# Interval Estimation of Actuarial Risk Measures

Thomas Kaiser* and Vytaras Brazauskas†

## Abstract

This article investigates performance of interval estimators of various actuarial risk measures. We consider the following risk measures: proportional hazards transform (PHT), Wang transform (WT), value-at-risk (VaR), and conditional tail expectation (CTE). Confidence intervals for these measures are constructed by applying nonparametric approaches (empirical and bootstrap), the strict parametric approach (based on the maximum likelihood estimators), and robust parametric procedures (based on trimmed means).

Using Monte Carlo simulations, we compare the average lengths and proportions of coverage (of the true measure) of the intervals under two data-generating scenarios: "clean" data and "contaminated" data. In the "clean" case, data sets are generated by the following (similar shape) parametric families: exponential, Pareto, and lognormal. Parameters of these distributions are selected so that all three families are equally risky with respect to a fixed risk measure. In the "contaminated" case, the "clean" data sets from these distributions are mixed with a small fraction of unusual observations (outliers). It is found that *approximate* knowledge of the underlying distribution combined with a sufficiently robust estimator (designed for that distribution) yields intervals with satisfactory performance under both scenarios.

## 1. Introduction

When determining the price of an insurance risk, the central problem is to quantify the "riskiness" of the underlying distribution of losses. The loss variables, for which accurate predictions are difficult to obtain (e.g., those with large variance and/or heavy right-tail), are deemed more risky, and therefore they necessitate a higher price. Various risk measures have been proposed in the actuarial literature to solve this problem. These include the β-*factor, conditional tail expectation* and variants, *proportional hazards transform, ruin probability, value-at-risk* and variants, *Wang transform,* and others (see Albrecht 2004; Dowd 2004; Hardy and Wirch 2004; Jones and Zitikis 2003; Wang 1998, 2000, 2002; Wang, Young, and Panjer 1997; Wirch and Hardy 1999). However, the quality of statistical estimators of the risk measures is an insufficiently explored—yet very important in practice—issue. In this paper we construct asymptotic interval estimators for several frequently used risk measures and, using Monte Carlo simulations, investigate their performance under two data-generating scenarios: "clean" data and "contaminated" data. (For motivation and how to formalize such scenarios effectively see Section 2.3.)

There is substantial literature devoted to the question what a "reasonable" risk measure is and what conditions it should satisfy. As discussed by Albrecht (2004), a number of axiomatic systems are available in the literature for characterization of risk measures (though most of them have some overlap). A quite influential one was proposed by Artzner and his collaborators (see Artzner 1999 and the ref-

---
* Thomas Kaiser is Associate at Towers Perrin, Tillinghast, Neue Weyerstrasse 6, 50676 Cologne, Germany, thomas.kaiser@towersperrin.com. This work was completed while the author was a student at the University of Ulm.
† Vytaras Brazauskas is Associate Professor in the Department of Mathematical Sciences, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, WI 53201, vytaras@uwm.edu.

erences cited therein) who advocate the use of *coherent* measures. These are defined as follows. For loss variables $X$ and $Y$, a coherent risk measure RISK[·] should satisfy the following axioms:

- *Translation invariance:* RISK$[X + a]$ = RISK$[X]$ + $a$, for $-\infty < a < +\infty$
- *Scale invariance:* RISK$[bX]$ = $b$ RISK$[X]$, for $b > 0$
- *Subadditivity:* RISK$[X + Y]$ ≤ RISK$[X]$ + RISK$[Y]$
- *Monotonicity:* If $X \leq Y$ (with probability 1), then RISK$[X]$ ≤ RISK$[Y]$.

For the purpose of estimation (which is one of the main objectives of this paper), it is useful to note that many of risk measures can be defined as the expectation with respect to *distorted probabilities*. That is, for a loss variable $X \geq 0$ with distribution function $F$, a risk measure $R$ can be defined as

$$R(F) = \int_0^\infty g(1 - F(x)) \, dx, \tag{1.1}$$

where the *distortion function* $g(\cdot)$ is an increasing function with $g(0) = 0$ and $g(1) = 1$. In addition, if $g$ is also differentiable, then

$$R(F) = \int_0^1 F^{-1}(s)\psi(s) \, ds, \tag{1.2}$$

where $\psi(s) = g'(1 - s)$ and $F^{-1}$ is the quantile function (the inverse of $F$) of variable $X$.

While expression (1.1) has interpretive advantages, the second representation (1.2) is more convenient for developing empirical estimators of the risk measure $R$. Indeed, if in equation (1.2) we replace $F$ by the empirical distribution function $\hat{F}_n$, then the integral $\int_0^1 \hat{F}_n^{-1}(s)\psi(s) \, ds$ becomes $\sum_{i=1}^n X_{(i)} [\int_{(i-1)/n}^{i/n} \psi(s) \, ds]$, where $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the ordered values of data $X_1, \ldots, X_n$. Hence, the empirical estimator of a risk measure $R(F)$ is given by

$$R(\hat{F}_n) = \sum_{i=1}^n c_{in} X_{(i)} \tag{1.3}$$

with $c_{in} = \int_{(i-1)/n}^{i/n} \psi(s) \, ds$. Note that $R(\hat{F}_n)$ as defined in equation (1.3) is an $L$-statistic (linear combination of order statistics). Using this fact, Jones and Zitikis (2003) employ asymptotic theory for $L$-statistics to prove that, for underlying distributions with a sufficient number of finite moments and under certain regularity conditions on function $\psi$, the empirical estimator $R(\hat{F}_n)$ of a risk measure $R(F)$ is *strongly consistent* and *asymptotically normal* with the mean $R(F)$ and the variance $Q(\psi, \psi)/n$, where

$$Q(\psi, \psi) = \int_{-\infty}^\infty \int_{-\infty}^\infty [F(\min(x, y)) - F(x)F(y)] \, \psi(F(x))\psi(F(y)) \, dx \, dy. \tag{1.4}$$

To apply these results in practice, one needs to estimate $Q(\psi, \psi)$ (this approach was suggested by Jones and Zitikis 2003). Alternatively, instead of relying on asymptotic normality of the empirical estimator $R(\hat{F}_n)$, one can turn to resampling methods and construct *bootstrap* confidence intervals. Both approaches are treated in this paper.

We will also consider the case when in equation (1.2) function $F$ is estimated using parametric and robust parametric approaches. In this case we assume a specific form of $F$ and then estimate its parameters using either *maximum likelihood estimators* (this leads to the so-called strict parametric approach) or *trimmed mean*–type estimators (this is known as the robust parametric approach). Since both techniques lead to estimators that have an asymptotically normal distribution, confidence intervals for the parameters can be easily constructed. Finally, these intervals are transformed according to equation (1.2) to get asymptotic confidence intervals for $R(F)$.

Section 2 provides some preliminary definitions, motivation, and the study design. In Section 3 we construct asymptotic confidence intervals for the risk measures by applying (1) nonparametric approaches (empirical and bootstrap), (2) the strict parametric approach (based on the maximum like-

lihood estimators), and (3) robust parametric procedures (based on trimmed means). The next section shows how these intervals are constructed when data are available. The illustration is based on the Norwegian Fire Claims (for the year 1975) data. In Section 5 we investigate—via simulation—the performance of interval estimators based on the three methodologies (1)–(3). Finally, conclusions are drawn, and a final discussion is provided in Section 6.

## 2. Preliminaries

In this section we first introduce the risk measures under study (Section 2.1), then define general and specific objectives of interval estimation (Section 2.2), and finally discuss the design of our Monte Carlo simulation study (Section 2.3).

## 2.1 Risk Measures

For a loss variable $X \geq 0$ having a *continuous* distribution function $F$ (equivalently, continuous quantile function $F^{-1}$), we consider the following risk measures.

### 2.1.1 Proportional Hazards Transform (PHT)

The PHT measure is defined by the distortion function $g(s) = s^r$ or, equivalently, since function $g$ is differentiable, by function $\psi(s) = r(1 - s)^{r-1}$, and thus is given by

$$\text{PHT}(F, r) = \int_0^\infty [1 - F(u)]^r \, du$$

$$= r \int_0^1 F^{-1}(t)(1 - t)^{r-1} \, dt, \tag{2.1}$$

where constant $r$ $(0 < r \leq 1)$ represents the degree of distortion (small $r$ corresponds to high distortion). Note that $\text{PHT}(F, r = 1)$ is the expected value of $X$, and $\text{PHT}(F, 1/2) - \text{PHT}(F, 1)$ is the right-tail deviation of Wang (1998). The name—proportional hazards transform—is motivated by the fact that the *hazard function* of the distorted distribution is proportional to the hazard function of $F$. Also, the PHT measure is coherent and satisfies the desirable axioms for a risk measure studied by Wang, Young, and Panjer (1997).

### 2.1.2 Wang Transform (WT)

For the WT measure the distortion function $g(s) = \Phi(\Phi^{-1}(s) + \lambda)$ is also differentiable; therefore, the measure can be defined by either function $g$ or $\psi(s) = e^{\lambda \Phi^{-1}(s) - \lambda^2/2}$ and is given by

$$\text{WT}(F, \lambda) = \int_0^\infty \Phi(\Phi^{-1}(1 - F(u)) + \lambda) \, du$$

$$= \int_0^1 F^{-1}(t) e^{\lambda \Phi^{-1}(t) - \lambda^2/2} \, dt, \tag{2.2}$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$, respectively, denote the cdf and the inverse of the standard normal distribution, and parameter $\lambda$ reflects the level of systematic risk and is called the *market price of risk*. The distortion function $g(s) = \Phi(\Phi^{-1}(s) + \lambda)$ was introduced by Wang (2000, 2002) as a tool for pricing both liabilities (insurance losses) and asset-returns (gains) and, therefore, is valid on interval $(-\infty, \infty)$. Here we focus on insurance losses; thus the definition of $\text{WT}(F, \lambda)$ as in equation (2.2) is sufficient. Further, the measure based on $g(s) = \Phi(\Phi^{-1}(s) + \lambda)$ is coherent, and, for normally distributed asset-returns, it recovers the *Capital Asset Pricing Model* and the *Black-Scholes* formula. Finally, although theoretically the risk parameter $\lambda$ can be any real number, we will consider the range of $\lambda$ between $-1$ and $1$ that was used in several examples presented by Wang (2002).

### 2.1.3 Value-at-Risk (VaR)

As the *Encyclopedia of Actuarial Science* defines it (Dowd 2004), the value-at-risk on a portfolio is the *maximum loss* we might expect over a given period, at a given *level of confidence* (e.g., $\beta = 0.05$). In mathematical terms the VaR measure is nothing else but the $(1 - \beta)$–level quantile of the distribution function $F$:

$$\text{VaR}(F, \beta) = F^{-1}(1 - \beta). \tag{2.3}$$

Although quantiles of the cdf is a well-understood and thoroughly investigated concept in statistics, the term value-at-risk came into widespread use in finance and insurance literature only in the early 1990s when J. P. Morgan revealed its RiskMetrics model to the public; the VaR was the key component of this model. There is vast literature devoted to the methods of estimation and usefulness of this quantity in applications; perhaps equally vast literature is also devoted to criticisms of VaR. For a general and recent overview about VaR, the reader should consult Dowd (2004) and the references therein. Interestingly, this measure is not coherent nor is it the distortion measure (though mathematically it can be expressed as equation [1.1] by choosing $g(s) = 0$, for $0 \le s < \beta$, and $= 1$, for $\beta \le s \le 1$). We include the VaR measure in this study because of its mathematical simplicity and wide popularity among practitioners.

### 2.1.4 Conditional Tail Expectation (CTE)

The CTE measure (also known as Tail-VaR or expected shortfall) is the conditional expectation of a loss variable $X$ given that $X$ exceeds a specified quantile (e.g., $\text{VaR}(F, \beta)$). In other words, it measures the *expected maximum loss* in the $100\beta\%$ worst cases, over a given period. Thus, by choosing the distortion function $g(s) = s/\beta$, for $0 \le s < \beta$, and $= 1$, for $\beta \le s \le 1$, or function $\psi(s) = 0$, for $0 \le s \le 1 - \beta$, and $= 1/\beta$, for $1 - \beta < s \le 1$, we have

$$\text{CTE}(F, \beta) = \int_0^{\text{VaR}(F,\beta)} 1 \ du + \frac{1}{\beta} \int_{\text{VaR}(F,\beta)}^{\infty} [1 - F(u)] \ du$$

$$= \frac{1}{\beta} \int_{1-\beta}^1 F^{-1}(t) \ dt. \tag{2.4}$$

The CTE is a coherent and intuitively appealing measure. And because of these properties it has become a popular risk-measuring tool in insurance and finance industries. For example, use of the CTE for determining liabilities associated with variable life insurance and annuity products with guarantees is recommended in the United States (American Academy of Actuaries 2002) and required in Canada (Canadian Institute of Actuaries 2002).

