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

VYTARAS BRAZAUSKAS[1]
University of Wisconsin-Milwaukee

BRUCE L. JONES
University of Western Ontario

RIČARDAS ZITIKIS
University of Western Ontario

Abstract. Many useful quantities in non-life insurance depend on the claim severity distribution, which is usually modeled by assuming a parametric form. Estimating the model parameters is therefore very important in obtaining good estimates of the quantities of interest. The maximum likelihood technique yields efficient estimators with a number of other desirable properties. However, these estimators are generally not robust. Since outliers are common in insurance loss data, it is beneficial to use a method that allows one to achieve a balance between efficiency and robustness. In this paper, we suggest a method we call the method of trimmed moments (MTM). The resulting MTM estimators can achieve various degrees of robustness and allow the actuary to easily see the actions of the estimators on the data. We illustrate these points with detailed analyses of the MTM estimators for location-scale families, as well as for the lognormal, Pareto and other loss distributions. The performance of the estimators is illustrated in a simulation study as well as on a real data set on hurricane damage in the United States.

Keywords: Loss models; Premium calculations; Robust statistics.

[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

1 Introduction

Loss severity distributions are required for a variety of premium and reserve calculations in non-life insurance. Both nonparametric and parametric estimators can be used in these calculations. While nonparametric estimators of the distribution functions are convenient in many circumstances, a parametric model is often preferred for a number of reasons. First, parametric models yield smooth estimates of distribution function values; this is important because the lack of smoothness can lead to undesirable progressions of premium rates for different deductible values or anomalous orderings of premiums by rating category. Second, parametric models allow one to extrapolate probabilities to ranges of values that are beyond the range of the actual data; while the researcher has to be always cautious with extrapolating, it may be necessitated by a quantity to be estimated, such as the premium for a high deductible excess of loss policy. Third, 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 wrong, but the 'risk' might be worth taking. Fourth, there may be specific aspects of the structure of the distribution that are either desirable or believed to exist (e.g., a mode at 0); parametric models allow the researcher to impose such structures even when they are not apparent from the data. Finally, parametric models often have fairly simple forms that make them appealing and easy to use. For details on families of 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. Here we only note in passing that, for example, nested families are particularly useful to facilitate model selection; scale families are generally used for modeling insurance loss amounts since, for example, inflation or changes in currency can be handled by simply changing the scale parameter.

A variety of methods exist for estimating the parameters of parametric models. The most popular method is maximum likelihood estimation, due to, for example, its desirable asymptotic properties, invariance under parameter transformations, etc. However, the maximum likelihood estimators are not robust, and the influence of outlying observations can be substantial. Such observations can certainly occur in the context of insurance losses. The circumstances surrounding a particular loss might be quite unusual and atypical of the other losses in the portfolio; think, for example, about a loss that receives an extensive media attention. Also, an insurer may wish to use the loss data from one insurance portfolio to analyze a different portfolio for which little or no data 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.

There are many methods proposed in the literature for robust estimation of parameters. Practically all of them can be found as special cases of some general classes of statistics, such as M-, L-, or R-statistics (see, e.g., Serfling, 1980, Chapters 7–9), with the class of M-statistics being arguably the most popular choice, which is mostly due to a close relationship between the objective function of M-statistic and its influence function. The latter function is an important tool for studying robustness properties of the estimators. Recent examples of successful implementations of robust procedures in practice include extreme-value applications in finance and economics (see, e.g., Dupuis and Victoria-Feser, 2006, Cowell and Victoria-Feser, 2006, 2007) and geopedology (Vandewalle *et al.*, 2007), robust fitting of non-standard regression models (Marazzi and Yohai, 2004, 2006), and a general minimum-distance modeling approach of Scott (2001). While all these methods deliver what they are designed to deliver (i.e., accurate, meaningful model fits), their main drawbacks are: computational complexity, as they require numerical solution of one or more equations, and the lack of transparency. Avoidance of such complexities is very appealing in practice.

To resolve a number of issues addressed above, in the current paper we introduce and develop a new general method for estimating the parameters of claim severity distributions, which we call the *method of trimmed moments*, or MTM for short. 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 and k sample *trimmed* moments, and then we solve the resulting system of equations with respect to the unknown parameters. The proposed approach has several desirable features. First, the trimmed moments *always exist* irrespectively of the underlying distribution. Consequently, the estimators based on them exist provided that the system of equations (2.3) has a solution. Second, the estimators obtained via this technique can achieve various degrees of robustness, and it can be *easily specified* by the user by setting appropriate trimming proportions. Third, the method 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 and in detail presented in Section 2, where we also derive asymptotic properties of the MTM estimators. Examples of MTM estimators for several most common claim severity distributions are developed in Section 3. In Section 4 we suggest an extension of the MTM method. 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. Main findings of the paper are summarized in Section 7.

2 Method of trimmed moments

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables, whose distribution function (cdf) we denote by F. Assume that the cdf is given in a parametric form, and let it depend on, say, k unknown parameters, which we denote by $\theta_1, \ldots, \theta_k$; the number k is fixed. Denote the order statistics of $X_1, \ldots, X_n$ by $X_{1:n} \leq \cdots \leq X_{n:n}$. The MTM estimators of $\theta_1, \ldots, \theta_k$ are found by the following four step procedure:

1. Compute k sample trimmed moments

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

   for all $j = 1, \ldots, k$, where $h_j : \mathbf{R} \to \mathbf{R}$ is a specially chosen function (to be discussed below), and $m_n(j)$ and $m_n^*(j)$ are integers such that $0 \leq m_n(j) < n - m_n^*(j) \leq n$, and $m_n(j)/n \to a_j$ and $m_n^*(j)/n \to b_j$ as $n \to \infty$, where the proportions a_j and b_j are chosen by the researcher (to be discussed below).

2. Derive the corresponding population trimmed moments

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

for all $j = 1, \dots, k$, where F^{-1} denotes the quantile function of F, defined by in the usual way as $F^{-1}(u) = \inf\{x : \; F(x) \geq u\}$ for all $u \in (0,1)$. Note that when $a_j = b_j = 0$, then $\mu_j = \mathbf{E}[h_j(X)]$.

3. Match the population and sample trimmed moments and solve the following system of equations with respect to $\theta_1, \dots, \theta_k$:

$$\left\{ \begin{array}{rcl} \mu_1(\theta_1, \dots, \theta_k) & = & \widehat{\mu}_1, \\ & \vdots & \\ \mu_k(\theta_1, \dots, \theta_k) & = & \widehat{\mu}_k. \end{array} \right. \tag{2.3}$$

4. The above obtained solutions, which we denote by

$$\left\{ \begin{array}{rcl} \widehat{\theta}_1 & = & g_1(\widehat{\mu}_1, \dots, \widehat{\mu}_k), \\ \vdots & & \\ \widehat{\theta}_k & = & g_k(\widehat{\mu}_1, \dots, \widehat{\mu}_k), \end{array} \right. \tag{2.4}$$

are, by definition, the MTM estimators of the parameters $\theta_1, \dots, \theta_k$. Note that the functions g_j are such that $g_j(\mu_1, \dots, \mu_k) = \theta_j$.

