Robust Fitting of Claim Severity Distributions and the Method of Trimmed Moments

Vytaras Brazauskas[1]
University of Wisconsin-Milwaukee

Bruce L. Jones[2]
University of Western Ontario

Ričardas Zitikis[3]
University of Western Ontario

Abstract. Many quantities arising in non-life insurance depend on claim severity distributions, which are usually modeled assuming a parametric form. Obtaining good estimates of the quantities, therefore, reduces to having good estimates of the model parameters. However, the notion of 'good estimate' depends on the problem at hand. For example, the maximum likelihood estimators (MLE) are efficient, but they generally lack robustness. Since outliers are common in insurance loss data, it is therefore important to have a method that allows one to balance between efficiency and robustness. Guided by this philosophy, in the present paper we suggest a general estimation method that we call the method of trimmed moments (MTM). This method is appropriate for various model-fitting situations including those for which a close fit in one or both tails of the distribution is not required. The MTM estimators can achieve various degrees of robustness, and they also allow the decision maker to easily see the actions of the estimators on the data, which makes them particularly appealing. We illustrate these features with detailed theoretical analyses and simulation studies of the MTM estimators in the case of location-scale families and several loss distributions such as lognormal and Pareto. As a further illustration, we analyze a real data set concerning hurricane damages in the United States from 1925 to 1995.

[1]Corresponding author: Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, U.S.A. E-mail address: `vytaras@uwm.edu`

[2]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada. E-mail address: `jones@stats.uwo.ca`

[3]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada. E-mail address: `zitikis@stats.uwo.ca`

1 Introduction

Loss severity distributions are required for a variety of premium and reserve calculations in non-life insurance. Both non-parametric and parametric estimators can be used in the calculations. While non-parametric estimators are convenient in many circumstances, parametric models are often preferred for a number of reasons:

- They yield smooth estimates of distribution function values. This is important because lack of smoothness can lead to undesirable progressions of premium rates for different deductible values or to anomalous orderings of premiums by rating category.

- Parametric models allow us to extrapolate probabilities to ranges of values that are beyond the actual data. While we always have to be cautious when extrapolating, the procedure may nevertheless be required by a quantity to be estimated as in the case of the premium for a high deductible excess of loss policy.

- By imposing a specific parametric form involving a small number of parameters, the parameters can be estimated more efficiently. Of course, a bias is introduced when the assumed form is not correct, but the 'risk' might be offset by other considerations.

- There may be specific aspects of the form, or structure, of the distribution that are either desirable or believed to exist, such as a mode at 0. Parametric models allow us to impose such structures even when they are not apparent in the data.

- Parametric models often have fairly simple forms that make them appealing and easy to use. For details on parametric loss models as well as for methods of classifying and creating distributions, we refer to Kleiber and Kotz (2003), Klugman *et al.* (2004), and references therein. For example, nested families are particularly useful for model selection. Scale families are generally used for modeling insurance loss amounts since, for example, inflation or changes in currency can be easily handled by changing the scale parameter.

Many methods exist for estimating the parameters. The most popular among them is arguably the maximum likelihood method in view of its desirable asymptotic properties, invariance under parameter transformations, etc. However, the MLE's are not robust and the influence of outliers on them can be substantial. It should be noted in this regard that outliers do occur frequently in the context of insurance losses, and the circumstances surrounding a particular loss might be quite unusual and atypical if compared to other losses in a portfolio (e.g., think about a loss that receives extensive media attention). Furthermore, an insurer may wish to use the loss data from one insurance portfolio to analyze another one for which few or no observations exist. In this case the uncertainty about the similarity of the two portfolios may lead the insurer to prefer a method that reduces the influence of outlying observations.

Practically, most of the methods available in the literature on robust estimation can be viewed as special cases of some general classes such as M-, L-, or R-statistics (see, e.g., Serfling, 1980, Chapters 7–9), and the class of M-statistics among them is arguably the most popular one, which is mostly due to a close relationship between the objective function of the M-statistic and its influence function. Recent examples of successful implementation of robust procedures include extreme-value applications in finance and economics (see, e.g., Dupuis and Victoria-Feser, 2006, Cowell and Victoria-Feser, 2006, 2007), geopedology (Vandewalle *et al.*, 2007), robust fitting of non-standard regression models (Marazzi and Yohai, 2004, 2006), a general minimum-distance modeling approach by Scott (2001). While all these methods are intuitively appealing to experts and deliver what they are designed for, their popularity among practitioners have been limited due to reasons such as computational complexity and lack of transparency.

To resolve a number of issues noted above, in the current paper we introduce and develop a general method for estimating the parameters of claim severity distributions, which we call the method of trimmed moments (MTM). The method utilizes the underlying principle of the classical method of moments. Namely, in order to estimate k unknown parameters, we equate k population trimmed moments with the corresponding k sample trimmed moments and then we solve the resulting system of k equations with respect to the unknown parameters. In practice, the MTM estimation is appropriate for various model-fitting situations including those where a close fit in one or both tails of the distribution is not required. The herein proposed MTM has several desirable features:

- Trimmed moments always exist irrespectively of the underlying distribution. Consequently, the

MTM estimators exist whenever a system of equations (analogous to that of the classical method of moments) has a solution, which can easily be checked using available software.

- The MTM estimators can achieve various degrees of robustness, which can easily be specified by the user by appropriately setting trimming proportions (to be explained later in this paper).

- The MTM is transparent in the sense that its actions on the data are relatively easy to understand and the asymptotic properties of the MTM estimators are readily available.

The rest of the paper is organized as follows. The MTM idea is rigorously presented in Section 2, where we also derive asymptotic properties of the MTM estimators. MTM estimators for several common claim severity distributions are developed in Section 3. In Section 4 we extend the MTM idea by developing a combined MTM, which we call CMTM for short. In Section 5 we investigate small-sample properties of the MTM estimators in a simulation study. Practical performance of the MTM estimators is illustrated on real data in Section 6. We conclude the paper with a brief summary of main findings in Section 7.

2 Method of trimmed moments

In this section we describe the method of trimmed moments, derive asymptotic properties of the obtained estimators, and conclude with a discussion on how to choose estimating functions and trimming proportions.

2.1 Definition

Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables, and let F denote the cumulative distribution function (cdf) of each X_j. Assume that the cdf F is given in a parametric form, and let $k \geq 1$ be the (fixed) number of unknown parameters, which we denote by $\theta_1, \dots, \theta_k$. Denote the order statistics of $X_1, \dots, X_n$ by $X_{1:n} \leq \cdots \leq X_{n:n}$. The MTM estimators of $\theta_1, \dots, \theta_k$ are derived in three steps:

- Compute the sample trimmed moments

$$\widehat{\mu}_j = \frac{1}{n - m_n(j) - m_n^*(j)} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(X_{i:n}), \quad 1 \leq j \leq k, \tag{2.1}$$

where h_j are specially chosen functions (to be discussed below) and $m_n(j)$ and $m_n^*(j)$ are integers $0 \leq m_n(j) < n - m_n^*(j) \leq n$ such that $m_n(j)/n \to a_j$ and $m_n^*(j)/n \to b_j$ when $n \to \infty$, where the 'proportions' a_j and b_j are chosen by the researcher (to be discussed below).

- Derive the corresponding population trimmed moments

$$\mu_j := \mu_j(\theta_1, \ldots, \theta_k) = \frac{1}{1 - a_j - b_j} \int_{a_j}^{1-b_j} h_j(F^{-1}(u))\, \mathrm{d}u, \quad 1 \leq j \leq k, \tag{2.2}$$

where $F^{-1}(u) = \inf\{x : \; F(x) \geq u\}$ is the quantile function. (Note that when $a_j = b_j = 0$, then $\mu_j = \mathbf{E}[h_j(X)]$.)

- Match the population and sample trimmed moments and solve the system of equations

$$\begin{cases} \mu_1(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_1 \\ & \vdots & \\ \mu_k(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_k \end{cases} \tag{2.3}$$

with respect to $\theta_1, \ldots, \theta_k$. The solutions, which we denote by $\widehat{\theta}_j = g_j(\widehat{\mu}_1, \ldots, \widehat{\mu}_k)$, $1 \leq j \leq k$, are, by definition, the MTM estimators of $\theta_1, \ldots, \theta_k$. Note that the functions g_j are such that $g_j(\mu_1, \ldots, \mu_k) = \theta_j$.

Note 2.1 Equations (2.3) can also be written as $\mu_j(\theta_1, \ldots, \theta_k) - \widehat{\mu}_j = 0$, $j = 1, \ldots, k$, and thus the MTM estimator $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ can be viewed as an M-estimator. □

Note 2.2 It is possible that equations (2.3) do not have a solution, or that they are difficult to solve even numerically when k is large due to stability problems. To alleviate the problem, the functions h_j have to be chosen thoughtfully. On the other hand, most claim severity distributions have a small number k of parameters, frequently not exceeding three (see Klugman *et al.*, 2004, Appendix A), and thus solving equations (2.3) does not usually present computational challenges. □

2.2 Asymptotic properties

The sample trimmed moments in equation (2.1) can be viewed as special cases of trimmed L-statistics, whose joint asymptotic normality is analyzed in detail by Brazauskas, Jones and Zitikis (2007). It follows from the latter work that the k-variate vector $(\sqrt{n}\,(\widehat{\mu}_1 - \mu_1), \ldots, \sqrt{n}\,(\widehat{\mu}_k - \mu_k))$ converges in distribution to the k-variate normal random vector with the mean $\mathbf{0} = (0, \ldots, 0)$ and the covariance-variance matrix $\boldsymbol{\Sigma} := \big[\sigma_{ij}^2\big]_{i,j=1}^k$ with the entries

$$\sigma_{ij}^2 = \frac{1}{(1-a_i-b_i)(1-a_j-b_j)} \int_{a_i}^{1-b_i} \int_{a_j}^{1-b_j} \big(\min\{u,v\} - uv\big)\, \mathrm{d}h_j\big(F^{-1}(v)\big)\, \mathrm{d}h_i\big(F^{-1}(u)\big). \qquad (2.4)$$