## 2.2 Interval Estimation: Motivation and Objectives

In point estimation problems one usually reports the value of an estimator, called the *point estimate,* along with its estimated standard deviation (also known as the *standard error*). Then the interval "point estimate ± standard error" is supposed to give some idea about the variability of the estimator and is interpreted as the range where the true parameter is *potentially* located. Although such a procedure does yield an interval, it fails to provide the rate of success (also known as the *confidence level*) of how often the true parameter will be covered. Hence, it makes more sense to focus on the construction of *confidence intervals* because these do not possess the above-mentioned deficiency.

In general, the key objective of (confidence) interval estimation is to identify statistical procedures that yield the shortest interval while maintaining the desired (high) confidence level. However, as is known from classical statistical inference, for a fixed method of estimation (and a fixed sample size), the length and the confidence level of the interval are two competing criteria. This fact suggests, therefore, that simultaneous improvement with respect to both criteria is possible only when one considers confidence intervals that are constructed using different techniques of estimation.

In this article confidence intervals for the risk measures of Section 2.1 are constructed by employing nonparametric (empirical and bootstrap), parametric (based on the maximum likelihood estimators), and robust parametric (based on trimmed means) methodologies. Performances of the intervals, that is, the lengths and the coverage probabilities of the true measure, then are compared. As will be seen in Section 3, most of the intervals are asymptotic, and all of them are based on various underlying assumptions. Thus in this setting, it is of great interest to investigate the following issues:

a. Convergence rates: How fast do the proposed (asymptotic) intervals attain the intended confidence level?
b. Performance at the model: Under strict distributional assumptions, how much do we gain or lose if, instead of nonparametric intervals, parametric or robust parametric confidence intervals are used?
c. Sensitivity to assumptions: How bad are the consequences if the underlying assumptions necessary for the theoretical statements to hold are ignored or cannot be verified?

The questions in (a) and (b) are answered for the PHT measure by Brazauskas and Kaiser (2004) and for the CTE measure by Brazauskas et al. (2006), and similar findings could be anticipated for the other two measures. Thus, (a) and (b) will be less emphasized in the present paper, but the conclusions from the two references regarding these issues will be included in the discussion of Section 6. Our main objective here is issue (c), which will be approached using Monte Carlo simulations.

## 2.3 Simulation Study: Motivation and Design

Performance of the nonparametric, parametric, and robust parametric intervals (all defined precisely in Section 3) is investigated under two data-generating scenarios: "clean" data and "contaminated" data. Such scenarios are formalized by employing $\varepsilon$-contamination neighborhoods:

$$G_{F,\varepsilon} = (1 - \varepsilon)F + \varepsilon H, \tag{2.5}$$

where $F$ is the "central" (assumed) model, $H$ is a "contaminating" distribution (or a mixture of distributions) that generates outliers, and $\varepsilon$ represents the probability that a sample observation comes from the distribution $H$ instead of $F$. For $\varepsilon = 0$, family $G_{F,\varepsilon}$ generates "clean" data, and, for $\varepsilon > 0$, it generates "contaminated" data.

### REMARK 1

An "outlier" should be understood as not necessarily just another large observation, but rather the observation that violates the distributional assumptions. In the situations when $F$ itself is supposed to produce large data points with relatively high probability it is impossible to distinguish between outliers and representative data. □

From a practical standpoint it is of interest to see how accurately we can estimate a risk measure when the underlying scenarios of data generation are very similar (but not identical). Thus, for the central model $F$ in equation (2.5), we consider the following (similar shape) parametric families:

• *Pareto* with the cdf given by $F_1(x) = 1 - (x_0/x)^{\gamma}$, for $x > x_0$ and $\gamma > 0$
• *Lognormal* with the cdf given by $F_2(x) = \Phi(\log(x - x_0) - \mu)$, for $x > x_0$ and $-\infty < \mu < \infty$, where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution
• *Exponential* with the cdf given by $F_3(x) = 1 - e^{-(x-x_0)/\theta}$, for $x > x_0$ and $\theta > 0$.

Here the parameter $x_0$ can be interpreted as a deductible or a retention level and, thus, assumed to be known. (Note that, due to $x_0$, distributions $F_1$, $F_2$, and $F_3$ have the same support.) The remaining parameters $\gamma$, $\mu$, and $\theta$ are unknown, and we must estimate them from the data. To achieve greater similarity among these distributions and, thus, to generate data under more realistic scenarios, we select parameters $\gamma$, $\mu$, $\theta$ so that families $F_1$, $F_2$, $F_3$ are equally risky with respect to a fixed risk measure: that is, they satisfy the following equation:

$$R(F_1) = R(F_2) = R(F_3), \tag{2.6}$$

where $R(\cdot)$ represents any of the four risk measures of Section 2.1. Evaluation of integrals (2.1)–(2.4) for distributions $F_1$, $F_2$, $F_3$ yields the following expressions of equation (2.6):

- For the PHT measure:

$$x_0 + \frac{x_0}{\gamma r - 1} = x_0 + C_r^{(1)} e^\mu = x_0 + \frac{\theta}{r}, \tag{2.7}$$

  where, for fixed $r$, the integral

$$C_r^{(1)} = \int_{-\infty}^{\infty} [1 - \Phi(z)]^r e^z \, dz$$

  is found numerically. For example, as reported by Brazauskas and Kaiser (2004), $C_{0.55}^{(1)} = 3.896$, $C_{0.70}^{(1)} = 2.665$, $C_{0.85}^{(1)} = 2.030$, $C_{0.95}^{(1)} = 1.758$.
- For the WT measure:

$$x_0 + \frac{x_0 C_{\gamma,\lambda}^{(2)}}{\gamma} = x_0 + e^{\lambda + \mu + 0.5} = x_0 + \theta C_\lambda^{(3)}, \tag{2.8}$$

  where, for fixed $\lambda$, the integrals

$$C_{\gamma,\lambda}^{(2)} = \int_0^1 \Phi[\Phi^{-1}(z) + \lambda] \, z^{(-\gamma-1)/\gamma} \, dz,$$

  with $\Phi^{-1}(\cdot)$ denoting the inverse of the standard normal cdf, and

$$C_\lambda^{(3)} = 1/\sqrt{2\pi} \int_{-\infty}^{\infty} \Phi(z)/\Phi(z - \lambda) \, e^{-(z-\lambda)^2/2} \, dz$$

  are found numerically. For example:

| $\lambda$ | $-1$ | $-\dfrac{1}{2}$ | $-\dfrac{1}{4}$ | $0$ | $\dfrac{1}{4}$ | $\dfrac{1}{2}$ | $1$ |
|---|---|---|---|---|---|---|---|
| $C_{\gamma=5.5,\lambda}^{(2)}$ | 0.3984 | 0.7151 | 0.9401 | 1.2222 | 1.5739 | 2.0101 | 3.2147 |
| $C_\lambda^{(3)}$ | 0.3593 | 0.6185 | 0.7923 | 1.0000 | 1.2449 | 1.5301 | 2.2318 |

- For the VaR measure:

$$x_0 \beta^{-1/\gamma} = x_0 + e^{\mu + \Phi^{-1}(1-\beta)} = x_0 - \theta \log(\beta). \tag{2.9}$$

- For the CTE measure:

$$\frac{x_0 \gamma}{\gamma - 1} \beta^{-1/\gamma} = x_0 + \frac{1}{\beta} e^{\mu + 0.5} \Phi(1 - \Phi^{-1}(1 - \beta))$$

$$= x_0 - \theta(\log(\beta) - 1). \tag{2.10}$$

To see how similar the shapes of these parametric distributions are, in Figure 1 we plot the density functions of families $F_1$, $F_2$, $F_3$ that satisfy condition (2.10) for $\beta = 0.05$. As one can see, these functions are virtually identical in the upper tail (right panel) and slightly differ for smaller values of $x$ (left panel). (Note that the scale of the vertical axis of the right panel is 1,000 times smaller than that of the left panel.) Of course, after investing more time and creativity, a determined actuary/statistician may discover the existing differences. Then he or she would have to figure out how to handle those differences. The point we are trying to make here is that the distributions are indeed very similar, straightforward diagnostics most likely will not detect those differences, and more sophisticated approaches are time and energy consuming. Hence, instead of doing this "micro-analysis," one can adopt the approach proposed in this paper and achieve reliable results with a reasonable investment of effort.

For the contaminating distribution $H$ in equation (2.5), we choose the *uniform distribution* on the interval $(10x_0, 50x_0)$, denoted $U(10x_0, 50x_0)$, with the probability density function given by

Figure 1

**Density Functions of Three Equally Risky (According to the CTE Measure) Distributions: Pareto($x_0 = 1$, $\gamma = 5.5$), Lognormal($x_0 = 1$, $\mu = -2.044$, $\sigma = 1$), Exponential($x_0 = 1$, $\theta = 0.277$), with CTE($F_1$, $\beta = 0.05$) = CTE($F_2$, $\beta = 0.05$) = CTE($F_3$, $\beta = 0.05$) = 2.107**
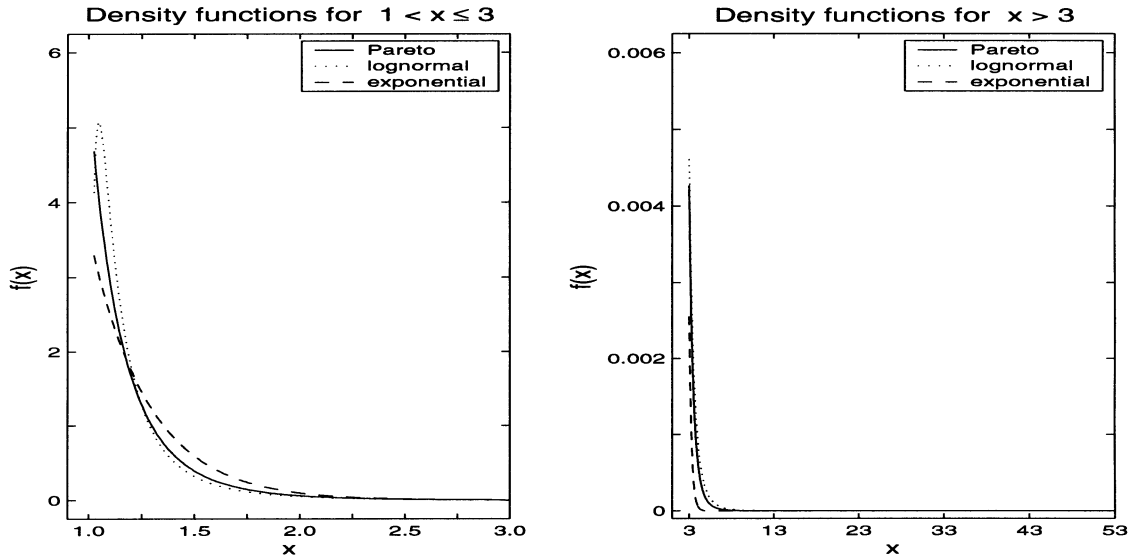


$$h(x) = \begin{cases} (40x_0)^{-1}, & \text{for } 10x_0 < x < 50x_0, \\ 0, & \text{elsewhere.} \end{cases}$$

There are countless possibilities to contaminate the central model in equation (2.5). The choice of $U(10x_0, 50x_0)$ 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. Further, although the resulting probability distribution $G_{F,\varepsilon}$ is functionally different from $F$, their shapes are so similar that practically no violations can be diagnosed. To see this, in Figure 2 we plot the density functions of model $G_{F,\varepsilon}$ for $\varepsilon = 0.00$ (left panel) and $\varepsilon = 0.05$ with $H = U(10, 50)$ (right panel) and for the same choice of $F_1$, $F_2$, $F_3$ as in Figure 1. Clearly, the $U(10, 50)$ density is supposed to introduce a ''bump'' in all curves, for $10 < x < 50$; however, the bump is so small that cannot be visually detected.