We next discuss choosing the functions h_j and the trimming proportions a_j and b_j, as well as the flexibility that the MTM can offer. To start with, note that the choice $h_j(t) = t^j$ for all $j = 1, \dots, k$ leads to the matching of 'genuine' trimmed moments, which justifies the name of the 'method of trimmed moments'. Functions of this type work well for the location-scale families, where we choose $h_1(t) = t$ and $h_2(t) = t^2$ (see Section 4.1). However, for claim severity distributions, which are typically asymmetric and have non-zero densities only on the real half-line, other choices of h_j may be more appropriate. Fortunately, a number of loss models can be converted – through the logarithmic transformation – to a location-scale family, thus suggesting the use of $h_1(t) = \log t$ and $h_2(t) = (\log t)^2$. Examples of such distributions include lognormal, log-t, log-logistic, and Weibull, which after the logarithmic transformation become normal, Student's t, logistic, and Gumbel (extreme-value), respectively. Further, for the single-parameter Pareto distribution, $h(t) = \log t$ is also an appropriate choice because log-Pareto is shifted-exponential with 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 choose $h_1(t) = h_2(t) = t$

but use two different pairs of the trimming proportions $(a_1,\, b_1)$ and $(a_2,\, b_2)$. The same idea also works for the two-parameter Pareto family. In general, we note that if h_j is some sort of a power function, mathematical computations in (2.2) and (2.3) are more straightforward when the exponent is chosen as small as possible. Note also that the choice of $h_j(t) = t^j$ combined with $a_j = b_j = 0$ for all $j = 1, \dots, k$ yields the classical method of moments. As to the choice of the trimming proportions a_j and b_j, there is no rigorous answer to how much robustness one needs, and typically an argument is made on the basis of how much efficiency at the assumed model F one is willing to sacrifice. It also requires experience and intuition, and is probably more art than science. However, one thing is certain and has to be kept in mind, namely, robustness of the entire estimation procedure is determined by the least trimmed moments. That is, the MTM estimators will remain resistant against the proportion $a_* = \min\{a_1, \dots, a_k\}$ of lowest observations and against the proportion $b_* = \min\{b_1, \dots, b_k\}$ of highest observations. In conclusion, we see that the MTM is fairly flexible by allowing the researcher to work with various combinations of h_j, a_j, and b_j, depending on the problem at hand.

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

$$\sigma_{ij}^2 = \frac{1}{(1-a_i-b_i)(1-a_j-b_j)} \int_{a_i}^{1-b_i} \int_{a_j}^{1-b_j} \big(\min\{u,v\} - uv\big)\, \mathrm{d}h_j\big(F^{-1}(v)\big)\, \mathrm{d}h_i\big(F^{-1}(u)\big). \tag{2.5}$$

Following Serfling (1980, p. 20), this result can be neatly summarized as saying that

$$(\widehat{\mu}_1, \dots, \widehat{\mu}_k) \quad \text{is} \quad \mathcal{AN}\big((\mu_1, \dots, \mu_k),\, n^{-1}\,\mathbf{\Sigma}\big). \tag{2.6}$$

Using statement (2.6), we now easily derive the asymptotic distribution, which is normal, of the MTM estimators in equations (2.4). Indeed, the delta method (cf., e.g., Serfling, 1980, Section 3.3) implies that the vector of MTM estimators $(\widehat{\theta}_1, \dots, \widehat{\theta}_k)$ is asymptotically normal with the mean vector $(\theta_1, \dots, \theta_k)$ and the covariance-variance matrix $n^{-1}\,\mathbf{D\Sigma D}'$, where $\mathbf{D} = [d_{ij}]_{i,j=1}^{k}$ is the Jacobian of the transformations $g_1, \ldots, g_k$ evaluated at the vector $(\mu_1, \ldots, \mu_k)$, that is, $d_{ij} = \partial g_i / \partial \widehat{\mu}_j |_{(\mu_1,\ldots,\mu_k)}$. In summary,

$$\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right) \quad \text{is} \quad \mathcal{AN}\left((\theta_1, \ldots, \theta_k),\, n^{-1}\,\mathbf{D\Sigma D}'\right). \tag{2.7}$$

Statement (2.7) can be utilized in several ways. First, it of course implies that $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ are consistent estimators of $(\theta_1, \ldots, \theta_k)$. Secondly, it can be easily applied to the hypothesis testing problems and constructing of confidence intervals, or regions. Thirdly, it can also be incorporated in further (parametric) actuarial modeling, such as estimation of risk measures, calculation of credibility premiums, pricing of reinsurance contracts, solvency analysis, etc. Some of these applications are studied using specialized derivations of robust estimators by Brazauskas and Serfling (2000a, 2003), Kaiser and Brazauskas (2006), and Dornheim and Brazauskas (2007).

3 Examples: three parametric models

In this section, we present three examples of MTM estimators and derive their asymptotic properties. The first example treats general (i.e., not necessarily symmetric) location-scale families, whereas the other two examples are more specialized and focus on the Pareto and lognormal distributions. In these 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) of the MTM estimators with respect to the maximum likelihood estimator (MLE). We use the following ratio as the definition of the univariate ARE:

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

In the multivariate case, we replace the variances by the corresponding generalized variances, which are the determinant of the asymptotic covariance-variance matrices of vector-estimators, and then raise the ratio to the power $1/k$, where k is the dimension of the vector-parameter. 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 from a location-scale family:

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

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 $F^{-1}(t) = \theta + \sigma F_0^{-1}(t)$. The cdf F has two unknown parameters, and so we employ two trimmed moments. There are no restrictions on the location parameter θ, and so choose h as $h_1(t) = t$. To ensure that the estimator of σ 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 specified by the researcher proportions $m_n(1)/n = m_n(2)/n \to a$ and $m_n^*(1)/n = m_n^*(2)/n \to b$.

Note 3.1 We can, of course, consider more general trimming, such as $m_n(1) \neq m_n(2)$ or $m_n^*(1) \neq m_n^*(2)$ or both. 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 case when $m_n(1)/n = m_n(2)/n$ and $m_n^*(1)/n = m_n^*(2)/n$.

