May 2017

Robust and Computationally Efficient Methods for Fitting Loss Models and Pricing Insurance Risks

Qian Zhao
University of Wisconsin-Milwaukee

ROBUST AND COMPUTATIONALLY EFFICIENT METHODS FOR FITTING LOSS MODELS AND PRICING INSURANCE RISKS

by

Qian Zhao

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Mathematics

at

The University of Wisconsin-Milwaukee
May 2017

ABSTRACT

ROBUST AND COMPUTATIONALLY EFFICIENT METHODS FOR FITTING LOSS MODELS AND PRICING INSURANCE RISKS

by

Qian Zhao

The University of Wisconsin-Milwaukee, 2017
Under the Supervision of Professor Vytaras Brazauskas and Professor Jugal Ghorai

Robustness and efficiency of estimators are two critical factors for any parametric model. The former is concerned with the model risk and stability and the latter is used to assess the predictive power and model accuracy. It is well known that, under certain regularity conditions, the maximum likelihood estimator (MLE) is the most efficient method for estimating unknown parameters, but it generally lacks robustness and is sensitive to outliers. To balance the robustness and efficiency, Brazauskas et al. (2009) introduced a new general approach that is based on the L-statistics, Method of Trimmed Moments (MTM), and demonstrated its effectiveness in actuarial analysis. A primary objective of this dissertation is to improve the MTM approach by replacing data trimming with data Winsorizing, which we call the Method of Winsorized Moments (MWM), and to compare the performance of these two approaches in different actuarial models. For applications, we specifically consider log-location-scale families because they are frequently used for measuring and pricing insurance risks, modeling income inequality in economics, and in many other areas of application involving positive random variables. Specifically, we focus on Pareto, log-logistic and log-Laplace distributions, and present robust and computationally efficient methods for estimation of their parameters based on the MTM and MWM approaches, respectively. Large-sample properties of the new MWM estimators are established, and their small-sample performances are investigated through simulations and compared to those of the MTM and MLE. Also, the effect of model choice and parameter estimation method on risk pricing are illustrated using actual data that represent hurricane damages in the United States from 1925 to 1995. In particular, the estimated pure premiums for an insurance contract are computed when the log-logistic, log-Laplace, and lognormal models are fitted to the data using the MTM, MWM and MLE methods.

To

my grandfather,

and my parents

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ARE	Asymptotic Relative Efficiency
CI	Confidence Interval
MAD	Median Absolute Deviation
MLE	Maximum Likelihood Estimator
MTM	Method of Trimmed Moments
MWM	Method of Winsorized Moments
RE	Relative Efficiency

Acknowledgements

I would like to express my very great appreciation to Professor Vytaras Brazauskas and Professor Jugal Ghorai, my thesis advisors, for their patient guidances, enthusiastic encouragements and useful critiques of this research work. Your advice on both research as well as on my career have been priceless. I would also like to thank Professor Jay Beder, Professor Wei Wei and Professor Dashan Fan for serving as my committee members. You provide brilliant comments and suggestions and let my defense be an enjoyable moment.

I want to thank Professor Yang Ho, Dr. Xianwei Van Harpen, and all the students in the actuarial science program, especially Ranadeera Samanthi, Daoping Yu, Hua Huang, and Gilbert Nat-Narh for their career advise and for sharing their experience on both study and work.

A special thanks to my family. Words cannot express how grateful I am to my grandfather, my mother, and my father for all of the sacrifices that you have made on my behalf. Your support for me was what sustained me thus far. I would also like to thank all of my friends, Lijuan Bi, Sami Cheong, Wen Yang, and Hung C. Tran who supported me in writing, and encouraged me to strive towards my goal.

Qian Zhao

Milwaukee, Wisconsin

March 2017

1 Introduction

1.1 Motivation

Financial risk management relies on models for asset prices, exchange rates and other market variables. Insurance industry relies on models for loss variables, payments, reserves and so on. In practice, however, all models are inevitably dependent on simplifying assumptions, imperfect parameter estimates and other inputs, thus creating model risk. Model risk can be defined as the risk that a financial institution or other organization incurs losses because its models are misspecified or because some of the assumptions underlying these models are not met in practice (see McNeil et al., 2005). Take loss severity distribution, for instance, we might work with a lognormal distribution to model losses whereas the true underlying distribution is heavy-tailed.

1.2 Literature Review

In the financial literature, the concept of model risk has been studied by Cont (2006). In economics, Hansen and Sargent (2008) are the leading developers of robust macroeconomic models. In actuarial science, estimators' robustness has been studied by Brazauskas and Serfling (2000, 2003), Marceau and Rioux (2001), Serfling (2002), Kaiser and Brazauskas (2006), Dornheim and Brazauskas (2007), and others.

To reduce the model risk and improve the stability of parameter estimates, Huber (1964) and Hampel (1971) introduced the concept of estimator's robustness. Robust statistics, loosely speaking, is concerned with the fact that many assumptions commonly made in statistics are at most approximations to reality. Robust statistical methods have been developed for many common problems, such as estimating location, scale and regression parameters. Primary motivation is to produce statistical methods that are not unduly affected by outliers. Another motivation is to provide methods with good performance when there are small departures from parametric distributions (for details, see, Maronna et al., 2005).

In actuarial science, insurance risks usually follow highly skewed and heavy tailed distributions that are supposed to generate outliers (i.e., extremely large observations). However, not all large observations are outliers and not all outliers are large. Therefore, in this context robust statistical methods play a crucial role as they provide tools for quantifying influence of each data point on an estimator.

Huber (1964) was the first author to study robustness formally, although many ad-hoc robust procedures existed for centuries. Huber's idea was to robustify the likelihood function of the model, which had led to the introduction of a class of M-estimators. Further, Hampel (1968, 1971, 1974) introduced the infinitesimal approach to robustness based on influence functions, which also contained an important global robustness aspect, namely the breakdown point. All the main findings and the theory and techniques of robust statistics are summarized in Hampel et al. (1986), Maronna et al. (2006), and Huber et al. (2009).

For pricing insurance risks, a more natural and computationally efficient approach is based on L-statistics and is called the Method of Trimmed Moments (MTM). It was introduced by Brazauskas et al. (2009) who developed its large-and small-sample properties and demonstrated its effectiveness in actuarial analysis (see also Brazauskas, 2009). A primary objective of this dissertation is to improve the MTM approach by replacing data trimming with data Winsorizing, which we call the Method of Winsorized Moments (MWM), and compare the performance of these two approaches under different parametric models.

1.3 Thesis Organization

The rest of this dissertation is organized as follows. In Chapter 2, we present the MTM idea, along with the asymptotic properties of the MTM estimators. Examples of estimators for location-scale families and several loss models–exponential, logistic and Laplace are provided in details. In Chapter 3, we introduce and develop the MWM approach by replacing data trimming with data Winsorizing, and then derive the explicit formulas of parameter estimators and observe the trade-off between robustness and efficiency with the same loss models. In the simulation study of Chapter 4, we compare the finite-sample performance of MTM and MWM estimators with those of the MLEs and demonstrate how large the sample size should be for the estimators to achieve asymptotic unbiasedness. In Chapter 5, the effect of model choice and parameter estimation method on risk pricing is illustrated using actual data that represent hurricane damages in the United States from 1925 to 1995. In particular, the estimated pure premiums for an insurance contract are computed when the log-logistic, log-Laplace, and lognormal models are fitted to the data using the MTM, MWM and MLE methods. Finally, in Chapter 6, the conclusions are drawn and future research is discussed.

2 Method of Trimmed Moments

In this chapter we describe the MTM idea, along with the asymptotic properties of the obtained estimators, and conclude with some examples of MTM estimators for loss models. Throughout the chapter, we closely follow the MTM description presented by Brazauskas et al. (2009).

2.1 Definition

Let $X_1, \ldots, X_n$ be i.i.d. random variables with common parametric distribution F. Suppose there are k unknown parameters $\theta_1, \ldots, \theta_k$. 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 derived in three steps:

- Compute the sample trimmed moments

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

  where h_j are specially chosen functions 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.

- Derive the corresponding population trimmed moments

$$T_j = \frac{1}{1 - a_j - b_j} \int_{a_j}^{1-b_j} h_j(F^{-1}(u))du, \quad 1 \leq j \leq k, \tag{2.2}$$

  where $F^{-1}(u) = \inf\{x \in \mathbb{R} : u \leq F(x)\}$ is the quantile function. (Note that when $a_j = b_j = 0$, then $T_j = \mathbb{E}[h_j(X)]$.)

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

$$\begin{cases} T_1(\theta_1, \ldots, \theta_k) = \widehat{T}_1 \\ \quad \vdots \qquad\qquad \vdots \\ T_k(\theta_1, \ldots, \theta_k) = \widehat{T}_k \end{cases} \tag{2.3}$$

with respect to $\theta_1, \ldots, \theta_k$. The solutions, which we denote by $\widehat{\theta}_j = g_j(\widehat{T}_1, \ldots, \widehat{T}_k)$, $1 \leq j \leq k$, are, by definition, the MTM estimators of $\theta_1, \ldots, \theta_k$. Note that the functions g_j are such that $\theta_j = g_j(T_1, \ldots, T_k)$.