We used the following design for the Monte Carlo simulation study. Five thousand samples of size $n$ were generated from a distribution $G_{F,\varepsilon}$. For each sample and for each risk measure, a $(1 - \alpha)$–level confidence interval was constructed using the empirical, bootstrap, parametric, and robust parametric approaches. Then, based on these 5,000 intervals for each approach, the average length of the interval and the proportion of times the interval covers the true value of the risk measure was evaluated. This procedure was repeated five times, and the means and standard errors of the average length and of the proportion of coverage were recorded. The study was performed for the following (specific) choices of simulation parameters:

- Confidence level: $1 - \alpha = 0.95$
- Contamination level: $\varepsilon = 0.00, 0.05$
- Sample size: $n = 25, 50, 100, 250$
- Number of bootstrap samples: $B = 1,000$

Figure 2
**Density Functions of Distribution $G_{F,\varepsilon}$ for $F_1 = $ Pareto($x_0 = 1$, $\gamma = 5.5$),**
**$F_2 = $ Lognormal($x_0 = 1$, $\mu = -2.044$, $\sigma = 1$), and $F_3 = $ Exponential($x_0 = 1$, $\theta = 0.277$):**
**"Clean" Case ($\varepsilon = 0.00$) versus "Contaminated" Case ($\varepsilon = 0.05$ and $U(10x_0, 50x_0) = U(10, 50)$)**
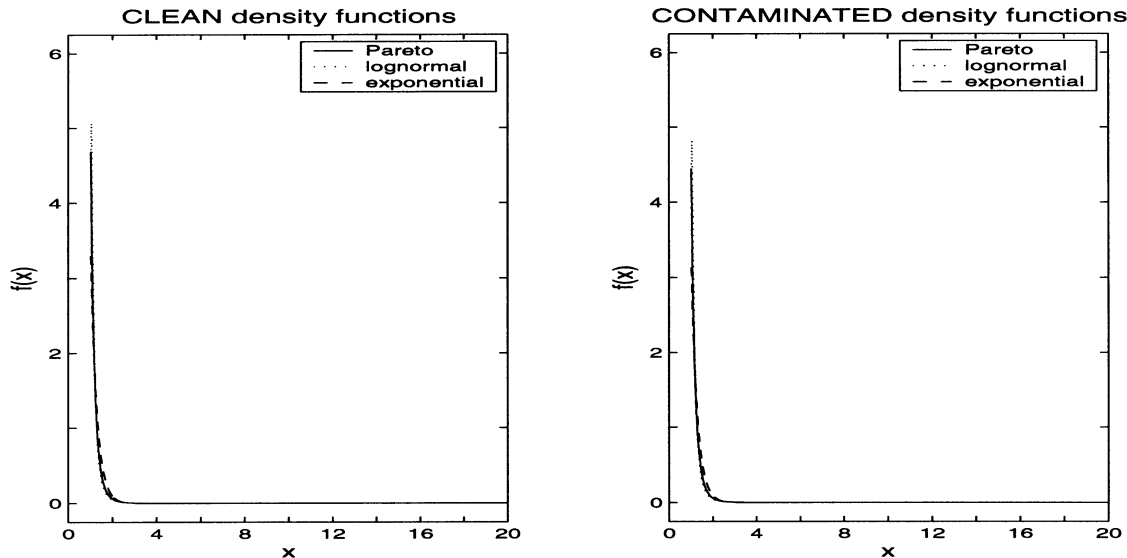


- Measure-related parameters:
  Distortion level (for PHT): $r = 0.85$
  Systematic risk (for WT): $\lambda = 0.25$
  Threshold level (for VaR and CTE): $\beta = 0.05$
- Target quantities (derived from eqs. (2.7)–(2.10)):
  - $1.272 = \text{PHT}(F_1, r = 0.85) = \text{PHT}(F_2, r = 0.85) = \text{PHT}(F_3, r = 0.85)$,
    where $F_1 = $ Pareto($x_0 = 1$, $\gamma = 5.5$), $F_2 = $ lognormal($x_0 = 1$, $\mu = -2.010$, $\sigma = 1$), and $F_3 = $ exponential($x_0 = 1$, $\theta = 0.231$).
  - $1.286 = \text{WT}(F_1, \lambda = 0.25) = \text{WT}(F_2, \lambda = 0.25) = \text{WT}(F_3, \lambda = 0.25)$,
    where $F_1 = $ Pareto($x_0 = 1$, $\gamma = 5.5$), $F_2 = $ lognormal($x_0 = 1$, $\mu = -2.001$, $\sigma = 1$), and $F_3 = $ exponential($x_0 = 1$, $\theta = 0.230$).
  - $1.724 = \text{VaR}(F_1, \beta = 0.05) = \text{VaR}(F_2, \beta = 0.05) = \text{VaR}(F_3, \beta = 0.05)$,
    where $F_1 = $ Pareto($x_0 = 1$, $\gamma = 5.5$), $F_2 = $ lognormal($x_0 = 1$, $\mu = -1.968$, $\sigma = 1$), and $F_3 = $ exponential($x_0 = 1$, $\theta = 0.242$).
  - $2.107 = \text{CTE}(F_1, \beta = 0.05) = \text{CTE}(F_2, \beta = 0.05) = \text{CTE}(F_3, \beta = 0.05)$,
    where $F_1 = $ Pareto($x_0 = 1$, $\gamma = 5.5$), $F_2 = $ lognormal($x_0 = 1$, $\mu = -2.044$, $\sigma = 1$), and $F_3 = $ exponential($x_0 = 1$, $\theta = 0.277$).

## 3. Interval Estimators

In this section we present three major methodologies—nonparametric (empirical and bootstrap), parametric, and robust parametric—for deriving confidence intervals for the PHT, WT, VaR, and CTE risk measures. The empirical nonparametric intervals are constructed by applying asymptotic theory for $L$-statistics, and the bootstrap intervals are a result of (nonparametric) data resampling. Parametric and robust parametric intervals are derived via the two-step procedure: (1) construct asymptotic confidence intervals for the parameters of the families $F_1$, $F_2$, $F_3$, and then (2) transform the end-points of these intervals according to an appropriate formula of equations (2.7)–(2.10).

**REMARK 2**

The second step in the procedure described above can also be accomplished by using the so-called Delta Method (see, e.g., Serfling 1980, Section 3.1). Asymptotically both approaches yield identical intervals; there are some differences, mainly in the length of intervals, in samples of fewer than 50 observations. However, as far as the sensitivity to assumptions is concerned (see the objectives (a), (b), (c) of Section 2.2), both techniques lead to the same conclusion. □

Throughout this section we will assume that for each distribution of Section 2.3 there is available a data set of independent and identically distributed observations $X_1, \ldots, X_n$. Also, $X_{(1)} \leq \cdots \leq X_{(n)}$ will denote the ordered values of data $X_1, \ldots, X_n$.

## 3.1 Nonparametric Approaches

### 3.1.1 Empirical Approach

For the PHT and WT measures, the distortion function $g$ (or, equivalently, function $\psi$) satisfies the regularity conditions established by Jones and Zitikis (2003). As is also discussed by these authors, it follows from asymptotic theory for $L$-statistics that, for underlying distributions with a sufficient number of finite moments (this is true for our choice of $F_1$, $F_2$, $F_3$ in Section 2.3), the empirical estimator $R(\hat{F}_n)$, defined in equation (1.3), is *asymptotically normal* with the mean $R(F)$ and the variance $Q(\psi, \psi)/n$, where $Q(\psi, \psi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(\min(x, y)) - F(x)F(y)]\, \psi(F(x))\psi(F(y))\, dx\, dy$. To construct confidence intervals for $R(F)$ that could be used in practice, the asymptotic variance of $R(\hat{F}_n)$ has to be estimated. To this end, Jones and Zitikis proposed the following (strongly consistent) estimator of $Q(\psi, \psi)$:

$$Q_n(\psi, \psi) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} c_n(i, j)\, \psi(i/n)\psi(j/n)[(X_{(i+1)} - X_{(i)})(X_{(j+1)} - X_{(j)})] \tag{3.1}$$

with $c_n(i, j) = \min\{i/n, j/n\} - (i/n)(j/n)$. Thus, combining equations (3.3) and (3.1) with the specific definition of function $\psi$ (of Section 2.1), we have the following empirical point and interval estimators of the PHT and WT risk measures.

#### 3.1.1.1 PHT MEASURE

The $100(1 - \alpha)\%$ confidence interval of $\text{PHT}(F, r)$ is given by

$$\text{PHT}(\hat{F}_n, r) \pm z_{\alpha/2} \sqrt{\frac{Q_n^{\text{PHT}}}{n}},$$

where

$$Q_n^{\text{PHT}} = \Sigma_{i=1}^{n-1} \Sigma_{j=1}^{n-1} c_n(i, j)[r(1 - i/n)^{r-1}][r(1 - j/n)^{r-1}][(X_{(i+1)} - X_{(i)})(X_{(j+1)} - X_{(j)})], \text{PHT}(\hat{F}_n, r)$$

$$= \Sigma_{i=1}^{n} X_{(i)}[(1 - (i - 1)/n)^r - (1 - i/n)^r]$$

is the point estimator of PHT, and $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution.

#### 3.1.1.2 WT MEASURE

The $100(1 - \alpha)\%$ confidence interval of $\text{WT}(F, \lambda)$ is given by

$$\text{WT}(\hat{F}_n, \lambda) \pm z_{\alpha/2} \sqrt{\frac{Q_n^{\text{WT}}}{n}},$$

where

$$Q_n^{\text{WT}} = \Sigma_{i=1}^{n-1} \Sigma_{j=1}^{n-1} c_n(i, j)e^{\lambda\Phi^{-1}(i/n)-\lambda^2/2}e^{\lambda\Phi^{-1}(j/n)-\lambda^2/2}[(X_{(i+1)} - X_{(i)})(X_{(j+1)} - X_{(j)})],$$

$$\text{WT}(\hat{F}_n, \lambda) = \Sigma_{i=1}^{n} X_{(i)}[(\Phi(\Phi^{-1}(i/n) + \lambda) - \Phi(\Phi^{-1}(j/n) + \lambda)]$$

is the point estimator of WT, and $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution.

For the VaR and CTE measures, the approach of Jones and Zitikis (2003) does not directly apply. However, these two measures are quite simple and can be estimated empirically by corresponding sample statistics, general asymptotic theory for which is available, for example, in Serfling (1980). Hence, we have the following empirical point and interval estimators of the VaR and CTE risk measures.

### 3.1.1.3 VaR MEASURE

Since VaR is a quantile of the distribution function, its empirical point estimator is the corresponding sample quantile: that is, $\text{VaR}(\hat{F}_n, \beta) = X_{(n-[n\beta])}$, where $[\cdot]$ denotes "greatest integer part." The $100(1 - \alpha)\%$ *distribution-free* confidence interval of $\text{VaR}(F, \beta)$ is given by

$$(X_{(k_{1n})}, X_{(k_{2n})}),$$

where sequences of integers $k_{1n}$ and $k_{2n}$ satisfy $1 \leq k_{1n} < k_{2n} \leq n$ and

$$k_{1n}/n - (1 - \beta) \approx -z_{\alpha/2} \sqrt{\beta(1 - \beta)/n}$$

and

$$k_{2n}/n - (1 - \beta) \approx z_{\alpha/2} \sqrt{\beta(1 - \beta)/n}, \text{ as } n \rightarrow \infty.$$

Here again $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution. For derivation of this interval and discussion, see Serfling (1980, Section 2.6).

### 3.1.1.4 CTE MEASURE

The $100(1 - \alpha)\%$ confidence interval of $\text{CTE}(F, \beta)$ is given by

$$\text{CTE}(\hat{F}_n, \beta) \pm z_{\alpha/2}\sqrt{V_\beta/[n\beta]},$$

where

$$\text{CTE}(\hat{F}_n, \beta) = [n\beta]^{-1} \Sigma_{i=n-[n\beta]+1}^n X_{(i)}$$

is the point estimator of $\text{CTE}(F, \beta)$,

$$V_\beta = s_\beta^2 + (1 - \beta)[\text{VaR}(\hat{F}_n, \beta) - \text{CTE}(\hat{F}_n, \beta)]^2$$

with

$$s_\beta^2 = ([n\beta] - 1)^{-1} \Sigma_{i=n-[n\beta]+1}^n [X_{(i)} - \text{CTE}(\hat{F}_n, \beta)]^2,$$

and $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution. In the actuarial literature, the variance function $V_\beta$ is derived by Manistre and Hancock (2005). More general CTE estimation problems are treated by Brazauskas et al. (2006).