Following Serfling (1980, p. 20), this asymptotic normality statement can be written concisely as

$$(\widehat{\mu}_1, \ldots, \widehat{\mu}_k) \ \sim\ \mathcal{AN}\big((\mu_1, \ldots, \mu_k),\, n^{-1}\,\boldsymbol{\Sigma}\big). \qquad (2.5)$$

Using statement (2.5), we can derive the asymptotic distribution of the MTM estimators. Indeed, the delta method (see, e.g., Serfling, 1980, Section 3.3) implies that the MTM estimator $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ is asymptotically normal with the mean $(\theta_1, \ldots, \theta_k)$ and the covariance-variance matrix $n^{-1}\,\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'$, where $\mathbf{D} = [d_{ij}]_{i,j=1}^k$ is the Jacobian of the transformations $g_1, \ldots, g_k$ evaluated at $(\mu_1, \ldots, \mu_k)$, that is, $d_{ij} = \partial g_i / \partial \widehat{\mu}_j \big|_{(\mu_1, \ldots, \mu_k)}$. In summary, we have that

$$\big(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\big) \ \sim\ \mathcal{AN}\big((\theta_1, \ldots, \theta_k),\, n^{-1}\,\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'\big). \qquad (2.6)$$

Statement (2.6) can be used for testing hypotheses, constructing confidence intervals or sets. It can be incorporated into further (parametric) actuarial modeling, including estimating risk measures, calculating credibility premiums, pricing reinsurance contracts, analyzing solvency. In contrast, using specialized derivations of robust estimators, some of the aforementioned applications have been studied by Brazauskas and Serfling (2000a, 2003), Kaiser and Brazauskas (2006), and Dornheim and Brazauskas (2007).

2.3 Choosing $\mathbf{h_j}$, $\mathbf{a_j}$, and $\mathbf{b_j}$

To start with, note that the choices $h_j(t) = t^j$, $j = 1, \ldots, k$, lead to matching the 'genuine' sample and population trimmed moments, thus justifying the name of 'method of trimmed moments'. Functions h_j of this type work well for the location-scale families where we choose $h_1(t) = t$ and $h_2(t) = t^2$. These choices yield the following explicit expressions

$$\begin{cases} \widehat{\theta}_{\mathrm{MTM}} &=& \widehat{\mu}_1 - c_1\,\widehat{\sigma}_{\mathrm{MTM}}, \\ \widehat{\sigma}_{\mathrm{MTM}} &=& \sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)} \end{cases} \qquad (2.7)$$

of the MTM estimators of θ (location) and σ (scale), where $\widehat{\mu}_j = (n - m_n(j) - m_n^*(j))^{-1} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} X_{i:n}^j$ and $c_j \equiv c_j(F_0, a_j, b_j) := (1 - a_j - b_j)^{-1} \int_{a_j}^{1-b_j} \left[F_0^{-1}(u)\right]^j \, \mathrm{d}u$, where F_0^{-1} is the standard (i.e., with $\theta = 0$ and $\sigma = 1$) quantile function of the underlying location-scale family. The constants c_j can be evaluated using, for example, the so-called trapezoidal rule. We refer to subsection 3.1 below for a complete derivation of estimators (2.7) and numerical evaluations of robustness-efficiency tradeoffs.

In the case of claim severity distributions, which are typically asymmetric and have non-zero densities only on the positive half-line, other choices of h_j may be more appropriate. For example, upon first observing that the logarithmic transformation converts a number of loss models into a location-scale family, we arrive at the choices $h_1(t) = \log t$ and $h_2(t) = (\log t)^2$. In this case, the only change needed in formulas (2.7) is the computation of empirical trimmed moments, which are now given by $\widehat{\mu}_j = (n - m_n(j) - m_n^*(j))^{-1} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} \left[\log X_{i:n}\right]^j$. The constants c_1 and c_2 remain unchanged. Examples of log-location-scale distributions include lognormal, log-t, log-logistic, and Weibull. Indeed, after the logarithmic transformation, these distributions become normal, Student's t, logistic, and Gumbel (extreme-value), respectively. Furthermore, using $h(t) = \log t$ is also appropriate for the single-parameter Pareto distribution because the log-Pareto distribution is shifted-exponential with a known location.

Of course, not all distributions can be transformed into a location-scale family and thus other ideas have to be explored. For example, in the case of the gamma distribution, we can use the same functions $h_1(t) = t$ and $h_2(t) = t$ but different trimming proportions (a_1, b_1) and (a_2, b_2). This gives a closed form expression for the MTM estimator of the scale parameter β but the shape parameter α requires solving the nonlinear equation

$$c_1(\alpha)\, \widehat{\mu}_2 - c_2(\alpha)\, \widehat{\mu}_1 = 0, \tag{2.8}$$

where $\widehat{\mu}_j = (n - m_n(j) - m_n^*(j))^{-1} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} X_{i:n}$ and $c_j(\alpha) = (1 - a_j - b_j)^{-1} \int_{a_j}^{1-b_j} F_\alpha^{-1}(u) \, \mathrm{d}u$ with $F_\alpha^{-1}(u)$ denoting the quantile function of the gamma distribution with $\beta = 1$. A solution of (2.8), which we denote by $\widehat{\alpha}_{\text{MTM}}$, is then used to estimate the parameter β, giving the estimator $\widehat{\beta}_{\text{MTM}} = \widehat{\mu}_1 / c_1(\widehat{\alpha}_{\text{MTM}})$. The same idea can also be employed for other claim severity distributions with nonlinear (in terms of parameters) representations of their quantile functions. Such distributions include log-gamma, two-parameter Pareto, and generalized Pareto. In general, if h_j is a power function, derivations of (2.2) and (2.3) are more straightforward when the exponent is as small as possible. The choices $h_j(t) = t^j$ and $a_j = b_j = 0$ for all $j = 1, \ldots, k$ lead of course to the classical method of moments.

As to choosing the trimming proportions a_j and b_j, there is no rigorous answer to this problem as it depends on how much robustness we need. Typically in such situations, we make a choice based on how much efficiency we are willing to sacrifice. Choosing proportions also requires some experience and intuition. The following general points provide useful hints:

- Robustness of the entire estimation procedure is determined by the least trimmed moments. That is, the MTM estimators remain resistant against the proportion $a_* = \min\{a_1, \ldots, a_k\}$ of lowest observations and against the proportion $b_* = \min\{b_1, \ldots, b_k\}$ of highest observations.

- For location-scale families and their variants such as log-location-scale families, the quantile-quantile (QQ-) plot may serve as an excellent guide when choosing a_j and b_j (see Section 6 for an illustration). The QQ-plot would not, however, be appropriate for other distributions.

- The trimming proportions can also be chosen based on the 'outside of the model' knowledge. For example, the actuary may employ the (re)insurance contract's specifications. For an illustration, see Section 6.

Hence, in view of the above, we see that the MTM is fairly flexible and allows us to work with various combinations of h_j, a_j, and b_j, depending on the problem at hand. Throughout the rest of the paper we shall analyze a number of specific distributions using the MTM and in this way gain further intuition on how to construct and analyze MTM estimators.

3 Examples: three parametric models

In this section, we analyze three examples: the first one concerns the general (i.e., not necessarily symmetric) location-scale family, and the other two examples focus on the Pareto and lognormal distributions. Specifically, in the three examples we show how to find MTM estimators and derive the entries of the asymptotic covariance-variance matrix. For the Pareto and lognormal distributions, we also evaluate the asymptotic relative efficiency (ARE)

$$\text{ARE(MTM, MLE)} = \frac{\text{asymptotic variance of MLE}}{\text{asymptotic variance of an MTM estimator}}$$

of the MTM estimators with respect to the maximum likelihood estimator (MLE). In the multivariate case, the ARE is defined by replacing the two variances by the corresponding generalized variances, which are the determinants of the asymptotic covariance-variance matrices of vector-estimators, and then raise the ratio to the power $1/k$. For details on these issues, we refer, for example, to Serfling (1980, Section 4.1).

3.1 Location-scale families

Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with the same distribution

$$\text{Location-scale:} \quad F(x) = F_0\left(\frac{x-\theta}{\sigma}\right), \qquad -\infty < x < \infty, \tag{3.1}$$

where $-\infty < \theta < \infty$ and $\sigma > 0$ are unknown parameters, and F_0 is the standard (i.e., with $\theta = 0$ and $\sigma = 1$) parameter-free version of F. The corresponding quantile function is $F^{-1}(t) = \theta + \sigma F_0^{-1}(t)$. Since the cdf F has two unknown parameters, we employ two trimmed moments. There is no restriction on the real values of θ and we so choose $h_1(t) = t$, but to ensure that the estimator of $\sigma > 0$ is positive, we choose $h_2(t) = t^2$. Following the procedure of Section 2, we have

$$\begin{aligned}
\widehat{\mu}_1 &= \frac{1}{n - m_n(1) - m_n^*(1)} \sum_{i=m_n(1)+1}^{n-m_n^*(1)} X_{i:n}, \\
\widehat{\mu}_2 &= \frac{1}{n - m_n(2) - m_n^*(2)} \sum_{i=m_n(2)+1}^{n-m_n^*(2)} X_{i:n}^2
\end{aligned}$$

with $m_n(1)/n = m_n(2)/n \to a$ and $m_n^*(1)/n = m_n^*(2)/n \to b$.