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 established by Brazauskas et al. (2007). It follows from the latter work that the k-variate vector $\big(\sqrt{n}(\widehat{T}_1 - T_1), \ldots, \sqrt{n}(\widehat{T}_k - T_k)\big)$ converges in distribution to the k-variate normal random vector with the mean $\mathbf{0} = (0, \ldots, 0)$ and the covariance-variance matrix $\mathbf{\Sigma} :=[\sigma^2_{ij}]_{i,j=1}^k$ with the entries

$$\sigma^2_{ij} = \frac{1}{(1-a_i-b_i)(1-a_j-b_j)} \int_{a_i}^{1-b_i} \int_{a_j}^{1-b_j} (\min\{u,v\} - uv)\, dh_j(F^{-1}(v)) dh_i(F^{-1}(u)). \tag{2.4}$$

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

$$(\widehat{T}_1, \ldots, \widehat{T}_k) \sim \mathcal{AN}\big((T_1, \ldots, T_k), n^{-1}\mathbf{\Sigma}\big). \tag{2.5}$$

By delta method (see, e.g., Serfling, 1980, Section 3.3), 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 $n^{-1}\mathbf{D}\mathbf{\Sigma}\mathbf{D}'$, where $\mathbf{D}:=[d_{ij}]_{i,j=1}^k$ is the Jacobian of the transformations $g_1, \ldots, g_k$ evaluated at $(T_1, \ldots, T_k)$, that is, $d_{ij} = \partial g_i / \partial \widehat{T}_j \big|_{(T_1, \ldots, T_k)}$. In summary, we have that

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

Statement (2.6) can be used for testing hypotheses, constructing confidence intervals or sets.

2.3 Examples

In this section, we analyze the robustness and efficiency for general (i.e., not necessarily symmetric) location-scale family, and take exponential, logistic and Laplace distributions as three examples to illustrate. Specifically, we present how to find MTM estimators and obtain the entries of the asymptotic covariance-variance matrix for one- and two-parameter distributions. Later, we evaluate the asymptotic relative efficiency (ARE)

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

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

2.3.1 Location-Scale Families

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

$$\text{Location-scale:} \quad F(x) = F_0\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < x < \infty, \tag{2.7}$$

where location $-\infty < \mu < \infty$ and scale $\sigma > 0$ are unknown parameters, and F_0 is the standard (i.e., with μ=0 and σ=1) parameter-free version of F. The corresponding quantile function is $F^{-1}(t) = \mu + \sigma F_0^{-1}(t)$. Since F has two unknown parameters, we employ two trimmed moments. Choosing $h_1(t) = t$ and $h_2(t) = t^2$, then following the procedure of Section 2.1, we have

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

$$\widehat{T}_2 = \frac{1}{n - m_n(2) - m_n^*(2)} \sum_{i=m_n(2)+1}^{n-m_n^*(2)} X_{i:n}^2$$

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

Note 2.1. For log-location-scale families, such as lognormal, log-logistic and Weibull distributions, we choose $h_1(t) = \log(t)$ and $h_2(t) = (\log(t))^2$, because the logarithmic transformation converts these loss models into a location-scale family (to be discussed in Chapter 6). $\square$

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

$$T_1 := T_1(\mu, \sigma) = \frac{1}{1-a-b} \int_a^{1-b} F^{-1}(u) du = \mu + \sigma c_1,$$

$$T_2 := T_2(\mu, \sigma) = \frac{1}{1-a-b} \int_a^{1-b} [F^{-1}(u)]^2 du = \mu^2 + 2\mu\sigma c_1 + \sigma^2 c_2.$$

where

$$c_k = c_k(F_0, a, b) = \frac{1}{1-a-b} \int_a^{1-b} [F_0^{-1}(u)]^k du.$$

Equating $\widehat{T_1}$ to T_1 and $\widehat{T_2}$ to T_2, and then solving the resulting system of equations with respect to μ and σ, we obtain the MTM estimators

$$\begin{cases} \widehat{\mu}_{\text{MTM}} = \widehat{T_1} - c_1 \widehat{\sigma}_{MTM} =: g_1(\widehat{T_1}, \widehat{T_2}); \\ \\ \widehat{\sigma}_{\text{MTM}} = \sqrt{(\widehat{T_2} - \widehat{T_1}^2)/(c_2 - c_1^2)} =: g_2(\widehat{T_1}, \widehat{T_2}). \end{cases} \tag{2.8}$$

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

$$\begin{aligned} \sigma_{11}^2 &= \frac{\sigma^2}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} (\min\{u,v\} - uv)\, dF_0^{-1}(v) dF_0^{-1}(u) \\ &= \sigma^2 c_1^*; \end{aligned}$$

$$\begin{aligned} \sigma_{12}^2 = \sigma_{21}^2 &= \frac{2\mu\sigma^2}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} (\min\{u,v\} - uv)\, dF_0^{-1}(v) dF_0^{-1}(u) \\ &+ \frac{2\sigma^3}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} (\min\{u,v\} - uv) F_0^{-1}(u)\, dF_0^{-1}(v) dF_0^{-1}(u) \\ &= 2\mu\sigma^2 c_1^* + 2\sigma^2 c_2^*; \end{aligned} \tag{2.9}$$

$$
\begin{aligned}
\sigma_{22}^2 &= \frac{4\mu^2\sigma^2}{(1-a-b)^2}\int_a^{1-b}\int_a^{1-b}(\min\{u,v\}-uv)\,dF_0^{-1}(v)dF_0^{-1}(u)\\
&+\frac{8\mu\sigma^3}{(1-a-b)^2}\int_a^{1-b}\int_a^{1-b}(\min\{u,v\}-uv)F_0^{-1}(u)\,dF_0^{-1}(v)dF_0^{-1}(u)\\
&+\frac{4\sigma^4}{(1-a-b)^2}\int_a^{1-b}\int_a^{1-b}(\min\{u,v\}-uv)F_0^{-1}(u)F_0^{-1}(v)\,dF_0^{-1}(v)dF_0^{-1}(u)\\
&= 4\mu^2\sigma^2c_1^* + 8\mu\sigma^3c_2^* + 4\sigma^4c_3^*
\end{aligned}
$$

with $c_k^* = c_k^*(F_0, a, b)$, which is defined above. The entries of the matrix $\mathbf{D}$ are found by differentiating the functions g_i (see equations (2.8)):

$$
\begin{aligned}
d_{11} &= \left.\frac{\partial g_1}{\partial \widehat{T}_1}\right|_{(T_1,T_2)} = 1 - c_1\left.\frac{\partial g_2}{\partial \widehat{T}_1}\right|_{(T_1,T_2)} = \frac{c_1\mu + c_2\sigma}{\sigma(c_2 - c_1^2)}\\
d_{12} &= \left.\frac{\partial g_1}{\partial \widehat{T}_2}\right|_{(T_1,T_2)} = -c_1\left.\frac{\partial g_2}{\partial \widehat{T}_2}\right|_{(T_1,T_2)} = \frac{-0.5c_1}{\sigma(c_2 - c_1^2)}\\
d_{21} &= \left.\frac{\partial g_2}{\partial \widehat{T}_1}\right|_{(T_1,T_2)} = \left.\frac{-\widehat{T}_1}{\sqrt{(c_2-c_1^2)(\widehat{T}_2 - \widehat{T_1^2})}}\right|_{(T_1,T_2)} = \frac{-\mu - c_1\sigma}{\sigma(c_2 - c_1^2)}\\
d_{22} &= \left.\frac{\partial g_2}{\partial \widehat{T}_2}\right|_{(T_1,T_2)} = \left.\frac{0.5}{\sqrt{(c_2-c_1^2)(\widehat{T}_2 - \widehat{T_1^2})}}\right|_{(T_1,T_2)} = \frac{0.5}{\sigma(c_2 - c_1^2)}
\end{aligned}
\tag{2.10}
$$

Consequently,

$$
\begin{aligned}
\mathbf{D\Sigma D'} &= \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22}\end{bmatrix}\begin{bmatrix} \sigma^2c_1^* & 2\mu\sigma^2c_1^* + 2\sigma^3c_2^* \\ 2\mu\sigma^2c_1^* + 2\sigma^3c_2^* & 4\mu^2\sigma^2c_1^* + 8\mu\sigma^3c_2^* + 4\sigma^4c_3^*\end{bmatrix}\begin{bmatrix} d_{11} & d_{21} \\ d_{12} & d_{22}\end{bmatrix}\\
&= \frac{\sigma^2}{(c_2-c_1^2)^2}\begin{bmatrix} c_1^*c_2^2 - 2c_1c_2c_2^* + c_1^2c_3^* & -c_1^*c_1c_2 + c_2c_2^* + c_1^2c_2^* - c_1c_3^* \\ -c_1^*c_1c_2 + c_2c_2^* + c_1^2c_2^* - c_1c_3^* & c_1^*c_1^2 - 2c_1c_2^* + c_3^*\end{bmatrix}.
\end{aligned}
\tag{2.11}
$$

We summarize the above findings by saying that

$$(\widehat{\mu}_{\mathrm{MTM}}, \widehat{\sigma}_{\mathrm{MTM}}) \sim \mathcal{AN}\left((\mu, \sigma), \frac{\sigma^2}{n}\mathbf{S}\right) \qquad \text{with} \quad \mathbf{S} = \sigma^{-2}\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'. \tag{2.12}$$

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.