We next calculate the corresponding population trimmed moments

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

where

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

The numerical values of c_k are provided in Tables A.1 and A.2 for several choices of F_0 and various values of a and b. 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, σ, we have the MTM estimators

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

of θ and σ, respectively. Furthermore, the entries of the matrix $\mathbf{\Sigma}$ are

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

and

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

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 (cf. Appendix A for detail). The entries of the matrix $\mathbf{D}$

are found by differentiating the functions g_i given by equations (3.2), with the following results:

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

As a consequence, we have that

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

We summarize the above findings by saying that the MTM estimator

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

Note that the matrix $\mathbf{S}$ does not depend on any unknown parameters and can be expressed only 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 and can be interpreted as, for example, 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 The hint about the logarithmic function comes from the fact that Pareto(x_0, α) is equivalent – through the logarithmic transformation of the variable – to a shifted-exponential variable. That is, 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 $G(z) = 1 - e^{-(z-z_0)/\theta}$, $z > z_0$, where $z_0 = \log x_0$ and $\theta = \alpha^{-1}$. Hence, α^{-1} represents the unknown part of the mean of Z and therefore could naturally be estimated with a trimmed mean. The relationship between the Pareto and shifted-exponential distributions has effectively been used by Brazauskas and Serfling (2000a,b, 2003) to develop a robust estimation technique for the Pareto(x_0, α) model.

Having thus chosen the function h, we now 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 specified proportions $m_n(1)/n \to a$ 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

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

The entries of the matrices $\mathbf{\Sigma}$ and $\mathbf{D}$, which are one dimensional, are as follows. For $\mathbf{\Sigma}$, we have

$$\begin{aligned}
\sigma_{11}^2 &= \frac{1}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} \big(\min\{u,v\} - uv \big)\, \mathrm{d}\log\big(F^{-1}(v)/x_0\big)\, \mathrm{d}\log\big(F^{-1}(u)/x_0\big) \\
&= \frac{1/\alpha^2}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} \frac{\min\{u,v\} - uv}{(1-u)(1-v)}\, \mathrm{d}v\, \mathrm{d}u \\
&= \frac{1/\alpha^2}{(1-a-b)^2}\, J\big((a, 1-b),\, (a, 1-b)\big)
\end{aligned}$$

with the obvious notation for the function J. The Jacobian $\mathbf{D}$ is found by differentiating the function in (3.6) and evaluating the derivative at μ_1, which gives

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

Hence,

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

Summarizing the above findings, we have that the MTM estimator

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

Finally, we investigate the amount of efficiency lost because of using $\widehat{\alpha}_{\text{MTM}}$ instead of the MLE of α. The latter estimator is $\widehat{\alpha}_{\text{MLE}} = n/\sum_{i=1}^{n}\log(X_i/x_0)$, and we know (cf., e.g., Brazauskas and Serfling, 2000a) that it is $\mathcal{AN}(\alpha,\,\alpha^2/n)$. Hence, $\text{ARE}(\widehat{\alpha}_{\text{MTM}},\,\widehat{\alpha}_{\text{MLE}}) = 1/C$. Note in passing that when $m_n(1) = m_n^*(1) = 0$, then the MTM estimator (3.6) becomes $\widehat{\alpha}_{\text{MLE}}$; also, since $C \to 1$ as $a = b \to 0$, the MLE's asymptotic distribution follows from (3.7). Numerical values of these AREs are provided in Table 3.1 for several values of the trimming proportions a and b. Specifically, we see from Table 3.1 that when

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

b is fixed, then the MTM estimators with no lower trimming ($a = 0$) and with symmetric trimming ($a = b$) are almost equivalent. When $b = 0$ (i.e., no upper trimming), then the efficiency decreases slowly. Moreover, the MTM approach 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 $a = 0.10$ and $b = 0.85$, and $a = 0.85$ and $b = 0.10$ 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. In summary, all this discussion just confirms the well-known fact that most information about α is contained in the upper tail of Pareto(x_0, α).

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, and Φ is the standard normal cdf. The parameter x_0 is assumed to be known, and it can be interpreted, for example, as in the Pareto case discussed above. Note that since the logarithmic transformation makes this distribution normal, which is a member of the location-scale family, the results of Section 3.1 apply, though with two obvious modifications: we now use $h_1(t) = \log(t - x_0)$ and $h_2(t) = \log^2(t - x_0)$. Hence, the MTM estimators of θ and σ are

$$\left\{ \begin{array}{lcl} \widehat{\theta}_{\mathrm{MTM}} & = & \widehat{\mu}_1 - c_1\, \widehat{\sigma}_{\mathrm{MTM}}, \\[2ex] \widehat{\sigma}_{\mathrm{MTM}} & = & \sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)}, \end{array} \right. \tag{3.9}$$

where numerical values of c_1 and c_2 are available in, respectively, Tables A.1 and A.2 when F_0 is the standard normal. 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 specified proportions $m_n(1)/n = m_n(2)/n \to a$ and $m_n^*(1)/n = m_n^*(2)/n \to b$. As follows from (3.4), the MTM estimator

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

where $\mathbf{D\Sigma D}'$ is given by (3.3) with the standard normal F_0.

Finally, we examine how much efficiency is lost because of using $(\widehat{\theta}_{\text{MTM}},\, \widehat{\sigma}_{\text{MTM}})$ instead of the MLE of $(\theta,\, \sigma)$:

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

We know (cf., e.g., Serfling, 2002) that

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

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, equals $\left(0.5/\det(\mathbf{S})\right)^{1/2}$. We note in passing that for $m_n(1) = m_n^*(1) = m_n(2) = m_n^*(2) = 0$, the MTM estimators (3.9) become $(\widehat{\theta}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}})$. Also, since $\mathbf{S} \to \mathbf{S_0}$ when $a = b \to 0$, 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 the proportions a and b. Note from Table 3.2 that since the logarithmic transformation of lognormal

	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 highliting the case $a = b$.

data makes the data normal, which is a symmetric distribution, we can anticipate an equivalent performance from the MTM estimators with similar trimming schemes. For example, the efficiencies are identical for the MTM estimators with reversed trimming proportions; e.g., $a = 0.05$ and $b = 0.25$ has ARE $= 0.678$, and $a = 0.25$ and $b = 0.05$ also has ARE $= 0.678$. In fact, we notice that the MTM estimators based on the same fraction of actual observations used (i.e., having same $a + b$) have similar AREs. For example, compare these three cases: first, $a = 0.10$ and $b = 0.15$ has ARE $= 0.722$, and $a = 0.25$ and $b = 0.00$ has ARE $= 0.722$; second, $a = 0.15$ and $b = 0.15$ has ARE $= 0.676$, and $a = 0.05$ and $b = 0.25$ has ARE $= 0.678$; and third, $a = 0.85$ and $b = 0.10$ has ARE $= 0.097$, and $a = 0.25$ and $b = 0.70$ has ARE $= 0.113$. Next, trimming schemes that exclusively focus 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 $= 0.872$ (not bad); $a = b = 0.25$ has ARE $= 0.507$ (so-so); $a = b = 0.49$ has ARE $= 0.074$ (very poor). Compare the latter estimator with the median estimator of θ which is almost 10 times more efficient, as it has ARE $= 0.637$.