Note 3.1 We can of course consider more general trimming such as $m_n(1) \neq m_n(2)$ and/or $m_n^*(1) \neq m_n^*(2)$. Recall, however, that robustness of the entire procedure is determined by the least trimmed moments. Thus, for example, choosing $m_n(2) > m_n(1)$ and $m_n^*(2) > m_n^*(1)$ yields MTM estimators with the breakdown points (i.e., degrees of resistance to outliers) of $m_n(1)/n$ and $m_n^*(1)/n$, respectively, but with usually smaller efficiency than in the cases $m_n(1)/n = m_n(2)/n$ and $m_n^*(1)/n = m_n^*(2)/n$. $\square$

As our next step in deriving MTM estimators, we calculate the population trimmed moments using equation (2.2) and obtain

$$\mu_1 := \mu_1(\theta,\, \sigma) = \frac{1}{1-a-b} \int_a^{1-b} F^{-1}(u)\, \mathrm{d}u = \theta + \sigma\, c_1,$$
$$\mu_2 := \mu_2(\theta,\, \sigma) = \frac{1}{1-a-b} \int_a^{1-b} \big[F^{-1}(u)\big]^2\, \mathrm{d}u = \theta^2 + 2\theta\sigma\, c_1 + \sigma^2\, c_2,$$

where

$$c_k \equiv c_k(F_0,\, a,\, b) := \frac{1}{1-a-b} \int_a^{1-b} \big[F_0^{-1}(u)\big]^k\, \mathrm{d}u.$$

Equating $\widehat{\mu}_1$ to μ_1 and $\widehat{\mu}_2$ to μ_2, and then solving the resulting system of equations with respect to θ and σ, we obtain the MTM estimators

$$\begin{cases} \widehat{\theta}_{\mathrm{MTM}} & = & \widehat{\mu}_1 - c_1\, \widehat{\sigma}_{\mathrm{MTM}} & =: & g_1(\widehat{\mu}_1, \widehat{\mu}_2) \\ \widehat{\sigma}_{\mathrm{MTM}} & = & \sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)} & =: & g_2(\widehat{\mu}_1, \widehat{\mu}_2). \end{cases} \tag{3.2}$$

The entries of the covariance-variance matrix $\mathbf{\Sigma}$ calculated using equation (2.4) are

$$\begin{aligned}
\sigma_{11}^2 &= \frac{\sigma^2}{(1-a-b)^2} \int_a^{1-b}\!\! \int_a^{1-b} \big(\min\{u,v\} - uv\big)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&= \sigma^2\, c_1^*, \\
\sigma_{12}^2 &= \frac{2\theta\sigma^2}{(1-a-b)^2} \int_a^{1-b}\!\! \int_a^{1-b} \big(\min\{u,v\} - uv\big)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&\quad + \frac{2\sigma^3}{(1-a-b)^2} \int_a^{1-b}\!\! \int_a^{1-b} \big(\min\{u,v\} - uv\big)\, F_0^{-1}(u)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&= 2\theta\sigma^2\, c_1^* + 2\sigma^3\, c_2^*, \\
\sigma_{22}^2 &= \frac{4\theta^2\sigma^2}{(1-a-b)^2} \int_a^{1-b}\!\! \int_a^{1-b} \big(\min\{u,v\} - uv\big)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&\quad + \frac{8\theta\sigma^3}{(1-a-b)^2} \int_a^{1-b}\!\! \int_a^{1-b} \big(\min\{u,v\} - uv\big)\, F_0^{-1}(u)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&\quad + \frac{4\sigma^4}{(1-a-b)^2} \int_a^{1-b}\!\! \int_a^{1-b} \big(\min\{u,v\} - uv\big)\, F_0^{-1}(u)\, F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&= 4\theta^2\sigma^2\, c_1^* + 8\theta\sigma^3\, c_2^* + 4\sigma^4\, c_3^*
\end{aligned}$$

with the obvious notation for $c_k^* \equiv c_k^*(F_0,\, a,\, b)$, which can be written in terms of the earlier introduced constants c_k (see Appendix A for details). The entries of the matrix $\mathbf{D}$ are found by differentiating the functions g_i (see equations (3.2)):

$$\begin{aligned}
d_{11} &= \left.\frac{\partial g_1}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = 1 - c_1 \left.\frac{\partial g_2}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = \frac{c_1\theta + c_2\sigma}{\sigma(c_2 - c_1^2)}, \\
d_{12} &= \left.\frac{\partial g_1}{\partial \widehat{\mu}_2}\right|_{(\mu_1,\mu_2)} = -c_1 \left.\frac{\partial g_2}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = \frac{-0.5c_1}{\sigma(c_2 - c_1^2)}, \\
d_{21} &= \left.\frac{\partial g_2}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = \left.\frac{-\widehat{\mu}_1}{\sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)}}\right|_{(\mu_1,\mu_2)} = \frac{-\theta - c_1\sigma}{\sigma(c_2 - c_1^2)}, \\
d_{22} &= \left.\frac{\partial g_2}{\partial \widehat{\mu}_2}\right|_{(\mu_1,\mu_2)} = \left.\frac{0.5}{\sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)}}\right|_{(\mu_1,\mu_2)} = \frac{0.5}{\sigma(c_2 - c_1^2)}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbf{D\Sigma D}' &= \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} \sigma^2 c_1^* & 2\theta\sigma^2 c_1^* + 2\sigma^3 c_2^* \\ 2\theta\sigma^2 c_1^* + 2\sigma^3 c_2^* & 4\theta^2\sigma^2 c_1^* + 8\theta\sigma^3 c_2^* + 4\sigma^4 c_3^* \end{bmatrix} \begin{bmatrix} d_{11} & d_{21} \\ d_{12} & d_{22} \end{bmatrix} \\
&= \frac{\sigma^2}{(c_2 - c_1^2)^2} \begin{bmatrix} c_1^* c_2^2 - 2c_1c_2c_2^* + c_1^2 c_3^* & -c_1^* c_1 c_2 + c_2 c_2^* + c_1^2 c_2^* - c_1 c_3^* \\ -c_1^* c_1 c_2 + c_2 c_2^* + c_1^2 c_2^* - c_1 c_3^* & c_1^* c_1^2 - 2c_1 c_2^* + c_3^* \end{bmatrix}.
\end{aligned} \tag{3.3}$$

We summarize the above findings by saying that

$$\left(\widehat{\theta}_{\text{MTM}},\, \widehat{\sigma}_{\text{MTM}}\right) \sim \mathcal{AN}\left((\theta, \sigma),\, \frac{\sigma^2}{n}\mathbf{S}\right) \quad \text{with} \quad \mathbf{S} = \sigma^{-2}\mathbf{D\Sigma D}'. \tag{3.4}$$

Note that the matrix $\mathbf{S}$ does not depend on any unknown parameters and can be expressed in terms of F_0, a and b, which are specified by the researcher.

3.2 Pareto model

Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with the same Pareto distribution

$$\text{Pareto}(x_0, \alpha): \quad F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\alpha}, \quad x > x_0, \tag{3.5}$$

where $\alpha > 0$ is an unknown parameter and $x_0 > 0$ is assumed to be known. (The x_0 can, for example, be interpreted as a deductible or a retention level.) The corresponding quantile function is $F^{-1}(t) = x_0(1-t)^{-1/\alpha}$. Since F has only one unknown parameter, we need only one trimmed moment. As mentioned in Section 2, the choice $h_1(t) = \log(t/x_0)$ is appropriate in this case.

Note 3.2 A hint about using the logarithmic function comes from the fact that if X is Pareto(x_0, α) with the cdf defined by equation (3.5), then $Z = \log X$ is a shifted-exponential variable with the cdf $G(z) = 1 - e^{-(z-z_0)/\theta}, \quad z > z_0$, where $z_0 = \log x_0$ and $\theta = \alpha^{-1}$. Hence, α^{-1} represents the unknown part of the mean of Z and can thus be naturally estimated with a trimmed mean. The relationship between the Pareto and shifted-exponential distributions has been utilized by Brazauskas and Serfling (2000a,b, 2003) when developing a robust estimation technique for the Pareto(x_0, α) model. $\square$

Having thus chosen the function $h_1(t) = \log(t/x_0)$, we follow the procedure of Section 2 and have

$$\widehat{\mu}_1 = \frac{1}{n - m_n(1) - m_n^*(1)} \sum_{i=m_n(1)+1}^{n-m_n^*(1)} \log(X_{i:n}/x_0)$$

with $m_n(1)/n \to a$ and $m_n^*(1)/n \to b$. The corresponding population trimmed moment is

$$\begin{aligned}
\mu_1 := \mu_1(\alpha) &= \frac{1}{1-a-b} \int_a^{1-b} \log(F^{-1}(u)/x_0)\, \mathrm{d}u \\
&= \frac{-1/\alpha}{1-a-b} \int_a^{1-b} \log(1-u)\, \mathrm{d}u \\
&= \frac{-1/\alpha}{1-a-b}\, I_0(a,\, 1-b)
\end{aligned}$$

with the obvious notation for the function I_0. Equating $\widehat{\mu}_1$ to μ_1 and then solving the equation with respect to α yields the MTM estimator

$$\widehat{\alpha}_{\text{MTM}} = \left(\frac{-I_0(a,\, 1-b)}{1-a-b} \right) \frac{1}{\widehat{\mu}_1} =: g_1(\widehat{\mu}_1). \tag{3.6}$$

The entries of the matrices $\mathbf{\Sigma}$ and $\mathbf{D}$, which are one-dimensional, are as follows. For calculating $\mathbf{\Sigma}$ we use equation (2.4) and obtain

$$\begin{aligned}
\sigma_{11}^2 &= \frac{1/\alpha^2}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} \frac{\min\{u,v\} - uv}{(1-u)(1-v)}\, \mathrm{d}v\, \mathrm{d}u \\
&= \frac{1/\alpha^2}{(1-a-b)^2}\, J\big((a, 1-b),\, (a, 1-b)\big)
\end{aligned}$$

with the obvious notation for the function J. The Jacobian $\mathbf{D}$ is found by differentiating the function g_1 (see equation (3.6)) and then evaluating its derivative at μ_1:

$$\left. \frac{\partial g_1}{\partial \widehat{\mu}_1} \right|_{\mu_1} = \left(\frac{I_0(a,\, 1-b)}{1-a-b} \right) \left. \frac{1}{\widehat{\mu}_1^2} \right|_{\mu_1} = \left(\frac{1-a-b}{I_0(a,\, 1-b)} \right) \alpha^2.$$

Hence,

$$\mathbf{D\Sigma D}' = \frac{J\big((a, 1-b),\, (a, 1-b)\big)}{\big[I_0(a,\, 1-b)\big]^2}\, \alpha^2.$$

Summarizing the above findings, we have that

$$\widehat{\alpha}_{\mathrm{MTM}} \;\sim\; \mathcal{AN}\left(\alpha,\, \frac{\alpha^2}{n}\, C\right) \quad \text{with} \quad C = \frac{J\big((a, 1-b),\, (a, 1-b)\big)}{\big[I_0(a,\, 1-b)\big]^2}. \tag{3.7}$$

We conclude this Paretian subsection with an investigation of how much efficiency we lose because of using $\widehat{\alpha}_{\mathrm{MTM}}$ and not the MLE when estimating α. Since the MLE is $\widehat{\alpha}_{\mathrm{MLE}} = n / \sum_{i=1}^{n} \log(X_i/x_0)$, which is $\mathcal{AN}(\alpha,\, \alpha^2/n)$ (see, e.g., Brazauskas and Serfling, 2000a), we have that $\mathrm{ARE}(\widehat{\alpha}_{\mathrm{MTM}},\, \widehat{\alpha}_{\mathrm{MLE}}) = 1/C$.