2.3.2 Exponential and Pareto Models

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

$$\text{Exponential}(\lambda): \quad F(x) = 1 - e^{-\lambda x}, \quad x \geq 0, \tag{2.13}$$

where location $\lambda > 0$ is unknown parameter. Obviously, it is a member of the location-scale family. The corresponding quantile function is

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda}.$$

Since F has only one unknown parameter, we need only one trimmed moment. Choosing $h_1(t) = t$ and following the procedure of Section 2.3.1, we have

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

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

$$\begin{aligned}
T_1 := T_1(\lambda) &= \frac{1}{1-a-b}\int_a^{1-b} -\frac{\log(1-u)}{\lambda}\,du \\
&= \frac{-1/\lambda}{1-a-b}\int_a^{1-b} \log(1-u)\,du \\
&= \frac{-1/\lambda}{1-a-b} I(a, 1-b),
\end{aligned}$$

with the obvious notation for the function I. Equating $\widehat{T_1}$ with T_1 and solving the equation with respect to λ yields the MTM estimator

$$\widehat{\lambda}_{\text{MTM}} = \frac{-I(a, 1-b)}{1-a-b} \frac{1}{\widehat{T_1}} =: g_1(\widehat{T_1}). \tag{2.14}$$

The entries of matrix $\mathbf{\Sigma}$, which is one-dimensional, follow from equation (2.4):

$$\begin{aligned}
\sigma_{11}^2 &:= \frac{1/\lambda^2}{(1-a-b)^2} \int_a^{1-b} \int_a^{1-b} \frac{\min\{u, v\} - uv}{(1-u)(1-v)} \, du \, dv \\
&= \frac{1/\lambda^2}{(1-a-b)^2} J((a, 1-b), (a, 1-b)),
\end{aligned}$$

with the obvious notation for the function J. The Jacobian $\mathbf{D}$ is found by differentiating the function g_1 in equation (2.14) and then evaluating its derivative at T_1:

$$\left. \frac{\partial g_1}{\partial \widehat{T_1}} \right|_{T_1} = \left. \frac{I(a, 1-b)}{1-a-b} \frac{1}{\widehat{T_1^2}} \right|_{T_1} = \frac{1-a-b}{I(a, 1-b)} \lambda^2.$$

Hence,

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

Summarizing the above findings, we have

$$\widehat{\lambda}_{\text{MTM}} \sim \mathcal{AN}\left(\lambda, \frac{\lambda^2}{n} \mathbf{C}\right) \qquad \text{with} \quad \mathbf{C} = \frac{J((a, 1-b), (a, 1-b))}{[I(a, 1-b)]^2}.$$

Now we investigate how much efficiency we lose because of using MTM approach instead of the MLE when estimating λ. Since the $\widehat{\lambda}_{\text{MLE}} = n / \sum_{i=1}^n X_i$, which is $\mathcal{AN}\left(\lambda, \frac{\lambda^2}{n}\right)$, (see, e.g, Brazauskas and Serfling, 2000a), we have that $\text{ARE}(\widehat{\lambda}_{\text{MTM}}, \widehat{\lambda}_{\text{MLE}}) = 1/\mathbf{C}$.

Numerical values of these AREs are provided in Table 2.1 for several trimming proportions a and b. We find that when a is fixed, the ARE decreases as b increases except for the extreme case (i.e., $a = 0.85$). And when b is fixed, the MTM estimator with no lower trimming (i.e., $a = 0$) and with symmetric trimming (i.e., $a = b$) are almost equivalent. The efficiency decreases slowly when there is no upper trimming (i.e., $b = 0$).

Table 2.1: Exponential distribution–ARE($\widehat{\lambda}_{\text{MTM}}, \widehat{\lambda}_{\text{MLE}}$) for selected a and b, with the boxed numbers highlighting the case $a = b$.

a	b							
	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{1.000}$	0.917	0.847	0.783	0.666	0.423	0.238	0.115
0.05	1.000	$\boxed{0.918}$	0.848	0.783	0.667	0.425	0.241	0.122
0.10	1.000	0.918	$\boxed{0.848}$	0.785	0.669	0.430	0.250	0.135
0.15	0.999	0.918	0.849	$\boxed{0.787}$	0.672	0.437	0.260	-
0.25	0.995	0.918	0.851	0.790	$\boxed{0.679}$	0.452	0.284	-
0.49	0.958	0.897	0.839	0.786	0.688	$\boxed{0.487}$	-	-
0.70	0.857	0.824	0.781	0.738	0.659	-	-	-
0.85	0.681	0.688	0.663	-	-	-	-	-

Moreover, the MTM allows some very extreme scenarios yielding valid though inefficient estimators. For example, the MTM based on the middle 2% (i.e., $a = b = 0.49$) of actual data leads to about 49% efficiency when estimating the parameter λ. Interestingly, $\widehat{\lambda}_{\text{MTM}}$, 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 are located in the sample. Certainly, this note just confirms the fact that most information about the parameter λ is contained in the upper tail of Exponential(λ).

The single-parameter Pareto distribution

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

shares the common AREs with exponential distribution in Table 2.1 because the log-Pareto distribution is shifted exponential (see equation (2.13)) with a known location $x_0 > 0$. As mentioned in Section 2.3.1, we can switch the chosen function to $h_1(t) = \log(t/x_0)$ in this case and obtain the identical MTM formula from equation (2.14) for parameter α (to be discussed in Chapter 6).

2.3.3 Logistic and Log-Logistic Models

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

$$\text{Logistic}(\mu, \sigma): \quad F(x) = \frac{1}{1 + e^{-\frac{x-\mu}{\sigma}}}, \quad -\infty < x < \infty, \tag{2.15}$$

where location $-\infty < \mu < \infty$ and scale $\sigma > 0$ are unknown parameters. Obviously, it is a member of the location-scale family and the standard parameter-free version of (2.15) is

$$F_0(x) = \frac{1}{1 + e^{-x}}. \tag{2.16}$$

The corresponding quantile function is

$$F_0^{-1}(u) = -\log(1/u - 1),$$

and the derivative

$$\frac{d}{dt}\left[F_0^{-1}(u)\right] = 1/(u - u^2).$$

Applying all the steps of Section 2.3.1, the MTM estimators are obtained as the same form as equation (2.8) and following from statement (2.12),

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

where $\mathbf{D\Sigma D}'$ is given by equation (2.11) but now with the standard logistic F_0 in equation (2.16) instead of the therein used standardized location-scale distribution.

We next examine how much efficiency is lost because of using $(\widehat{\mu}_{\text{MTM}}, \widehat{\sigma}_{\text{MTM}})$ instead of the MLE. The MLE of the logistic distribution here is in implicit form and it is derived from the log-likelihood:

$$l = \log \prod_{i=1}^{n} f(x_i) = -\sum_{i=1}^{n} \frac{x_i - \mu}{\sigma} - n \log \sigma - 2 \sum_{i=1}^{n} \log\left(1 + e^{-\frac{x_i - \mu}{\sigma}}\right).$$

Take derivative with respect to μ and σ, respectively, and set each derivative to 0, then the MLE estimators are the solution(s) of the non-linear equations:

$$\begin{cases} p_1(\mu, \sigma) = \sum_{i=1}^n \frac{1}{e^{\Delta_i}+1} - \frac{n}{2} = 0 \\ \\ p_2(\mu, \sigma) = \sum_{i=1}^n \Delta_i - 2\sum_{i=1}^n \frac{\Delta_i}{e^{\Delta_i}+1} - n = 0 \end{cases} \tag{2.17}$$

where $\Delta_i = \frac{x_i - \mu}{\sigma}$.

The Newton method is used to solve this system. The $(i+1)$th Newton iteration is given by

$$\begin{bmatrix} \widehat{\mu}_{i+1} \\ \\ \widehat{\sigma}_{i+1} \end{bmatrix} = \begin{bmatrix} \widehat{\mu}_i \\ \\ \widehat{\sigma}_i \end{bmatrix} - J_i^{-1} \begin{bmatrix} p_1(\widehat{\mu}_i, \widehat{\sigma}_i) \\ \\ p_2(\widehat{\mu}_i, \widehat{\sigma}_i) \end{bmatrix}, \tag{2.18}$$

where J_i is the Jacobian matrix for the functions p_1 and p_2 and $i = 1, 2, \ldots$. Define the error of ith iteration by $\sqrt{(p_1(\widehat{\mu}_i, \widehat{\sigma}_i))^2 + (p_2(\widehat{\mu}_i, \widehat{\sigma}_i))^2}$ and the error tolerance can be set by the researcher.

When estimating $(\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}})$, we know (see e.g., deCani and Stine, 1986) that

$$(\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}) \sim \mathcal{AN}\Big((\mu, \sigma), \frac{\sigma^2}{n}\mathbf{S_0}\Big) \quad \text{with} \quad \mathbf{S_0} = \begin{bmatrix} 3 & 0 \\ \\ 0 & 9/(3+\pi^2) \end{bmatrix}. \tag{2.19}$$

Hence, it follows that $\text{ARE}((\widehat{\mu}_{\text{MTM}}, \widehat{\sigma}_{\text{MTM}}), (\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$ is equivalent to $(\det(\mathbf{S_0})/\det(\mathbf{S}))^{1/2}$.