We conclude this section with a brief discussion of a similar type of study conducted by Serfling (2002), where the generalized median (GM) estimators are proposed in the case of the lognormal distribution. For a fixed breakdown point (in our case, fixed a and b), the GM estimators are more efficient than the MTM estimators. However, the MTM estimators are more flexible (e.g., they can provide asymmetric protection against upper and lower outliers) and are not demanding computationally. The flexibility problem of GM estimators can be alleviated by considering generalized quantile (GQ) estimators (GM's are special cases of GQ's), as seen from the research by Brazauskas and Serfling (2000b) in the Pareto and exponential cases. The same reference also contains suggestions on how to reduce the computational burden of GQ and GM estimators.

4 Combined MTM estimators

In this section, we discuss a natural extension of the above proposed MTM estimation procedure, which we call the *combined* MTM, or CMTM for short. We next formulate the general idea of the CMTM. Using trimmed moments of various orders, we can generate a set of different estimators for the same parameter θ. Denote the estimators by $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$. Then, as follows from Section 2 (cf. equation (2.7)), the vector $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ is $\mathcal{AN}((\theta, \ldots, \theta),\, n^{-1}\mathbf{\Sigma}_*)$, where the covariance-variance matrix $\mathbf{\Sigma}_* = \mathbf{D\Sigma D}'$ may depend on θ. Next we combine the above estimators of θ in some way hoping to get a better (i.e., with a smaller variance) estimator of θ than the best one among $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$. Consider, for example, the linear combination of the estimators (cf., 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 $\mathcal{AN}(\theta,\, n^{-1}\mathbf{w\Sigma}_*\mathbf{w}')$. Since our criterion is based on minimizing the variance, we choose the weights w_i so that the quadratic form $\mathbf{w\Sigma}_*\mathbf{w}'$, is minimized subject to the constraint $\sum_{i=1}^{k} w_i = 1$. Assuming without loss of generality that $\mathbf{\Sigma}_*$ is nonsingular, a standard variational technique implies that the minimum is attained, and it equals

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

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

Note 4.1 More general situations addressing the case when the matrix $\mathbf{\Sigma}_*$ 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 objective to improve their efficiency. Further, we note that in the case when θ is a scale parameter, the covariance-variance matrix $\mathbf{\Sigma}_*$ is of the form $\theta^2\mathbf{\Sigma}_{**}$, where $\mathbf{\Sigma}_{**}$ does not depend on θ. Thus, the vector $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ is $\mathcal{AN}((\theta, \ldots, \theta),\, n^{-1}\theta^2\mathbf{\Sigma}_{**})$, and the task of estimating θ is within the framework of problems studied, though in a different context, by Brazauskas and Ghorai (2007) who 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 equations (4.1) is preferred in the current project.

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 it 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 would select k estimators, *each* with $0.25 \leq a < 1 - b \leq 0.75$, and combine them according to (4.1). We shall next provide a numerical illustration of this approach in the case of the single-parameter Pareto distribution. Namely, following subsection 3.2, we start with the k sample trimmed moments

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

with specified proportions $m_n(j)/n \to a_j$ and $m_n^*(j)/n \to b_j$. We equate these empirical moments to the corresponding population trimmed moments, which are

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

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:

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

The entries of the covariance-variance matrix $\mathbf{\Sigma}$ and the Jacobian $\mathbf{D}$ are found by straightforward computations; we have 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 an explicit formula of 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, $i = 1, \ldots, k$. In summary, we have that

$$\left(\widehat{\alpha}_{\text{MTM},\,1}, \ldots, \widehat{\alpha}_{\text{MTM},\,k}\right) \quad \text{is} \quad \mathcal{AN}\left((\alpha, \ldots, \alpha),\, \frac{\alpha^2}{n} \mathbf{\Delta} \right).$$

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

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

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 $\text{ARE}(\widehat{\widehat{\alpha}}, \, \widehat{\alpha}_{\text{MLE}}) = (1, \ldots, 1)\mathbf{\Delta}^{-1}(1, \ldots, 1)'$. Numerical values of these AREs are provided in Table 4.1 for various combined estimators, and for eight

Combined	*Breakdown Points*		k							
Estimator	*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: $\text{ARE}(\widehat{\widehat{\alpha}}, \, \widehat{\alpha}_{\text{MLE}})$ for various combined estimators and selected k.

values of k. Specifically, the first estimator $\widehat{\widehat{\alpha}}_1$ 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. To see how important is the choice of component estimators, we consider a second estimator $\widehat{\widehat{\alpha}}_2$ with the same robustness properties as the first one but 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 remaining 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 $b = 0.15 : 0.015 : 0.15 + 0.015(k-1)$; the fourth $\widehat{\widehat{\alpha}}_4$ combines $\widehat{\alpha}_{\text{MTM}}$'s with $a = 0.25 : 0.01 : 0.25 + 0.01(k-1)$ and $b = a$; and the fifth $\widehat{\widehat{\alpha}}_5$ 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 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 which 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, using a combined estimator with $k = 3$ instead of the non-combined ($k = 1$), the ARE is improved by 3% for $\widehat{\widehat{\alpha}}_1$, 8% for $\widehat{\widehat{\alpha}}_3$, 9% for $\widehat{\widehat{\alpha}}_4$, and by 15% for $\widehat{\widehat{\alpha}}_5$.

5 Simulations

In this section we augment the asymptotic results for selected MTM and CMTM estimators with finite-sample performance investigations. Our objective is to see how large the sample size n is needed for the proposed 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 ARE except that we now want to account for possible bias; therefore, all entries of the covariance-variance matrix are replaced with the corresponding mean-squared errors.

Using Monte Carlo simulations, we generate 10,000 samples of length n from a specified distribution, say F. 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 evaluated and reported. (Such repetitions are useful for assessing standard errors of the estimated means and RE's; hence, the study 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 they are estimating. The standard error is standardized in a similar manner. The study was performed for the following choices of simulation parameters:

- Pareto($x_0 = 1$, $\alpha = 0.50$):
  - *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 $k = 9$ ($a = b = 0.25 : 0.02 : 0.41$).

- LN($x_0 = 1$, $\theta = 5$, $\sigma = 3$):
  - *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 $b = 0.00$.