### 3.1.2 Bootstrap

Let us resample (with replacement) $n$ observations from the given data $X_1, \ldots, X_n$ assuming that each data point has an equal chance, $1/n$, to appear in the new sample, called the bootstrap sample and denoted $X_1^{(i)}, \ldots, X_n^{(i)}$. (Such a resampling scheme is known as *nonparametric bootstrap*.) Next, evaluation of the empirical estimator $R(\hat{F}_n)$, based on the sample $X_1^{(i)}, \ldots, X_n^{(i)}$, yields estimate $R^{(i)}(\hat{F}_n)$. (Note that $R(\hat{F}_n)$ represents any of the point estimators of Section 3.1.1: $\text{PHT}(\hat{F}_n, r)$, $\text{WT}(\hat{F}_n, \lambda)$, $\text{VaR}(\hat{F}_n, \beta)$, $\text{CTE}(\hat{F}_n, \beta)$.) After repeating this process $B$ number of times one gets $B$ estimates $R^{(1)}(\hat{F}_n), \ldots, R^{(B)}(\hat{F}_n)$. Then the $100(1 - \alpha)\%$ *bootstrap* confidence interval of $R(F)$ is given by

$$(R_*^{([B\alpha/2])}(\hat{F}_n), R_*^{([B(1-\alpha/2)])}(\hat{F}_n)),$$

where $[\cdot]$ denotes "greatest integer part" and $R_*^{(1)}(\hat{F}_n) \leq \cdots \leq R_*^{(B)}(\hat{F}_n)$ denote the ordered values of $R^{(1)}(\hat{F}_n), \ldots, R^{(B)}(\hat{F}_n)$. For further reading on bootstrap techniques, see Efron and Tibshirani (1993).

## 3.2 Parametric Methods

In the strict parametric modeling, parameters of the families $F_1$, $F_2$, $F_3$ are estimated using the maximum likelihood approach. Straightforward derivations yield the following formulas for the MLEs of the parameters $\gamma$, $\mu$, and $\theta$, respectively:

$$\hat{\gamma}_{\text{MLE}} = \left[ \frac{1}{n} \sum_{i=1}^{n} \log(X_i/x_0) \right]^{-1},$$

$$\hat{\mu}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} \log(X_i - x_0),$$

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - x_0).$$

Further, as is well known from asymptotic theory for the maximum likelihood procedures, the estimators $\hat{\gamma}_{\text{MLE}}$, $\hat{\mu}_{\text{MLE}}$, and $\hat{\theta}_{\text{MLE}}$ are each asymptotically normal with, respectively, the means $\gamma$, $\mu$, and $\theta$ and the variances $\gamma^2/n$, $1/n$, and $\theta^2/n$. Therefore, the $100(1 - \alpha)\%$ confidence intervals for the parameters $\gamma$, $\mu$, and $\theta$ are given by

$$\hat{\gamma}_{\text{MLE}}(1 \pm z_{\alpha/2}\sqrt{1/n}),$$

$$\hat{\mu}_{\text{MLE}} \pm z_{\alpha/2}\sqrt{1/n},$$

$$\hat{\theta}_{\text{MLE}}(1 \pm z_{\alpha/2}\sqrt{1/n}), \tag{3.2}$$

where $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution.

Finally, to get the *parametric* confidence intervals for the risk measures of Section 2.1, one just has to transform the end-points of each interval in expressions (3.2) according to the corresponding formula of equations (2.7)–(2.10) and keep in mind that $x_0$ is known.

## 3.3 Robust Parametric Procedures

In the robust parametric modeling, parameters of the families $F_1$, $F_2$, $F_3$ are estimated using the *trimmed-mean*–type estimators. For Pareto and exponential distributions, such procedures and their applications to actuarial problems were extensively studied by Brazauskas and Serfling (2000, 2003) and by Brazauskas (2003). (In this section a similar approach will be applied to define trimmed-mean–type estimators for the lognormal distribution parameter $\mu$.) Besides trimmed means, these authors also considered the robust estimators based on *generalized medians*. For the lognormal distribution, the latter-type estimators were developed by Serfling (2002). The generalized-median–types estimators, though possessing favorable theoretical properties, are much more complex computationally and will not be included in the present study.

For specified $\delta_1$ and $\delta_2$ satisfying $0 \leq \delta_1, \delta_2 < \frac{1}{2}$, a trimmed mean (TM) is formed by discarding the proportion $\delta_1$ of the lowermost observations and the proportion $\delta_2$ of the uppermost observations and averaging the remaining ones in some sense. More specifically, for the Pareto distribution $F_1$, the TM estimator of $\gamma$ is given by

$$\hat{\gamma}_{\text{TM}} = \left( \sum_{i=1}^{n} d_{ni} (\log X_{(i)} - \log x_0) \right)^{-1},$$

with $d_{ni} = 0$, for $1 \leq i \leq [n\delta_1]$ and $n - [n\delta_2] + 1 \leq i \leq n$, and $= 1/d$, for $[n\delta_1] + 1 \leq i \leq n - [n\delta_2]$, where $[\cdot]$ denotes "greatest integer part," and

$$d = d(\delta_1, \delta_2, n) = \sum_{j=[n\delta_1]+1}^{n-[n\delta_2]} \sum_{k=0}^{j-1} (n - k)^{-1}.$$

For the exponential distribution $F_3$, the TM estimator of $\theta$ is given by

$$\hat{\theta}_{TM} = \sum_{i=1}^{n} d_{ni}\,(X_{(i)} - x_0),$$

with the same choice of coefficients $d_{ni}$ as above. In a similar vein, for the lognormal distribution $F_2$, the TM estimator of $\mu$ (with $\delta_1 = \delta_2 = \delta$) is given by

$$\hat{\mu}_{TM} = \frac{1}{n - 2[n\delta]} \sum_{i=[n\delta]+1}^{n-[n\delta]} \log(X_{(i)} - x_0).$$

Note that, for $\delta_1 = \delta_2 = \delta = 0$, the estimators $\hat{\gamma}_{TM}$, $\hat{\mu}_{TM}$, $\hat{\theta}_{TM}$ become $\hat{\gamma}_{MLE}$, $\hat{\mu}_{MLE}$, $\hat{\theta}_{MLE}$, respectively.

Further, since all these TM estimators fall in a general class of $L$-statistics, their asymptotic distribution is readily available (see, e.g., Serfling 1980, Chapter 8). That is, the estimators $\hat{\gamma}_{TM}$, $\hat{\mu}_{TM}$, and $\hat{\theta}_{TM}$ are each asymptotically normal with, respectively, the means $\gamma$, $\mu$, and $\theta$ and the variances $C_{\delta_1,\delta_2}\,\gamma^2/n$, $K_\delta/n$, and $C_{\delta_1,\delta_2}\,\theta^2/n$, where constants $C_{\delta_1,\delta_2}$ and $K_\delta$ represent the loss of efficiency of the TM estimator relative to the MLE, at the model, and are equal to

| $\delta_1 = \delta_2 = \delta$ | 0.00 | 0.05 | 0.15 | 0.45 |
|---|---|---|---|---|
| $C_{\delta_1,\delta_2}$ | 1.00 | 1.090 | 1.271 | 1.946 |
| $K_\delta$ | 1.00 | 1.026 | 1.100 | 1.474 |

Thus, the $100(1 - \alpha)\%$ confidence intervals for the parameters $\gamma$, $\mu$, and $\theta$ are given by

$$\hat{\gamma}_{TM}(1 \pm z_{\alpha/2}\sqrt{C_{\delta_1,\delta_2}/n}),$$

$$\hat{\mu}_{TM} \pm z_{\alpha/2}\sqrt{K_\delta/n},$$

$$\hat{\theta}_{TM}\,(1 \pm z_{\alpha/2}\sqrt{C_{\delta_1,\delta_2}/n}), \tag{3.3}$$

where $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution.

Finally, to get the *robust parametric* confidence intervals for the risk measures of Section 2.1, one just has to transform the end-points of each interval in expression (3.3) according to the corresponding formula of equations (2.7)–(2.10).

## 4. ILLUSTRATION

As an illustration of the ideas discussed in the previous sections, we show how the interval estimators of all risk measures work on real data. For this, we use the Norwegian fire claims data, taken from Beirlant, Teugels, and Vynckier (1996), which has been studied in the actuarial literature. It represents the total damage done by 142 fires in Norway for the year 1975, which exceed 500,000 Norwegian krones. (For convenience the data are presented in Appendix, Table A.1.) To avoid clustering of data due to rounding, we degroup the data by following the degrouping method of Brazauskas and Serfling (2003). It is based on the assumption that claims are *uniformly distributed* around an integer observation. For example, instead of four observations of ''650'' we take the expected value of four-order statistics of the random variable distributed uniformly on the interval (649.5; 650.5): that is, the observations 650, 650, 650, 650 are replaced with

649.700, 649.900, 650.100, 650.300.

All diagnostics and estimation in this section are done using the degrouped data (measured in millions of Norwegian krones).

In Figure 3 we illustrate the results of preliminary diagnostics—histogram and quantile-quantile plots (QQ-plots)—for the data set at hand. While the histogram suggests that any of the three distributions of Section 2.3 ($F_1$, $F_2$, $F_3$, all with $x_0 = 0.5$) might be appropriate for the Norwegian fire claims, the QQ-plot approach reveals the following:

- A truncated exponential distribution is *inappropriate* for this data set

Figure 3
**Preliminary Diagnostics for the Norwegian Fire Claims (1975) Data**
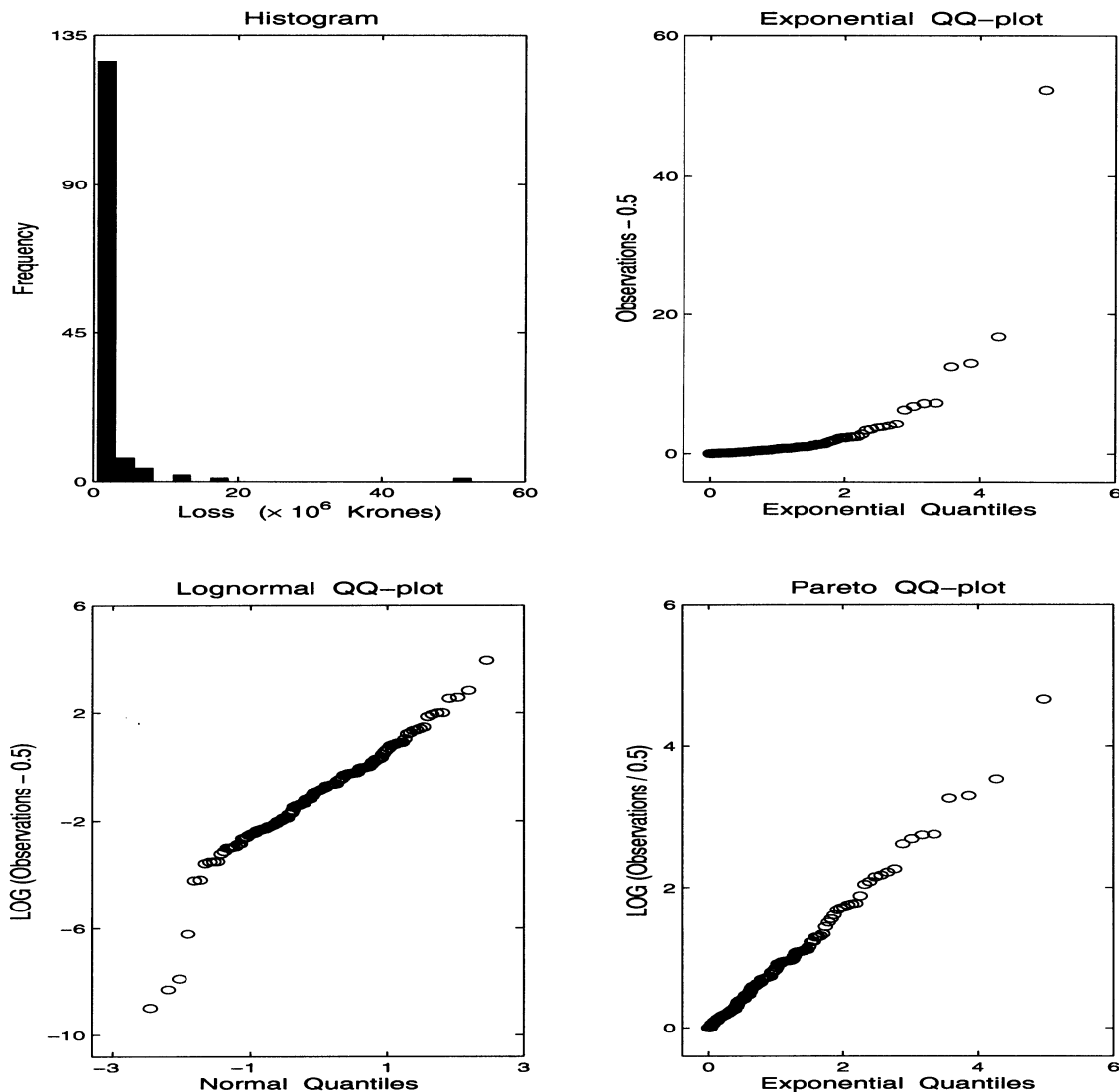


- In comparison with the exponential case, the lognormal QQ-plot shows *mild improvement* and
- The Pareto QQ-plot exhibits *nearly perfect fit* for the Norwegian fire claims data.