Numerical values of the AREs are provided in Table 2.2 for selected values of a and b. Note from Table 2.2 that since logistic distribution is symmetric, we can expect similar performance of the MTM estimators with similar trimming schemes. For example, the AREs are identical for the MTM estimators with reversed trimming proportions: e.g., $(a, b) = (0.05, 0.25)$ has ARE $= 0.768$ and $(a, b) = (0.25, 0.05)$ also has ARE $= 0.768$.

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 μ and σ, we observe that inefficiency of σ estimators dominates the overall ARE: $a = b = 0.05$ has ARE $= 0.936$ (good);

Table 2.2: Log-logistic distribution–ARE$((\widehat{\mu}_{\text{MTM}}, \widehat{\sigma}_{\text{MTM}}), (\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$ for selected a and b, with the boxed numbers highlighting the case $a = b$.

a	b							
	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{0.893}$	0.884	0.834	0.782	0.681	0.449	0.258	0.127
0.05	0.884	$\boxed{0.936}$	0.903	0.861	0.768	0.529	0.311	0.146
0.10	0.834	0.903	$\boxed{0.874}$	0.835	0.745	0.507	0.283	0.106
0.15	0.782	0.861	0.835	$\boxed{0.797}$	0.709	0.473	0.245	-
0.25	0.681	0.768	0.745	0.709	$\boxed{0.625}$	0.391	0.138	-
0.49	0.449	0.529	0.507	0.473	0.391	$\boxed{0.095}$	-	-
0.70	0.258	0.311	0.283	0.245	0.138	-	-	-
0.85	0.127	0.146	0.106	-	-	-	-	-

$a = b = 0.25$ has ARE $= 0.625$ (moderate); $a = b = 0.49$ has ARE $= 0.095$ (very poor). In addition, instead of the case $a = b = 0$, the largest ARE appears at $a = b = 0.05$.

After the modification of estimating functions $h_1(t) = \log(t)$ and $h_2(t) = (\log(t))^2$, log-logistic distribution shares the common AREs with logistic distribution in Table 2.2, and we will use the log-logistic model for real-data illustrations of Chapter 5.

2.3.4 Laplace and Log-Laplace Models

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

$$\text{Laplace}\,(\mu, \sigma): \quad F(x) = \begin{cases} \frac{1}{2} e^{\frac{x-\mu}{\sigma}}, & \text{if } x < \mu, \\ 1 - \frac{1}{2} e^{-\frac{x-\mu}{\sigma}}, & \text{if } x \geq \mu, \end{cases} \qquad -\infty < x < \infty \tag{2.20}$$

$$= \frac{1}{2} + \frac{1}{2}\,\text{sgn}\,(x-\mu)(1 - e^{-\frac{|x-\mu|}{\sigma}}),$$

where location $-\infty < \mu < \infty$ and scale $\sigma > 0$ are unknown parameters and sgn is the sign function given by

$$\text{sgn}(x) = \begin{cases} -1, & \text{if } x < 0; \\ 0, & \text{if } x = 0; \\ 1, & \text{if } x > 0. \end{cases}$$

Laplace distribution is a location-scale family with the standard parameter-free version of (2.20) given by

$$F_0(x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{sgn}(x)(1 - e^{-|x|}). \tag{2.21}$$

The corresponding quantile function is

$$F_0^{-1}(u) = -\mathrm{sgn}\left(u - \frac{1}{2}\right)\log\left(1 - 2\left|u - \frac{1}{2}\right|\right),$$

and its derivative is

$$\frac{d}{du}\left[F_0^{-1}(u)\right] = -2\left[\delta\left(u - \frac{1}{2}\right)\log\left(1 - 2\left|u - \frac{1}{2}\right|\right) - \frac{\mathrm{sgn}^2(u - \frac{1}{2})}{1 - 2|u - \frac{1}{2}|}\right],$$

where δ represents the Dirac delta function

$$\delta(x) = \begin{cases} +\infty, & \text{if } x = 0; \\ 0, & \text{if } x \neq 0. \end{cases}$$

Following the procedure of Section 2.3.1, the MTM estimators have the same form as equation (2.8) and as follows from statement (2.12),

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

where $\mathbf{D\Sigma D'}$ depends only on F_0 in (2.21) and the chosen proportions a and b.

The MLE of Laplace distribution parameter has explicit form (see Norton, 1984)

$$\begin{cases} \widehat{\mu}_{\mathrm{MLE}} = \mathrm{Median}(x_1, \ldots, x_n) \\ \widehat{\sigma}_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \widehat{\mu}_{\mathrm{MLE}}| \end{cases} \tag{2.22}$$

And by the central limit theorem and the information matrix (see Kotz, et al., 2001)

$$(\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}) \sim \mathcal{AN}\bigg((\mu, \sigma), \frac{\sigma^2}{n}\mathbf{S_0}\bigg) \quad \text{with} \quad \mathbf{S_0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.23}$$

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

Table 2.3: Log-Laplace distribution–$\text{ARE}((\widehat{\mu}_{\text{MTM}}, \widehat{\sigma}_{\text{MTM}}), (\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$ for selected a and b, with the boxed numbers highlighting the case $a = b$.

a	b							
	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{0.633}$	0.644	0.621	0.595	0.539	0.363	0.194	0.092
0.05	0.644	$\boxed{0.723}$	0.718	0.706	0.671	0.490	0.256	0.112
0.10	0.621	0.718	$\boxed{0.714}$	0.702	0.669	0.495	0.243	0.083
0.15	0.595	0.706	0.702	$\boxed{0.689}$	0.653	0.486	0.218	-
0.25	0.539	0.671	0.669	0.653	$\boxed{0.607}$	0.435	0.110	-
0.49	0.363	0.490	0.495	0.486	0.435	$\boxed{0.128}$	-	-
0.70	0.194	0.256	0.243	0.218	0.110	-	-	-
0.85	0.092	0.112	0.083	-	-	-	-	-

Numerical values of the AREs are provided in Table 2.3 for chosen a and b. Since the Laplace distribution is symmetric as well, we can see the similar performance of the MTM estimators with similar trimming schemes. For example, the AREs are identical for the MTM estimators with reversed trimming proportions: $(a, b) = (0.1, 0.25)$ has ARE $= 0.669$ and $(a, b) = (0.25, 0.1)$ also has ARE $= 0.669$. Additionally, the largest ARE locates at $a = b = 0.05$ and the estimator that uses only 50% of the actual data observations ($a = b = 0.25$) yields a 60.7% efficiency.

Like other logarithmic transformations, log-Laplace distribution has the common ARE with Laplace distribution and it is used for illustrations in Chapter 5.

3 Method of Winsorized Moments

A primary objective of this dissertation is to develop a Method of Winsorized Moments (MWM) by following the MTM approach. Winsorization is the transformation of data by limiting extreme values in the data set to reduce the effect of possibly spurious outliers. It is named after the engineer-turned-biostatistician Charles P. Winsor (1895-1951). The effect is the same as clipping in signal processing. The distribution of many statistics can be heavily influenced by outliers. A typical strategy is to set all outliers to a specified percentile of the data; for example, a 90% symmetric Winsorization would set all data below the 5th percentile to the 5th percentile, and data above the 95th percentile to the 95th percentile. Winsorized estimators are usually more robust to outliers than their standard counterparts and they can achieve a similar (if not better) effect as the earlier discussed trimmed estimators.

The MTM estimator in Chapter 2 is a type of two-sided truncation, which focuses on the data between the trimming points (dark blue area in Figure 3.1). The MWM, on the other hand, is a censored form that takes into account the upper and lower outside values (light blue area) as well. Figure 3.1 shows the difference between the former and the latter by plotting a quantile function.

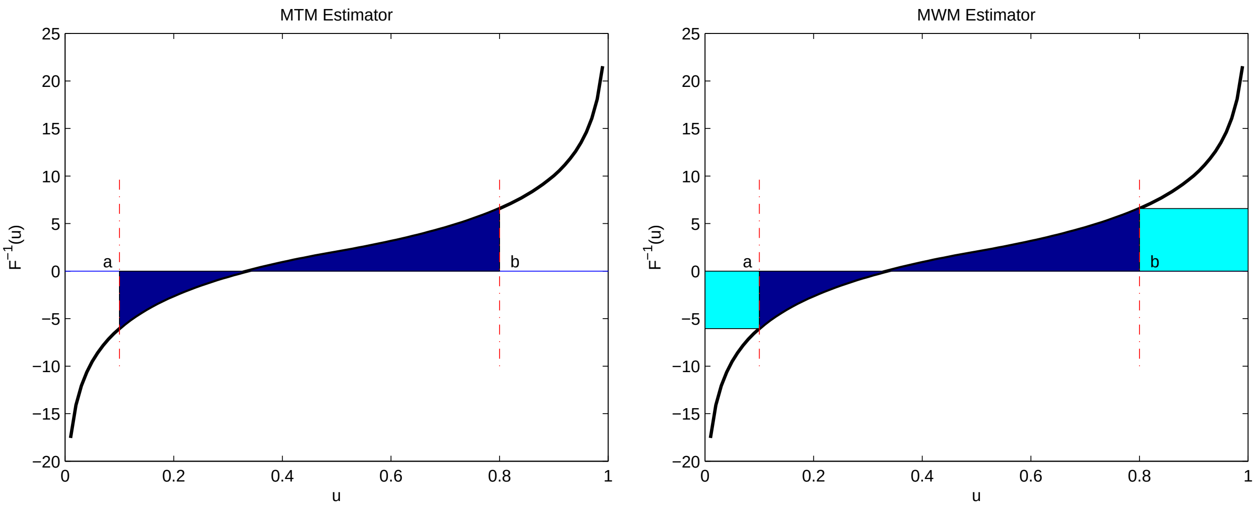


Figure 3.1: MTM (left panel) and MWM (right panel).