The entries of the last column of Table 5.1 are included 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 α quite fast (mostly) overestimating it by a couple of percentage points for $n = 50$ and 100. The bias is practically eliminated for $n \geq 500$. Note how fast the asymptotic unbiasedness is reached by the combined estimators. The RE's converge to their asymptotic counterparts slower. The relative efficiencies of MLE (i.e., $a = b = 0$) achieve their ARE level only for $n \geq 500$. Other estimators' RE's practically reach their ARE levels at $n \geq 250$, some even at $n = 100$. We also note how identical are the performances of two combined estimators, which supports the theoretical finding that there is essentially nothing to gain by combining more than three MTM estimators. But, of course, they are both performing uniformly better than the corresponding non-combined estimator with $a = b = 0.25$.

The lognormal distribution has two parameters. Thus, there are more entries to report and the case $n \to \infty$ has therefore been dropped. For this distribution, we observe that all the estimators are quite successful in estimating θ (i.e., location), practically becoming unbiased for sample sizes as small as $n = 50$. Estimation of σ, however, is a different story. Although most estimators have less than 1% relative bias for $n \geq$

Statistic	*Estimator* a	b	n = 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: Values of the standardized mean and RE for selected a, b, and n in the Pareto case. The given entries are mean values (with standard errors in parentheses), based on 100,000 samples.

100, the median-based estimator (i.e., $a = b = 0.49$) performs very poor, having the relative bias of +71% for $n = 50$, −13% for $n = 100$, +24% for $n = 250$, +3% for $n = 500$. Nonetheless, the RE's remain practically unaffected by this and attain their corresponding ARE levels for $n \geq 100$ for all estimators.

6 Real-data illustration

In this section, we illustrate the MTM approach using real data on hurricane damages. Pielke and Landsea (1998) present normalized damage amounts from the 30 most damaging hurricanes in the United States from 1925 to 1995. These amounts are shown in Table 6.1 and a histogram is provided in Figure 6.1. The shape of the distribution is similar to many insurance loss distributions faced by actuaries, and there is one observation that is much larger than the others. The lognormal distribution provides a satisfactory

Estimator a	b	$n=50$ θ	σ	$n=100$ θ	σ	$n=250$ θ	σ	$n=500$ 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: Values of the standardized MEAN and RE for selected a, b, and n in the lognormal case. The given entries are mean values (with standard errors in parentheses), based on 100,000 samples.

overall fit to the data, though other distributions may fit quite well over certain ranges (see Figure 6.2, for QQ-plots of lognormal and Weibull distributions).

Damage amounts in billions									
72.303	22.602	13.795	10.965	9.380	7.069	6.313	5.368	3.108	2.399
33.094	16.864	12.434	10.705	9.066	7.039	6.293	4.056	3.000	2.396
26.619	16.629	12.048	10.232	8.308	6.536	5.838	3.338	2.435	2.266

Table 6.1: Top 30 damaging hurricanes in the United States, normalized to 1995 dollars by inflation, personal property increases, and coastal county population changes (1925-1995).

To illustrate the usefulness of MTM estimators in a specific application, we consider 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

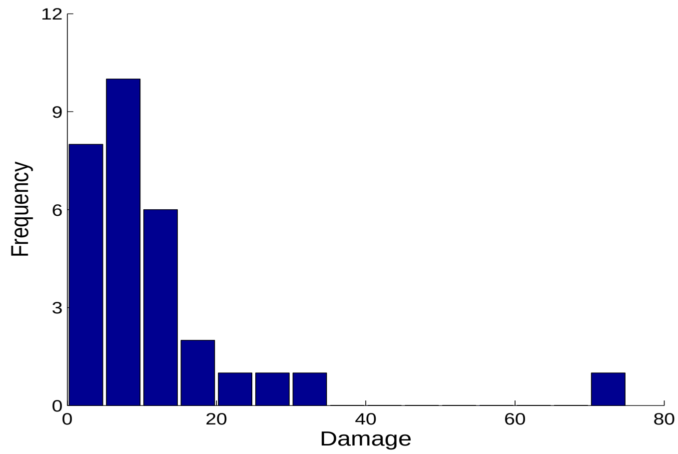

Figure 6.1: The histogram of the top 30 damaging hurricanes.

of 20. That is, if the hurricane damage is X with distribution function F, we seek

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

Clearly, then it is most important that our fitted distribution capture the behavior of the underlying damage distribution between 5 and 25. The MTM approach is therefore quite natural.

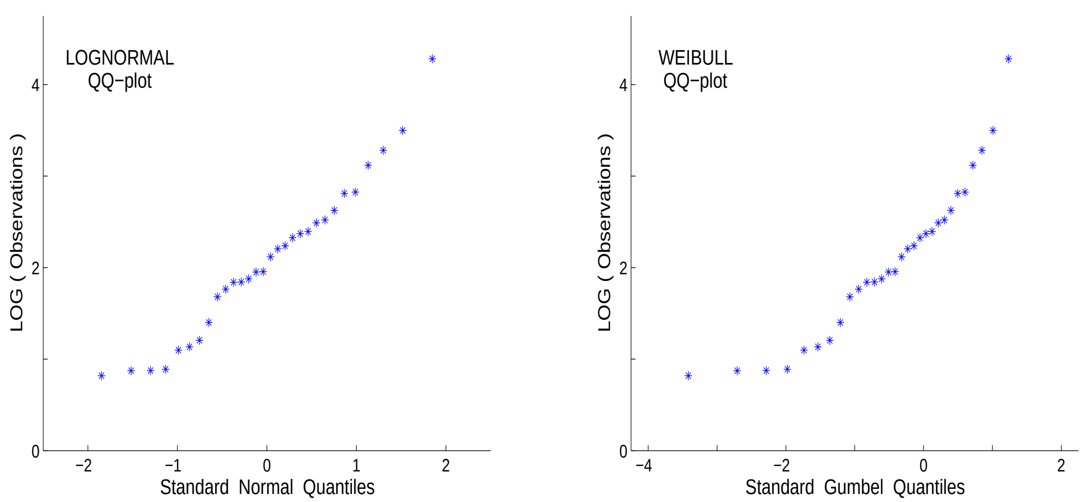

Figure 6.2: Lognormal and Weibull QQ-plots of the top 30 damaging hurricanes.

As mentioned above, the lognormal distribution provides a reasonably good fit to the data. There is also evidence that the Weibull distribution may fit well when one ignores the tails. We therefore consider the MTM fits obtained using a lognormal distribution given by (3.8), and $x_0 = 0$ and a Weibull distribution

$$F(x) = 1 - \exp\{-(x/\beta)^{\alpha}\}, \quad x > 0.$$

As is the case for the lognormal distribution, the logarithm of a Weibull random variable is a member of the location-scale family, the Gumbel (extreme-value) distribution. We can therefore use the results of Section 3.1 with $h_1(t) = \log t$ and $h_2(t) = (\log t)^2$.