Fitting a Pareto model (with $x_0 = 0.5$) to these claims, using the MLE and TM($\delta_1 = \delta_2$) procedures, yields the following point and 95% confidence interval estimates for $\gamma$: $\hat{\gamma}_{MLE} = 1.218$ and $(1.017; 1.418)$; $\hat{\gamma}_{TM, .05} = 1.220$ and $(1.017; 1.430)$; $\hat{\gamma}_{TM, .15} = 1.236$ and $(1.007; 1.465)$; $\hat{\gamma}_{TM, .45} = 1.173$ and $(0.904; 1.442)$. The narrowness of the range of $\hat{\gamma}$ (ranging from 1.173 to 1.236) points to a very good fit between the data and the Pareto model, which was initially suggested by the QQ-plot. Also, such low values of $\hat{\gamma}$ indicate that the underlying model is heavy-tailed, and, intuitively, we would expect the claim data to be "risky." In Table 1 we provide a summary of formal estimates of the "riskiness" of this data set.

Clearly, the "wrong" models (i.e., lognormal and exponential) lead to quite different risk evaluations from those based on Pareto distribution or nonparametric approaches. It is also interesting to note that the robust estimators designed for $F_2$ and $F_3$ do not make significant corrections. Thus, preliminary

Table 1

**Empirical (EMP), Bootstrap (BOOT) with *B* = 1,000, Parametric (MLE), and Robust (TM) Parametric Point and 95% Confidence Interval Estimates of Selected Risk Measures**

| Method of Estimation | PHT($r$ = 0.85) | | WT($\lambda$ = 0.25) | | VaR($\beta$ = 0.05) | | CTE($\beta$ = 0.05) | |
|---|---|---|---|---|---|---|---|---|
| | Point | Interval | Point | Interval | Point | Interval | Point | Interval |
| EMP | 2,736 | (1,463; 4,010) | 2,769 | (1,474; 4,100) | 4,810 | (3,860; 13,000) | 15,769 | (2,812; 28,726) |
| BOOT | 2,681 | (1,660; 4,061) | 2,760 | (1,702; 4,253) | 5,658 | (2,953; 7,834) | 15,888 | (6,318; 31,663) |
| MLE ($F_1$) | 14,816 | (2,937; $\infty$) | 5,148 | (4,491; 6,063) | 5,855 | (4,136; 9,503) | 32,767 | (14,035; $0.6 \times 10^6$) |
| TM (0.05, $F_1$) | 13,897 | (2,821; $\infty$) | 5,068 | (4,398; 6,015) | 5,822 | (4,063; 9,685) | 32,239 | (13,512; $0.9 \times 10^6$) |
| TM (0.15, $F_1$) | 10,399 | (2,538; $\infty$) | 4,668 | (4,016; 5,617) | 5,645 | (3,864; 9,802) | 29,576 | (12,172; $1.5 \times 10^6$) |
| TM (0.45, $F_1$) | $\infty$ | — | 6,867 | (5,679; 8,763) | 6,432 | (3,993; 13,762) | 43,661 | (13,030; $\infty$) |
| MLE ($F_2$) | 1,199 | (1,093; 1,325) | 1,229 | (1,118; 1,359) | 2,284 | (2,013; 2,603) | 3,445 | (2,998; 3,971) |
| TM (0.05, $F_2$) | 1,260 | (1,143; 1,398) | 1,292 | (1,171; 1,436) | 2,439 | (2,141; 2,790) | 3,701 | (3,209; 4,281) |
| TM (0.15, $F_2$) | 1,251 | (1,132; 1,393) | 1,284 | (1,159; 1,431) | 2,417 | (2,113; 2,778) | 3,665 | (3,164; 4,261) |
| TM (0.45, $F_2$) | 1,322 | (1,174; 1,504) | 1,357 | (1,202; 1,547) | 2,598 | (2,218; 3,062) | 3,964 | (3,337; 4,729) |
| MLE ($F_3$) | 2,286 | (1,992; 2,580) | 2,390 | (2,079; 2,701) | 5,047 | (4,299; 5,795) | 6,565 | (5,568; 7,563) |
| TM (0.05, $F_3$) | 1,527 | (1,351; 1,703) | 1,587 | (1,400; 1,773) | 3,115 | (2,666; 3,564) | 3,988 | (3,390; 4,588) |
| TM (0.15, $F_3$) | 1,282 | (1,137; 1,427) | 1,327 | (1,174; 1,480) | 2,490 | (2,121; 2,859) | 3,155 | (2,662; 3,647) |
| TM (0.45, $F_3$) | 1,189 | (1,031; 1,347) | 1,229 | (1,062; 1,396) | 2,254 | (1,851; 2,656) | 2,839 | (2,302; 3,375) |

*Note:* MLE and TM estimates are computed for Pareto ($F_1$), lognormal ($F_2$), and exponential ($F_3$) distributions.

diagnostics is necessary to get some (*approximate*) knowledge about the underlying distribution because robust procedures designed for an inappropriate model do not fix the situation. Further, as expected, the empirical and bootstrap estimates are reasonably close. However, they both differ (substantially, for the PHT and CTE measures, and somewhat, for the WT and VaR measures) from the Pareto-based estimates, which suggest that the data set is very risky (in some cases, infinitely risky) according to all risk measures. The infinite estimates occur because of the theoretical relationships between the risk measure parameters and Pareto $\gamma$. Indeed, it follows from equations (2.7) and (2.10) that PHT($F_1$, $r$) = $\infty$, for $\gamma \le 1/r$, and CTE($F_1$, $\beta$) = $\infty$, for $\gamma \le 1$.

## 5. Comparisons

In this section we summarize findings of the Monte Carlo simulation study with a special focus on the sensitivity to assumptions issue (see the objectives (a), (b), (c) of Section 2.2).

Table 2 provides comparisons between the interval estimators of the WT($\lambda$ = 0.25) = 1.286 measure, under the "clean" and "contaminated" data scenarios, where $n$ = 25, 50, 100, 250. As one can see, under the "clean" data scenario, coverage proportions of the nonparametric intervals are below the nominal level in small samples (0.81–0.91 for $n$ = 25) for all distributions. These proportions increase as sample size gets larger (0.85–0.92 for $n$ = 50; 0.88–0.93 for $n$ = 100; 0.91–0.94 for $n$ = 250). Overall, coverages of the nonparametric intervals get reasonably close to the intended 95% confidence level for $n \ge 250$. Parametric and robust parametric intervals, on the other hand, perform very well with respect to the coverage criterion having coverage proportions of at least 0.92 for all distributions and in samples as small as $n$ = 25. Further, as predicted by asymptotic theory, for large sample size (e.g., $n \ge 100$) the MLE intervals dominate empirical, bootstrap, and TM counterparts with respect to the length criterion. These conclusions change, however, when we consider the "contaminated" data scenario. First, the length of nonparametric intervals increases by more than 20 times for all distributions and for all sample sizes. Although the coverage proportions of nonparametric intervals are not bad for $n$ = 25 (between 0.80 and 0.88), they decrease dramatically (0.57–0.73 for $n$ = 50; 0.17–0.27 for $n$ = 100) becoming virtually 0 for $n$ = 250. Second, the parametric intervals, though significantly less inflated (e.g., the largest increase of length is by a factor of about 7, for $F_3$ and all $n$), have very poor coverage proportions that converge to 0–7% coverage for $n$ = 250. Third, the robust parametric intervals are also affected by outliers, but overall they perform quite well. Except for the TM(0.05) estimator, for $F_3$ and $n \le 50$, the lengths of TM intervals do not change much (by a factor of less than

Table 2

**Length and Proportion of Coverage of 95% Empirical (EMP), Bootstrap (BOOT), Parametric (MLE), Robust Parametric (TM) Confidence Intervals of WT($\lambda = 0.25$) = 1.286, for Selected Sample Size $n$, When Data Are Generated by ''Clean'' and ''Contaminated'' Pareto ($F_1$), Lognormal ($F_2$), and Exponential ($F_3$) Distributions (Standard Errors Given in Parentheses)**