Like the MTM estimators, MWM estimators provide a flexible framework for achieving balance between robustness and efficiency via the AREs as well. For specific distributions, we can see the trade-offs between efficiency and robustness in various MWM estimators and then with the fixed trimming proportions, the difference of efficiency between the MWM estimators and MTM estimators will also be compared in this chapter.

3.1 Definition

Let $X_1, \ldots, X_n$ be i.i.d. random variables, which follow a parametric distribution F with k unknown parameters $\theta_1, \ldots, \theta_k$. Denote the order statistics of $X_1, \ldots, X_n$ by $X_{1:n} \leq \cdots \leq X_{n:n}$. Like the MTM estimators, the MWM estimators of $\theta_1, \ldots, \theta_k$ are also derived in three steps:

- Compute the sample Winsorized moments

$$\widehat{W}_j = \frac{1}{n}\Bigg[m_n(j)h_j(X_{m_n(j):n}) + \sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(X_{i:n}) + m_n^*(j)h_j(X_{n-m_n^*(j)-1})\Bigg], \quad 1 \leq j \leq k, \tag{3.1}$$

  where h_j are specially chosen functions 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.

- Derive the corresponding population Winsorized moments

$$W_j = W_j(\theta_1, ..., \theta_k) = a_j h_j(F^{-1}(a_j)) + \int_{a_j}^{1-b_j} h_j(F^{-1}(u))\, du + b_j h_j(F^{-1}(1-b_j)), \quad 1 \leq j \leq k, \tag{3.2}$$

  where $F^{-1}(u) = \inf\{x \in \mathbb{R} : u \leq F(x)\}$ is the quantile function. (Note that when $a_j = b_j = 0$, then $W_j = \mathbb{E}[h_j(X)]$.)

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

$$\begin{cases} W_1(\theta_1, \dots, \theta_k) = \widehat{W}_1 \\ \quad \vdots \qquad\qquad \vdots \\ W_k(\theta_1, \dots, \theta_k) = \widehat{W}_k \end{cases} \tag{3.3}$$

with respect to $\theta_1, \dots, \theta_k$. The solutions, which we denote by $\widehat{\theta}_j = g_j(\widehat{W}_1, \dots, \widehat{W}_k)$, $1 \leq j \leq k$, are, by definition, the MWM estimators of $\theta_1, \dots, \theta_k$. Note that the functions g_j are such that $\theta_j = g_j(W_1, \dots, W_k)$.

3.2 Asymptotic Properties

The sample Winsorized moments in equation (3.1) can be viewed as the combination of trimmed moments, defined by equation (2.1), and the trimming points censored on two sides, which is a special case of linear combination of ordered statistics (see Serfling, 1980, Chapter 8). Then equation (3.1) can be written as

$$\widehat{W}_n = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{i}{n+1}\right) h(X_{i:n}) + \sum_{m=1}^{2} c_m h(X_{[np_m],n})$$

where

$$K(x) = \begin{cases} 1, & \text{if } np_1 \leq x \leq np_2; \\ 0, & \text{otherwise;} \end{cases} \tag{3.4}$$

and np_m is the censored point.

Chernoff et al. (1967) prove that $\widehat{W}_n$ is $\mathcal{AN}\left(\mu, \frac{\sigma^2}{n}\right)$, where the mean

$$\mu = \int_0^1 K(u)H(u)du + \sum_{m=1}^{2} c_m h(p_m) \tag{3.5}$$

and the variance

$$\sigma^2 = \int_0^1 \alpha^2(u)du$$

where $H = h \circ F^{-1}$, p_m is the Winsorizing fraction a or $1-b$, and

$$\alpha(u) = \frac{1}{1-u}\int_u^1 K(w)H^{'}(w)(1-w)dw + \sum_{p_m \geq u} c_m(1-p_m)H^{'}(p_m).$$

Thus, the k-variate vector $\big(\sqrt{n}(\widehat{W}_1 - W_1), \ldots, \sqrt{n}(\widehat{W}_k - W_k)\big)$ in equation (3.3) converges in distribution to the k-variate normal random vector with the mean $\mathbf{0} = (0, \ldots, 0)$ and the covariance-variance matrix $\mathbf{\Sigma} :=[\sigma^2_{ij}]_{i,j=1}^k$ with the entries

$$\begin{aligned}
\sigma^2_{ij} &= \int_0^1 \alpha_i(u)\alpha_j(u)du \\
&= \int_0^1 \Bigg\{\frac{1}{1-u}\bigg[\int_u^1 K_i(w)H_i^{'}(w)(1-w)dw + \sum_{p_i \geq u} c_i(1-p_i)H_i^{'}(p_i)\bigg] \\
&\quad \times \frac{1}{1-u}\bigg[\int_u^1 K_j(v)H_j^{'}(v)(1-v)dv + \sum_{p_j \geq u} c_j(1-p_j)H_j^{'}(p_j)\bigg]\Bigg\}du \\
&= \int_0^1 \frac{1}{(1-u)^2}\bigg[\int_u^1 K_i(w)H_i^{'}(w)(1-w)dw \int_u^1 K_j(v)H_j^{'}(v)(1-v)dv\bigg]du \qquad (3.6) \\
&+ \int_0^1 \frac{1}{(1-u)^2}\bigg[\int_u^1 K_i(w)H_i^{'}(w)(1-w)dw \times \sum_{p_j \geq u} c_j(1-p_j)H_j^{'}(p_j)\bigg]du \\
&+ \int_0^1 \frac{1}{(1-u)^2}\bigg[\int_u^1 K_j(v)H_j^{'}(v)(1-v)dv \times \sum_{p_i \geq u} c_i(1-p_i)H_i^{'}(p_i)\bigg]du \\
&+ \int_0^1 \frac{1}{(1-u)^2}\bigg[\sum_{p_i \geq u} c_i(1-p_i)H_i^{'}(p_i) \times \sum_{p_j \geq u} c_j(1-p_j)H_j^{'}(p_j)\bigg]du
\end{aligned}$$

Following Chernoff et al. (1967), the asymptotic normality statement is

$$(\widehat{W}_1, \ldots, \widehat{W}_k) \sim \mathcal{AN}\big((W_1, \ldots, W_k), n^{-1}\mathbf{\Sigma}\big),$$

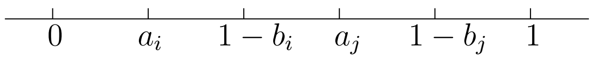
and the MWM estimator $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ is asymptotically normal with the mean $(\theta_1, \ldots, \theta_k)$ and the covariance-variance matrix $n^{-1}\mathbf{D\Sigma D'}$, where $\mathbf{D}:=[d_{ij}]_{i,j=1}^k$ is the Jacobian of the transformations $g_1, \ldots, g_k$ evaluated at $(W_1, \ldots, W_k)$.

3.2.1 Covariance-Variance Matrix: General Cases

From equations (3.4) and (3.5), we know that the entries of covariance-variance matrix in equation (3.6) actually depend on the proportions (a_i, b_i) and (a_j, b_j). Since there are six different combinations of Winsorizing fractions, we will analyze the specific entries of equation (3.6) in six separate cases.

Equation (3.6) consists of four terms and the first term is always equivalent to equation (2.4) (i.e., pure trimming), hence we need to figure out the second, the third and the fourth term for each case. In the following, we present a complete solution for the first combination of Winsorizing fractions. The remaining five combinations are provided in Appendix A.