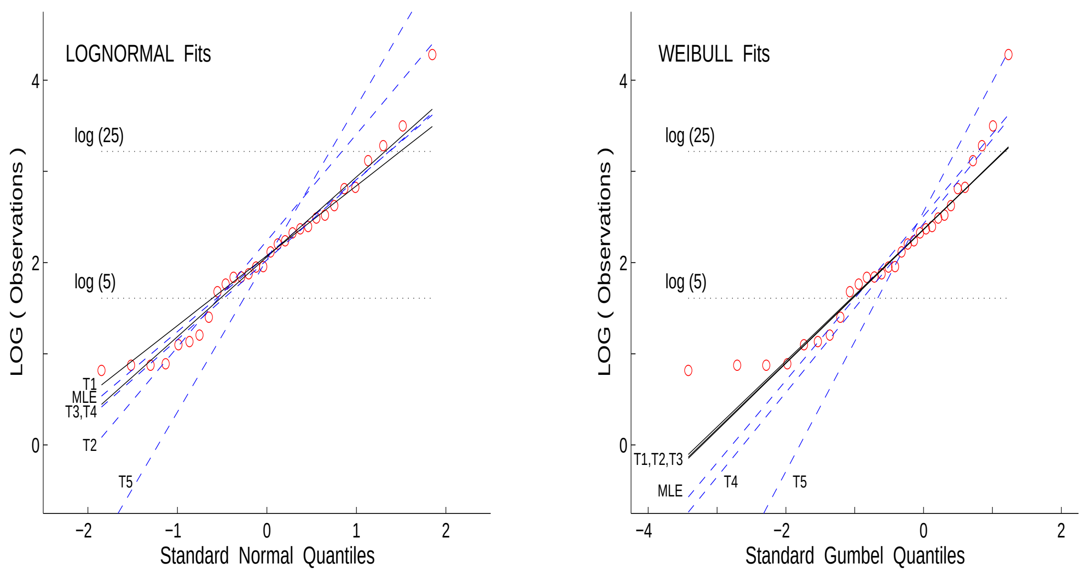

Figure 6.3: Lognormal and Weibull fits to the top 30 damaging hurricanes.

Both models were fit using the MTM approach and several pairs of trimming proportions, a and b. For comparison, we also fit these models using the MLE approach. The resulting fits are illustrated in Figure 6.3, where dotted lines mark the range of interest (i.e., damages between 5 and 25), and the models fitted by the MTM approach are labeled by T1, ..., T5 with the numerical part representing the order of estimators in Table 6.2 (i.e., T1 corresponds to MTM with $a = 8/30$ and $b = 3/30$, T2 to MTM with $a = 3/30$ and $b = 21/30$, etc.). The resulting parameter estimates and goodness-of-fit measurements appear in Table 6.2. The goodness of model fit is measured by the mean absolute deviation between the log-fitted and log-observed data over the range of interest: $\frac{1}{19}\sum_{j=8+1}^{30-3} \left| \log \widehat{F}^{-1}(X_{j:30}) - \log X_{j:30} \right|$. It is clear that the parameter estimates and model fits are greatly dependent on a and b and hence the trimming proportions should be chosen carefully. Table 6.2 also provides the premium calculated using (6.1) for each fitted model. These premiums can be compared with empirical premium calculated using (6.1) with F replaced by the empirical distribution function.

Estimator		LOGNORMAL Model				WEIBULL Model			
a	b	$\widehat{\theta}$	$\widehat{\sigma}$	Fit	Premium	$\widehat{\alpha}$	$\widehat{\beta}$	Fit	Premium
8/30	3/30	2.075	0.766	0.041	5.3355	1.383	10.665	0.056	5.1981
3/30	21/30	2.240	1.167	0.268	7.5671	1.370	10.610	0.055	5.1706
3/30	3/30	2.028	0.872	0.062	5.4117	1.362	10.630	0.055	5.1936
8/30	0/30	2.063	0.876	0.054	5.6573	1.080	11.321	0.102	5.9492
14/30	14/30	2.037	1.675	0.383	7.3466	0.702	12.920	0.317	7.0363
MLE		2.077	0.834	0.048	5.6037	1.109	12.303	0.143	6.5100
EMPIRICAL Premium = 5.4161									

Table 6.2: Parameter estimates, goodness-of-fit measurements, and premiums

We observe that, for this calculation and regardless of the assumed underlying model, the MTM estimates with appropriate trimming proportions lead to premium estimates that are closer to the empirical estimate than those obtained with over- or under-trimming, or MLE. This of course does not justify the MTM approach, as one could use an estimation method that matches the model premium with the empirical premium. However, it exemplifies the idea that MTM estimation is an appropriate choice when one does not require a close fit in one or both tails of the distribution.

7 Summary and conclusions

In this paper we have introduced and developed two new general methods for estimating the parameters of claim severity distributions, which we call the method of trimmed moments, or MTM for short, and the combined MTM, or CMTM for short. The methods utilize the underlying principle of the classical method of moments and is therefore easy to understand. The proposed methods produce estimators that can achieve various easily controlled by the user degrees of robustness, and thus help to reach a desired balance between the robustness and efficiency of the estimators. A general asymptotic theory for the new estimators is provided, thus making them readily applicable for constructing confidence intervals, or sets, and for testing hypotheses. The performance of the MTM and CMTM estimators for various sample sizes has been investigated in the case of several parametric families of loss distributions. The application of MTM estimators have been illustrated on a real-life data set providing damage amounts caused by major hurricanes. The calculation of premiums for a layer of insurance coverage is a task for which MTM estimates are a natural choice.

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).

[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 tern 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

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

Numerical values of the constants $c_k \equiv c_k(F_0,\, a,\, b)$ are provided (up to 4 decimal digits) in Tables A.1–A.4 at the end of this appendix, for several choices of F_0 and various values of a and b.