Note 3.3 When $m_n(1) = m_n^*(1) = 0$, then the MTM estimator (3.6) becomes $\widehat{\alpha}_{\mathrm{MLE}}$. Furthermore, note that since $C \to 1$ when $a = b \to 0$, the MLE's asymptotic distribution follows from (3.7). $\square$

Numerical values of these ARE's are provided in Table 3.1 for several trimming proportions a and b. For example, we see from Table 3.1 that when b is fixed, then the MTM estimators with no

	b							
a	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{1}$	.918	.847	.783	.666	.423	.238	.116
0.05	1.00	$\boxed{.918}$	.848	.783	.667	.425	.242	.122
0.10	1.00	.918	$\boxed{.848}$	.785	.669	.430	.250	.135
0.15	.999	.919	.850	$\boxed{.787}$	.672	.437	.261	–
0.25	.995	.918	.851	.790	$\boxed{.679}$	.452	.285	–
0.49	.958	.897	.839	.786	.688	$\boxed{.487}$	–	–
0.70	.857	.824	.781	.738	.659	–	–	–
0.85	.681	.688	.663	–	–	–	–	–

Table 3.1: $\mathrm{ARE}(\widehat{\alpha}_{\mathrm{MTM}},\, \widehat{\alpha}_{\mathrm{MLE}})$ for selected a and b, with the boxed numbers highlighting the case $a = b$.

lower trimming (i.e., $a = 0$) and with symmetric trimming (i.e., $a = b$) are almost equivalent. When $b = 0$ (i.e., no upper trimming), then the efficiency decreases slowly. Moreover, the MTM allows some very extreme trimming scenarios yielding valid though inefficient estimators. For instance, $\widehat{\alpha}_{\mathrm{MTM}}$ with $a = b = 0.49$ is practically a percentile (i.e., median) matching estimator described by Klugman, Panjer and Willmot (2004, Section 12.1), which has about 49% efficiency. This estimator uses actual values of only 2% of the data. Interestingly, $\widehat{\alpha}_{\mathrm{MTM}}$ with $(a, b) = (0.10, 0.85)$ and $(a, b) = (0.85, 0.10)$ are both based on 5% of actual observations but have dramatically different efficiencies: 0.135 and

0.663, respectively. This implies that the accuracy of estimators depends not only on the fraction of the used actual data but also where the data is located in the sample. Certainly, this note just confirms the well-known fact that most information about the parameter α is contained in the upper tail of Pareto$(x_0,\, \alpha)$.

3.3 Lognormal model

Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with the same lognormal distribution

$$\mathrm{LN}(x_0,\, \theta,\, \sigma): \quad F(x) = \Phi\left(\frac{\log(x - x_0) - \theta}{\sigma}\right), \quad x > x_0, \tag{3.8}$$

where $-\infty < \theta < \infty$ and $\sigma > 0$ are unknown parameters with Φ denoting the standard normal cdf. The parameter x_0 is assumed to be known and it can be interpreted as in the Pareto case noted above. Since the logarithmic transformation makes this distribution normal, which is a member of the location-scale family, results of Section 3.1 apply with two modifications: $h_1(t) = \log(t - x_0)$ and $h_2(t) = \log^2(t - x_0)$. Hence, the MTM estimators of θ and σ are

$$\begin{cases} \widehat{\theta}_{\mathrm{MTM}} &= \widehat{\mu}_1 - c_1\, \widehat{\sigma}_{\mathrm{MTM}}, \\ \widehat{\sigma}_{\mathrm{MTM}} &= \sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)}. \end{cases} \tag{3.9}$$

The sample trimmed moments are

$$\widehat{\mu}_1 = \frac{1}{n - m_n(1) - m_n^*(1)} \sum_{i=m_n(1)+1}^{n-m_n^*(1)} \log(X_{i:n} - x_0),$$

$$\widehat{\mu}_2 = \frac{1}{n - m_n(2) - m_n^*(2)} \sum_{i=m_n(2)+1}^{n-m_n^*(2)} \big[\log(X_{i:n} - x_0)\big]^2$$

with $m_n(1)/n = m_n(2)/n \to a$ and $m_n^*(1)/n = m_n^*(2)/n \to b$. As follows from statement (3.4),

$$\left(\widehat{\theta}_{\mathrm{MTM}},\, \widehat{\sigma}_{\mathrm{MTM}}\right) \sim \mathcal{AN}\left((\theta, \sigma),\, \frac{\sigma^2}{n}\,\mathbf{S}\right) \quad \text{with} \quad \mathbf{S} = \sigma^{-2}\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}', \tag{3.10}$$

where $\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'$ is given by equation (3.3) but now with the standard normal F_0 instead of the therein used standardized location-scale distribution.

We next examine how much efficiency is lost because of using $(\widehat{\theta}_{\mathrm{MTM}},\, \widehat{\sigma}_{\mathrm{MTM}})$ instead of the MLE's

$$\begin{cases} \widehat{\theta}_{\mathrm{MLE}} &= n^{-1} \sum_{i=1}^{n} \log(X_i - x_0), \\ \widehat{\sigma}_{\mathrm{MLE}} &= \sqrt{n^{-1} \sum_{i=1}^{n} \left(\log(X_i - x_0) - \widehat{\theta}_{\mathrm{MLE}}\right)^2} \end{cases}$$

when estimating (θ, σ). We know (see, e.g., Serfling, 2002) that

$$\left(\widehat{\theta}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}\right) \sim \mathcal{AN}\left((\theta,\sigma), \frac{\sigma^2}{n}\mathbf{S_0}\right) \quad \text{with} \quad \mathbf{S_0} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

Hence, it follows that $\text{ARE}((\widehat{\theta}_{\text{MTM}}, \widehat{\sigma}_{\text{MTM}}), (\widehat{\theta}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$, which is $(\det(\mathbf{S_0})/\det(\mathbf{S}))^{1/2}$ by definition, is equal to $\left(0.5/\det(\mathbf{S})\right)^{1/2}$.

Note 3.4 When $m_n(1) = m_n^*(1) = m_n(2) = m_n^*(2) = 0$, then the MTM estimators (3.9) become $(\widehat{\theta}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}})$. Also note that since $\mathbf{S} \to \mathbf{S_0}$ when $a = b \to 0$, then the MLE's asymptotic distribution follows from statement (3.10). □

Numerical values of the ARE's are provided in Table 3.2 for selected values of a and b. Note

	b							
a	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{1}$	.932	.874	.821	.722	.502	.312	.169
0.05	.932	$\boxed{.872}$	.820	.771	.678	.470	.286	.142
0.10	.874	.820	$\boxed{.769}$	.722	.633	.430	.248	.097
0.15	.821	.771	.722	$\boxed{.676}$	.590	.390	.208	–
0.25	.722	.678	.633	.590	$\boxed{.507}$	.312	.113	–
0.49	.502	.470	.430	.390	.312	$\boxed{.074}$	–	–
0.70	.312	.286	.248	.208	.113	–	–	–
0.85	.169	.142	.097	–	–	–	–	–

Table 3.2: $\text{ARE}((\widehat{\theta}_{\text{MTM}}, \widehat{\sigma}_{\text{MTM}}), (\widehat{\theta}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$ for selected a and b, with the boxed numbers highlighting the case $a = b$.

from Table 3.2 that since the logarithmic transformation of the lognormal random variables makes the variable normal, which is symmetric, we can expect similar performance of the MTM estimators with similar trimming schemes. For example, the ARE's are identical for the MTM estimators with reversed trimming proportions: e.g., $(a, b) = (0.05, 0.25)$ has ARE $= 0.678$ and $(a, b) = (0.25, 0.05)$ also has ARE $= 0.678$. In fact, we notice that the MTM estimators based on the same fraction of actual observations (i.e., having same $a + b$) have similar ARE's:

- $(a, b) = (0.10, 0.15)$ has ARE $= 0.722$ and $(a, b) = (0.25, 0.00)$ has ARE $= 0.722$.
- $(a, b) = (0.15, 0.15)$ has ARE $= 0.676$ and $(a, b) = (0.05, 0.25)$ has ARE $= 0.678$.

- $(a, b) = (0.85, 0.10)$ has ARE $= 0.097$ and $(a, b) = (0.25, 0.70)$ has ARE $= 0.113$.

The trimming schemes that focus exclusively on data in the center (i.e., when $a = b$) are known to be efficient for estimating the location/center but not necessarily for estimating the scale. For the joint estimation of θ and σ, we observe that inefficiency of σ estimators dominates the overall ARE: $a = b = 0.05$ has ARE $= 0.872$ (good); $a = b = 0.25$ has ARE $= 0.507$ (moderate); $a = b = 0.49$ has ARE $= 0.074$ (very poor). In comparison with the latter case, note that the median estimator of θ is almost 10 times more efficient: ARE $= 0.637$.