Case 1: $0 \leq a_i \leq 1 - b_i < a_j \leq 1 - b_j \leq 1$



Let us introduce the following notation:

$$A_i = a_i(1 - a_i)H_i^{'}(a_i), \quad B_i = b_i^2 H_i^{'}(1 - b_i),$$

$$A_j = a_j(1 - a_j)H_j^{'}(a_j), \quad B_j = b_j^2 H_j^{'}(1 - b_j),$$

$$W(w) = H_i^{'}(w)(1 - w), \quad V(v) = H_j^{'}(v)(1 - v),$$

$$W_*(w) = \frac{W(w)}{(1 - u)^2}, \quad V_*(v) = \frac{V(v)}{(1 - u)^2}.$$

The second term is

$$= \int_0^1 \frac{1}{(1 - u)^2} \left[\int_u^1 K_i(w) H_i^{'}(w)(1 - w) dw \times \sum_{p_j \geq u} c_j (1 - p_j) H_j^{'}(p_j) \right] du$$

$$= \int_0^{a_i} \frac{A_j + B_j}{(1 - u)^2} \int_{a_i}^{1 - b_i} W(w)\, dw\, du + \int_{a_i}^{1 - b_i} \frac{A_j + B_j}{(1 - u)^2} \int_u^{1 - b_i} W(w)\, dw\, du$$

$$= \frac{a_i}{1 - a_i} \left[A_j + B_j \right] \int_{a_i}^{1 - b_i} W(w)\, dw + \left[A_j + B_j \right] \int_{a_i}^{1 - b_i} \int_u^{1 - b_i} W_*(w)\, dw\, du.$$

The third term is

$$= \int_0^1 \frac{1}{(1-u)^2} \left[\int_u^1 K_j(v) H_j^{'}(v)(1-v) dv \times \sum_{p_i \geq u} c_i (1-p_i) H_i^{'}(p_i) \right] du$$

$$= \int_0^{a_i} \frac{A_i + B_i}{(1-u)^2} \int_{a_j}^{1-b_j} V(v)\, dv\, du + \int_{a_i}^{1-b_i} \frac{B_i}{(1-u)^2} \int_{a_j}^{1-b_j} V(v)\, dv\, du$$

$$= \left[\frac{a_i}{1-a_i} A_i + \frac{1-b_i}{b_i} B_i \right] \int_{a_j}^{1-b_j} V(v)\, dv.$$

The fourth term is

$$= \int_0^1 \frac{1}{(1-u)^2} \left[\sum_{p_i \geq u} c_i (1-p_i) H_i^{'}(p_i) \times \sum_{p_j \geq u} c_j (1-p_j) H_j^{'}(p_j) \right] du$$

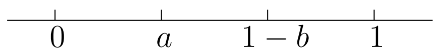
$$= \left[A_i + B_i\right]\left[A_j + B_j\right] \int_0^{a_i} \frac{1}{(1-u)^2}\, du + \left[B_i\right]\left[A_j + B_j\right] \int_{a_i}^{1-b_i} \frac{1}{(1-u)^2}\, du$$

$$= \frac{a_i}{1-a_i}\left[A_i + B_i\right]\left[A_j + B_j\right] + \frac{1-a_i-b_i}{(1-a_i)b_i}\left[B_i\right]\left[A_j + B_j\right].$$

For cases 2-6, the covariance-variance entries are derived in Appendix A.

3.2.2 Covariance-Variance Matrix: Special Case

Suppose the lower and upper Winsorizing proportions are identical, then we have the special case: $0 \leq a = a_i = a_j < 1-b_i = 1-b_j = 1-b \leq 1$



Now σ_{ij}^2 is the sum of the following four terms:

The first term reduces to:

$$= \int_a^{1-b} \int_a^{1-b} \{\min(w, v) - v\, w\}\, dh_j(F^{-1}(v))\, dh_i(F^{-1}(w)).$$

The second term is:

$$= \int_0^a \frac{1}{(1-u)^2} \int_a^{1-b} H_i^{'}(w)(1-w) dw \times \left[a(1-a)H_j^{'}(a) + b^2 H_j^{'}(1-b)\right] du$$

$$+\int_a^{1-b}\frac{1}{(1-u)^2}\int_u^{1-b}H_i^{'}(w)(1-w)dw\times\big[b^2H_j^{'}(1-b)\big]du$$

$$=\frac{a}{1-a}\big[a(1-a)H_j^{'}(a)+b^2H_j^{'}(1-b)\big]\int_a^{1-b}H_i^{'}(w)(1-w)\,dw$$

$$+\,b^2H_j^{'}(1-b)\int_a^{1-b}\int_u^{1-b}\frac{H_i^{'}(w)(1-w)}{(1-u)^2}\,dw\,du.$$

The third term is:

$$=\int_0^{a}\frac{1}{(1-u)^2}\int_a^{1-b}H_j^{'}(w)(1-w)dw\times\big[a(1-a)H_i^{'}(a)+b^2H_i^{'}(1-b)\big]du$$

$$+\int_a^{1-b}\frac{1}{(1-u)^2}\int_u^{1-b}H_j^{'}(w)(1-w)dw\times\big[b^2H_i^{'}(1-b)\big]du$$

$$=\frac{a}{1-a}\big[a(1-a)H_i^{'}(a)+b^2H_i^{'}(1-b)\big]\int_a^{1-b}H_j^{'}(v)(1-v)\,dv$$

$$+\,b^2H_i^{'}(1-b)\int_a^{1-b}\int_u^{1-b}\frac{H_j^{'}(v)(1-v)}{(1-u)^2}\,dv\,du.$$

The fourth term is:

$$=\int_0^{a}\frac{1}{(1-u)^2}\times\big[a(1-a)H_i^{'}(a)+b^2H_i^{'}(1-b)\big]\big[a(1-a)H_j^{'}(a)+b^2H_j^{'}(1-b)\big]du$$

$$+\int_a^{1-b}\frac{1}{(1-u)^2}\times\big[b^2H_i^{'}(1-b)\big]\big[b^2H_j^{'}(1-b)\big]du$$

$$=\frac{a}{1-a}\big[a(1-a)H_i^{'}(a)+b^2H_i^{'}(1-b)\big]\big[a(1-a)H_j^{'}(a)+b^2H_j^{'}(1-b)\big]$$

$$+\,\frac{1-a-b}{(1-a)b}\big[b^2H_i^{'}(1-b)\,b^2H_j^{'}(1-b)\big].$$

Similar to Chapter 2, we only consider these special Winsorizing proportions in the following examples.

3.3 Examples

In this section, we analyze the robustness and efficiency of MWM estimators for general location-scale family, and take exponential, logistic and Laplace distributions as three examples to illustrate. Specifically, we present how to find MWM estimators and derive the entries of the asymptotic covariance-variance matrix for one- and two-parameter distributions. Later, we evaluate the AREs of the MWM estimators with respect to the MLEs and compare the performance of MTM and MWM approaches in different cases.

3.3.1 Location-Scale Families

Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with the common distribution function defined by equation (2.7). Choosing $h_1(t) = t$ and $h_2(t) = t^2$, and following the procedure of Section 3.1, we have

$$\widehat{W}_1 = \frac{1}{n}\left[m_n(1) \cdot X_{m_n(1):n} + \sum_{i=m_n(1)+1}^{n-m_n^*(1)} X_{i:n} + m_n^*(1) \cdot X_{n-m_n^*(1)+1:n}\right],$$

$$\widehat{W}_2 = \frac{1}{n}\left[m_n(2) \cdot X^2_{m_n(2):n} + \sum_{i=m_n(2)+1}^{n-m_n^*(2)} X^2_{i:n} + m_n^*(2) \cdot X^2_{n-m_n^*(2)+1:n}\right]$$

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

As our next step in deriving MWM estimators, we calculate the population Winsorized moments using equation (3.2) and obtain

$$\begin{aligned}
W_1 := W_1(\mu, \sigma) &= \int_a^{1-b} F^{-1}(u)\, du + a\, F^{-1}(a) + b\, F^{-1}(1-b) \\
&= \int_a^{1-b} \left[\mu + \sigma\, F_0^{-1}(u)\right] du + a\left[\mu + \sigma\, F_0^{-1}(a)\right] + b\left[\mu + \sigma\, F_0^{-1}(1-b)\right] \\
&= \mu + \sigma\left[\int_a^{1-b} F_0^{-1}(u)\, du + a\, F_0^{-1}(a) + b\, F_0^{-1}(1-b)\right] \\
&= \mu + \sigma\, \tilde{c}_1,
\end{aligned}$$

$$
\begin{aligned}
W_2 := W_2(\mu,\sigma) &= \int_a^{1-b} [F^{-1}(u)]^2\,du + a\left[F^{-1}(a)\right]^2 + b\left[F^{-1}(1-b)\right]^2 \\
&= \int_a^{1-b} \left[\mu + \sigma\,F_0^{-1}(u)\right]^2 du + a\left[\mu + \sigma\,F_0^{-1}(a)\right]^2 + b\left[\mu + \sigma\,F_0^{-1}(1-b)\right]^2 \\
&= \mu^2 + 2\mu\sigma\left\{\int_a^{1-b} F_0^{-1}(u)\,du + a\,F_0^{-1}(a) + b\,F_0^{-1}(1-b)\right\} \\
&\quad + \sigma^2\left\{\int_a^{1-b} \left[F_0^{-1}(u)\right]^2 du + a\left[F_0^{-1}(a)\right]^2 + b\left[F_0^{-1}(1-b)\right]^2\right\} \\
&= \mu^2 + 2\mu\sigma\,\tilde{c}_1 + \sigma^2\,\tilde{c}_2,
\end{aligned}
$$

where

$$
\tilde{c}_k = \tilde{c}_k(F_0,a,b) = \int_a^{1-b} \left[F_0^{-1}(u)\right]^k du + a\left[F_0^{-1}(a)\right]^k + b\left[F_0^{-1}(1-b)\right]^k.
$$