| Data | Method of Estimation | n = 25 Length | n = 25 Coverage | n = 50 Length | n = 50 Coverage | n = 100 Length | n = 100 Coverage | n = 250 Length | n = 250 Coverage |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | ''Clean'' data scenario: $\varepsilon = 0.00$ | | | | |
| $F_1$ | EMP | 0.24 (0.001) | 0.84 (0.004) | 0.19 (0.001) | 0.87 (0.002) | 0.14 (0.000) | 0.90 (0.004) | 0.09 (0.000) | 0.92 (0.003) |
| | BOOT | 0.24 (0.001) | 0.87 (0.005) | 0.18 (0.001) | 0.89 (0.002) | 0.14 (0.002) | 0.91 (0.009) | 0.09 (0.000) | 0.93 (0.003) |
| | MLE | 0.25 (0.000) | 0.96 (0.002) | 0.17 (0.000) | 0.95 (0.002) | 0.12 (0.000) | 0.95 (0.002) | 0.07 (0.000) | 0.95 (0.002) |
| | TM (0.05) | 0.27 (0.000) | 0.96 (0.002) | 0.18 (0.000) | 0.95 (0.002) | 0.12 (0.000) | 0.95 (0.002) | 0.08 (0.000) | 0.95 (0.003) |
| | TM (0.15) | 0.31 (0.001) | 0.96 (0.002) | 0.20 (0.000) | 0.95 (0.003) | 0.13 (0.000) | 0.95 (0.002) | 0.08 (0.000) | 0.95 (0.002) |
| | TM (0.45) | 0.43 (0.001) | 0.96 (0.002) | 0.26 (0.001) | 0.95 (0.002) | 0.17 (0.000) | 0.95 (0.002) | 0.10 (0.000) | 0.95 (0.001) |
| $F_2$ | EMP | 0.25 (0.002) | 0.81 (0.005) | 0.19 (0.001) | 0.85 (0.005) | 0.15 (0.001) | 0.88 (0.003) | 0.10 (0.000) | 0.91 (0.004) |
| | BOOT | 0.24 (0.003) | 0.84 (0.006) | 0.19 (0.002) | 0.87 (0.005) | 0.15 (0.001) | 0.89 (0.007) | 0.10 (0.001) | 0.92 (0.003) |
| | MLE | 0.23 (0.004) | 0.95 (0.002) | 0.16 (0.002) | 0.95 (0.001) | 0.11 (0.001) | 0.95 (0.002) | 0.07 (0.001) | 0.95 (0.002) |
| | TM (0.05) | 0.23 (0.000) | 0.95 (0.002) | 0.16 (0.000) | 0.95 (0.001) | 0.11 (0.000) | 0.95 (0.002) | 0.07 (0.000) | 0.95 (0.001) |
| | TM (0.15) | 0.24 (0.000) | 0.95 (0.001) | 0.17 (0.000) | 0.95 (0.001) | 0.12 (0.000) | 0.95 (0.002) | 0.07 (0.000) | 0.95 (0.002) |
| | TM (0.45) | 0.28 (0.000) | 0.96 (0.001) | 0.20 (0.000) | 0.95 (0.002) | 0.14 (0.000) | 0.95 (0.002) | 0.09 (0.000) | 0.95 (0.003) |
| $F_3$ | EMP | 0.20 (0.001) | 0.88 (0.004) | 0.15 (0.001) | 0.91 (0.005) | 0.11 (0.000) | 0.93 (0.003) | 0.07 (0.000) | 0.94 (0.003) |
| | BOOT | 0.20 (0.000) | 0.91 (0.007) | 0.15 (0.002) | 0.92 (0.002) | 0.11 (0.001) | 0.93 (0.004) | 0.07 (0.001) | 0.94 (0.002) |
| | MLE | 0.22 (0.003) | 0.93 (0.003) | 0.16 (0.000) | 0.94 (0.001) | 0.11 (0.000) | 0.95 (0.002) | 0.07 (0.000) | 0.95 (0.002) |
| | TM (0.05) | 0.23 (0.000) | 0.93 (0.003) | 0.17 (0.000) | 0.94 (0.002) | 0.12 (0.000) | 0.94 (0.002) | 0.07 (0.000) | 0.95 (0.002) |
| | TM (0.15) | 0.25 (0.000) | 0.93 (0.002) | 0.18 (0.000) | 0.94 (0.002) | 0.13 (0.000) | 0.95 (0.002) | 0.08 (0.000) | 0.95 (0.002) |
| | TM (0.45) | 0.31 (0.001) | 0.92 (0.003) | 0.22 (0.001) | 0.93 (0.002) | 0.16 (0.000) | 0.94 (0.003) | 0.10 (0.000) | 0.95 (0.002) |
| | | | | ''Contaminated'' data scenario: $\varepsilon = 0.05$, $H = U(10, 50)$ | | | | | |
| $F_1$ | EMP | 5.91 (0.066) | 0.87 (0.003) | 4.99 (0.031) | 0.73 (0.005) | 3.84 (0.007) | 0.27 (0.004) | 2.52 (0.003) | 0.00 (0.001) |
| | BOOT | 5.30 (0.028) | 0.82 (0.004) | 4.78 (0.046) | 0.59 (0.007) | 3.80 (0.018) | 0.17 (0.007) | 2.50 (0.003) | 0.00 (0.000) |
| | MLE | 0.47 (0.003) | 0.29 (0.006) | 0.32 (0.002) | 0.15 (0.006) | 0.22 (0.001) | 0.03 (0.003) | 0.13 (0.000) | 0.00 (0.000) |
| | TM (0.05) | 0.41 (0.002) | 0.58 (0.006) | 0.26 (0.002) | 0.50 (0.009) | 0.16 (0.000) | 0.48 (0.006) | 0.10 (0.000) | 0.26 (0.000) |
| | TM (0.15) | 0.36 (0.001) | 0.85 (0.003) | 0.22 (0.001) | 0.84 (0.006) | 0.15 (0.000) | 0.79 (0.004) | 0.09 (0.000) | 0.66 (0.000) |
| | TM (0.45) | 0.47 (0.001) | 0.92 (0.002) | 0.28 (0.001) | 0.90 (0.004) | 0.18 (0.000) | 0.88 (0.002) | 0.11 (0.000) | 0.83 (0.000) |
| $F_2$ | EMP | 5.88 (0.045) | 0.86 (0.004) | 5.00 (0.024) | 0.72 (0.005) | 3.84 (0.011) | 0.27 (0.006) | 2.52 (0.004) | 0.00 (0.000) |
| | BOOT | 5.34 (0.025) | 0.82 (0.004) | 4.76 (0.044) | 0.61 (0.006) | 3.79 (0.014) | 0.18 (0.005) | 2.50 (0.003) | 0.00 (0.000) |
| | MLE | 0.31 (0.001) | 0.67 (0.004) | 0.21 (0.001) | 0.54 (0.007) | 0.15 (0.000) | 0.32 (0.005) | 0.09 (0.000) | 0.07 (0.003) |
| | TM (0.05) | 0.28 (0.001) | 0.78 (0.003) | 0.20 (0.001) | 0.73 (0.004) | 0.13 (0.000) | 0.70 (0.005) | 0.08 (0.000) | 0.51 (0.006) |
| | TM (0.15) | 0.27 (0.000) | 0.89 (0.004) | 0.19 (0.000) | 0.89 (0.003) | 0.13 (0.000) | 0.85 (0.003) | 0.08 (0.000) | 0.75 (0.004) |
| | TM (0.45) | 0.31 (0.001) | 0.93 (0.002) | 0.21 (0.000) | 0.92 (0.003) | 0.15 (0.000) | 0.90 (0.002) | 0.09 (0.000) | 0.85 (0.003) |
| $F_3$ | EMP | 5.90 (0.029) | 0.88 (0.002) | 4.99 (0.031) | 0.73 (0.004) | 3.85 (0.008) | 0.27 (0.005) | 2.52 (0.003) | 0.00 (0.001) |
| | BOOT | 5.28 (0.032) | 0.80 (0.004) | 4.76 (0.027) | 0.57 (0.004) | 3.78 (0.023) | 0.17 (0.005) | 2.51 (0.006) | 0.00 (0.000) |
| | MLE | 1.60 (0.009) | 0.24 (0.005) | 1.15 (0.001) | 0.07 (0.007) | 0.82 (0.003) | 0.01 (0.001) | 0.52 (0.001) | 0.00 (0.000) |
| | TM (0.05) | 0.85 (0.008) | 0.59 (0.006) | 0.49 (0.005) | 0.51 (0.006) | 0.23 (0.002) | 0.52 (0.005) | 0.12 (0.001) | 0.30 (0.006) |
| | TM (0.15) | 0.34 (0.002) | 0.91 (0.003) | 0.20 (0.000) | 0.93 (0.002) | 0.14 (0.000) | 0.88 (0.003) | 0.09 (0.000) | 0.75 (0.006) |
| | TM (0.45) | 0.33 (0.000) | 0.95 (0.002) | 0.24 (0.001) | 0.95 (0.003) | 0.17 (0.000) | 0.94 (0.003) | 0.11 (0.000) | 0.90 (0.004) |

2 in all cases). Moreover, as the level of trimming ($\delta_1 = \delta_2$) increases, the robust estimators maintain stable interval lengths and coverage proportions, under both data-generating scenarios, for all $F$ and for all $n$. For instance, for the TM(0.45) intervals, for $F_2$, the length vector changes from 0.28, 0.20, 0.14, 0.09 (''clean'' scenario) to 0.31, 0.21, 0.15, 0.09 (''contaminated'' scenario), and the coverage proportions change from 0.96, 0.95, 0.95, 0.95 (''clean'' scenario) to 0.93, 0.92, 0.90, 0.85 (''contaminated'' scenario).

Very similar patterns were observed for the other three risk measures: PHT($r = 0.85$) = 1.272, VaR($\beta = 0.05$) = 1.724, CTE($\beta = 0.05$) = 2.107. To make this claim more credible, we include in Appendix (see Table A.2) an equivalent table for the VaR($\beta = 0.05$) = 1.724 measure. There we notice that the inflation of lengths of nonparametric intervals is even more dramatic; coverage proportions, however, are slightly better than in the WT($\lambda = 0.25$) = 1.286 case.

Furthermore, in the analysis of the Norwegian Fire Claims data we observed that application of robust procedures designed for the wrong model yields erroneous estimates. In the simulation study we examined this problem in more detail. In Table 3 we present a summary of this investigation for the

Table 3

**Length and Proportion of Coverage of the 95% Empirical (EMP), Bootstrap (BOOT), Parametric (MLE), Robust Parametric (TM) Confidence Intervals of CTE($\beta$ = 0.05) = 2.107, for Sample Size *n* = 100, When Data Are Generated by "Clean" and "Contaminated" Pareto ($F_1$), Lognormal ($F_2$), and Exponential ($F_3$) Distributions (Standard Errors Given in Parentheses)**

| Distributional Assumption | Method of Estimation | Distribution of Generated Data | | | | | |
| | | Pareto ($F_1$) | | Lognormal ($F_2$) | | Exponential ($F_3$) | |
| | | Length | Coverage | Length | Coverage | Length | Coverage |
| "Clean" data scenario: $\varepsilon$ = 0.00 | | | | | | | |
| | EMP | 0.92 (0.006) | 0.81 (0.003) | 1.07 (0.011) | 0.79 (0.005) | 0.64 (0.005) | 0.85 (0.002) |
| | BOOT | 0.89 (0.009) | 0.79 (0.005) | 1.04 (0.008) | 0.78 (0.009) | 0.60 (0.004) | 0.82 (0.006) |
| $F_1$ | MLE | 0.68 (0.003) | 0.95 (0.003) | 0.58 (0.002) | 0.85 (0.004) | 1.03 (0.002) | 0.41 (0.004) |
| | TM (0.05) | 0.73 (0.003) | 0.95 (0.000) | 0.67 (0.002) | 0.96 (0.001) | 1.22 (0.002) | 0.24 (0.003) |
| | TM (0.15) | 0.80 (0.004) | 0.95 (0.001) | 0.74 (0.002) | 0.97 (0.001) | 1.46 (0.003) | 0.21 (0.003) |
| | TM (0.45) | 1.06 (0.005) | 0.95 (0.002) | 0.98 (0.004) | 0.98 (0.001) | 2.17 (0.003) | 0.27 (0.003) |
| $F_2$ | MLE | 0.38 (0.001) | 0.65 (0.001) | 0.44 (0.001) | 0.95 (0.003) | 0.53 (0.001) | 0.53 (0.005) |
| | TM (0.05) | 0.41 (0.001) | 0.81 (0.011) | 0.45 (0.001) | 0.95 (0.003) | 0.58 (0.001) | 0.30 (0.005) |
| | TM (0.15) | 0.45 (0.001) | 0.88 (0.003) | 0.46 (0.001) | 0.95 (0.002) | 0.64 (0.001) | 0.17 (0.003) |
| | TM (0.45) | 0.56 (0.001) | 0.88 (0.004) | 0.54 (0.001) | 0.95 (0.004) | 0.79 (0.001) | 0.14 (0.006) |
| $F_3$ | MLE | 0.35 (0.001) | 0.35 (0.006) | 0.33 (0.001) | 0.23 (0.004) | 0.43 (0.001) | 0.94 (0.005) |
| | TM (0.05) | 0.34 (0.001) | 0.20 (0.005) | 0.32 (0.001) | 0.08 (0.003) | 0.45 (0.001) | 0.94 (0.002) |
| | TM (0.15) | 0.36 (0.001) | 0.16 (0.004) | 0.34 (0.000) | 0.05 (0.002) | 0.49 (0.001) | 0.94 (0.001) |
| | TM (0.45) | 0.42 (0.001) | 0.21 (0.004) | 0.41 (0.001) | 0.09 (0.004) | 0.61 (0.002) | 0.94 (0.001) |
| "Contaminated" data scenario: $\varepsilon$ = 0.05, $H = U(10, 50)$ | | | | | | | |
| | EMP | 38.55 (0.167) | 0.42 (0.006) | 38.36 (0.177) | 0.42 (0.010) | 38.42 (0.196) | 0.42 (0.004) |
| | BOOT | 31.69 (0.049) | 0.18 (0.005) | 31.76 (0.073) | 0.19 (0.002) | 31.74 (0.042) | 0.18 (0.003) |
| $F_1$ | MLE | 3.35 (0.038) | 0.03 (0.002) | 2.96 (0.036) | 0.06 (0.003) | 4.71 (0.054) | 0.00 (0.001) |
| | TM (0.05) | 1.41 (0.018) | 0.48 (0.006) | 1.31 (0.014) | 0.55 (0.008) | 2.21 (0.024) | 0.04 (0.003) |
| | TM (0.15) | 0.99 (0.004) | 0.80 (0.007) | 0.90 (0.001) | 0.91 (0.004) | 1.81 (0.008) | 0.06 (0.005) |
| | TM (0.45) | 1.24 (0.004) | 0.89 (0.005) | 1.12 (0.002) | 0.96 (0.001) | 2.60 (0.013) | 0.13 (0.007) |
| $F_2$ | MLE | 0.51 (0.002) | 0.62 (0.009) | 0.58 (0.001) | 0.33 (0.007) | 0.69 (0.002) | 0.07 (0.005) |
| | TM (0.05) | 0.48 (0.002) | 0.75 (0.010) | 0.51 (0.001) | 0.67 (0.008) | 0.66 (0.002) | 0.10 (0.005) |
| | TM (0.15) | 0.50 (0.001) | 0.81 (0.007) | 0.50 (0.000) | 0.85 (0.004) | 0.69 (0.002) | 0.06 (0.007) |
| | TM (0.45) | 0.60 (0.002) | 0.77 (0.006) | 0.57 (0.001) | 0.90 (0.006) | 0.85 (0.003) | 0.06 (0.005) |
| $F_3$ | MLE | 2.60 (0.012) | 0.00 (0.001) | 2.59 (0.020) | 0.01 (0.001) | 2.70 (0.010) | 0.01 (0.001) |
| | TM (0.05) | 0.71 (0.009) | 0.39 (0.004) | 0.69 (0.010) | 0.30 (0.013) | 0.83 (0.004) | 0.52 (0.007) |
| | TM (0.15) | 0.40 (0.001) | 0.49 (0.010) | 0.38 (0.000) | 0.30 (0.006) | 0.55 (0.001) | 0.88 (0.003) |
| | TM (0.45) | 0.46 (0.001) | 0.39 (0.010) | 0.44 (0.001) | 0.24 (0.007) | 0.65 (0.002) | 0.94 (0.002) |

CTE($\beta$ = 0.05) = 2.107 measure, for sample size $n$ = 100. (Conclusions remain the same for the other three measures and for other choices of $n$.) It is of interest to see what happens to all procedures when diagnostics step is inconclusive or is simply ignored. For example, if data were generated according to the Pareto model $F_1$ but the actuary estimates the CTE($\beta$ = 0.05) measure, which is equal to 2.107 for *all three* distributions, by assuming the lognormal distribution $F_2$, the coverage proportions of the MLE and TM estimators are not too bad: 0.65, 0.81, 0.88, 0.88. They are even better when one assumes $F_1$ for the $F_2$ data: 0.85, 0.96, 0.97, 0.98. However, this "success story" Pareto-for-lognormal and lognormal-for-Pareto seems to be accidental because none of the parametric or robust estimators designed for $F_1$ or $F_2$ work when data were generated by the exponential distribution $F_3$. In addition, consideration of the "contaminated" data scenarios eliminates MLE and TM(0.05) estimators from the list of reasonably good performers. Hence, only sufficiently robust estimators, that is, TM(0.15) and TM(0.45), designed for the right model work consistently well.

