997	0.999	1.000	1.000	1.000	1.000	1.000
1.7	0.928	0.983	0.995	0.998	0.999	1.000	1.000	1.000	1.000	1.000
1.8	0.935	0.986	0.996	0.999	1.000	1.000	1.000	1.000	1.000	1.000
1.9	0.942	0.988	0.997	0.999	1.000	1.000	1.000	1.000	1.000	1.000
2.0	0.948	0.990	0.997	0.999	1.000	1.000	1.000	1.000	1.000	1.000
3.0	0.981	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5.0	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
∞	1	1	1	1	1	1	1	1	1	1

differentiate it and perform straightforward integration. (Recall that now we are focusing on the case $b_i > 0$, which implies that $c_\infty \in (-b_i, \infty)$.) Hence, we arrive at the following: for the parameter $\mathbf{E}(T_i|\theta_i) = \gamma\theta_i$ estimated by T_i, which is defined by (2.3), the influence function is given by

$$\mathrm{IF}_{T_i}(x;\,\gamma\theta_i) = \gamma\theta_i\,\frac{\psi_1(x/(\gamma\theta_i))\mathbf{1}\{-b_i < c_\infty \le a_i\} + \psi_2(x/(\gamma\theta_i))\mathbf{1}\{c_\infty > a_i\}}{\Gamma^*(w_i\gamma(c_\infty + b_i);\, w_i\gamma + 1) - \mathbf{1}\{c_\infty > a_i\}\Gamma^*(w_i\gamma(c_\infty - a_i);\, w_i\gamma + 1)}, \tag{2.7}$$

where functions ψ_1, ψ_2, and Γ^* are defined as in Section 2.3.1.

Now we are in the position of deriving formulas for the robustness measures. By definition, the gross error sensitivity of estimator T_i is the supremum of $|\mathrm{IF}_{T_i}(x;\,\gamma\theta_i)|$ with respect to x. This is equivalent to finding maximum of the numerator of (2.7). Thus, we have

$$\mathrm{GES}_{T_i} = \gamma\theta_i\,\frac{\max\{c_\infty\,\mathbf{1}\{-b_i < c_\infty \le a_i\} + a_i\,\mathbf{1}\{c_\infty > a_i\},\, b_i\}}{\Gamma^*(w_i\gamma(c_\infty + b_i);\, w_i\gamma + 1) - \mathbf{1}\{c_\infty > a_i\}\Gamma^*(w_i\gamma(c_\infty - a_i);\, w_i\gamma + 1)}. \tag{2.8}$$

REMARK 2 *Important Special Cases*

a. *No lower trimming* ($a_i = 1$). As follows from Remark 1, for $a_i = 1$ and finite $b_i > 0$, $c_\infty \in (0, 1)$. Therefore, (2.8) can be rewritten as $\mathrm{GES}_{T_i} = \gamma\theta_i \max\{c_\infty, b_i\}/\Gamma^*(w_i\gamma(c_\infty + b_i);\, w_i\gamma + 1)$. In the case of no lower trimming, we are only concerned about having protection against large positive outliers (i.e., $x \to \infty$); so, instead of considering GES, we focus on the upper gross error sensitivity GES^+. To this end, $\max\{c_\infty, b_i\}$ is replaced by b_i, and, consequently, we have

$$\mathrm{GES}^+_{T_i} = \gamma\theta_i\,\frac{b_i}{\Gamma^*(w_i\gamma(c_\infty + b_i);\, w_i\gamma + 1)}.$$

b. *Standard estimator* $\bar{X}_i$ ($a_i = 1$, $b_i \to \infty$). Clearly, letting $a_i = 1$ and $b_i \to \infty$ in (2.8), implies that $\mathrm{GES}_{\bar{X}_i} = \infty$; thus, the standard estimator $\bar{X}_i$ is nonrobust with respect to the GES criterion. □

The general expression for the asymptotic breakdown point of an M-estimator of scale is available in Huber (1981, Section 5.2), that is,

$$\mathrm{BP} = \min\{\mathrm{BP}^+, \mathrm{BP}^-\} = \min\left\{\frac{-\psi(0;\, c_\infty)}{\psi(\infty;\, c_\infty) - \psi(0;\, c_\infty)},\, \frac{\psi(\infty;\, c_\infty)}{\psi(\infty;\, c_\infty) - \psi(0;\, c_\infty)}\right\},$$

where function ψ is defined by (2.3). Direct evaluation of these expressions yields

$$\mathrm{BP}^+_{T_i} = \frac{a_i}{a_i + b_i}\,\mathbf{1}\{c_\infty > a_i\} + \frac{c_\infty}{c_\infty + b_i}\,\mathbf{1}\{-b_i < c_\infty \le a_i\}, \tag{2.9}$$

$$\mathrm{BP}^-_{T_i} = \frac{b_i}{a_i + b_i}\,\mathbf{1}\{c_\infty > a_i\} + \frac{b_i}{c_\infty + b_i}\,\mathbf{1}\{-b_i < c_\infty \le a_i\}. \tag{2.10}$$

REMARK 3 *Important Special Cases*

If there is no lower trimming (i.e., $a_i = 1$ and, consequently, $c_\infty \in (0, 1)$), we focus on the BP^+ version, and from (2.9) we get

$$\mathrm{BP}^+_{T_i} = \frac{c_\infty}{c_\infty + b_i}.$$

In addition, it is easily seen that the standard estimator $\bar{X}_i$ (i.e., T_i with $a_i = 1$, $b_i \to \infty$) is nonrobust with respect to the BP criterion because $\mathrm{BP}_{\bar{X}_i} = 0$. □

2.3.3 Asymptotic Relative Efficiency

To find asymptotic relative efficiency of T_i with respect to $\bar{X}_i$, we have to evaluate asymptotic variances of these two estimators. Let us start with estimator T_i. It follows from Huber (1981, Section 3.2) that, under certain regularity conditions (which are satisfied for our choice of the underlying distribution f and function ψ), estimator T_i is asymptotically normal with mean $\gamma\theta_i$ and variance $V(T_i, f)/\tau_i$, where τ_i is the sample size (number of observation periods for a risk i) and

$$V(T_i, f) = \int \mathrm{IF}^2_{T_i}(x; \gamma\theta_i) f(x|\theta_i)\, dx$$

with function IF_{T_i} defined in (2.7). Next, straightforward (but lengthy) integration leads to

$$\begin{aligned} V(T_i, f) = & \left(\frac{\gamma\theta_i}{\Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma + 1) - \mathbf{1}\{c_\infty > a_i\}\Gamma^*(w_i\gamma(c_\infty - a_i); w_i\gamma + 1)} \right)^2 \\ & \times \Bigg\{ b_i^2 + \frac{w_i\gamma + 1}{w_i\gamma} \Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma + 2) - 2c_\infty\Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma + 1) \\ & + (c_\infty^2 - b_i^2)\Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma) - \mathbf{1}\{c_\infty > a_i\} \Bigg[\frac{w_i\gamma + 1}{w_i\gamma} \Gamma^*(w_i\gamma(c_\infty - a_i); w_i\gamma + 2) \\ & - 2c_\infty\Gamma^*(w_i\gamma(c_\infty - a_i); w_i\gamma + 1) + (c_\infty^2 - a_i^2)\Gamma^*(w_i\gamma(c_\infty - a_i); w_i\gamma) \Bigg] \Bigg\}, \end{aligned} \tag{2.11}$$

where function Γ^* is defined as in Section 2.3.1.

REMARK 4 *Important Special Cases*

a. *No lower trimming* ($a_i = 1$). For $a_i = 1$ and finite $b_i > 0$, $c_\infty \in (0, 1)$. Thus, expression (2.11) can be significantly simplified:

$$\begin{aligned} V(T_i, f) = & \left(\frac{\gamma\theta_i}{\Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma + 1)} \right)^2 \Bigg[b_i^2 + \frac{w_i\gamma + 1}{w_i\gamma} \Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma + 2) \\ & - 2c_\infty\Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma + 1) + (c_\infty^2 - b_i^2)\Gamma^*(w_i\gamma(c_\infty + b_i); w_i\gamma) \Bigg]. \end{aligned}$$

b. *Standard estimator* $\bar{X}_i$ ($a_i = 1$, $b_i \to \infty$). It follows from the central limit theorem that estimator $\bar{X}_i$ is asymptotically normal with mean $\gamma\theta_i$ and variance $V(\bar{X}_i, f)/\tau_i = (\gamma\theta_i^2/w_i)/\tau_i$. Thus it is no surprise that we get $V(T_i, f) \to V(\bar{X}_i, f)$, by letting $a_i = 1$ and $b_i \to \infty$ in (2.11). □

Finally, we have all the necessary components for computation of asymptotic relative efficiency of estimator T_i with respect to $\bar{X}_i$, that is, $\mathrm{ARE}(T_i, \bar{X}_i) = \dfrac{V(\bar{X}_i, f)/\tau_i}{V(T_i, f)/\tau_i} = \dfrac{V(\bar{X}_i, f)}{V(T_i, f)} = \dfrac{\gamma\theta_i^2/w_i}{w_i/V(T_i, f)}$.

2.3.4 Summary

In Table 2 we provide numerical illustrations of (asymptotic) robustness and efficiency properties of estimator T_i for various choices of model parameters.

2.4 Small-Sample Properties

So far we have studied the asymptotic behavior of our robust estimators and have a pretty good idea of how each estimator performs under specified scenarios (i.e., for specific (known) values of $w_i\gamma$). These properties, however, are valid for large number of observation periods, for instance, $\tau_i = 30$ or 50 years, and have some bias for $\tau_i \leq 10$ (a realistic span of individual experience for a risk). Since for insurance applications unbiasedness is indispensable, small-sample corrections for the asymptotic properties of the estimators are necessary.