We conclude this section with a note on a similar type of study by Serfling (2002) where a generalized median (GM) estimator is proposed in the case of the lognormal distribution. For a fixed breakdown point (fixed a and b in our case), the GM estimator is more efficient than the MTM estimator. However, the latter estimator is more flexible (e.g., it can provide asymmetric protection against upper and lower outliers) and is not computationally demanding. The flexibility issue of the GM estimator can be alleviated by considering a generalized quantile (GQ) estimator of which the GM is a special case. This has been accomplished by Brazauskas and Serfling (2000b) in the Pareto and exponential cases. There we also find suggestions on how to reduce the computational burden of the GQ and GM estimators.

4 Combined MTM estimators

Here we suggest a natural extension of the MTM, which we call the combined method of trimmed moments (CMTM). The main idea is as follows. Using trimmed moments of various orders, we generate a set of different estimators for the same parameter θ. Denote the estimators by $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$. It follows from Section 2 (see equation (2.6)) that $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k) \sim \mathcal{AN}((\theta, \ldots, \theta),\, n^{-1}\mathbf{\Sigma}_*)$, where the covariance-variance matrix $\mathbf{\Sigma}_* = \mathbf{D\Sigma D}'$ may depend on θ. Next we combine these estimators of θ in some way hoping to get an estimator of θ with a smaller variance than that of the best one estimator among $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$. For example, consider the linear combination of the estimators (see, e.g., Serfling, 1980, Section 3.4.3)

$$\widehat{\widehat{\theta}} = \sum_{i=1}^{k} w_i \widehat{\theta}_i, \tag{4.1}$$

where the weight vector $\mathbf{w} = (w_1, \ldots, w_k)$ is such that $w_1 + \cdots + w_k = 1$. The estimator $\widehat{\widehat{\theta}}$ is $\mathcal{AN}(\theta,\, n^{-1}\mathbf{w}\boldsymbol{\Sigma}_*\mathbf{w}')$. Since our criterion is based on minimizing the variance, we choose the weights w_i so that the quadratic form $\mathbf{w}\boldsymbol{\Sigma}_*\mathbf{w}'$ is minimized subject to the constraint $\sum_{i=1}^{k} w_i = 1$. Assuming without loss of generality that $\boldsymbol{\Sigma}_*$ is nonsingular, a standard variational technique implies that the minimum is attained and is equal to

$$\mathbf{w_0}\boldsymbol{\Sigma}_*\mathbf{w_0}' = \frac{1}{(1,\ldots,1)\boldsymbol{\Sigma}_*^{-1}(1,\ldots,1)'} \quad \text{with} \quad \mathbf{w_0} = \frac{(1,\ldots,1)\boldsymbol{\Sigma}_*^{-1}}{(1,\ldots,1)\boldsymbol{\Sigma}_*^{-1}(1,\ldots,1)'}\,, \tag{4.2}$$

where $(1, \ldots, 1)$ is the $1 \times k$ vector of ones.

Note 4.1 General situations addressing the case when the matrix $\boldsymbol{\Sigma}_*$ depends on θ and when the parameter θ itself is a vector are discussed by Soong (1969), who studied combined classical (i.e., untrimmed) moment estimators with the aim of improving their efficiency. $\square$

Note 4.2 When θ is a scale parameter, the covariance-variance matrix $\boldsymbol{\Sigma}_*$ is of the form $\theta^2\boldsymbol{\Sigma}_{**}$, where $\boldsymbol{\Sigma}_{**}$ does not depend on θ. Thus, $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k) \sim \mathcal{AN}((\theta, \ldots, \theta),\, n^{-1}\theta^2\boldsymbol{\Sigma}_{**})$. This implies that estimating θ is within the framework of problems studied (though in a different context) by Brazauskas and Ghorai (2007). Namely, the authors consider a number of asymptotically (when $k \to \infty$) efficient estimators of θ and examine their sensitivity to various model misspecification scenarios. The estimators presented by these authors are asymptotically optimal and can be applied in the current research but they are nonlinear and require a large k to achieve the asymptotic normality and unbiasedness of the combined estimator. Hence, the CMTM estimator given by equation (4.1) is preferred in the current project. $\square$

The main benefit of using the combined estimator $\widehat{\widehat{\theta}}$ instead of the most efficient but typically least robust estimator among $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ is that the combined estimator allows us to estimate θ more efficiently while still maintaining the chosen degree of robustness. For example, if 25% resistance against both upper and lower outliers is desirable, we select k estimators, each with $0.25 \leq a < 1-b \leq 0.75$, and then combine them according to (4.1). We shall next provide a numerical illustration of how this works in the case of the single-parameter Pareto distribution. Namely, following subsection 3.2, we start with the sample trimmed moments

$$\widehat{\mu}_j = \frac{1}{n - m_n(j) - m_n^*(j)} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} \log(X_{i:n}/x_0), \quad j = 1, \ldots, k,$$

where $m_n(j)/n \to a_j$ and $m_n^*(j)/n \to b_j$. We equate these empirical moments to the corresponding population trimmed moments

$$\mu_j := \mu_j(\alpha) = \frac{-1/\alpha}{1 - a_j - b_j} I_0(a_j, 1 - b_j), \quad j = 1, \ldots, k,$$

where an explicit formula for I_0 is given in Appendix A. Solving the resulting system of equations yields the following MTM estimators of α:

$$\widehat{\alpha}_{\mathrm{MTM},\,j} = \left(\frac{-I_0(a_j, 1 - b_j)}{1 - a_j - b_j} \right) \frac{1}{\widehat{\mu}_j} =: g_j(\widehat{\mu}_j), \quad j = 1, \ldots, k.$$

The entries of the covariance-variance matrix $\mathbf{\Sigma}$ and the Jacobian $\mathbf{D}$ are found by straightforward computations and we arrive at the equation

$$\mathbf{D\Sigma D}' = \alpha^2 \left[\frac{J\big((a_j, 1 - b_j), (a_i, 1 - b_i)\big)}{I_0(a_i, 1 - b_i)\, I_0(a_j, 1 - b_j)} \right]_{i,j=1}^{k} =: \alpha^2 \mathbf{\Delta},$$

where J is given in Appendix A depending on the order of the four proportions a_i, a_j, $1 - b_i$, $1 - b_j$. Thus, the matrix $\mathbf{\Delta}$ is completely specified and can be expressed in terms of the proportions a_i and b_i. In summary, we have that

$$\big(\widehat{\alpha}_{\mathrm{MTM},\,1}, \ldots, \widehat{\alpha}_{\mathrm{MTM},\,k}\big) \;\sim\; \mathcal{AN}\left((\alpha, \ldots, \alpha),\, \frac{\alpha^2}{n}\, \mathbf{\Delta} \right).$$

An application of equations (4.1) to the estimators $\widehat{\alpha}_{\mathrm{MTM},\,1}, \ldots, \widehat{\alpha}_{\mathrm{MTM},\,k}$ yields that

$$\widehat{\widehat{\alpha}} = \sum_{i=1}^{k} w_i^*\, \widehat{\alpha}_{\mathrm{MTM},\,i} \;\sim\; \mathcal{AN}\left(\alpha,\, \frac{\alpha^2}{n}\, \frac{1}{(1, \ldots, 1)\mathbf{\Delta}^{-1}(1, \ldots, 1)'} \right),$$

where

$$(w_1^*, \ldots, w_k^*) = \frac{(1, \ldots, 1)\mathbf{\Delta}^{-1}}{(1, \ldots, 1)\mathbf{\Delta}^{-1}(1, \ldots, 1)'}.$$

From this we have that $\mathrm{ARE}(\widehat{\widehat{\alpha}}, \widehat{\alpha}_{\mathrm{MLE}}) = (1, \ldots, 1)\mathbf{\Delta}^{-1}(1, \ldots, 1)'$.

We shall next numerically investigate (see Table 4.1 below) the ARE's for various combined estimators and values of k, as follows:

- $\widehat{\widehat{\alpha}}_1$ combines k symmetrically trimmed estimators with $b = a = 0.05 : 0.02 : 0.05 + 0.02(k-1)$ for $k = 1, \ldots, 5,\ 10,\ 15,\ 20$.

For example, when $k = 3$, this estimator combines $\widehat{\alpha}_{\text{MTM}}$'s with $a_1 = b_1 = 0.05$, $a_2 = b_2 = 0.07$, and $a_3 = b_3 = 0.09$. Thus, its overall lower and upper breakdown points are $a_* = \min\{a_1, a_2, a_3\} = 0.05$ and $b_* = \min\{b_1, b_2, b_3\} = 0.05$, respectively. Next, to see how important is the choice of component estimators, we consider

- $\widehat{\widehat{\alpha}}_2$ (it has same robustness properties as $\widehat{\widehat{\alpha}}_1$) which combines $\widehat{\alpha}_{\text{MTM}}$'s with $a = 0.05 : 0.02 : 0.05 + 0.02(k-1)$ and $b_1 = \cdots = b_k = 0.05$ for $k = 1, \ldots, 5,\ 10,\ 15,\ 20$.

The other three combined estimators are defined as follows:

- $\widehat{\widehat{\alpha}}_3$ combines $\widehat{\alpha}_{\text{MTM}}$'s with $a = 0.10 : 0.015 : 0.10 + 0.015(k-1)$ and $b = 0.15 : 0.015 : 0.15 + 0.015(k-1)$.

- $\widehat{\widehat{\alpha}}_4$ combines $\widehat{\alpha}_{\text{MTM}}$'s with $a = 0.25 : 0.01 : 0.25 + 0.01(k-1)$ and $b = a$.

- $\widehat{\widehat{\alpha}}_5$ combines $\widehat{\alpha}_{\text{MTM}}$'s with $a = 0.10 : 0.01 : 0.10 + 0.01(k-1)$ and $b = 0.49 : 0.01 : 0.49 + 0.01(k-1)$.