    Finally, the last conclusion becomes even more evident when we graphically summarize the simulation study by plotting coverage proportions of all estimators for all four risk measures, under the "clean" and "contaminated" data scenarios. In each plot of Figure 4, estimators are displayed as points (asterisks), dashed lines represent the nominal 0.95 levels of coverage, and the distribution marked inside the plot shows which family was used to generate either "clean" (vertical axis) or "contami-

Figure 4

## Proportions of Coverage of the 95% Confidence Intervals of Selected Risk Measures, under "Clean" ($\varepsilon = 0$) and "Contaminated" ($\varepsilon = 0.05$, $H = U(10, 50)$) Data Scenarios, When $n = 250$
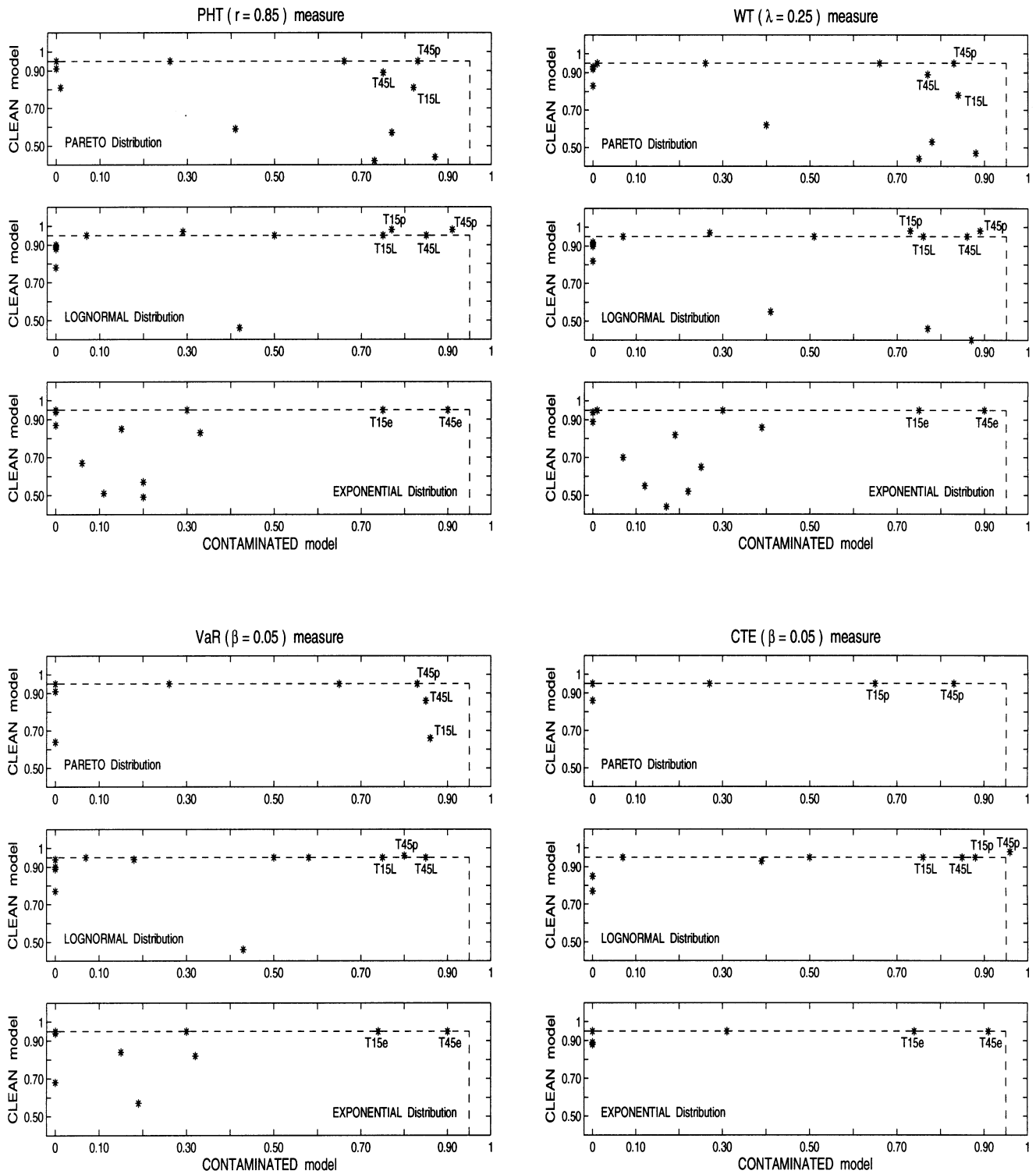
nated'' (horizontal axis) data. Only most favorable estimators being located in the upper right corner (or northeast direction) of the plot, reflecting high coverages under both scenarios, are labeled. For convenience, the following notation for the estimators is used: TM(0.15) and TM(0.45), designed for $F_1$, are denoted as T15p and T45p, respectively; TM(0.15) and TM(0.45), designed for $F_2$, are denoted as T15L and T45L, respectively; TM(0.15) and TM(0.45), designed for $F_3$, are denoted as T15e and T45e, respectively. Although we only present plots for sample size $n = 250$, other choices of $n$ do not change the conclusion.

## 6. DISCUSSION

The findings of this paper combined with the conclusions of Brazauskas and Kaiser (2004) and Brazauskas et al. (2006) suggest the following answers to the questions raised in Section 2.2.

a. *Convergence rates*
   Convergence of the proportion of coverage of the nonparametric intervals is slow and depends on the function $\psi$. For ''light'' $\psi$ (e.g., for PHT with $0.85 \leq r < 1$), the coverage levels of these intervals get reasonably close to the nominal level for $n \geq 100$ and for all distributions $F$ that we considered. For ''severe'' $\psi$ (e.g., for the PHT measure with $r \leq 0.70$), however, their performances are unacceptable even for $n = 1,500$. On the other hand, parametric and robust intervals attain the intended confidence levels for all $\psi$ and $F$, and for sample sizes as small as $n = 50$.
b. *Performance at the model*
   At the assumed model $F$, robust and parametric intervals perform better than nonparametric intervals with respect to the coverage criterion. Also, for $n \geq 250$, parametric intervals dominate robust and nonparametric counterparts with respect to the length criterion. Superior overall performance of parametric intervals should not come as a surprise because they are equipped with additional information, namely, the knowledge of underlying $F$.
c. *Sensitivity to assumptions*
   When the assumed model $F$ is contaminated or cannot be identified with the 100% accuracy (this is almost always true in practice!), both parametric and nonparametric procedures perform poorly. In such situations, only the *sufficiently* robust estimators, designed for that model, yield intervals with *consistently* satisfactory performance. Also, it is quite obvious that preliminary diagnostics is *always* necessary because even highly robust estimators, if designed for the wrong model, will not save a flawed analysis, which, in turn, will lead to erroneous risk estimates.

Further, let us summarize what we wanted to accomplish with this work. The objective of this paper is not to discredit nonparametric methods as an approach for measuring the riskiness of portfolios or for solving some other actuarial problem. Rather, we want the reader to be aware that under some idealistic scenarios (when the assumed distribution is *exactly* correct, but it is impossible to prove that in practice) or under some realistic scenarios (when the assumed distribution is *approximately* correct, and that's what one can check in practice) there are better ways for solving problems. In particular, the robust parametric approach is quite general and relatively simple:

• Use preliminary diagnostics and goodness-of-fit techniques to decide what distribution approximately fits the data, and
• Use robust intervals (designed for that distribution) to estimate the quantity of interest.

The resulting intervals will have (approximately) the intended coverage level.

Finally, let us mention a problem for further research, which goes beyond the scope of the present paper but is of practical interest. In practice, an actuary has one data set that is updated over time. Thus, instead of seeking general solutions, he or she may be satisfied with a methodology that works for that particular data set. if that's the case, then the so-called cross-validation techniques are appropriate, and they work as follows. One splits the data into two parts: one part for finding a solution and the other for its verification/comparison. To find the solution, nonparametric, parametric, or robust parametric methods can be employed. Thus, it would be interesting to see whether, in this context, the robust parametric approach still maintains an edge over the other methodologies.

# APPENDIX

### Table A.1
### Norwegian Fire Claims (1975) Data (×1,000 Norwegian Krones)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 500 | 552 | 600 | 650 | 798 | 948 | 1,180 | 1,479 | 2,497 | 7,371 |
| 500 | 557 | 605 | 672 | 800 | 957 | 1,243 | 1,485 | 2,690 | 7,772 |
| 500 | 558 | 610 | 674 | 800 | 1,000 | 1,248 | 1,491 | 2,760 | 7,834 |
| 502 | 570 | 610 | 680 | 800 | 1,002 | 1,252 | 1,515 | 2,794 | 13,000 |
| 515 | 572 | 613 | 700 | 826 | 1,009 | 1,280 | 1,519 | 2,886 | 13,484 |
| 515 | 574 | 615 | 725 | 835 | 1,013 | 1,285 | 1,587 | 2,924 | 17,237 |
| 528 | 579 | 620 | 728 | 862 | 1,020 | 1,291 | 1,700 | 2,953 | 52,600 |
| 530 | 583 | 622 | 736 | 885 | 1,024 | 1,293 | 1,708 | 3,289 | |
| 530 | 584 | 632 | 737 | 900 | 1,033 | 1,298 | 1,820 | 3,860 | |
| 530 | 586 | 635 | 740 | 900 | 1,038 | 1,300 | 1,822 | 4,016 | |
| 540 | 593 | 635 | 748 | 910 | 1,041 | 1,305 | 1,848 | 4,300 | |
| 544 | 596 | 640 | 752 | 912 | 1,104 | 1,327 | 1,906 | 4,397 | |
| 550 | 596 | 650 | 756 | 927 | 1,108 | 1,387 | 2,110 | 4,585 | |
| 550 | 600 | 650 | 756 | 940 | 1,137 | 1,455 | 2,251 | 4,810 | |
| 551 | 600 | 650 | 777 | 940 | 1,143 | 1,475 | 2,362 | 6,855 | |

*Source:* Beirlant, Teugels, and Vynckier (1996), Appendix I.