Table 2

Values of $BP^+_{T_i}$, $GES^* = GES^+_{T_i}/(\gamma\theta_i)$, $ARE(T_i, \bar{X}_i)$, for $a_i = 1$ and Selected $w_i\gamma > 0$ and $b_i > 0$

b_i	BP$^+$	GES*	ARE	BP$^+$	GES*	ARE	BP$^+$	GES*	ARE
	$w_i\gamma = 1$			$w_i\gamma = 2$			$w_i\gamma = 3$		
0.2	0.717	0.063	0.420	0.771	0.102	0.523	0.791	0.130	0.589
0.4	0.619	0.226	0.560	0.667	0.345	0.684	0.684	0.422	0.755
0.6	0.551	0.464	0.652	0.592	0.676	0.780	0.606	0.802	0.847
0.8	0.499	0.759	0.720	0.533	1.064	0.844	0.544	1.231	0.903
1.0	0.457	1.098	0.772	0.484	1.487	0.889	0.493	1.682	0.938
1.2	0.422	1.473	0.814	0.444	1.931	0.920	0.450	2.139	0.961
1.4	0.391	1.873	0.847	0.409	2.385	0.943	0.414	2.594	0.976
1.6	0.365	2.293	0.874	0.380	2.842	0.959	0.383	3.043	0.985
1.8	0.342	2.728	0.896	0.354	3.297	0.971	0.356	3.483	0.991
2.0	0.322	3.172	0.914	0.331	3.749	0.979	0.333	3.915	0.994
5.0	0.167	9.826	0.995	0.167	9.995	0.999	0.167	9.999	1.000
	$w_i\gamma = 4$			$w_i\gamma = 5$			$w_i\gamma = 6$		
0.2	0.802	0.151	0.636	0.808	0.169	0.673	0.813	0.183	0.703
0.4	0.693	0.478	0.803	0.698	0.521	0.837	0.702	0.555	0.863
0.6	0.613	0.888	0.888	0.617	0.950	0.915	0.619	0.997	0.935
0.8	0.549	1.336	0.936	0.551	1.406	0.956	0.553	1.455	0.969
1.0	0.496	1.794	0.964	0.498	1.862	0.977	0.498	1.907	0.986
1.2	0.452	2.248	0.980	0.454	2.309	0.989	0.454	2.344	0.994
1.4	0.416	2.693	0.988	0.416	2.743	0.995	0.416	2.768	0.997
1.6	0.384	3.127	0.994	0.384	3.166	0.997	0.385	3.183	0.998
1.8	0.357	3.552	0.997	0.357	3.580	0.999	0.357	3.592	0.999
2.0	0.333	3.970	0.998	0.333	3.989	0.999	0.333	3.996	1.000
5.0	0.167	10.000	1.000	0.167	10.000	1.000	0.167	10.000	1.000
	$w_i\gamma = 7$			$w_i\gamma = 8$			$w_i\gamma = 9$		
0.2	0.816	0.196	0.727	0.818	0.207	0.748	0.820	0.217	0.766
0.4	0.705	0.584	0.883	0.707	0.608	0.899	0.708	0.628	0.912
0.6	0.621	1.033	0.949	0.622	1.062	0.959	0.622	1.085	0.967
0.8	0.554	1.491	0.978	0.554	1.517	0.984	0.555	1.536	0.988
1.0	0.499	1.936	0.991	0.500	1.956	0.994	0.500	1.969	0.996
1.2	0.454	2.366	0.996	0.454	2.379	0.998	0.454	2.387	0.999
1.4	0.416	2.783	0.999	0.417	2.790	0.999	0.416	2.795	1.000
1.6	0.385	3.192	1.000	0.385	3.196	1.000	0.385	3.198	1.000
1.8	0.357	3.596	1.000	0.357	3.598	1.000	0.357	3.599	1.000
2.0	0.333	3.998	1.000	0.333	3.999	1.000	0.333	4.000	1.000
5.0	0.167	10.000	1.000	0.167	10.000	1.000	0.167	10.000	1.000
∞	0	∞	1	0	∞	1	0	∞	1

In this context the most important (and perhaps the most difficult) problem is to find Fisher consistency factors $c_{\tau_i}(w_i\gamma,\ a_i = 1,\ b_i)$. It is infeasible to determine these factors analytically. However, using Monte Carlo simulations, regression fitting, and numerous attempts of trial and error, we were able to find a simple (yet remarkably accurate) approximation:

$$c_{\tau_i}(w_i\gamma,\ a_i = 1,\ b_i) \approx \frac{1}{\tau_i} + \left(1 - \frac{1}{\tau_i}\right) c_\infty(w_i\gamma,\ a_i = 1,\ b_i), \tag{2.12}$$

where $c_\infty(w_i\gamma,\ a_i = 1,\ b_i)$ is tabulated in Table 1. Note that approximation (2.12) satisfies all known (or, at least easily verifiable) limiting and special cases, for example: $c_{\tau_i=1}(w_i\gamma,\ a_i = 1,\ b_i) \equiv 1$, $c_{\tau_i}(w_i\gamma,\ a_i = 1,\ b_i) \to c_{\tau_i}(w_i\gamma,\ a_i = 1,\ \infty) \equiv 1$ as $b_i \to \infty$, and $c_{\tau_i}(w_i\gamma,\ a_i = 1,\ b_i) \to c_\infty(w_i\gamma,\ a_i = 1,\ b_i)$ as $\tau_i \to \infty$.

To get some idea on how accurate the approximation is, in Table 3 we present (an estimate of) the standardized bias of the estimators that are defined by (2.3) with $c_{\tau_i}(w_i\gamma,\ a_i = 1,\ b_i)$ given by (2.12). The bias is estimated using simulations, and the standardized bias is computed by taking the estimated bias and dividing it by the target parameter $\gamma\theta_i$. As one can see from the table, the standardized bias is small (though not negligible), for $2 \leq \tau_i \leq 5$, and virtually vanishes, for $\tau_i \geq 10$. This behavior is even more evident for larger values of b_i and $w_i\gamma$. Also, note that, while some rows and columns of

Table 3

Values of $[\widehat{E(T_i)} - \gamma\theta_i]/(\gamma\theta_i)$, for $a_i = 1$ and Selected τ_i, $w_i\gamma > 0$, and $b_i > 0$

	τ_i						τ_i					
b_i	2	3	4	5	7	10	2	3	4	5	7	10
0.2	−0.160	−0.163	−0.099	−0.098	−0.074	−0.054	−0.088	−0.079	−0.052	−0.043	−0.031	−0.025
0.4	−0.099	−0.011	−0.038	−0.027	−0.023	−0.019	−0.036	−0.008	−0.013	−0.009	−0.008	−0.006
0.6	−0.028	−0.004	−0.002	−0.001	−0.002	−0.001	0.007	0.002	0.003	0.002	0.002	0.001
0.8	0.051	0.007	0.017	0.010	0.009	0.006	0.031	0.009	0.009	0.006	0.006	0.004
1.0	0.084	0.020	0.022	0.019	0.013	0.008	0.031	0.016	0.012	0.010	0.007	0.006
1.2	0.070	0.039	0.025	0.022	0.015	0.011	0.024	0.019	0.014	0.011	0.008	0.006
1.4	0.055	0.042	0.025	0.023	0.015	0.013	0.017	0.016	0.013	0.009	0.008	0.005
1.6	0.039	0.046	0.029	0.021	0.016	0.010	0.011	0.013	0.010	0.008	0.005	0.004
1.8	0.035	0.044	0.029	0.021	0.014	0.011	0.006	0.008	0.008	0.006	0.006	0.004
2.0	0.027	0.036	0.028	0.021	0.014	0.009	0.006	0.008	0.006	0.004	0.006	0.004
5.0	0.005	0.000	0.002	0.002	0.000	0.003	−0.001	0.001	0.000	0.000	−0.001	0.001
	$w_i\gamma = 3$						$w_i\gamma = 5$					
0.2	−0.055	−0.045	−0.028	−0.024	−0.018	−0.013	−0.035	−0.023	−0.014	−0.012	−0.010	−0.006
0.4	−0.015	−0.003	−0.005	−0.004	−0.004	−0.002	0.001	0.002	0.002	0.001	0.001	0.000
0.6	0.014	0.004	0.005	0.004	0.004	0.002	0.013	0.006	0.004	0.003	0.003	0.002
0.8	0.023	0.009	0.007	0.008	0.006	0.003	0.008	0.005	0.004	0.004	0.003	0.003
1.0	0.014	0.012	0.008	0.008	0.004	0.004	0.005	0.005	0.004	0.003	0.002	0.001
1.2	0.008	0.009	0.008	0.006	0.004	0.003	0.003	0.004	0.003	0.003	0.002	0.000
1.4	0.005	0.006	0.005	0.005	0.004	0.003	0.002	0.001	0.001	0.001	0.001	0.001
1.6	0.003	0.005	0.005	0.004	0.002	0.002	−0.001	−0.001	0.000	0.000	0.001	0.000
1.8	0.003	0.003	0.004	0.003	0.001	0.002	0.002	0.001	0.001	0.000	0.001	0.001
2.0	0.001	0.002	0.002	0.002	0.002	0.000	0.000	−0.001	0.000	0.000	0.000	0.001
5.0	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: For each τ_i, simulations are based on 10,000 samples of the specified size. Standard errors for all entries are between 0.0007 and 0.0085; their estimates are based on 10 simulation runs.

Table 3 show monotonic behavior, the other ones do not. This is due to the nature of approximation (2.12), which does not uniformly underestimate or overestimate the true factor $c_{\tau_i}(w_i\gamma, a_i = 1, b_i)$ with respect to all its variables, but rather tries to capture the overall pattern.