We see from Table 4.1 that for a combined estimator with fixed robustness properties (i.e., fixed minimum a and b) the ARE as a function of k becomes flat as k increases. In almost all cases, the

	Breakdown points		k							
Estimators	Lower(a_*)	Upper(b_*)	1	2	3	4	5	10	15	20
$\widehat{\widehat{\alpha}}_1$	0.05	0.05	.918	.945	.945	.945	.945	.945	.945	.945
$\widehat{\widehat{\alpha}}_2$	0.05	0.05	.918	.918	.918	.918	.919	.920	.921	.922
$\widehat{\widehat{\alpha}}_3$	0.10	0.15	.785	.844	.845	.845	.845	.845	.845	.845
$\widehat{\widehat{\alpha}}_4$	0.25	0.25	.679	.741	.742	.742	.742	.742	.742	.742
$\widehat{\widehat{\alpha}}_5$	0.10	0.49	.430	.492	.493	.493	.493	.493	.493	.493

Table 4.1: ARE($\widehat{\widehat{\alpha}}$, $\widehat{\alpha}_{\text{MLE}}$) for various combined estimators and selected k.

main improvement occurs when a combined estimator with $k = 2$ is used instead of a non-combined estimator ($k = 1$), and there is no additional benefit (up to 3 decimal places) when using $\widehat{\widehat{\alpha}}$ with $k \geq 4$ instead of $k = 3$. A comparison of $\widehat{\widehat{\alpha}}_1$ and $\widehat{\widehat{\alpha}}_2$ shows that the choice of component estimators is important. The $\widehat{\alpha}_{\text{MTM}}$ estimators included in $\widehat{\widehat{\alpha}}_2$ all have $b = 0.05$. As we see from Table 4.1, all MTM estimators with $b = 0.05$ have extremely similar and high ARE's. This means that they contribute virtually identical information to the combined estimator and, therefore, the ARE of $\widehat{\widehat{\alpha}}$ does not get better. This provides a hint on how to select component estimators. Namely, they should satisfy the initially specified robustness requirements and they should also have as different ARE's as possible. Finally, the strategy of combining MTM estimators is most beneficial for highly robust but inefficient estimators and less so for less robust but highly efficient estimators. Indeed, when we use a combined estimator with $k = 3$ instead of the non-combined ($k = 1$), the ARE is improved by 3% for $\widehat{\widehat{\alpha}}_1$, by 8% for $\widehat{\widehat{\alpha}}_3$, by 9% for $\widehat{\widehat{\alpha}}_4$, and by 15% for $\widehat{\widehat{\alpha}}_5$.

5 Simulations

Here we supplement our theoretical results concerning the MTM and CMTM estimators with their finite-sample performance evaluations. The objective is to see how large the sample size n should be for the estimators to achieve (asymptotic) unbiasedness and for their finite-sample relative efficiency (RE) to reach the corresponding ARE level. The univariate and multivariate RE definitions are similar to those of the ARE except that we now want to account for possible bias, which we do by replacing all entries in the covariance-variance matrix by the corresponding mean-squared errors.

From a specified distribution F (i.e., Pareto and lognormal) we generate 10,000 samples of a specified length n using Monte Carlo. For each sample we estimate the parameters of F using various MTM estimators and then compute the average mean and RE of those 10,000 estimates. This process is repeated 10 times and the 10 average means and the 10 RE's are again averaged and their standard deviations are reported. (Such repetitions are useful for assessing standard errors of the estimated means and RE's. Hence, our findings are essentially based on 100,000 samples.) The standardized MEAN that we report is defined as the average of 100,000 estimates divided by the true value of the parameter that we are estimating. The standard error is standardized in a similar manner.

We start our simulation study with the Pareto distribution Pareto($x_0 = 1$, $\alpha = 0.50$) using the following parameters:

- *Sample size*: $n = 50, 100, 250, 500, 1000$.

- *Estimators of* α:

– MLE (corresponds to MTM with $a = b = 0$).

– MTM with: $a = b = 0.05$; $a = b = 0.10$; $a = b = 0.25$; $a = b = 0.49$; $a = 0.10$ and $b = 0.70$; $a = 0.25$ and $b = 0.00$.

– Combined MTM with: $k = 3$ $(a = b = 0.25 : 0.02 : 0.29)$ and $k = 9$ $(a = b = 0.25 : 0.02 : 0.41)$.

Simulation results are recorded in Table 5.1. The entries of the last column of the table are included

			n					
	a	b	50	100	250	500	1000	∞
MEAN	0	0	1.02(.000)	1.01(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1
	0.05	0.05	0.99(.000)	1.01(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1
	0.10	0.10	1.01(.000)	1.01(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1
	0.25	0.25	1.01(.001)	1.01(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1
	0.49	0.49	1.03(.001)	1.01(.000)	1.01(.000)	1.00(.000)	1.00(.000)	1
	0.10	0.70	1.04(.001)	1.02(.001)	1.01(.000)	1.00(.000)	1.00(.000)	1
	0.25	0.00	1.03(.000)	1.01(.000)	1.01(.000)	1.00(.000)	1.00(.000)	1
	Combined 3		1.00(.001)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1
	Combined 9		1.00(.001)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1
RE	0	0	0.92(.005)	0.96(.004)	0.98(.004)	1.00(.004)	1.00(.002)	1
	0.05	0.05	0.90(.005)	0.92(.006)	0.92(.004)	0.92(.003)	0.92(.003)	0.918
	0.10	0.10	0.80(.005)	0.83(.005)	0.84(.004)	0.85(.002)	0.85(.003)	0.848
	0.25	0.25	0.65(.004)	0.65(.004)	0.68(.003)	0.68(.002)	0.68(.002)	0.679
	0.49	0.49	0.43(.004)	0.45(.002)	0.47(.002)	0.48(.002)	0.49(.001)	0.487
	0.10	0.70	0.21(.002)	0.23(.001)	0.24(.001)	0.25(.001)	0.25(.001)	0.250
	0.25	0.00	0.87(.005)	0.95(.004)	0.97(.004)	0.99(.004)	0.99(.002)	0.995
	Combined 3		0.71(.004)	0.73(.004)	0.74(.004)	0.74(.002)	0.74(.002)	0.742
	Combined 9		0.71(.004)	0.73(.004)	0.74(.004)	0.74(.002)	0.74(.002)	0.742

Table 5.1: Standardized MEAN and RE in the Pareto case. The entries are mean values (with standard errors in parentheses) based on 100,000 samples.

for completeness and are found in Sections 3 and 4, not from simulations. We see that the MEAN of all Pareto α estimators converges to the parameter α very fast, mostly overestimating it by a few percentage points for $n = 50$ and 100. The bias practically disappears when $n \geq 500$. Note how fast the asymptotic unbiasedness is reached by the combined estimators. The RE's converge to their asymptotic counterparts slower. The relative efficiency of the MLE (i.e., $a = b = 0$) achieves its ARE level only for $n \geq 500$. Other estimators' RE's practically reach their ARE levels at $n \geq 250$, some even at $n = 100$. Note also how similar is the performance of the two combined estimators which supports the earlier made note that there is essentially nothing to gain by combining more than three MTM estimators. But, of course, the combined estimators are performing uniformly better than the corresponding non-combined estimator with $a = b = 0.25$.

We continue our simulation study with the lognormal distribution LN($x_0 = 1$, $\theta = 5$, $\sigma = 3$) using the following parameters:

- *Sample size*: $n = 50$, 100, 250, 500.
- *Estimators of* θ, σ:
  - MLE (corresponds to MTM with $a = b = 0$).
  - MTM with: $a = b = 0.05$; $a = b = 0.10$; $a = b = 0.25$; $a = b = 0.49$; $a = 0.10$ and $b = 0.70$; $a = 0.25$ and $b = 0.00$.

Simulation results are recorded in Table 5.2. The lognormal distribution has two parameters. Since

		$n = 50$		$n = 100$		$n = 250$		$n = 500$	
a	b	θ	σ	θ	σ	θ	σ	θ	σ
MEAN: Mean values of $\widehat{\theta}/\theta$ and of $\widehat{\sigma}/\sigma$									
0	0	1.00(.000)	0.98(.000)	1.00(.000)	0.99(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)
0.05	0.05	1.00(.000)	1.04(.000)	1.00(.000)	1.01(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)
0.10	0.10	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)
0.25	0.25	1.00(.000)	1.05(.001)	1.00(.000)	1.01(.001)	1.00(.000)	1.01(.000)	1.00(.000)	1.00(.000)
0.49	0.49	1.00(.000)	1.71(.006)	1.00(.000)	0.87(.003)	1.00(.000)	1.24(.002)	1.00(.000)	1.03(.002)
0.10	0.70	1.01(.001)	1.02(.001)	1.01(.001)	1.01(.001)	1.00(.000)	1.01(.001)	1.00(.000)	1.00(.000)
0.25	0.00	0.99(.000)	0.99(.000)	1.00(.000)	0.99(.000)	1.00(.000)	1.00(.000)	1.00(.000)	1.00(.000)
RE: Finite-sample efficiencies of MTMs relative to MLEs									
0	0	0.99(.003)		1.00(.002)		1.00(.003)		1.00(.003)	
0.05	0.05	0.82(.002)		0.87(.002)		0.87(.003)		0.87(.003)	
0.10	0.10	0.77(.002)		0.77(.001)		0.77(.003)		0.77(.003)	
0.25	0.25	0.48(.001)		0.50(.001)		0.50(.002)		0.51(.002)	
0.49	0.49	0.04(.000)		0.06(.000)		0.07(.000)		0.07(.000)	
0.10	0.70	0.24(.001)		0.25(.001)		0.25(.001)		0.25(.001)	
0.25	0.00	0.73(.002)		0.72(.002)		0.72(.002)		0.72(.003)	

Table 5.2: Standardized MEAN and RE in the lognormal case. The entries are mean values (with standard errors in parentheses) based on 100,000 samples.

there are more entries to report, we have thus dropped the limiting case $n \to \infty$. We see from Table 5.2

that all estimators in the lognormal case successfully estimate the location θ. Indeed, they practically become unbiased for sample sizes as small as $n = 50$. Estimation of σ, however, reveals a different story. Although most estimators have less than 1% relative bias for $n \geq 100$, the median-based estimator (i.e., $a = b = 0.49$) performs very poorly: it has the relative bias of $+71\%$ for $n = 50$, -13% for $n = 100$, $+24\%$ for $n = 250$, $+3\%$ for $n = 500$. Nonetheless, the RE remains practically unaffected by this and attains its corresponding ARE level for $n \geq 100$ for all estimators.