Equating $\widehat{W_1}$ to W_1 and $\widehat{W_2}$ to W_2, and then solving the resulting system of equations with respect to μ and σ, we obtain the MWM estimators

$$
\begin{cases}
\widehat{\mu}_{\text{MWM}} = \widehat{W_1} - \tilde{c}_1\widehat{\sigma}_{\text{MWM}} =: g_1(\widehat{W_1},\widehat{W_2}); \\
\widehat{\sigma}_{\text{MWM}} = \sqrt{(\widehat{W_2} - \widehat{W_1}^{\,2})/(\tilde{c}_2 - \tilde{c}_1^2)} =: g_2(\widehat{W_1},\widehat{W_2}).
\end{cases}
\tag{3.7}
$$

The entries of the covariance-variance matrix $\mathbf{\Sigma}$ are calculated using the formulas of Section 3.2.2. We can obtain (see details in Appendix B),

$$
\begin{aligned}
\sigma_{11}^2 &= \sigma^2\,l_1^*; \\
\sigma_{12}^2 &= \sigma_{21}^2 = 2\mu\sigma^2\,l_1^* + 2\sigma^3\,l_2^*; \\
\sigma_{22}^2 &= 4\mu^2\sigma^2\,l_1^* + 8\mu\sigma^3\,l_2^* + 4\sigma^4\,l_3^*.
\end{aligned}
$$

Here, the entries are in the same format as (2.9), with constants c_k^* replaced by l_k^*. For calculating the matrix $\mathbf{D}$, we differentiate the functions g_i in (3.7), and the results have again the same structure as those in equation (2.10) except constants c_k are now replaced by $\tilde{c}_k$. Consequently,

$$\mathbf{D\Sigma D}' = \frac{\sigma^2}{(\tilde{c}_2 - \tilde{c}_1^2)^2} \begin{bmatrix} l_1^*\tilde{c}_2^2 - 2\tilde{c}_1\tilde{c}_2 l_2^* + \tilde{c}_1^2 l_3^* & -l_1^*\tilde{c}_1\tilde{c}_2 + \tilde{c}_2 l_2^* + \tilde{c}_1^2 l_2^* - \tilde{c}_1 l_3^* \\ -l_1^*\tilde{c}_1\tilde{c}_2 + \tilde{c}_2 l_2^* + \tilde{c}_1^2 l_2^* - \tilde{c}_1 l_3^* & l_1^* c_1^2 - 2\tilde{c}_1 l_2^* + l_3^* \end{bmatrix}. \tag{3.8}$$

We summarize the above findings by saying that

$$(\widehat{\mu}_{\text{MWM}}, \widehat{\sigma}_{\text{MWM}}) \sim \mathcal{AN}\left((\mu, \sigma), \frac{\sigma^2}{n}\tilde{\mathbf{S}}\right) \qquad \text{with} \quad \tilde{\mathbf{S}} = \sigma^{-2}\mathbf{D\Sigma D}'. \tag{3.9}$$

Note that the matrix $\tilde{\mathbf{S}}$ 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.

3.3.2 Exponential and Pareto Models

Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with the common exponential distribution function defined by equation (2.13). Since the distribution function F has only one unknown parameter, we need only one Winsorized moment. Following the procedure of Section 3.3.1 and choosing $h(t) = t$, we have

$$\widehat{W_1} = \frac{1}{n}\left[m_n(1) \cdot X_{m_n(1):n} + \sum_{i=m_n(1)+1}^{n-m_n^*(1)} X_{i:n} + m_n^*(1) \cdot X_{m_n^*(1):n}\right]$$

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

$$\begin{aligned} W_1 : = W_1(\lambda) &= -\frac{1}{\lambda}\left[\int_a^{1-b} \log(1-u)\,du + a\log(1-a) + b\log(b)\right] \\ &= \frac{-1}{\lambda}\tilde{I}(a, 1-b) \end{aligned}$$

with the obvious notation for the function $\tilde{I}$. Equating $\widehat{W_1}$ with W_1 and solving the equation with respect to λ yields the MWM estimator

$$\widehat{\lambda}_{\mathrm{MWM}} = -\tilde{I}(a, 1-b)\frac{1}{\widehat{W_1}} =: g_1(\widehat{W_1}). \tag{3.10}$$

The one-dimensional matrix $\mathbf{\Sigma}$ is derived from equations of Section 3.2.2. Here, the chosen functions are:

$$H(u) = h \circ F^{-1}(u) = F^{-1}(u) = -\frac{\log(1-u)}{\lambda}, \quad H^{'}(u) = \frac{1}{\lambda\,(1-u)}.$$

Hence,

$$\begin{aligned}
\sigma_{11}^2 :=& \frac{1}{\lambda^2}\int_a^{1-b}\int_a^{1-b}\frac{\min\{u,w\}-uw}{(1-u)(1-w)}\,dw\,du + 2\Bigg\{\frac{a^2(1-a)\cdot 1/[\lambda(1-a)] + ab^2\cdot 1/[\lambda b]}{1-a} \\
&\int_a^{1-b}\frac{1}{\lambda(1-w)}(1-w)\,dw + b^2\frac{1}{\lambda b}\int_a^{1-b}\int_a^{1-b}\frac{1/[\lambda(1-w)]\cdot(1-w)}{(1-u)^2}\,dw\,du\Bigg\} \\
&+ \frac{a}{1-a}\left[a(1-a)\cdot\frac{1}{\lambda(1-a)} + b^2\cdot\frac{1}{\lambda b}\right]^2 + \frac{1-a-b}{(1-a)b}\left[b^4\cdot\frac{1}{\lambda b}\right]^2 \\
=& \frac{1}{\lambda^2}\Bigg[\int_a^{1-b}\int_a^{1-b}\frac{\min\{u,w\}-uw}{(1-u)(1-w)}\,dw\,du + \frac{2(1-a-b)}{1-a}(a^2+ab-b) - 2b\log\left(\frac{b}{1-a}\right) \\
&+ \frac{a^3}{1-a} + 2a^2b^2\frac{1}{(1-a)b} + b(1-b)\Bigg] \\
=& \frac{1}{\lambda^2}\tilde{J}((a,1-b),(a,1-b))
\end{aligned}$$

with the obvious notation for the function $\tilde{J}$. The Jacobian $\mathbf{D}$ is found by differentiating the function g_1 in equation (3.10) and then evaluating its derivative at W_1:

$$\left.\frac{\partial g_1}{\partial \widehat{W_1}}\right|_{W_1} = \frac{1}{\tilde{I}(a,1-b)}\lambda^2.$$

Hence,

$$\mathbf{D\Sigma D'} = \frac{\tilde{J}((a,1-b),(a,1-b))}{[\tilde{I}(a,1-b)]^2}\lambda^2.$$

Summarizing the above findings, we have

$$\widehat{\lambda}_{\text{MWM}} \sim \mathcal{AN}\left(\lambda, \frac{\lambda^2}{n}\tilde{\mathbf{C}}\right) \qquad \text{with} \quad \tilde{\mathbf{C}} = \frac{\tilde{J}((a,1-b),(a,1-b))}{[\tilde{I}(a,1-b)]^2}.$$

As mentioned in Chapter 2, we use the ARE to measure the efficiency loss when estimating λ via MWM instead of MLE, which is $\text{ARE}(\widehat{\lambda}_{\text{MWM}}, \widehat{\lambda}_{\text{MLE}}) = 1/\tilde{\mathbf{C}}$.

Table 3.1: Exponential distribution–$\text{ARE}(\widehat{\lambda}_{\text{MWM}}, \widehat{\lambda}_{\text{MLE}})$ for selected a and b, with the boxed numbers highlighting the case $a = b$.

a	b							
	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{1.000}$	0.950	0.900	0.850	0.750	0.510	0.300	0.150
0.05	1.000	$\boxed{0.950}$	0.900	0.850	0.750	0.510	0.300	0.150
0.10	1.000	0.950	$\boxed{0.900}$	0.850	0.750	0.510	0.300	0.150
0.15	0.999	0.949	0.899	$\boxed{0.849}$	0.749	0.509	0.299	-
0.25	0.994	0.944	0.894	0.844	$\boxed{0.745}$	0.506	0.297	-
0.49	0.952	0.904	0.856	0.808	0.714	$\boxed{0.490}$	-	-
0.70	0.859	0.818	0.778	0.738	0.660	-	-	-
0.85	0.720	0.692	0.663	-	-	-	-	-

Numerical values of these AREs are provided in Table 3.1 for several Winsorizing proportions a and b. Generally, the MWM estimator keeps the characteristics similar to those of the MTM approach. For instance, when b is fixed, the MWM estimator with no lower Winsorizing (i.e., $a = 0$) and with symmetric Winsorizing (i.e., $a = b$) are almost equivalent. And the accuracy of estimators depends on both the fractions a, b and the location of the data in the sample. Undoubtedly, MWM estimators also possess their own specificity. The small lower Winsorizing proportions (i.e., $a = 0.05, 0.10$) do not decrease the efficiency at all. At last, comparing with MTM, MWM performs better (at least not worse) in most cases, with few exceptions when a is large.

After the modifying the estimating function $h(t) = t$ to $h(t) = \log(t)$, we find that single parameter Pareto distribution shares the common AREs with those of the exponential distribution in Table 3.1.