Next, we use the same simulation study to estimate finite-sample relative efficiencies (for definition see Section 1.1). In Table 4 we present REs of estimator T_i (with $c_{\tau_i}(w_i\gamma, a_i = 1, b_i)$ given by (2.12)) with respect to $\bar{X}_i$ for the number of observation periods τ_i ranging from 2 to 10. We notice that, for a fixed sample size τ_i and varying b_i and $w_i\gamma$, REs of estimator T_i follow similar patterns as in the asymptotic case $\tau_i = \infty$, that is, they increase as b_i and $w_i\gamma$ become larger. Also, for fixed b_i and $w_i\gamma$, REs approach corresponding AREs from above, which means that in small samples robust estimators T_i perform even better than asymptotically with respect to the efficiency criterion.

Further, we need to evaluate the finite-sample upper breakdown point BP^+. These BPs will tell us how much protection a particular estimator T_i can provide. We start by recalling that estimator T_i is very similar to the estimator of Künsch (1992, equation (2.7)). Therefore, an equivalent result (with necessary modifications) to Künsch's Lemma 3.2 holds. Specifically we have

$$\mathrm{BP}^+_{T_i} = \tau_i^{-1} \max_{0 \le m \le \tau_i} \left\{ m \text{ (integer)}: m \le \tau_i \frac{c_{\tau_i}}{c_{\tau_i} + b_i} \right\}, \tag{2.13}$$

where c_{τ_i} is given by (2.12). Note that m satisfying (2.13) represents the number of outliers in a sample of size τ_i that estimator T_i can tolerate without breaking down. In Table 5 we provide numerical values of finite-sample BP^+, given by (2.13), for various choices of model parameters.

Finally, having all this information about finite-sample REs and BPs, we need to know how to select an estimator for a particular risk in the portfolio with specified parameters. The strategy is simple. First, an expert opinion has to be obtained regarding a portfolio's exposure to extremes (i.e., the probability that any of the risks can produce an outlier claim). Second, we look for estimators that provide that much protection against extremes (i.e., those that have sufficiently high BP^+). Third,

Table 4

Values of Finite-Sample (Estimated) RE, for $a_i = 1$ and Selected τ_i, $w_i\gamma > 0$, and $b_i > 0$

	τ_i									τ_i								
b_i	2	3	4	5	6	7	8	9	10	2	3	4	5	6	7	8	9	10
	$w_i\gamma = 1$									$w_i\gamma = 2$								
0.2	0.74	0.62	0.59	0.55	0.52	0.51	0.50	0.49	0.48	0.74	0.64	0.62	0.61	0.59	0.58	0.57	0.57	0.56
0.4	0.77	0.71	0.67	0.65	0.63	0.62	0.61	0.61	0.60	0.80	0.78	0.74	0.74	0.73	0.72	0.72	0.71	0.71
0.6	0.79	0.75	0.72	0.71	0.70	0.69	0.69	0.68	0.68	0.86	0.84	0.82	0.81	0.81	0.81	0.80	0.80	0.80
0.8	0.81	0.79	0.78	0.76	0.76	0.75	0.75	0.74	0.74	0.91	0.88	0.87	0.87	0.86	0.86	0.86	0.86	0.86
1.0	0.84	0.82	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.94	0.91	0.91	0.90	0.90	0.90	0.89	0.89	0.89
1.2	0.87	0.84	0.83	0.83	0.83	0.83	0.82	0.82	0.82	0.96	0.94	0.93	0.93	0.93	0.93	0.92	0.92	0.92
1.4	0.90	0.87	0.86	0.86	0.86	0.85	0.85	0.85	0.85	0.97	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.95
1.6	0.92	0.89	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.98	0.97	0.97	0.97	0.96	0.96	0.96	0.96	0.96
1.8	0.93	0.91	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.99	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97
2.0	0.95	0.93	0.92	0.92	0.92	0.91	0.92	0.91	0.91	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98
5.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$w_i\gamma = 3$									$w_i\gamma = 5$								
0.2	0.75	0.68	0.67	0.65	0.64	0.63	0.63	0.62	0.62	0.78	0.74	0.73	0.72	0.71	0.71	0.71	0.70	0.70
0.4	0.84	0.83	0.80	0.80	0.79	0.78	0.78	0.77	0.77	0.89	0.88	0.87	0.86	0.86	0.85	0.85	0.85	0.85
0.6	0.90	0.88	0.88	0.87	0.87	0.86	0.86	0.86	0.86	0.95	0.94	0.93	0.93	0.93	0.92	0.92	0.92	0.92
0.8	0.95	0.93	0.92	0.92	0.91	0.91	0.91	0.91	0.91	0.98	0.97	0.97	0.96	0.96	0.96	0.96	0.96	0.96
1.0	0.97	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98
1.2	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
1.4	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.6	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.8	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: For each τ_i, simulations are based on 10,000 samples of the specified size. Standard errors for all entries are between 0.0001 and 0.0125; their estimates are based on 10 simulation runs.

Table 5

Values of Finite-Sample $BP^{+}_{T_i}$, for $a_i = 1$ and Selected $\tau_i, w_i\gamma > 0$, and $b_i > 0$

	τ_i									τ_i								
b_i	2	3	4	5	6	7	8	9	10	2	3	4	5	6	7	8	9	10
	$w_i\gamma = 1$									$w_i\gamma = 2$								
0.2	1/2	2/3	3/4	3/5	4/6	5/7	5/8	6/9	7/10	1/2	2/3	3/4	3/5	4/6	5/7	6/8	7/9	7/10
0.4	1/2	1/3	2/4	3/5	3/6	4/7	5/8	5/9	6/10	1/2	2/3	2/4	3/5	4/6	4/7	5/8	6/9	6/10
0.6	1/2	1/3	2/4	2/5	3/6	3/7	4/8	5/9	5/10	1/2	1/3	2/4	2/5	3/6	4/7	4/8	5/9	5/10
0.8	1/2	1/3	2/4	2/5	3/6	3/7	4/8	4/9	5/10	1/2	1/3	2/4	2/5	3/6	3/7	4/8	4/9	5/10
1.0	0	1/3	1/4	2/5	2/6	3/7	3/8	4/9	4/10	0	1/3	1/4	2/5	2/6	3/7	3/8	4/9	4/10
1.2	0	1/3	1/4	2/5	2/6	2/7	3/8	3/9	4/10	0	1/3	1/4	2/5	2/6	3/7	3/8	4/9	4/10
1.4	0	1/3	1/4	1/5	2/6	2/7	3/8	3/9	3/10	0	1/3	1/4	2/5	2/6	2/7	3/8	3/9	4/10
1.6	0	1/3	1/4	1/5	2/6	2/7	2/8	3/9	3/10	0	1/3	1/4	1/5	2/6	2/7	3/8	3/9	3/10
1.8	0	1/3	1/4	1/5	2/6	2/7	2/8	3/9	3/10	0	1/3	1/4	1/5	2/6	2/7	2/8	3/9	3/10
2.0	0	0	1/4	1/5	1/6	2/7	2/8	2/9	3/10	0	0	1/4	1/5	1/6	2/7	2/8	2/9	3/10
5.0	0	0	0	0	0	1/7	1/8	1/9	1/10	0	0	0	0	0	1/7	1/8	1/9	1/10
	$w_i\gamma = 3$									$w_i\gamma = 5$								
0.2	1/2	2/3	3/4	4/5	4/6	5/7	6/8	7/9	7/10	1/2	2/3	3/4	4/5	4/6	5/7	6/8	7/9	8/10
0.4	1/2	2/3	2/4	3/5	4/6	4/7	5/8	6/9	6/10	1/2	2/3	2/4	3/5	4/6	4/7	5/8	6/9	7/10
0.6	1/2	1/3	2/4	3/5	3/6	4/7	4/8	5/9	6/10	1/2	1/3	2/4	3/5	3/6	4/7	4/8	5/9	6/10
0.8	1/2	1/3	2/4	2/5	3/6	3/7	4/8	4/9	5/10	1/2	1/3	2/4	2/5	3/6	3/7	4/8	4/9	5/10
1.0	0	1/3	1/4	2/5	2/6	3/7	3/8	4/9	4/10	0	1/3	1/4	2/5	2/6	3/7	3/8	4/9	4/10
1.2	0	1/3	1/4	2/5	2/6	3/7	3/8	4/9	4/10	0	1/3	1/4	2/5	2/6	3/7	3/8	4/9	4/10
1.4	0	1/3	1/4	2/5	2/6	2/7	3/8	3/9	4/10	0	1/3	1/4	2/5	2/6	2/7	3/8	3/9	4/10
1.6	0	1/3	1/4	1/5	2/6	2/7	3/8	3/9	3/10	0	1/3	1/4	1/5	2/6	2/7	3/8	3/9	3/10
1.8	0	1/3	1/4	1/5	2/6	2/7	2/8	3/9	3/10	0	1/3	1/4	1/5	2/6	2/7	2/8	3/9	3/10
2.0	0	0	1/4	1/5	1/6	2/7	2/8	2/9	3/10	0	1/3	1/4	1/5	2/6	2/7	2/8	3/9	3/10
5.0	0	0	0	0	1/6	1/7	1/8	1/9	1/10	0	0	0	0	1/6	1/7	1/8	1/9	1/10

among these sufficiently robust estimators, we choose the one with the highest efficiency. To get a better picture of how this strategy works, let us take a look at a specific example.