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 deriving the MTM estimators for the Pareto parameter α, as well as their asymptotic properties. For convenience, we next present explicit expressions of 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 x_2, y_2, x_1, y_1:

- 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

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

Constants c_k

a	b = 0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
				F_0: Normal				
0	0	−0.1086	−0.1950	−0.2743	−0.4237	−0.7820	−1.1590	−1.5544
0.05	0.1086	0	−0.0851	−0.1625	−0.3066	−0.6428	−0.9782	−1.3002
0.10	0.1950	0.0851	0	−0.0769	−0.2189	−0.5447	−0.8610	−1.1532
0.15	0.2743	0.1625	0.0769	0	−0.1410	−0.4602	−0.7636	–
0.25	0.4237	0.3066	0.2189	0.1410	0	−0.3117	−0.5983	–
0.49	0.7820	0.6428	0.5447	0.4602	0.3117	0	–	–
0.70	1.1590	0.9782	0.8610	0.7636	0.5983	–	–	–
0.85	1.5544	1.3002	1.1532	–	–	–	–	–
				F_0: Student's t with 5 degrees of freedom				
0	0	−0.1521	−0.2558	−0.3476	−0.5175	−0.9301	−1.4007	−1.9699
0.05	0.1521	0	−0.1008	−0.1888	−0.3481	−0.7171	−1.1030	−1.5101
0.10	0.2558	0.1008	0	−0.0871	−0.2430	−0.5955	−0.9501	−1.3059
0.15	0.3476	0.1888	0.0871	0	−0.1544	−0.4968	−0.8316	–
0.25	0.5175	0.3481	0.2430	0.1544	0	−0.3315	−0.6414	–
0.49	0.9301	0.7171	0.5955	0.4968	0.3315	0	–	–
0.70	1.4007	1.1030	0.9501	0.8316	0.6414	–	–	–
0.85	1.9699	1.5101	1.3059	–	–	–	–	–
				F_0: Logistic				
0	0	−0.2089	−0.3612	−0.4973	−0.7498	−1.3587	−2.0362	−2.8180
0.05	0.2089	0	−0.1489	−0.2802	−0.5197	−1.0749	−1.6494	−2.2419
0.10	0.3612	0.1489	0	−0.1302	−0.3650	−0.8972	−1.4289	−1.9525
0.15	0.4973	0.2802	0.1302	0	−0.2327	−0.7507	−1.2544	–
0.25	0.7498	0.5197	0.3650	0.2327	0	−0.5024	−0.9706	–
0.49	1.3587	1.0749	0.8972	0.7507	0.5024	0	–	–
0.70	2.0362	1.6494	1.4289	1.2544	0.9706	–	–	–
0.85	2.8180	2.2419	1.9525	–	–	–	–	–
				F_0: Gumbel (extreme-value)				
0	−0.5772	−0.6791	−0.7700	−0.8585	−1.0367	−1.5218	−2.1217	−2.8579
0.05	−0.3980	−0.4956	−0.5810	−0.6632	−0.8263	−1.2543	−1.7494	−2.2954
0.10	−0.2773	−0.3735	−0.4567	−0.5360	−0.6921	−1.0938	−1.5441	−2.0201
0.15	−0.1747	−0.2706	−0.3524	−0.4300	−0.5814	−0.9652	−1.3855	–
0.25	0.0033	−0.0935	−0.1743	−0.2500	−0.3956	−0.7555	−1.1358	–
0.49	0.3757	0.2688	0.1849	0.1087	−0.0335	−0.3666	–	–
0.70	0.7103	0.5805	0.4865	0.4041	0.2556	–	–	–
0.85	1.0165	0.8450	0.7338	–	–	–	–	–

Table A.1: Constant $c_1 \equiv c_1(F_0, a, b)$.

a	b = 0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
				F_0: Normal				
0	1	0.8214	0.7501	0.7157	0.7142	0.9803	1.6077	2.6109
0.05	0.8214	0.6230	0.5358	0.4859	0.4515	0.6095	1.0507	1.7201
0.10	0.7501	0.5358	0.4377	0.3779	0.3242	0.4271	0.7871	1.3349
0.15	0.7157	0.4859	0.3779	0.3096	0.2400	0.3010	0.6045	–
0.25	0.7142	0.4515	0.3242	0.2400	0.1427	0.1372	0.3599	–
0.49	0.9803	0.6095	0.4271	0.3010	0.1372	0.0002	–	–
0.70	1.6077	1.0507	0.7871	0.6045	0.3599	–	–	–
0.85	2.6109	1.7201	1.3349	–	–	–	–	–
				F_0: Student's t with 5 degrees of freedom				
0	1.6646	1.2569	1.1621	1.1297	1.1638	1.6320	2.7081	4.6958
0.05	1.2569	0.8039	0.6769	0.6121	0.5747	0.7864	1.3675	2.3381
0.10	1.1621	0.6769	0.5341	0.4554	0.3910	0.5210	0.9688	1.7137
0.15	1.1297	0.6121	0.4554	0.3655	0.2808	0.3554	0.7205	–
0.25	1.1638	0.5747	0.3910	0.2808	0.1622	0.1560	0.4137	–
0.49	1.6320	0.7864	0.5210	0.3554	0.1560	0.0002	–	–
0.70	2.7081	1.3675	0.9688	0.7205	0.4137	–	–	–
0.85	4.6958	2.3381	1.7137	–	–	–	–	–
				F_0: Logistic				
0	3.2894	2.5792	2.3639	2.2777	2.3170	3.2249	5.3302	9.0227
0.05	2.5792	1.7900	1.5156	1.3710	1.2836	1.7511	3.0393	5.1416
0.10	2.3639	1.5156	1.2070	1.0322	0.8859	1.1776	2.1858	3.8300
0.15	2.2777	1.3710	1.0322	0.8323	0.6406	0.8092	1.6377	–
0.25	2.3170	1.2836	0.8859	0.6406	0.3721	0.3578	0.9473	–
0.49	3.2249	1.7511	1.1776	0.8092	0.3578	0.0005	–	–
0.70	5.3302	3.0393	2.1858	1.6377	0.9473	–	–	–
0.85	9.0227	5.1416	3.8300	–	–	–	–	–
				F_0: Gumbel (extreme-value)				
0	1.9779	1.9822	2.0412	2.1294	2.3818	3.4855	5.5883	9.2067
0.05	1.1939	1.1550	1.1688	1.2080	1.3467	2.0304	3.3313	5.3734
0.10	0.8908	0.8317	0.8262	0.8451	0.9387	1.4669	2.5015	4.0964
0.15	0.7022	0.6276	0.6081	0.6129	0.6756	1.1017	1.9699	–
0.25	0.4881	0.3876	0.3467	0.3304	0.3491	0.6378	1.2939	–
0.49	0.3979	0.2352	0.1517	0.0975	0.0438	0.1346	–	–
0.70	0.6497	0.4005	0.2706	0.1803	0.0670	–	–	–
0.85	1.1190	0.7309	0.5415	–	–	–	–	–

Table A.2: Constant $c_2 \equiv c_2(F_0,\, a,\, b)$.