6 Real-data illustration

In this section we apply the MTM to analyze the normalized damage amounts from the 30 most damaging hurricanes in the United States from 1925 to 1995, as recorded by Pielke and Landsea (1998). The shape of the distribution (see Figure 6.1) is similar to many insurance loss distributions.

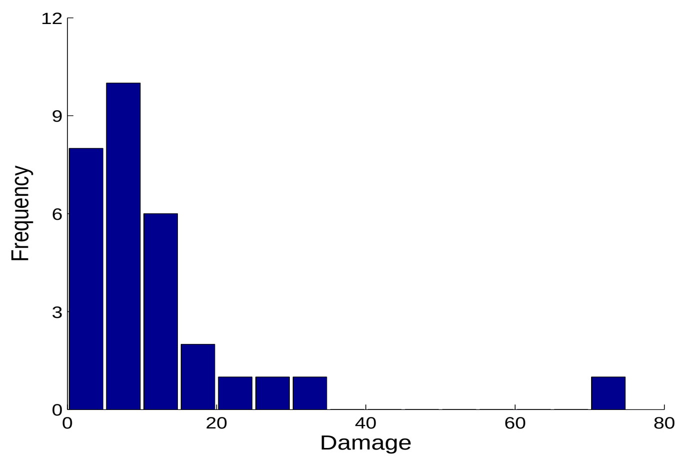

Figure 6.1: The histogram of the top 30 damaging hurricanes.

A preliminary diagnostics, which we have based on a histogram of log-claims and the lognormal QQ-plot, shows that the lognormal distribution provides a satisfactory, though not perfect, overall fit to the data. Nevertheless, we view the current exercise as purely illustrative and fit the lognormal distribution to the data using the MTM approach with two pairs of trimming proportions a and b. For comparison, we also fit this model using the MLE approach. The resulting fits are illustrated in the left panel of Figure 6.3 where the models fitted using the MTM are labeled using T1 (which corresponds to $a = b = 14/30$) and T2 (which corresponds to $a = b = 1/30$). The parameter estimates and goodness-of-fit measurements appear in Table 6.1. The goodness-of-fit is measured

Estimators	Original data			Modified data		
	$\widehat{\theta}$	$\widehat{\sigma}$	Fit	$\widehat{\theta}$	$\widehat{\sigma}$	Fit
MLE	2.077	0.834	0.104	2.154	1.098	0.293
T1 $\left(a = \frac{14}{30}, b = \frac{14}{30}\right)$	2.037	1.675	0.662	2.037	1.675	0.651
T2 $\left(a = \frac{1}{30}, b = \frac{1}{30}\right)$	2.043	0.852	0.101	2.043	0.852	0.178

Table 6.1: Parameter estimates and goodness-of-fit measurements of the lognormal model to the original and modified hurricane data.

using the mean absolute deviation $(1/30)\sum_{j=1}^{30} |\log \widehat{F}^{-1}((j-0.5)/30) - \log X_{j:30}|$ between the log-fitted and log-observed data. It is clear that the parameter estimates and model fits are strongly dependent on the trimming proportions a and b and thus the proportions should be chosen carefully. In particular, the T1 estimator is highly robust but very inefficient, with the ARE being only 14%. The MLE procedure, being most efficient but non-robust, yields a reasonable overall fit, especially when compared to that of T1. A closer examination of the QQ-plot reveals, however, that only the smallest and the largest observations do not follow the straight line pattern. Therefore, symmetric trimming of one observation at each tail leads to the T2 fit which is practically identical to the MLE fit. To see benefits of robust fitting, we have slightly modified the original data set by replacing the largest observation 72.303 with 723.03. The resulting fits are illustrated in the right panel of Figure 6.2. The new parameter estimates and goodness-of-fit measurements appear in Table 6.1. As we see, the T1 and T2 parameter estimates are not affected by the data modification whereas the new MLE fit is significantly different.

Consider now estimation of the severity component of the pure premium for an insurance benefit equal to the amount by which a hurricane's damage exceeds 5 (billion) with a maximum benefit of 20. That is, if the hurricane damage is X with distribution function F, we seek

$$\Pi[F] = \int_5^{25} (x-5)\, \mathrm{d}F(x) + 20(1 - F(25)). \tag{6.1}$$

Since it is now most important that our fitted distribution captures the behavior of the underlying damage distribution between 5 and 25, the MTM estimator is most natural with the choices $a = 8/30$ (which corresponds to the proportion of observations below 5) and $b = 3/30$ (which corresponds to

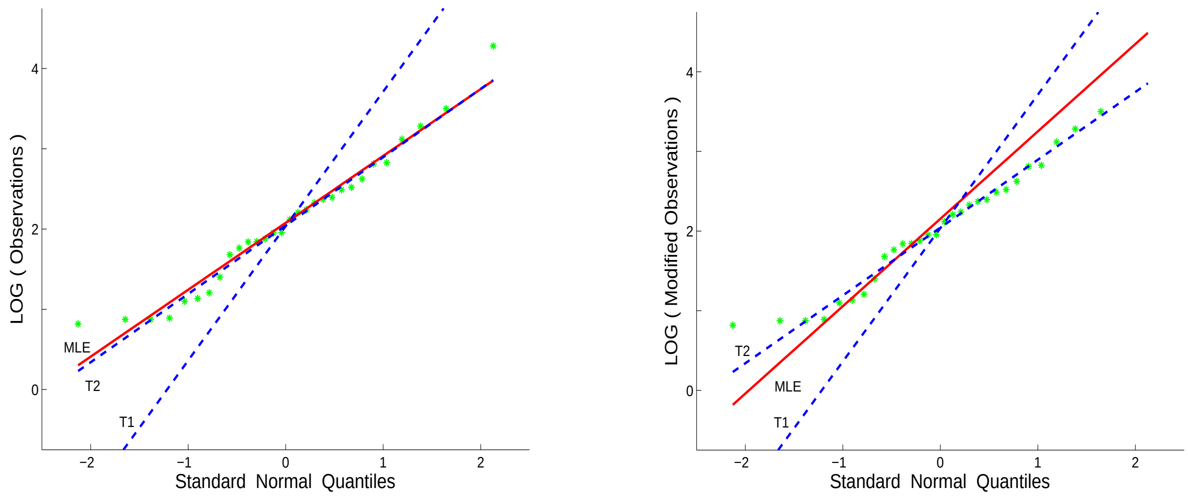

Figure 6.2: Lognormal fits to the original (left panel) and modified (right panel) hurricane data.

the proportion of observations above 25). We denote this MTM estimator by T3. As we see from Figure 6.3, the overall T3 fit is very similar to those of T2 and MLE but it yields a closer fit than

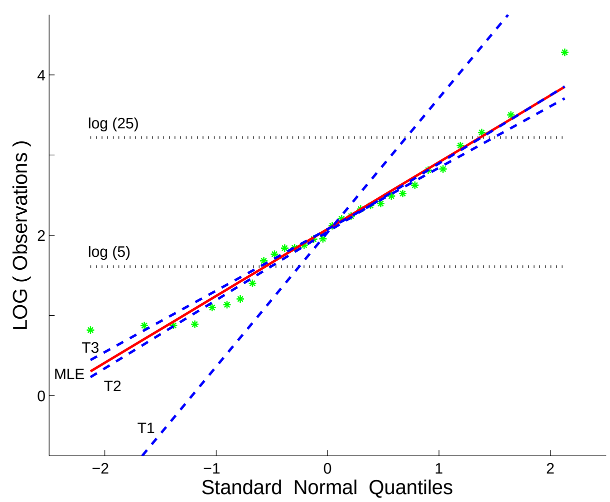

Figure 6.3: Lognormal fits to the top 30 damaging hurricanes with the two dotted lines marking the layer [5, 25].

the latter two procedures over the layer of interest, which is [5, 25]. In Table 6.2 we also provide the premiums calculated using (6.1) for each fitted model, and we can compare these premiums with the empirical premium $\Pi[\widehat{F}_n]$, where $\widehat{F}_n$ denotes the empirical distribution function. In addition, Table 6.2 also contains 95% confidence intervals for the premium $\Pi[F]$. For parametric confidence intervals, we

Estimators	Statistical quantities			Actuarial premiums	
	$\widehat{\theta}$	$\widehat{\sigma}$	Restricted fit	Estimate	Confidence interval
MLE	2.077	0.834	0.054	5.604	(3.368; 7.839)
T1 $\left(a = \frac{14}{30},\, b = \frac{14}{30}\right)$	2.037	1.675	0.413	7.347	(2.533; 12.161)
T2 $\left(a = \frac{1}{30},\, b = \frac{1}{30}\right)$	2.043	0.852	0.057	5.437	(3.168; 7.705)
T3 $\left(a = \frac{8}{30},\, b = \frac{3}{30}\right)$	2.075	0.766	0.042	5.336	(3.065; 7.606)
Empirical	—	—	—	5.416	(3.111; 7.722)

Table 6.2: Parameter estimates, goodness-of-fit measurements (restricted to the data in [5, 25]), point estimates, and 95% confidence intervals of the pure premium for the layer [5, 25].

use the delta method applied to the transformation of parameter estimators given by equation (6.1) together with the MTM and MLE asymptotic distributions, which have been discussed earlier. For constructing the empirical interval, we use the classical central limit theorem and have that

$$\Pi[\widehat{F}_n] \;\sim\; \mathcal{AN}\left(\Pi[F],\, n^{-1}V[F]\right),$$

where $V[F] = \int_5^{25}(x-5)^2\,\mathrm{d}F(x) + 400(1-F(25)) - (\Pi[F])^2$. We observe that the MTM estimates with appropriate trimming proportions (i.e., the estimators T2 and T3) lead to premium estimates that are closer to the empirical estimate than those obtained with overtrimming (i.e., T1) or undertrimming (i.e., MLE). Also, the main advantage of parametric procedures (MTM and MLE) over the empirical approach is that in general they produce shorter confidence intervals for the measures of interest, though the advantage is minimal in the current example. In summary, the illustration we have provided in this section exemplifies the idea that the MTM is an appropriate choice for various model-fitting situations including those when a close fit in one or both tails of the distribution is not required.