EXAMPLE 1 *Selection of an Appropriate Robust Estimator*

Suppose we have a portfolio of risks with claims following GAMMA($w_i\gamma$, θ_i/w_i) distribution, where parameter $\gamma = 1$. Our objective is to estimate the credibility premium for a risk i that has volumes $w_i = 1$ and has been observed for five years. (This implies that $w_i\gamma = 1$ and $\tau_i = 5$.) An expert evaluates that portfolio's exposure to extremes is 0.10. How many claims of this risk during the five years of experience can (potentially) be outliers? Although this question cannot be answered with 100% certainty, we can answer it statistically by employing a binomial model and by setting a very high standard for error. That is, we have five trials with the probability of success 0.10; therefore, as follows from the binomial model with parameters 5 and 0.10, the probability that there are at most two outliers is 0.991. (The chances that protection against two outliers in a sample of five observations will *not* be sufficient are only 0.009.) From Tables 5 and 4, we see that estimators T_i with $b_i \leq 1.2$ possess $\mathrm{BP}^+ \geq 2/5$ and that the most efficient among these is T_i with $b_i = 1.2$, which has RE = 0.83. Hence, our choice is estimator T_i with $b_i = 1.2$. □

3. PRACTICAL ISSUES AND EXAMPLES

In this section we first fill in the necessary details for the standard and robust credibility models to be applied in practice, that is, we present estimators of the structural parameters. Then we introduce a contamination model that will allow us to generate claims that are *approximately* gamma-distributed. Finally, we study practical performance of our methods under several data-generating scenarios.

3.1 Estimation of Structural Parameters

To apply the credibility models of Sections 1.2 and 2.1 in practice, one needs to estimate the structural parameters. In the standard credibility model, the following (nonparametric) estimators of μ, v, and σ^2 are typically used:

$$\hat{\mu} = \frac{\sum_{i=1}^{I} \hat{\alpha}_i \bar{X}_i}{\sum_{i=1}^{I} \hat{\alpha}_i}, \tag{3.1}$$

$$\hat{v} = \frac{\sum_{i=1}^{I} \sum_{t=1}^{\tau_i} w_{it}(X_{it} - \bar{X}_i)^2}{\sum_{i=1}^{I} (\tau_i - 1)} \equiv \frac{\sum_{i=1}^{I} w_i \sum_{t=1}^{\tau_i} (X_{it} - \bar{X}_i)^2}{\sum_{i=1}^{I} (\tau_i - 1)}, \tag{3.2}$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{w_{\bullet\bullet}}{w_{\bullet\bullet}^2 - \sum_{i=1}^{I} w_{i\bullet}^2} \left(\sum_{i=1}^{I} w_{i\bullet}(\bar{X}_i - \bar{X})^2 - \hat{v}(I - 1) \right) \\ &\equiv \frac{\sum_{i=1}^{I} \tau_i w_i}{(\sum_{i=1}^{I} \tau_i w_i)^2 - \sum_{i=1}^{I} (\tau_i w_i)^2} \left(\sum_{i=1}^{I} \tau_i w_i(\bar{X}_i - \bar{X})^2 - \hat{v}(I - 1) \right), \end{aligned} \tag{3.3}$$

where $\hat{\alpha}_i = (1 + \hat{v}/(w_{i\bullet}\hat{\sigma}^2))^{-1} \equiv (1 + \hat{v}/(\tau_i w_i \hat{\sigma}^2))^{-1}$, $\bar{X} = w_{\bullet\bullet}^{-1} \sum_{i=1}^{I} w_{i\bullet}\bar{X}_i = w_{\bullet\bullet}^{-1} \sum_{i=1}^{I} w_i \sum_{t=1}^{\tau_i} X_{it}$, $w_{i\bullet} = \sum_{t=1}^{\tau_i} w_{it} \equiv \tau_i w_i$, $w_{\bullet\bullet} = \sum_{i=1}^{I} \sum_{t=1}^{\tau_i} w_{it} \equiv \sum_{i=1}^{I} \tau_i w_i$. Due to the subtraction in (3.3), it is possible that $\hat{\sigma}^2$ could be negative. In such a case, it is customary to set $\hat{\sigma}^2 = 0$, which also implies $\hat{\alpha}_i = 0$. (If $\hat{\alpha}_i = \cdots = \hat{\alpha}_I = 0$, then $\hat{\mu} = I^{-1} \sum_{i=1}^{I} \bar{X}_i$.) For derivation of formulas (3.1)–(3.3) and further discussion, see Klugman, Panjer, and Willmot (2004, Section 16.5) and Goulet (1998).

To derive estimators for the robust structural parameters, we first have to understand how the procedure T_i defines the ordinary claim T_{it}. This is transparent from the definition of function ψ and its decomposition in terms of ψ_1 and ψ_2 (see Section 2.3.1); the ordinary claim is

$$T_{it} = T_{it}^{(1)} \, \mathbf{1}\{-b_i < c_{\tau_i} \leq a_i\} + T_{it}^{(2)} \, \mathbf{1}\{c_{\tau_i} > a_i\}, \tag{3.4}$$

where $T_{it}^{(1)} = X_{it}$, for $0 \le X_{it} < T_i(c_{\tau_i} + b_i)$, and $T_{it}^{(1)} = T_i(c_{\tau_i} + b_i)$, for $X_{it} \ge T_i(c_{\tau_i} + b_i)$; and $T_{it}^{(2)} = T_i(c_{\tau_i} - a_i)$, for $0 \le X_{it} < T_i(c_{\tau_i} - a_i)$, $T_{it}^{(2)} = X_{it}$, for $T_i(c_{\tau_i} - a_i) \le X_{it} < T_i(c_{\tau_i} + b_i)$, and $T_{it}^{(2)} = T_i(c_{\tau_i} + b_i)$, for $X_{it} \ge T_i(c_{\tau_i} + b_i)$.

REMARK 5 *Important Special Cases*

a. *No lower trimming* ($a_i = 1$). For $a_i = 1$ and finite $b_i > 0$, we have $c_{\tau_i} \in (0, 1)$. Consequently expression (3.4) for the ordinary claim reduces to

$$T_{it} = T_{it}^{(1)} = \begin{cases} X_{it}, \; 0 \le X_{it} < T_i(c_{\tau_i} + b_i); \\ T_i(c_{\tau_i} + b_i), \; X_{it} \ge T_i(c_{\tau_i} + b_i). \end{cases}$$

b. *Standard estimator* $\bar{X}_i$ ($a_i = 1$, $b_i \to \infty$). In this case T_{it} is simply X_{it} for $0 \le X_{it} < \infty$. □

Next, to get additional clues, let us see what the robust structural parameters represent at the assumed GAMMA model. Robust estimators T_i are designed to be unbiased in the following sense: $\mathbf{E}(T_i|\theta_i) = \mathbf{E}(\bar{X}_i|\theta_i) = \mathbf{E}(X_{it}|\theta_i) = \mu(\theta_i)$, which then implies

$$\mu_{\text{robust}} = \mathbf{E}(T_i) = \mathbf{E}(\mathbf{E}(T_i|\theta_i)) = \mathbf{E}(\mu(\theta_i)) = \mu \quad \text{and} \quad \sigma^2_{\text{robust}} = \mathbf{Var}(\mathbf{E}(T_i|\theta_i)) = \mathbf{Var}(\mu(\theta_i)) = \sigma^2.$$

These observations suggest that for estimation of parameters μ_{robust} and σ^2_{robust} we just have to mimic formulas (3.1) and (3.3), replacing there $\bar{X}_i$ with T_i and other quantities with their robust counterparts. Further, for estimation of parameter v_{robust} additional care has to be taken because, at this moment, it is not obvious how to estimate $\mathbf{Var}(T_{it}|\theta_i)$. Following the steps of Gisler and Reinhard (1993, Section 4.2) we first evaluate $\int \text{IF}^2_{T_i}(x; T_i)\, d\hat{F}_i(x)$, where $\hat{F}_i$ is the empirical distribution function. This (with some necessary modifications relevant to our situation) yields

$$\hat{v}^{(i)}_{\text{robust}} = \frac{w_i}{C_i^2(\tau_i - 1)} \sum_{t=1}^{\tau_i} [(T_{it}^{(1)} - c_{\tau_i}T_i)^2 + \mathbf{1}\{c_{\tau_i} > a_i\}[(T_{it}^{(2)} - c_{\tau_i}T_i)^2 - (T_{it}^{(1)} - c_{\tau_i}T_i)^2]], \tag{3.5}$$

where $C_i = c_{\tau_i} - ((c_{\tau_i} + b_i)/\tau_i) \sum_{t=1}^{\tau_i} \mathbf{1}\{X_{it} \ge T_i(c_{\tau_i} + b_i)\} - \mathbf{1}\{c_{\tau_i} > a_i\}\ ((c_{\tau_i} - a_i)/\tau_i) \sum_{t=1}^{\tau_i} \mathbf{1}\{0 \le X_{it} < T_i(c_{\tau_i} - a_i)\}$. Then, for estimation of v_{robust} we take the weighted average $\hat{v}_{\text{robust}} = \sum_{i=1}^{I} m_i \hat{v}^{(i)}_{\text{robust}}$ with the weights $m_i = C_i^2(\tau_i - 1)/\sum_{j=1}^{I} C_j^2(\tau_j - 1)$. Asymptotically, $\hat{v}_{\text{robust}}$ is an unbiased estimator of v_{robust}. Due to small-sample corrections of T_i, this estimator is *approximately* unbiased in small samples too.