a	b: 0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
				F_0: Normal				
0	0	−0.5107	−0.7101	−0.8431	−1.0400	−1.5642	−2.6363	−4.7777
0.05	0.5107	0	−0.1811	−0.2893	−0.4212	−0.6795	−1.2228	−2.3147
0.10	0.7101	0.1811	0	−0.1034	−0.2168	−0.3870	−0.7588	−1.5509
0.15	0.8431	0.2893	0.1034	0	−0.1056	−0.2253	−0.4948	–
0.25	1.0400	0.4212	0.2168	0.1056	0	−0.0683	−0.2176	–
0.49	1.5642	0.6795	0.3870	0.2253	0.0683	0	–	–
0.70	2.6363	1.2228	0.7588	0.4948	0.2176	–	–	–
0.85	4.7777	2.3147	1.5509	–	–	–	–	–
				F_0: Student's t with 5 degrees of freedom				
0	0	−1.9553	−2.3506	−2.6218	−3.0831	−4.5765	−7.7526	−14.8570
0.05	1.9553	0	−0.3036	−0.4638	−0.6498	−1.0359	−1.8731	−3.7103
0.10	2.3506	0.3036	0	−0.1507	−0.3028	−0.5328	−1.0512	−2.2601
0.15	2.6218	0.4638	0.1507	0	−0.1397	−0.2929	−0.6482	–
0.25	3.0831	0.6498	0.3028	0.1397	0	−0.0833	−0.2684	–
0.49	4.5765	1.0359	0.5328	0.2929	0.0833	0	–	–
0.70	7.7526	1.8731	1.0512	0.6482	0.2684	–	–	–
0.85	14.8570	3.7103	2.2601	–	–	–	–	–
				F_0: Logistic				
0	0	−4.0408	−5.1857	−5.9348	−7.1077	−10.5999	−17.9243	−33.6304
0.05	4.0408	0	−0.9746	−1.5072	−2.1315	−3.4069	−6.1540	−12.0577
0.10	5.1857	0.9746	0	−0.5032	−1.0210	−1.8019	−3.5505	−7.5476
0.15	5.9348	1.5072	0.5032	0	−0.4771	−1.0039	−2.2182	–
0.25	7.1077	2.1315	1.0210	0.4771	0	−0.2891	−0.9296	–
0.49	10.5999	3.4069	1.8019	1.0039	0.2891	0	–	–
0.70	17.9243	6.1540	3.5505	2.2182	0.9296	–	–	–
0.85	33.6304	12.0577	7.5476	–	–	–	–	–
				F_0: Gumbel (extreme-value)				
0	−5.4416	−5.8710	−6.2466	−6.6377	−7.5389	−11.0863	−18.5852	−34.2995
0.05	−1.6625	−1.9057	−2.0703	−2.2248	−2.5599	−3.8952	−6.8532	−12.8266
0.10	−0.7929	−0.9993	−1.1175	−1.2187	−1.4249	−2.2586	−4.2378	−8.3381
0.15	−0.3490	−0.5406	−0.6361	−0.7102	−0.8488	−1.4142	−2.8710	–
0.25	0.0801	−0.1083	−0.1851	−0.2340	−0.3051	−0.5861	−1.4783	–
0.49	0.4146	0.1644	0.0758	0.0305	−0.0044	−0.0496	–	–
0.70	0.7116	0.3106	0.1656	0.0868	0.0180	–	–	–
0.85	1.3365	0.6464	0.4020	–	–	–	–	–

Table A.3: Constant $c_3 \equiv c_3(F_0,\, a,\, b)$.

a \ b	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
				F_0: Normal				
0	3	1.9805	1.8392	1.8410	2.0118	2.9400	4.9884	9.5596
0.05	1.9805	0.8491	0.6329	0.5594	0.5594	0.8306	1.5169	3.1664
0.10	1.8392	0.6329	0.3897	0.2951	0.2544	0.3802	0.7651	1.8085
0.15	1.8410	0.5594	0.2951	0.1870	0.1249	0.1818	0.4172	–
0.25	2.0118	0.5594	0.2544	0.1249	0.0379	0.0365	0.1322	–
0.49	2.9400	0.8306	0.3802	0.1818	0.0365	0.0000	–	–
0.70	4.9884	1.5169	0.7651	0.4172	0.1322	–	–	–
0.85	9.5596	3.1664	1.8085	–	–	–	–	–
				F_0: Student's t with 5 degrees of freedom				
0	21.1772	11.8795	12.0361	12.5679	14.1347	20.7620	35.2833	69.9639
0.05	11.8795	1.5486	1.1067	0.9887	1.0133	1.5149	2.7729	6.0279
0.10	12.0361	1.1067	0.6096	0.4506	0.3942	0.5948	1.2010	2.9953
0.15	12.5679	0.9887	0.4506	0.2688	0.1775	0.2613	0.6029	–
0.25	14.1347	1.0133	0.3942	0.1775	0.0496	0.0477	0.1750	–
0.49	20.7620	1.5149	0.5948	0.2613	0.0477	0.0000	–	–
0.70	35.2833	2.7729	1.2010	0.6029	0.1750	–	–	–
0.85	69.9639	6.0279	2.9953	–	–	–	–	–
				F_0: Logistic				
0	45.3576	27.4235	26.5674	27.2513	30.3252	44.4682	75.5318	147.9604
0.05	27.4235	7.4969	5.4182	4.8230	4.9125	7.3339	13.4173	28.8875
0.10	26.5674	5.4182	3.0798	2.2890	1.9955	3.0047	6.0632	14.9416
0.15	27.2513	4.8230	2.2890	1.3852	0.9166	1.3468	3.1037	–
0.25	30.3252	4.9125	1.9955	0.9166	0.2606	0.2505	0.9172	–
0.49	44.4682	7.3339	3.0047	1.3468	0.2505	0.0000	–	–
0.70	75.5318	13.4173	6.0632	3.1037	0.9172	–	–	–
0.85	147.9604	28.8875	14.9416	–	–	–	–	–
				F_0: Gumbel (extreme-value)				
0	23.5114	24.5377	25.8527	27.3558	30.9947	45.5794	77.2711	150.2609
0.05	4.3079	4.3243	4.5277	4.7919	5.4673	8.3186	15.0495	31.2011
0.10	2.0269	1.9101	1.9753	2.0869	2.3982	3.8006	7.4704	17.0361
0.15	1.1440	0.9647	0.9712	1.0191	1.1784	1.9623	4.2819	–
0.25	0.5531	0.3060	0.2628	0.2597	0.2989	0.5726	1.6941	–
0.49	0.5222	0.1428	0.0544	0.0203	0.0036	0.0183	–	–
0.70	0.8854	0.2600	0.1081	0.0442	0.0049	–	–	–
0.85	1.7267	0.5839	0.3000	–	–	–	–	–

Table A.4: Constant $c_4 \equiv c_4(F_0,\, a,\, b)$.