7 Summary and conclusions

In this paper, we have introduced and developed two general methods for estimating the parameters of claim severity distributions: the method of trimmed moments (MTM) and the combined method of trimmed moments (CMTM). The two methods utilize the underlying principle of the classical method of moments and their actions on data are easily seen and understood, which is a most appealing feature from the practical point of view. Furthermore, we have illustrated with examples that the MTM and CMTM estimators can achieve various degrees of robustness that can easily be controlled by the user, thus allowing us to reach a desired balance between robustness and efficiency. A general asymptotic theory for the new estimators is provided, which makes them readily applicable for constructing confidence intervals and sets, and for testing hypotheses. The performance of the MTM and CMTM estimators for several parametric families of loss distributions and various sample sizes has been investigated in detail. The MTM estimators have also been applied on a real-life data set to analyze damage amounts caused by major hurricanes. Our considerations have also revealed that calculating premiums for layers of insurance coverage is a task for which the MTM is particularly natural.

Acknowledgments

The authors are grateful to two anonymous referees and an associate editor for suggestions and constructive criticism that helped to revise the paper. The second and third authors also gratefully acknowledge their research support by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] Brazauskas, V. and Ghorai, J. (2007). Estimating the common parameter of normal models with known coefficients of variation: a sensitivity study of asymptotically efficient estimators. *Journal of Statistical Computation and Simulation*, **77**(8), 663–681.

[2] Brazauskas, V., Jones, B.L., Zitikis, R. (2007). Robustification and performance evaluation of empirical risk measures and other vector-valued estimators. *Metron*, **LXV**(2), 175–199.

[3] Brazauskas, V. and Serfling, R. (2000a). 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), 123–126. Reply: **5**(3), 126–128.

[4] Brazauskas, V. and Serfling, R. (2000b). Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes*, **3**(3), 231–249.

[5] Brazauskas, V. and Serfling, R. (2003). Favorable estimators for fitting Pareto models: a study using goodness-of-fit measures with actual data. *ASTIN Bulletin*, **33**(2), 365–381.

[6] Cowell, F.A. and Victoria-Feser, M.P. (2006). Distributional dominance with trimmed data. *Journal of Business and Economic Statistics*, **24**, 291–300.

[7] Cowell, F.A. and Victoria-Feser, M.P. (2007). Robust stochastic dominance: a semiparametric approach. *Journal of Economic Inequality*, **7**, 21–37.

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

[9] Dupuis, D. and Victoria-Feser, M.P. (2006). Robust prediction error criterion for Pareto modeling of upper tails. *Canadian Journal of Statistics*, **34**(4), 639–658.

[10] Kaiser, T. and Brazauskas, V. (2006). Interval estimation of actuarial risk measures. *North American Actuarial Journal*, **10**(4), 249–268.

[11] Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, New York.

[12] Klugman, S.A., Panjer, H.H., and Willmot, G.E. (2004). *Loss Models: From Data to Decisions*, 2nd edition. Wiley, New York.

[13] Marazzi, A. and Yohai, V.J. (2004). Adaptively truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference*, **122**(1-2), 271–291.

[14] Marazzi, A. and Yohai, V.J. (2006). Robust Box-Cox transformations based on minimum residual autocorrelation. *Computational Statistics and Data Analysis*, **50**, 2752–2768.

[15] Pielke, Jr., R. A. and C. W. Landsea (1998). Normalized hurricane damages in the United States: 1925-1995, *Weather and Forecasting*, **13**, 621–631.

[16] Scott, D.W. (2001). Parametric statistical modeling by minimum integrated square error. *Technometrics*, **43**, 274–285.

[17] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

[18] Serfling, R. (2002). Efficient and robust fitting of lognormal distributions (with discussion). *North American Actuarial Journal*, **6**(4), 95–109. Discussion: **7**(3), 112–116. Reply: **7**(3), 116.

[19] Soong, T.T. (1969). An extension of the moment method in statistical estimation. *SIAM Journal on Applied Mathematics*, **17**(3), 560–568.

[20] Vandewalle, B., Beirlant, J., Christmann, A., Hubert, M. (2007). A robust estimator for the tail index of Pareto-type distributions. *Computational Statistics and Data Analysis*, **51**(12), 6252–6268.

A Appendix: Auxiliary results

Location-Scale Family

In subsection 3.1 we expressed the entries of the covariance-variance matrix $\mathbf{\Sigma}$ in terms of the constants $c_k^* \equiv c_k^*(F_0,\, a,\, b)$ and then noted that the latter ones can in turn be expressed in terms of the constants c_k. These expressions are as follows:

$$\begin{aligned}
c_1^* &= \frac{1}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} \big(\min\{u,v\} - uv \big)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&= \frac{1}{(1-a-b)^2} \Bigg\{ a(1-a)\big[F_0^{-1}(a)\big]^2 + b(1-b)\big[F_0^{-1}(1-b)\big]^2 - 2ab\, F_0^{-1}(a) F_0^{-1}(1-b) \\
&\quad - 2(1-a-b)\Big[aF_0^{-1}(a) + bF_0^{-1}(1-b)\Big]\, c_1 - (1-a-b)^2\, c_1^2 + (1-a-b)\, c_2 \Bigg\},
\end{aligned}$$

$$\begin{aligned}
c_2^* &= \frac{1}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} \big(\min\{u,v\} - uv \big)\, F_0^{-1}(u)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&= \frac{1}{2(1-a-b)^2} \Bigg\{ a(1-a)\big[F_0^{-1}(a)\big]^3 + b(1-b)\big[F_0^{-1}(1-b)\big]^3 \\
&\quad - ab\, F_0^{-1}(a) F_0^{-1}(1-b)\Big[F_0^{-1}(a) + F_0^{-1}(1-b)\Big] - (1-a-b)\Big[a\big[F_0^{-1}(a)\big]^2 + b\big[F_0^{-1}(1-b)\big]^2\Big]\, c_1 \\
&\quad - (1-a-b)\Big[aF_0^{-1}(a) + bF_0^{-1}(1-b)\Big]\, c_2 - (1-a-b)^2\, c_1\, c_2 + (1-a-b)\, c_3 \Bigg\},
\end{aligned}$$

and

$$\begin{aligned}
c_3^* &= \frac{1}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} \big(\min\{u,v\} - uv \big)\, F_0^{-1}(u)\, F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(v)\, \mathrm{d}F_0^{-1}(u) \\
&= \frac{1}{4(1-a-b)^2} \Bigg\{ a(1-a)\big[F_0^{-1}(a)\big]^4 + b(1-b)\big[F_0^{-1}(1-b)\big]^4 - 2ab\big[F_0^{-1}(a)\big]^2 \big[F_0^{-1}(1-b)\big]^2 \\
&\quad - 2(1-a-b)\Big[a\big[F_0^{-1}(a)\big]^2 + b\big[F_0^{-1}(1-b)\big]^2\Big]\, c_2 - (1-a-b)^2\, c_2^2 + (1-a-b)\, c_4 \Bigg\}.
\end{aligned}$$

Pareto Model

In subsection 3.2 we used the functions $I_0(x,y)$ and $J((x_2,y_2),\,(x_1,y_1))$ when obtaining the MTM estimators for the Pareto parameter α as well as for deriving their asymptotic properties. We next present explicit formulas for the functions depending on the values of their arguments. Specifically, for $0 < x < y < 1$, we have that

$$\begin{aligned}
I_0(x,y) &:= \int_x^y \log(1-u)\,\mathrm{d}u = (x-y) + (1-x)\log(1-x) - (1-y)\log(1-y), \\
I_1(x,y) &:= \int_x^y \frac{u}{1-u}\,\mathrm{d}u = (x-y) + \log\left(\frac{1-x}{1-y}\right), \\
I_2(x,y) &:= \int_x^y \frac{u^2}{1-u}\,\mathrm{d}u = I_1(x,y) - 0.5(y^2 - x^2).
\end{aligned}$$

As to the entries of the covariance-variance matrix $\mathbf{\Sigma}$, we have the following expressions for the function

$$J\big((x_2,y_2),\,(x_1,y_1)\big) = \int_{x_2}^{y_2}\int_{x_1}^{y_1} \frac{\min\{u,v\} - uv}{(1-u)(1-v)}\,\mathrm{d}v\,\mathrm{d}u$$

depending on the order of its arguments:

- When $0 < x_1 \le x_2 < y_2 \le y_1 < 1$, then $J((x_2,y_2),\,(x_1,y_1))$ is equal to

$$(y_2 - x_2)\big[x_1 + \log(1-x_1)\big] - I_0(x_2,y_2) + (y_1 - 1)I_1(x_2,y_2).$$

- When $0 < x_1 < y_1 \le x_2 < y_2 < 1$, then $J((x_2,y_2),\,(x_1,y_1))$ is equal to

$$(y_2 - x_2)I_1(x_1,y_1).$$

- When $0 < x_1 \le x_2 < y_1 \le y_2 < 1$, then $J((x_2,y_2),\,(x_1,y_1))$ is equal to

$$(y_1 - x_2)\big[x_1 + \log(1-x_1)\big] - I_0(x_2,y_1) + (y_1 - 1)I_1(x_2,y_1) + (y_2 - y_1)I_1(x_1,y_1).$$