REMARK 6 *Important Special Cases*

a. *No lower trimming* ($a_i = 1$). For $a_i = 1$ and finite $b_i > 0$, we have $c_{\tau_i} \in (0, 1)$. Thus, formula (3.5) becomes much simpler:

$$\hat{v}^{(i)}_{\text{robust}} = \frac{w_i}{C_i^2(\tau_i - 1)} \sum_{t=1}^{\tau_i} (T_{it}^{(1)} - c_{\tau_i}T_i)^2 = \frac{\dfrac{w_i}{\tau_i - 1} \sum_{t=1}^{\tau_i} (T_{it}^{(1)} - c_{\tau_i}T_i)^2}{\left[c_{\tau_i} - \dfrac{c_{\tau_i} + b_i}{\tau_i} \sum_{t=1}^{\tau_i} \mathbf{1}\{X_{it} \ge T_i(c_{\tau_i} + b_i)\}\right]^2}.$$

b. *Standard estimator* $\bar{X}_i$ ($a_i = 1$, $b_i \to \infty$). In this case, $c_{\tau_i} \to 1$, $T_{it} \to X_{it}$, $T_i \to \bar{X}_i$, and $C_i \to 1$. Thus, we have $\hat{v}^{(i)}_{\text{robust}} \to \hat{v}^{(i)} = (w_i/(\tau_i - 1)) \sum_{t=1}^{\tau_i} (X_{it} - \bar{X}_i)^2$. □

Furthermore, focusing only on large outliers, it seems natural to construct an estimator of μ_{extra} using the overshot of excess claims, $X_{it} - T_{it}^{(1)}$. However, at the assumed GAMMA model, we expect (on average) these overshots to be equal to 0, which is not possible unless all $T_{it}^{(1)} = X_{it}$ (see Remark 5a). Therefore, we have to evaluate and subtract the bias (from $X_{it} - T_{it}^{(1)}$):

$$\begin{aligned}\mathbf{E}(X_{it} - T_{it}^{(1)}|\theta_i) &= \mathbf{E}[\mathbf{1}\{X_{it} \geq T_i(c_{\tau_i} + b_i)\}[X_{it} - T_i(c_{\tau_i} + b_i)]|\theta_i] \\ &\overset{\text{for } \tau_i \text{ large}}{\approx} \mathbf{E}[\mathbf{1}\{X_{it} \geq \gamma\theta_i(c_\infty + b_i)\}[X_{it} - \gamma\theta_i(c_\infty + b_i)]|\theta_i] \\ &= \gamma\theta_i K_{w_i\gamma}(c_\infty + b_i) > 0,\end{aligned}$$

where $K_{w_i\gamma}(c_\infty + b_i) = \dfrac{e^{-w_i\gamma(c_\infty+b_i)}}{w_i\gamma} \sum_{j=0}^{w_i\gamma-1} \sum_{k=0}^{j} \dfrac{[w_i\gamma(c_\infty + b_i)]^k}{k!}$, with the assumption that $w_i\gamma \geq 1$ is an integer. In view of this, we replace $\gamma\theta_i$ with T_i, c_∞ with c_{τ_i} and define the (approximately unbiased) overshot of excess claims as $T_{it}^* = X_{it} - T_{it}^{(1)} - T_i K_{w_i\gamma}(c_{\tau_i} + b_i)$.

To summarize all this, we will use the following estimators of the robust structural parameters μ_{extra}, v_{robust}, μ_{robust}, and σ^2_{robust} (for the case of no lower trimming, i.e., $a_i = 1$):

$$\hat{\mu}_{\text{extra}} = w_{\bullet\bullet}^{-1} \sum_{i=1}^{I} w_i \sum_{t=1}^{\tau_i} T_{it}^*,$$

$$\hat{\mu}_{\text{robust}} = \frac{\sum_{i=1}^{I} \hat{\beta}_i T_i}{\sum_{i=1}^{I} \hat{\beta}_i}, \tag{3.6}$$

$$\hat{v}_{\text{robust}} = \frac{\sum_{i=1}^{I} w_i \sum_{t=1}^{\tau_i} (T_{it} - c_{\tau_i} T_i)^2}{\sum_{i=1}^{I} C_i^2(\tau_i - 1)}, \tag{3.7}$$

$$\hat{\sigma}^2_{\text{robust}} = \frac{w_{\bullet\bullet}}{w_{\bullet\bullet}^2 - \sum_{i=1}^{I} w_{i\bullet}^2} \left(\sum_{i=1}^{I} w_{i\bullet}(T_i - \bar{T})^2 - \hat{v}_{\text{robust}}(I - 1) \right), \tag{3.8}$$

where $T_{it}^* = X_{it} - T_{it}^{(1)} - T_i K_{w_i\gamma}(c_{\tau_i} + b_i)$ with $K_{w_i\gamma}(c_{\tau_i} + b_i) = e^{-w_i\gamma(c_{\tau_i}+b_i)}/w_i\gamma \sum_{j=0}^{w_i\gamma-1} \sum_{k=0}^{j} [w_i\gamma(c_{\tau_i} + b_i)]^k/k!$, $\hat{\beta}_i = (1 + \hat{v}_{\text{robust}}/(w_{i\bullet}\hat{\sigma}^2_{\text{robust}}))^{-1} \equiv (1 + \hat{v}_{\text{robust}}/(\tau_i w_i \hat{\sigma}^2_{\text{robust}}))^{-1}$, $C_i = c_{\tau_i} - c_{\tau_i} + b_i/\tau_i \sum_{t=1}^{\tau_i} \mathbf{1}\{X_{it} \geq T_i(c_{\tau_i} + b_i)\}$, and $\bar{T} = w_{\bullet\bullet}^{-1} \sum_{i=1}^{I} w_{i\bullet} T_i$. Note that, as $b_i \to \infty$, we have $\hat{\mu}_{\text{extra}} \to 0$, $\hat{\mu}_{\text{robust}} \to \hat{\mu}$, $\hat{v}_{\text{robust}} \to \hat{v}$, $\hat{\sigma}^2_{\text{robust}} \to \hat{\sigma}^2$. Also note that σ^2_{robust} could be estimated with $\tilde{\sigma}^2_{\text{robust}} = (I - 1)^{-1} \sum_{i=1}^{I} \hat{\beta}_i (T_i - \hat{\mu}_{\text{robust}})^2$, which is a robust version of the Bichsel-Straub estimator (see Goulet 1998, p. 36). The latter estimator, however, is an iterative estimator (via the credibility factors $\hat{\beta}_i$) and perhaps computationally intensive. At this point we prefer to use an estimator of σ^2_{robust} that is given by an explicit formula. Finally, when applying the estimators (3.6)–(3.8) in practice, one should bear in mind that they will inherit the robustness properties of the "weakest" (least robust) estimator among $T_1, \ldots, T_I$.

3.2 Contamination Model

To study the performance of our methods via simulations, we need a model that would allow us to generate claims (or loss ratios) from an approximate gamma distribution. This can be accomplished by employing an ε-contamination model:

$$G_{F_i,\varepsilon} = (1 - \varepsilon)F_i + \varepsilon F_{\text{entm},i}, \tag{3.9}$$

where F_i is the assumed model, that is, GAMMA($w_i\gamma$, θ_i/w_i), $F_{\text{entm},i}$ is a "contaminating" distribution (or a mixture of distributions) that generates outliers, and ε represents the probability that a sample observation comes from the distribution $F_{\text{entm},i}$ instead of F_i. For $\varepsilon = 0$, family $G_{F_i,\varepsilon}$ generates exact GAMMA data and, for $\varepsilon > 0$, it generates approximate (or "contaminated") GAMMA data.

For the contaminating distribution $F_{\text{entm},i}$ in equation (3.9), we choose the *uniform distribution* on interval (a_i, b_i), denoted $U(a_i, b_i)$, with the probability density function given by $f_{\text{entm},i}(x) = 1/(b_i - a_i)$, for $a_i < x < b_i$, and $= 0$, elsewhere. Parameters a_i and b_i are selected as follows:

$$a_i = 3\mathbf{E}(X_{it}|\theta_i) = 3\gamma\theta_i \quad \text{and} \quad b_i = 7\mathbf{E}(X_{it}|\theta_i) = 7\gamma\theta_i. \tag{3.10}$$

There are countless possibilities to contaminate the central model in equation (3.9). The choice of $U(a_i, b_i)$ is simple and reflects what one would encounter in practice. For example, insurance portfolios typically generate claims, most of which are relatively small and a few are very large; hence, the chosen uniform distribution ensures that a small fraction of large claims consistently appear in generated data sets of our study. However, data from the $U(a_i, b_i)$ distribution are not necessarily just the largest observations in a sample, they blend in with the genuine GAMMA observations, which makes it impossible to (consistently) distinguish outliers from the representative data. The rationale for choosing a_i and b_i according to (3.9) is the following: Parameter a_i represents a threshold that can be exceeded by the assumed GAMMA($w_i\gamma$, θ_i/w_i) variable with the probability 0.0498 (for $w_i\gamma = 1$), 0.0062 (for $w_i\gamma = 3$), 0.0009 (for $w_i\gamma = 5$); and parameter b_i ensures that the expected value of $F_{\text{cntm},i}$ is 5 times $\mathbf{E}(X_{it}|\theta_i)$, which is a reasonable factor. This explains why $U(a_i, b_i)$ data are quite likely to appear as generated by GAMMA($w_i\gamma$, θ_i/w_i).

3.3 Numerical Illustrations

In this section we illustrate how the proposed new methods work in practice and how they compare with the existing methodology of Gisler and Reinhard (1993). For a fixed individual risk, an estimate of the credibility premium is used for estimation of the true underlying premium. Then, having estimates of the true premium for each risk in the portfolio, we evaluate the performance of the method used to estimate the credibility premium by measuring the mean squared error (MSE) of estimates for the entire portfolio. Such performance studies are done under several data-generating scenarios and by employing expert judgment. In this context, the so-called experts are intended to symbolize the level of the actuary's prior knowledge and/or skill. Thus, our overall goal is to identify the best performing procedure/expert under several scenarios. (Here "best" means "best in the long run, on average.") Below are the specific settings of the study.