### Table A.2
### Length and Proportion of Coverage of 95% Empirical (EMP), Bootstrap (BOOT), Parametric (MLE), Robust Parametric (TM) Confidence Intervals of VaR($\beta = 0.05$) = 1.724, for Selected Sample Size $n$, When Data Are Generated by ''Clean'' and ''Contaminated'' Pareto ($F_1$), Lognormal ($F_2$), and Exponential ($F_3$) Distributions (Standard Errors Given in Parentheses)

| Data | Method of Estimation | $n = 25$ Length | $n = 25$ Coverage | $n = 50$ Length | $n = 50$ Coverage | $n = 100$ Length | $n = 100$ Coverage | $n = 250$ Length | $n = 250$ Coverage |
|---|---|---|---|---|---|---|---|---|---|
| | | ''Clean'' data scenario: $\varepsilon = 0.00$ | | | | | | | |
| $F_1$ | EMP | 0.69 (0.006) | 0.73 (0.004) | 0.88 (0.005) | 0.91 (0.003) | 0.66 (0.003) | 0.95 (0.001) | 0.37 (0.001) | 0.95 (0.001) |
| | BOOT | 0.78 (0.003) | 0.74 (0.005) | 0.54 (0.004) | 0.73 (0.003) | 0.48 (0.003) | 0.88 (0.006) | 0.33 (0.002) | 0.92 (0.003) |
| | MLE | 0.91 (0.004) | 0.95 (0.002) | 0.58 (0.002) | 0.95 (0.002) | 0.39 (0.001) | 0.95 (0.001) | 0.24 (0.000) | 0.95 (0.001) |
| | TM (0.05) | 1.05 (0.004) | 0.95 (0.003) | 0.64 (0.002) | 0.95 (0.002) | 0.42 (0.001) | 0.95 (0.002) | 0.25 (0.000) | 0.95 (0.003) |
| | TM (0.15) | 1.21 (0.006) | 0.96 (0.003) | 0.71 (0.002) | 0.95 (0.002) | 0.45 (0.001) | 0.95 (0.002) | 0.27 (0.000) | 0.95 (0.002) |
| | TM (0.45) | 1.98 (0.018) | 0.96 (0.003) | 0.98 (0.004) | 0.95 (0.002) | 0.59 (0.002) | 0.95 (0.002) | 0.34 (0.001) | 0.95 (0.002) |
| $F_2$ | EMP | 0.79 (0.011) | 0.72 (0.005) | 1.06 (0.008) | 0.91 (0.002) | 0.77 (0.005) | 0.95 (0.002) | 0.42 (0.002) | 0.95 (0.003) |
| | BOOT | 0.89 (0.015) | 0.74 (0.008) | 0.60 (0.007) | 0.73 (0.009) | 0.54 (0.005) | 0.88 (0.002) | 0.37 (0.002) | 0.92 (0.005) |
| | MLE | 0.58 (0.001) | 0.95 (0.004) | 0.41 (0.001) | 0.95 (0.002) | 0.29 (0.000) | 0.95 (0.002) | 0.18 (0.000) | 0.95 (0.003) |
| | TM (0.05) | 0.59 (0.001) | 0.95 (0.003) | 0.42 (0.001) | 0.95 (0.002) | 0.29 (0.000) | 0.95 (0.003) | 0.18 (0.001) | 0.95 (0.003) |
| | TM (0.15) | 0.61 (0.002) | 0.96 (0.003) | 0.43 (0.001) | 0.95 (0.002) | 0.30 (0.000) | 0.95 (0.002) | 0.19 (0.000) | 0.95 (0.002) |
| | TM (0.45) | 0.72 (0.002) | 0.95 (0.003) | 0.50 (0.001) | 0.95 (0.002) | 0.35 (0.000) | 0.95 (0.002) | 0.22 (0.000) | 0.95 (0.003) |
| $F_3$ | EMP | 0.50 (0.003) | 0.73 (0.006) | 0.59 (0.004) | 0.91 (0.003) | 0.47 (0.002) | 0.95 (0.002) | 0.28 (0.001) | 0.95 (0.002) |
| | BOOT | 0.59 (0.002) | 0.74 (0.005) | 0.42 (0.003) | 0.73 (0.003) | 0.36 (0.002) | 0.88 (0.006) | 0.25 (0.002) | 0.92 (0.005) |
| | MLE | 0.56 (0.001) | 0.93 (0.003) | 0.40 (0.001) | 0.94 (0.002) | 0.28 (0.000) | 0.94 (0.002) | 0.18 (0.000) | 0.95 (0.003) |
| | TM (0.05) | 0.58 (0.001) | 0.93 (0.002) | 0.42 (0.000) | 0.94 (0.002) | 0.30 (0.000) | 0.94 (0.003) | 0.19 (0.000) | 0.95 (0.003) |
| | TM (0.15) | 0.63 (0.001) | 0.94 (0.002) | 0.45 (0.001) | 0.94 (0.003) | 0.32 (0.000) | 0.95 (0.002) | 0.20 (0.000) | 0.95 (0.003) |
| | TM (0.45) | 0.78 (0.003) | 0.92 (0.005) | 0.56 (0.001) | 0.93 (0.003) | 0.40 (0.000) | 0.94 (0.002) | 0.25 (0.001) | 0.95 (0.003) |
| | | ''Contaminated'' data scenario: $\varepsilon = 0.05$, $H = U(10, 50)$ | | | | | | | |
| $F_1$ | EMP | 24.11 (0.072) | 0.83 (0.007) | 33.05 (0.056) | 0.78 (0.005) | 32.47 (0.063) | 0.61 (0.005) | 25.54 (0.026) | 0.21 (0.004) |
| | BOOT | 24.40 (0.042) | 0.92 (0.001) | 20.91 (0.022) | 0.90 (0.004) | 25.07 (0.043) | 0.70 (0.009) | 22.02 (0.023) | 0.30 (0.005) |
| | MLE | 4.41 (0.010) | 0.30 (0.007) | 2.14 (0.013) | 0.15 (0.004) | 1.27 (0.004) | 0.03 (0.002) | 0.73 (0.002) | 0.00 (0.001) |
| | TM (0.05) | 3.34 (0.090) | 0.57 (0.004) | 1.40 (0.010) | 0.49 (0.003) | 0.68 (0.002) | 0.49 (0.006) | 0.38 (0.001) | 0.26 (0.004) |
| | TM (0.15) | 1.86 (0.058) | 0.84 (0.004) | 0.86 (0.002) | 0.84 (0.002) | 0.54 (0.002) | 0.80 (0.004) | 0.32 (0.003) | 0.65 (0.005) |
| | TM (0.45) | 2.34 (0.022) | 0.92 (0.005) | 1.13 (0.003) | 0.91 (0.003) | 0.67 (0.004) | 0.90 (0.002) | 0.39 (0.005) | 0.83 (0.004) |
| $F_2$ | EMP | 24.29 (0.130) | 0.83 (0.004) | 33.18 (0.100) | 0.78 (0.003) | 32.26 (0.097) | 0.62 (0.005) | 25.51 (0.045) | 0.21 (0.003) |
| | BOOT | 24.57 (0.099) | 0.92 (0.004) | 20.84 (0.099) | 0.90 (0.002) | 24.98 (0.095) | 0.70 (0.007) | 22.02 (0.094) | 0.30 (0.007) |
| | MLE | 0.78 (0.003) | 0.67 (0.004) | 0.54 (0.001) | 0.54 (0.005) | 0.38 (0.008) | 0.34 (0.007) | 0.23 (0.002) | 0.07 (0.002) |
| | TM (0.05) | 0.72 (0.002) | 0.78 (0.003) | 0.50 (0.001) | 0.73 (0.003) | 0.33 (0.000) | 0.70 (0.006) | 0.21 (0.000) | 0.50 (0.005) |
| | TM (0.15) | 0.68 (0.002) | 0.89 (0.003) | 0.47 (0.001) | 0.89 (0.004) | 0.33 (0.000) | 0.86 (0.005) | 0.21 (0.000) | 0.75 (0.004) |
| | TM (0.45) | 0.77 (0.001) | 0.93 (0.002) | 0.54 (0.001) | 0.92 (0.003) | 0.37 (0.000) | 0.90 (0.004) | 0.23 (0.000) | 0.85 (0.004) |
| $F_3$ | EMP | 23.85 (0.100) | 0.83 (0.003) | 33.09 (0.100) | 0.78 (0.006) | 32.24 (0.100) | 0.62 (0.005) | 25.63 (0.084) | 0.20 (0.003) |
| | BOOT | 24.27 (0.160) | 0.92 (0.004) | 20.80 (0.097) | 0.90 (0.003) | 24.85 (0.093) | 0.71 (0.006) | 22.17 (0.090) | 0.29 (0.005) |
| | MLE | 3.83 (0.052) | 0.25 (0.008) | 2.79 (0.002) | 0.07 (0.005) | 1.96 (0.006) | 0.00 (0.001) | 1.25 (0.003) | 0.00 (0.000) |
| | TM (0.05) | 2.04 (0.033) | 0.60 (0.005) | 1.21 (0.015) | 0.51 (0.009) | 0.56 (0.005) | 0.53 (0.005) | 0.31 (0.001) | 0.30 (0.004) |
| | TM (0.15) | 0.84 (0.015) | 0.92 (0.004) | 0.51 (0.002) | 0.93 (0.002) | 0.36 (0.001) | 0.89 (0.003) | 0.22 (0.000) | 0.74 (0.003) |
| | TM (0.45) | 0.84 (0.002) | 0.95 (0.002) | 0.60 (0.002) | 0.95 (0.003) | 0.43 (0.002) | 0.95 (0.003) | 0.27 (0.000) | 0.90 (0.003) |

## Acknowledgment

### References

Albrecht, Peter. 2004. Risk Measures. In *Encyclopedia of Actuarial Science,* edited by B. Sundt and J. Teugels, vol. 3, pp. 1493–1501. London: John Wiley.

American Academy of Actuaries. 2002. Recommended Approach for Setting Regulatory Risk-Based Capital Requirements for Variable Products with Guarantees (Excluding Index Guarantees). Presented by the American Academy of Actuaries Life Capital Adequacy Subcommittee to the National Association of Insurance Commissioners, Life Risk-Based Capital Working Group, San Diego, CA, December 2002. www.actuary.org/pdf/life/rbc_16dec02.pdf.

Artzner, Philippe. 1999. Application of Coherent Risk Measures to Capital Requirements in Insurance. *North American Actuarial Journal* 3(2): 11–29.

Beirlant, Jan, Joseph L. Teugels, and Petra Vynckier. 1996. *Practical Analysis of Extreme Values.* Leuven: Leuven University Press.

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

Brazauskas, Vytaras, and Thomas Kaiser. 2004. Discussion: Empirical Estimation of Risk Measures and Related Quantities. *North American Actuarial Journal* 8(3): 114–17.

Brazauskas, Vytaras, and Robert 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.

Brazauskas, Vytaras, Bruce L. Jones, Madan L. Puri, and Ricardas Zitikis. 2006. Estimating Conditional Tail Expectations with Actuarial Applications in View. *Journal of Statistical Planning and Inference,* in press.

Canadian Institute of Actuaries. 2002. Report of the CIA Task Force on Segregated Fund Investment Guarantees. www.actuaries.ca/publications/2002/202012e.pdf.

Dowd, Kevin. 2004. Value-at-Risk. In *Encyclopedia of Actuarial Science,* edited by B. Sundt and J. Teugels, vol. 3, pp. 1740–48. London: Wiley.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Hardy, Mary R., and Julia L. Wirch. 2004. The Iterated CTE: A Dynamic Risk Measure. *North American Actuarial Journal* 8(4): 62–75.

Jones, Bruce L., Ricardas Zitikis. 2003. Empirical Estimation of Risk Measures and Related Quantities (with discussion). *North American Actuarial Journal* 7(4): 44–54. Discussion: 8(3): 114–17; reply: 8(3): 117–18.

Manistre, John B., and Geoffrey H. Hancock. 2005. Variance of the CTE Estimator. *North American Actuarial Journal* 9(2): 129–56.

Serfling, Robert J. 1980. *Approximation Theorems of Mathematical Statistics.* New York: John Wiley.

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

Wang, Shaun. 1998. An Actuarial Index of the Right-Tail Risk. *North American Actuarial Journal* 2(2): 88–101.

———. 2000. A Class of Distortion Operators for Pricing Financial and Insurance Risks. *Journal of Risk and Insurance* 67(1): 15–36.

———. 2002. A Universal Framework for Pricing Financial and Insurance Risks. *ASTIN Bulletin* 32(2): 213–34.

Wang, Shaun, Virginia Young, and Harry Panjer. 1997. Axiomatic Characterization of Insurance Prices. *Insurance: Mathematics and Economics* 21(2): 173–83.

Wirch, Julia, and Mary Hardy. 1999. A Synthesis of Risk Measures for Capital Adequacy. *Insurance: Mathematics and Economics* 25(3): 337–47.

*Discussions on this paper can be submitted until April 1, 2007. 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.*