- *Portfolio Structure*
  We generate a portfolio of $I = 45$ risks. It contains 15 risks with the observation period of $\tau_1 = \cdots = \tau_{15} = 2$ years of experience, 15 risks with $\tau_{16} = \cdots = \tau_{30} = 5$ years of experience, and 15 risks with $\tau_{31} = \cdots = \tau_{45} = 10$ years of experience. In each group of 15 risks, there are five risks with volumes $w_{i_1} = \cdots = w_{i_5} = 1$ (small volumes), five risks with volumes $w_{i_6} = \cdots = w_{i_{10}} = 3$ (medium volumes), and five risks with volumes $w_{i_{11}} = \cdots = w_{i_{15}} = 5$ (large volumes), where $1 \leq i_1 < i_2 < \cdots < i_{15} \leq 45$.
- *Underlying Scenarios*
  The hidden risk parameters $\theta_1, \ldots, \theta_{45}$ are a random sample generated by GAMMA(5, 1/2) distribution. The loss ratios X_{it}, given θ_i, are generated according to the ε-contamination model (3.9) with $\gamma = 1$ and three levels of contamination: $\varepsilon = 0$ (Scenario 1), $\varepsilon = 0.05$ (Scenario 2), and $\varepsilon = 0.10$ (Scenario 3).
- *Experts*
  We introduce three experts in the study: E1 ("careless"), E2 ("careful"), and E3 ("oracle"). The first expert is a gambler who *always* believes that a portfolio has no exposure to extremes, that is, $\varepsilon = 0$, and therefore always recommends summarizing risk's individual experience with the standard estimator $\bar{X}_i$ (equivalently, T_i with $a_i = 1$ and $b_i \to \infty$). The second expert is a cautious conservative whose philosophy is "better be safe than sorry." He or she always recommends using the most robust estimator available in a particular situation; if there are several such estimators, then he or she chooses the most efficient among those. For example, for a risk with volumes $w_i = 3$ (i.e., $w_i\gamma = 3$) and $\tau_i = 2$ years of experience, this expert would recommend estimator T_i with $a_i = 1$ and $b_i = 0.8$. The third expert has a "special talent" that allows him or her to figure out the true value of ε with 100% accuracy. Having such information, he or she then follows the strategy described in Example 1 to select an appropriate estimator. For instance, for a risk with volumes $w_i = 5$ (i.e., $w_i\gamma = 5$) and $\tau_i = 10$ years of experience, this expert would recommend estimator T_i with $a_i = 1$ and $b_i \to \infty$ (for Scenario 1), $b_i = 2.0$ (for Scenario 2), and $b_i = 1.4$ (for Scenario 3). (In all these cases, the probability that protection against the recommended number of outliers is insufficient is $\leq .01$.) Table 6 presents values of b_i recommended by the three experts for the described portfolio structure and underlying scenarios.

Table 6

Values of b_i Recommended by Experts E1, E2, E3 for Various Portfolio Structures and Underlying Scenarios

		Scenario 1			Scenario 2			Scenario 3		
τ_i	w_i	E1	E2	E3	E1	E2	E3	E1	E2	E3
2	1	∞	0.8	∞	∞	0.8	0.8	∞	0.8	0.8
	3	∞	0.8	∞	∞	0.8	0.8	∞	0.8	0.8
	5	∞	0.8	∞	∞	0.8	0.8	∞	0.8	0.8
5	1	∞	0.4	∞	∞	0.4	1.2	∞	0.4	1.2
	3	∞	0.2	∞	∞	0.2	1.4	∞	0.2	1.4
	5	∞	0.2	∞	∞	0.2	1.4	∞	0.2	1.4
10	1	∞	0.2	∞	∞	0.2	2.0	∞	0.2	1.2
	3	∞	0.2	∞	∞	0.2	2.0	∞	0.2	1.4
	5	∞	0.2	∞	∞	0.2	2.0	∞	0.2	1.4

Of course, one does not have to rely on the choice of estimators presented in the previous sections and could pursue further refinements of b_i, which would *slightly* modify the recommendations of experts E2 and E3. For illustrative purposes, however, this level of accuracy should be sufficient. Also, to compare our procedures/experts with the approach of Gisler and Reinhard (1993), we will include their approach as expert E4.

To get a feel for how things work, we generate a portfolio of 45 risks according to the above described specifications. In Table 7 we provide estimates of structural parameters, credibility weights, and MSEs based on recommendations of experts E1, E2, E3, E4, under Scenario 2. Complete information on the estimation process is available in Appendix, Table 10. There we present the generated portfolio (*loss ratios*), true underlying premiums (*true* $\mu(\theta_i)$), and estimates of individual experience (*estimate* T_i) and credibility premiums (*estimate* $\tilde{\mu}(\theta_i)$) for each risk i, based on recommendations of experts E1, E2, E3, E4, under Scenario 2. The interested reader can easily replicate our computations.

DISCUSSION OF TABLE 7

Scenario 2 represents a 5% contamination of the portfolio; therefore it is not surprising that the robust experts E2, E3, and E4 perform significantly better than the nonrobust E1. For instance, MSE of E1 is 84%, 71%, and 32% larger than that of E2, E4, and E3, respectively. Further, notice that the total

Table 7

Quantities of Interest in the Estimation Process, Based on Recommendations of Experts E1, E2, E3, E4, under Scenario 2

Expert	Structural Parameters		Credibility Weights			MSE
E1	$\hat{\mu} = 2.966$	$\hat{\mu}_{\text{extra}} \equiv 0$	$\hat{\alpha}_1 = 0.04$	$\hat{\alpha}_6 = 0.12$	$\hat{\alpha}_{11} = 0.18$	1.300
	$\hat{v} = 27.409$		$\hat{\alpha}_{16} = 0.10$	$\hat{\alpha}_{21} = 0.25$	$\hat{\alpha}_{26} = 0.35$	
	$\hat{\sigma}^2 = 0.599$		$\hat{\alpha}_{31} = 0.18$	$\hat{\alpha}_{36} = 0.40$	$\hat{\alpha}_{41} = 0.52$	
E2	$\hat{\mu}_{\text{robust}} = 2.483$	$\hat{\mu}_{\text{extra}} = 0.415$	$\hat{\beta}_1 = 0.24$	$\hat{\beta}_6 = 0.49$	$\hat{\beta}_{11} = 0.61$	0.705
	$\hat{v}_{\text{robust}} = 10.794$		$\hat{\beta}_{16} = 0.44$	$\hat{\beta}_{21} = 0.70$	$\hat{\beta}_{26} = 0.80$	
	$\hat{\sigma}^2_{\text{robust}} = 1.698$		$\hat{\beta}_{31} = 0.61$	$\hat{\beta}_{36} = 0.83$	$\hat{\beta}_{41} = 0.89$	
E3	$\hat{\mu}_{\text{robust}} = 2.792$	$\hat{\mu}_{\text{extra}} = 0.146$	$\hat{\beta}_1 = 0.09$	$\hat{\beta}_6 = 0.23$	$\hat{\beta}_{11} = 0.34$	0.982
	$\hat{v}_{\text{robust}} = 17.228$		$\hat{\beta}_{16} = 0.20$	$\hat{\beta}_{21} = 0.43$	$\hat{\beta}_{26} = 0.56$	
	$\hat{\sigma}^2_{\text{robust}} = 0.880$		$\hat{\beta}_{31} = 0.34$	$\hat{\beta}_{36} = 0.60$	$\hat{\beta}_{41} = 0.72$	
E4	$\hat{\mu}_{\text{robust}} = 2.562$	$\hat{\mu}_{\text{extra}} = 0.380$	$\hat{\alpha}_1 = 0.17$	$\hat{\alpha}_6 = 0.37$	$\hat{\alpha}_{11} = 0.50$	0.760
	$\hat{v}_{\text{robust}} = 11.733$		$\hat{\alpha}_{16} = 0.33$	$\hat{\alpha}_{21} = 0.60$	$\hat{\alpha}_{26} = 0.71$	
	$\hat{\sigma}^2_{\text{robust}} = 1.163$		$\hat{\alpha}_{31} = 0.50$	$\hat{\alpha}_{36} = 0.75$	$\hat{\alpha}_{41} = 0.83$	

[a] Expert E4 estimators are robust, but their formulas are different from those of E2 and E3.

estimated premium is very similar for all experts: 2.966 + 0 = 2.966 (E1), 2.483 + 0.415 = 2.898 (E2), 2.792 + 0.146 = 2.938 (E3), 2.562 + 0.380 = 2.942 (E4). Thus, the main reason for the differences in the accuracy of estimation lies in the estimation procedure of the structural parameters v and σ^2, mostly v. While the robust estimators of v remain stable in the presence of outliers, the nonrobust estimator of v (i.e., for E1) gets inflated, distorts estimates of the credibility weights, and, consequently, the whole procedure performs poorly (i.e., has high MSE). Among the robust estimators, we notice that the most robust E2 gives more credibility to individual experience (i.e., has the largest estimates of credibility weights) and yields a higher extraordinary premium than the other two robust estimators. □

To see that this was not an accidental ''success story'' of robust procedures and to establish a pattern showing that some method/expert performs better than others in the long run, we generate 100 portfolios according to the above described specifications and evaluate MSEs of experts E1, E2, E3, E4 under all three scenarios.

Let us start with Scenario 1. Under this scenario, experts E1 and E3 give identical recommendations, thus we have only three different experts, E1(E3), E2, E4. For a fixed portfolio, when we rank expert performances according to MSE, we can observe only 3! = 6 different orderings (we shall call them ''events'') of experts E1(E3), E2, E4. For example, the first event occurs when E1(E3) has the smallest MSE (rank ''1''), E2 has the second smallest MSE (rank ''2''), and E4 has the largest MSE (rank ''3'') among the three experts. Thus, for each of 100 generated portfolios, we record which event occurred and what was the MSE increase of experts with the ranks ''2'' and ''3'' compared to that of expert ranked ''1.'' In Table 8 we summarize the study by reporting the relative frequency of each event (with the corresponding average MSE increase in parentheses), for estimation of true premium.

DISCUSSION OF TABLE 8

As theory predicts, under Scenario 1, expert E1(E3) is optimal, which implies that most of the time (not necessarily *always*) this expert should have the smallest MSE. Our simulation study supports this fact. Indeed, E1(E3) gets rank ''1'' 67% of the time, whereas respective percentages for the other two experts are 7% (for E2) and 26% (for E4). One can also make pairwise comparisons. For example, E1(E3) ''beats'' E2 90% of the time (with the average MSE increase of E2 ranging from 0.149 to 0.347) and E4 72% of the time (with the average MSE increase of E4 ranging from 0.070 to 0.219). □

Next, for the other two scenarios the simulation study is summarized in a similar fashion, except that now we have four different experts, and thus we can observe 4! = 24 different orderings/events. The summary is presented in Table 9.

DISCUSSION OF TABLE 9

Under contamination of Scenario 2, expert E1 is not a top performer anymore, yielding the leader's position to the most robust E2. Indeed, E1 is ranked ''1'' only 14% of the time, whereas E2 43%, E3 22%, and E4 21%. We also arrive at similar conclusions by making pairwise comparisons: that is, we observe the following scores of ''winning'' proportions between the experts: 64 : 36 (E2 vs. E1), 54 : 46

Table 8

Performance of Experts E1(E3), E2, E4, under Scenario 1

Event No.	Expert with Rank "1"	Expert with Rank "2"	Expert with Rank "3"	Relative Frequency (Average MSE Increase Relative to "1")
1	E1(E3)	E2	E4	0.10 (0, 0.149, 0.219)
2	E1(E3)	E4	E2	0.57 (0, 0.070, 0.347)
3	E2	E1(E3)	E4	0.05 (0, 0.032, 0.184)
4	E2	E4	E1(E3)	0.02 (0, 0.063, 0.082)
5	E4	E1(E3)	E2	0.23 (0, 0.046, 0.279)
6	E4	E2	E1(E3)	0.03 (0, 0.195, 0.244)

Table 9

Performance of Experts E1, E2, E3, E4, under Scenarios 2 and 3

Event No.	Expert with Rank "1"	"2"	"3"	"4"	Scenario 2 Relative Frequency (Average MSE Increase Relative to "1")	Scenario 3 Relative Frequency (Average MSE Increase Relative to "1")
1	E1	E2	E3	E4	0.02 (0, 0.010, 0.046, 0.567)	0.04 (0, 0.110, 0.122, 0.398)
2	E1	E2	E4	E3	0 (0,0,0,0)	0 (0,0,0,0)
3	E1	E3	E2	E4	0.06 (0, 0.042, 0.137, 0.635)	0.05 (0, 0.093, 0.150, 0.523)
4	E1	E3	E4	E2	0.03 (0, 0.267, 0.349, 0.504)	0.05 (0, 0.072, 0.148, 0.272)
5	E1	E4	E2	E3	0 (0,0,0,0)	0 (0,0,0,0)
6	E1	E4	E3	E2	0.03 (0, 0.033, 0.092, 0.255)	0.02 (0, 0.081, 0.176, 0.589)
7	E2	E1	E3	E4	0.03 (0, 0.122, 0.156, 0.195)	0.15 (0,0.105, 0.147, 0.583)
8	E2	E1	E4	E3	0.02 (0, 0.156, 0.184, 0.223)	0.03 (0, 0.048, 0.078, 0.109)
9	E2	E3	E1	E4	0.14 (0, 0.157, 0.218, 0.733)	0.19 (0, 0.121, 0.186, 0.481)
10	E2	E3	E4	E1	0.04 (0, 0.080, 0.136, 0.240)	0.06 (0, 0.113, 0.181, 0.245)
11	E2	E4	E1	E3	0.04 (0, 0.187, 0.231, 0.349)	0.04 (0, 0.060, 0.139, 0.168)
12	E2	E4	E3	E1	0.16 (0, 0.128, 0.269, 0.359)	0.10 (0, 0.184, 0.296, 0.396)
13	E3	E1	E2	E4	0.08 (0, 0.055, 0.124, 0.425)	0.05 (0, 0.077, 0.100, 0.426)
14	E3	E1	E4	E2	0.02 (0, 0.075, 0.172, 0.341)	0.01 (0, 0.040, 0.174, 0.287)
15	E3	E2	E1	E4	0.07 (0, 0.059, 0.107, 0.639)	0.07 (0, 0.050, 0.094, 0.820)
16	E3	E2	E4	E1	0 (0,0,0,0)	0.01 (0, 0.054, 0.140, 0.181)
17	E3	E4	E1	E2	0.03 (0, 0.083, 0.104, 0.306)	0 (0,0,0,0)
18	E3	E4	E2	E1	0.02 (0, 0.051, 0.062, 0.291)	0.01 (0, 0.027, 0.040, 0.045)
19	E4	E1	E2	E3	0 (0,0,0,0)	0 (0,0,0,0)
20	E4	E1	E3	E2	0.04 (0, 0.093, 0.110, 0.347)	0 (0,0,0,0)
21	E4	E2	E1	E3	0.03 (0, 0.036, 0.365, 0.446)	0.01 (0, 0.033, 0.047, 0.092)
22	E4	E2	E3	E1	0.06 (0, 0.057, 0.148, 0.196)	0.04 (0, 0.129, 0.239, 0.262)
23	E4	E3	E1	E2	0.05 (0, 0.061, 0.128, 0.279)	0.03 (0, 0.065, 0.154, 0.177)
24	E4	E3	E2	E1	0.03 (0, 0.065, 0.092, 0.118)	0.04 (0, 0.021, 0.064, 0.095)

(E2 vs. E3), 66:34 (E2 vs. E4), 70:30 (E3 vs. E1), 55:45 (E3 vs. E4), 50:50 (E4 vs. E1). Finally, higher level of contamination under Scenario 3 just reinforces the conclusions and highlights the patterns that we found under Scenario 2, that is, expert E2 is even more dominant. The following percentages of rank "1" were observed: E1 16%, E2 57%, E3 15%, E4 22%. The scores based on pairwise comparisons are: 75:25 (E2 vs. E1), 66:34 (E2 vs. E3), 79:21 (E2 vs. E4), 61:39 (E3 vs. E1), 69:31 (E3 vs. E4), 34:66 (E4 vs. E1). □

Taking all the scenarios together, it may seem a bit surprising that the "best" expert under a particular scenario does not "beat" each competitor as frequently as one would hope. For example, the proportion of E1 outperforming E4 is 72:28 (under Scenario 1) and that of E2 outperforming E1 is 64:36 (under Scenario 2) and 75:25 (under Scenario 3). To verify these numbers analytically is hopeless, so we have to rely on simulations. One explanation suggesting that the simulation study is correct is that although performed for substantially different scenarios, all these proportions seem to be in the same range, that is, the "winner" is right about twice as often. Another explanation for seemingly low-winning proportions is that the size of each portfolio is fairly small ($I = 45$ risks).

4. DISCUSSION

In this article, we fully developed a class of robust-efficient estimators for the credibility premium when claims (loss ratios) are approximately gamma-distributed. Large- and small-sample properties of these estimators were studied and then used to select a modeling strategy for insurance portfolio. Practical performance of such procedures was illustrated under several simulated scenarios and by employing expert judgment. We found that although the expert's opinion is important, it is not crucial, as long as the expert is first safety oriented and recommends conservatively: that is, except for the "clean" data scenario (which in reality is possible but not very likely), one will have the most accurate estimates of true premiums for the whole portfolio by choosing the most efficient method among the most robust available for estimation of the individual experience of each risk in the portfolio.

A natural next step in this line of research would be to develop similar robust procedures for other distributions or, more ambitiously, for any arbitrary distribution F. Some ideas for the general case are already available in Künsch (1992) and Gisler and Reinhard (1993). However, it seems to us that, to have as comprehensive and systematic a study of robust credibility models as the one presented here, one needs to commit to a specific distribution (like we did to the gamma) because all the quantitative performance measures depend on F (see Table 2). Therefore, we conjecture that further extensions of the ideas presented in this article will be approached on a case-by-case basis.

APPENDIX

Table 10

Loss Ratios, True Premiums $\mu(\theta_i)$, and Estimates Based on Recommendations of Experts E1, E2, E3, E4, under Scenario 2

Risk		Estimate T_i				Estimate $\tilde{\mu}(\theta_i)$				True
i	Loss Ratios	E1	E2	E3	E4	E1	E2	E3	E4	$\mu(\theta_i)$
1	1.26, 0.44	0.85	0.95	0.95	0.85	2.88	2.53	2.77	2.66	2.61
2	3.57, 3.60	3.58	3.98	3.98	3.58	2.99	3.26	3.05	3.11	3.57
3	0.24, 0.82	0.53	0.59	0.59	0.53	2.86	2.44	2.73	2.61	2.56
4	0.38, 0.89	0.64	0.71	0.71	0.64	2.87	2.47	2.74	2.62	1.96
5	4.99, 0.45	2.72	3.02	3.02	2.72	2.96	3.03	2.96	2.97	4.26
6	4.57, 0.88	2.73	2.79	2.79	2.73	2.94	3.05	2.94	3.00	1.79
7	1.41, 1.91	1.66	1.70	1.70	1.66	2.81	2.52	2.68	2.61	4.84
8	1.69, 0.90	1.30	1.33	1.33	1.30	2.77	2.34	2.60	2.47	3.48
9	0.43, 0.49	0.46	0.47	0.47	0.46	2.68	1.92	2.39	2.16	2.75
10	4.26, 2.78	3.52	3.61	3.61	3.52	3.03	3.44	3.13	3.30	4.89
11	2.12, 0.89	1.51	1.52	1.52	1.51	2.70	2.31	2.51	2.42	1.59
12	3.99, 6.63	5.31	5.36	5.36	5.31	3.39	4.65	3.80	4.31	5.25
13	1.14, 0.55	0.84	0.85	0.85	0.84	2.58	1.90	2.28	2.09	1.49
14	1.34, 7.79	4.56	4.60	4.60	4.56	3.25	4.19	3.55	3.94	4.52
15	2.67, 2.66	2.66	2.69	2.69	2.66	2.91	3.02	2.90	2.99	4.40
16	0.93, 0.52, 0.01, 0.51, 0.24	0.44	0.52	0.49	0.44	2.72	2.03	2.47	2.24	0.77
17	1.54, 2.06, 5.92, 6.18, 0.39	3.22	2.93	3.58	3.22	2.99	3.10	3.10	3.16	2.36
18	4.21, 0.78, 0.27, 7.40, 0.71	2.67	1.30	2.49	2.63	2.94	2.37	2.88	2.97	2.81
19	2.61, 0.92, 2.01, 0.02, 0.29	1.17	0.91	1.30	1.17	2.79	2.20	2.63	2.48	1.63
20	0.84, 0.64, 0.11, 0.30, 1.03	0.58	0.76	0.65	0.58	2.73	2.14	2.50	2.29	1.50
21	0.53, 0.47, 0.77, 0.66, 0.91	0.67	0.80	1.67	0.67	2.40	1.72	2.02	1.81	1.04
22	1.58, 1.25, 1.74, 1.17, 2.24	1.59	1.90	1.61	1.59	2.63	2.49	2.42	2.36	2.50
23	1.25, 8.39, 1.44, 1.44, 1.48	2.80	1.86	2.19	1.87	2.93	2.46	2.68	2.53	2.04
24	5.53, 9.36, 8.39, 8.15, 6.53	7.59	9.43	7.66	7.59	4.11	7.78	5.05	5.95	6.87
25	1.77, 1.11, 0.61, 0.94, 1.75	1.24	1.32	1.25	1.24	2.54	2.08	2.27	2.15	1.51
26	2.74, 2.80, 1.32, 2.42, 2.84	2.42	2.77	2.43	2.42	2.77	3.13	2.73	2.84	3.19
27	2.38, 1.26, 0.97, 2.87, 2.22	1.94	1.94	1.94	1.94	2.60	2.47	2.46	2.50	2.80
28	0.79, 1.13, 2.48, 0.89, 6.96	2.45	1.26	2.04	1.64	2.78	1.93	2.52	2.29	1.30
29	3.74, 2.32, 1.99, 1.01, 5.24	2.86	2.40	2.86	2.81	2.93	2.83	2.98	3.12	3.35
30	4.87, 6.76, 2.53, 6.08, 3.13	4.67	4.74	4.68	4.67	3.57	4.70	4.00	4.45	4.98
31	2.24, 2.10, 0.16, 1.86, 3.45, 2.53, 1.82, 2.44, 1.18, 0.42	1.82	2.97	1.91	1.82	2.76	3.20	2.64	2.57	2.27
32	4.09, 0.92, 1.80, 0.28, 6.54, 0.22, 0.34, 2.22, 1.05, 3.26	2.07	1.58	2.16	1.95	2.81	2.35	2.72	2.64	1.76
33	2.83, 4.45, 2.58, 2.39, 0.12, 0.29, 2.30, 0.33, 2.73, 2.56	2.06	2.75	2.16	2.06	2.80	3.06	2.72	2.69	2.08
34	14.97, 3.96, 0.57, 5.72, 14.39, 1.61, 10.03, 0.71, 0.04, 1.28	5.33	2.36	5.59	5.27	3.39	2.82	3.88	4.29	2.81
35	0.08, 0.17, 0.58, 3.81, 4.53, 2.54, 2.31, 1.10, 4.61, 0.62	2.04	1.42	2.14	2.04	2.80	2.25	2.72	2.68	3.23
36	4.54, 1.56, 1.76, 1.37, 0.75, 7.61, 1.65, 3.68, 1.84, 0.86	2.56	2.01	2.57	2.24	2.81	2.51	2.80	2.70	2.52
37	8.27, 1.00, 1.05, 1.77, 1.01, 1.68, 1.76, 1.69, 2.01, 6.85	2.71	2.05	2.70	2.00	2.86	2.54	2.88	2.52	1.50
38	4.67, 2.60, 1.37, 8.29, 3.60, 0.94, 1.84, 2.99, 0.96, 1.45	2.87	2.25	2.88	2.55	2.93	2.71	2.99	2.93	3.08
39	1.93, 2.52, 1.35, 3.62, 0.86, 0.97, 9.10, 0.81, 1.75, 2.41	2.53	1.97	2.32	2.03	2.79	2.48	2.66	2.54	2.36
40	2.88, 0.93, 1.30, 1.93, 3.14, 2.59, 0.88, 1.40, 8.23, 2.05	2.53	2.18	2.45	2.14	2.79	2.65	2.73	2.62	2.88
41	1.70, 2.22, 2.61, 1.79, 4.01, 0.66, 3.73, 3.43, 1.11, 1.31	2.26	2.00	2.26	2.26	2.60	2.47	2.55	2.69	3.36
42	5.12, 1.66, 7.69, 1.29, 1.20, 23.32, 0.98, 4.09, 3.22, 2.00	5.06	2.17	3.89	3.03	4.06	2.62	3.73	3.33	4.12
43	3.26, 2.13, 1.98, 0.84, 5.99, 6.09, 2.97, 2.67, 2.42, 2.39	3.07	2.85	3.07	2.89	3.02	3.22	3.14	3.22	3.43
44	3.43, 5.55, 5.76, 9.05, 4.60, 6.33, 6.50, 1.57, 6.31, 2.04	5.11	5.22	5.11	5.11	4.09	5.33	4.61	5.07	6.92
45	4.75, 3.50, 2.11, 1.65, 3.63, 3.62, 3.43, 1.63, 20.88, 2.46	4.77	3.40	3.83	3.26	3.91	3.72	3.68	3.52	3.61

ACKNOWLEDGMENTS

The authors are very appreciative of valuable insights and comments provided by two anonymous referees which led to many improvements in the paper. Also, the second author gratefully acknowledges the support provided by a grant from the Casualty Actuarial Society and the Society of Actuaries, with administrative support from the Committee on Knowledge Extension Research.

REFERENCES

BAILEY, A. L. 1945. A Generalized Theory of Credibility. *Proceedings of the Casualty Actuarial Society* 32: 13–20.

———. 1950. Credibility Procedures—Laplace's Generalization of Bayes' Rule and the Combination of Collateral Knowledge with Observed Data. *Proceedings of the Casualty Actuarial Society* 37: 7–23.

BRAZAUSKAS, V. 2003. Influence Functions of Empirical Nonparametric Estimators of Net Reinsurance Premiums. *Insurance: Mathematics and Economics* 32(1): 115–33.

BRAZAUSKAS, V., AND R. SERFLING. 2000. Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution (with Discussion). *North American Actuarial Journal* 4(4): 12–27. Discussion, 5(3): 123–26. Reply, 5(3): 126–28.

———. 2003. Favorable Estimators for Fitting Pareto Models: A Study Using Goodness-of-Fit Measures with Actual Data. *ASTIN Bulletin* 33(2): 365–81.

BÜHLMANN, H. 1967. Experience Rating and Credibility. *ASTIN Bulletin* 4(3): 199–207.

———. 1969. Experience Rating and Credibility. *ASTIN Bulletin* 5(2): 157–65.

BÜHLMANN, H., AND E. STRAUB. 1970. Glaubgewürdigkeit für Schadensätze (Credibility for Loss Ratios). *Bulletin of the Swiss Association of Actuaries* 70: 111–33. (English translation by C. E. Brooks.)

GISLER, A., AND P. REINHARD. 1993. Robust Credibility. *ASTIN Bulletin* 23(1): 118–43.

GOULET, V. 1998. Principles and Application of Credibility Theory. *Journal of Actuarial Practice* 6(1–2): 5–62.

HAMPEL, F. R., E. M. RONCHETTI, P. J. ROUSSEEUW, AND W. A. STAHEL. 1986. *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

HUBER, P. J. 1981. *Robust Statistics.* New York: Wiley.

KEFFER, R. 1929. An Experience Rating Formula. *Transactions of the Actuarial Society of America* 30: 130–39.

KLUGMAN, S. A., H. H. PANJER, AND G. E. WILLMOT. 2004. *Loss Models: From Data to Decisions.* 2nd edition. New York: Wiley.

KÜNSCH, H. R. 1992. Robust Methods for Credibility. *ASTIN Bulletin* 22(1): 33–49.

MARCEAU, E., AND J. RIOUX. 2001. On Robustness in Risk Theory. *Insurance: Mathematics and Economics* 29(2): 167–85.

MOWBRAY, A. H. 1914. How Extensive a Payroll Exposure is Necessary to Give a Dependable Pure Premium? *Proceedings of the Casualty Actuarial Society* 1: 25–30.

NORBERG, R. 1979. The Credibility Approach to Experience Rating. *Scandinavian Actuarial Journal,* 181–221.

SERFLING, R. 2002. Efficient and Robust Fitting of Lognormal Distributions. *North American Actuarial Journal* 6(4): 95–109.

WHITNEY, A. W. 1918. The Theory of Experience Rating. *Proceedings of the Casualty Actuarial Society* 4: 275–93.

YOUNG, V. R. 1997. Credibility Using Semiparametric Models. *ASTIN Bulletin* 27(2): 273–85.

———. 1998. Robust Bayesian Credibility Using Semiparametric Models. *ASTIN Bulletin* 28(2): 187–203.

Discussions on this paper can be submitted until January 1, 2008. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.