3.3.3 Logistic and Log-Logistic Models

Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with the common logistic distribution function defined by equation (2.15). Applying all the steps of Section 3.3.1, the MWM estimators are obtained in the same form as in (3.7), and it follows from statement (3.9) that

$$(\widehat{\mu}_{\mathrm{MWM}}, \widehat{\sigma}_{\mathrm{MWM}}) \sim \mathcal{AN}\left((\mu, \sigma), \frac{\sigma^2}{n}\tilde{\mathbf{S}}\right) \qquad \text{with} \quad \tilde{\mathbf{S}} = \sigma^{-2}\mathbf{D\Sigma D}',$$

where $\mathbf{D\Sigma D}'$ is given by (3.8), but now with the standard logistic F_0, given by (2.16), instead of the therein used standardized location-scale distribution. To summarize, it follows that

$$\mathrm{ARE}((\widehat{\mu}_{\mathrm{MWM}}, \widehat{\sigma}_{\mathrm{MWM}}), (\widehat{\mu}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}})) = (\det(\mathbf{S_0})/\det(\tilde{\mathbf{S}}))^{1/2},$$

where $\mathbf{S_0}$ is from equation (2.19).

Numerical values of the AREs are provided in Table 3.2 for selected values of a and b. Like in the MTM case, we observe similar performance of the MWM estimators when similar Winsorizing schemes are used. For example, the AREs are identical with reversed Winsorizing proportions: $(a, b) = (0.05, 0.25)$ has ARE $= 0.801$ and $(a, b) = (0.25, 0.05)$ also has ARE $= 0.801$.

Table 3.2: Log-logistic distribution–$\mathrm{ARE}((\widehat{\mu}_{\mathrm{MWM}}, \widehat{\sigma}_{\mathrm{MWM}}), (\widehat{\mu}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}}))$ for selected a and b, with the boxed numbers highlighting the case $a = b$.

a	b							
	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{0.893}$	0.896	0.873	0.843	0.774	0.571	0.358	0.187
0.05	0.896	$\boxed{0.913}$	0.895	0.868	0.801	0.589	0.359	0.169
0.10	0.873	0.895	$\boxed{0.878}$	0.852	0.783	0.564	0.323	0.118
0.15	0.843	0.868	0.852	$\boxed{0.825}$	0.754	0.528	0.277	-
0.25	0.774	0.801	0.783	0.754	$\boxed{0.680}$	0.439	0.153	-
0.49	0.571	0.589	0.564	0.528	0.439	$\boxed{0.104}$	-	-
0.70	0.358	0.359	0.323	0.277	0.153	-	-	-
0.85	0.187	0.169	0.118	-	-	-	-	-

The Winsorizing 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 μ and σ, we observe that inefficiency of σ estimators dominates the overall ARE: $a = b = 0.05$ has ARE $= 0.913$ (good); $a = b = 0.25$ has ARE $= 0.680$ (moderate); $a = b = 0.49$ has ARE $= 0.104$ (very poor). In addition, instead of the case $a = b = 0$, the largest ARE appears at $a = b = 0.05$. All the findings above are similar to those of the MTM approach in Section 2.3.3.

Considering the performance of MWM and MTM approaches, MTM estimators only achieve higher efficiency at the location around the largest ARE. As seen in Tables 2.2 and 3.2, when $a = b = 0.05$, MTM has ARE $= 0.936$ and MWM has ARE $= 0.913$. And when $(a, b) = (0.05, 0.10)$ or $(0.10, 0.05)$, MTM has ARE $= 0.903$, which is slightly larger than that of MWM with 0.895. Beyond these Winsorizing fractions, MWM estimators dominate in any other case.

After modifying the estimating functions to $h_1(t) = \log(t)$ and $h_2(t) = (\log(t))^2$, we see that the log-logistic distribution shares the common AREs with those of the logistic distribution in Table 3.2. We will use the log-logistic model for real-data illustrations of Chapter 5.

3.3.4 Laplace and Log-Laplace Models

Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with the same Laplace distribution function defined by equation (2.20). Following the procedure of Section 3.3.1, we find that the MWM estimators have the same form as in (3.7). It follows from statement (3.9) that

$$(\widehat{\mu}_{\mathrm{MWM}}, \widehat{\sigma}_{\mathrm{MWM}}) \sim \mathcal{AN}\left((\mu, \sigma), \frac{\sigma^2}{n}\tilde{\mathbf{S}}\right) \qquad \text{with} \quad \tilde{\mathbf{S}} = \sigma^{-2}\mathbf{D\Sigma D}',$$

where $\mathbf{D\Sigma D}'$ depends only on F_0 in (2.21) and the chosen proportions a and b. It follows that $\mathrm{ARE}((\widehat{\mu}_{\mathrm{MWM}}, \widehat{\sigma}_{\mathrm{MWM}}), (\widehat{\mu}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}}))$, which is $(\det(\mathbf{S_0})/\det(\tilde{\mathbf{S}}))^{1/2}$, given by equation (2.23), and is equal to $(1/\det(\tilde{\mathbf{S}}))^{1/2}$.

Numerical values of the AREs are provided in Table 3.3 for chosen fractions a and b. We see similar performances of the MWM estimators when similar Winsorizing schemes are used, and the largest ARE locates at $a = b = 0.05$, which are identical to that of the MTM estimators in Section 2.3.4.

Table 3.3: Log-Laplace distribution–ARE$((\widehat{\mu}_{\text{MWM}}, \widehat{\sigma}_{\text{MWM}}), (\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$ for selected a and b, with the boxed numbers highlighting the case $a = b$.

a	b							
	0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
0	$\boxed{0.633}$	0.638	0.625	0.612	0.588	0.549	0.304	0.143
0.05	0.638	$\boxed{0.663}$	0.657	0.648	0.631	0.605	0.314	0.131
0.10	0.625	0.657	$\boxed{0.653}$	0.644	0.624	0.594	0.287	0.094
0.15	0.612	0.648	0.644	$\boxed{0.633}$	0.610	0.569	0.251	-
0.25	0.588	0.631	0.624	0.610	$\boxed{0.575}$	0.494	0.146	-
0.49	0.549	0.605	0.594	0.569	0.494	$\boxed{0.140}$	-	-
0.70	0.304	0.314	0.287	0.251	0.146	-	-	-
0.85	0.143	0.131	0.094	-	-	-	-	-

Interestingly, for extreme scenarios (i.e., $a \,\text{or}\, b = 0.49, 0.70, 0.85$), MWM estimators always perform better than MTM. However, the latter are more appropriate around the symmetric case (i.e., $a = b$). For example, when $a = b = 0.05$, MTM has ARE $= 0.723$ and MWM has ARE $= 0.663$. And when $(a, b) = (0.05, 0.15)$, MTM has ARE $= 0.706$ and MWM has ARE $= 0.648$.

Based on the MWM approach, log-Laplace distribution also has the common ARE with the Laplace distribution. The log-Laplace distribution will be used for illustrations in Chapter 5.

4 Simulation Study

Next we supplement our theoretical results concerning the MTM and MWM estimators with their finite-sample performance evaluations, respectively. 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 with the corresponding mean-squared errors, that means

$$\mathrm{RE}(\mathrm{Q},\, \mathrm{MLE}) = \frac{\text{asymptotic variance of MLE estimator}}{\text{small-sample variance of a competing estimator Q}},$$

where Q represents MTM or MWM. The denominator is given by

$$\begin{bmatrix} \mathbb{E}[(\widehat{\mu}-\mu)^2] & \mathbb{E}[(\widehat{\mu}-\mu)(\widehat{\sigma}-\sigma)] \\ \\ \mathbb{E}[(\widehat{\mu}-\mu)(\widehat{\sigma}-\sigma)] & \mathbb{E}[(\widehat{\sigma}-\sigma)^2] \end{bmatrix}.$$

From a specified distribution F (i.e., logistic or Laplace), 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 and MWM 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 REs are again averaged and their standard deviations are reported. (Such repetitions are useful for assessing standard errors of the estimated means and REs. 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.

4.1 Logistic Model

We start our simulation study with the logistic distribution LG($\mu = 5, \sigma = 2$) using the following parameters:

- *Sample size*: $n = 50, 100, 250, 500$.
- *Estimators of* μ, σ:
  - MLE
  - Q (MTM or MWM) 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 $b = 0.00$.

Let the error tolerance be 0.00001. The MLE estimator of logistic distribution can be obtained from equations (2.17) and (2.18) in Section 2.3.3. To guarantee the convergence of Newton iterations as n increases, we choose $(\widehat{\mu}_{\text{start}}, \widehat{\sigma}_{\text{start}}) = (4.9, 2.2)$ as initial values, which is sufficiently close to the root.

First, we summarize the simulation results of the MTM and MWM approaches in Table 4.1. We see that all MTM and MWM estimators in the logistic case successfully estimate the location μ. In most cases, MTM and MWM estimators become practically unbiased for sample sizes as small as $n = 50$. Estimation of σ, however, reveals a different story. For MTM, 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 +72% for $n = 50$, −13% for $n = 100$, +24% for $n = 250$, +3% for $n = 500$. For MWM, there is no relative bias for $n = 50$, 250, but it is equal to −49% for $n = 100$, and −10% for $n = 500$ in the median-based estimator. And in the other extreme scenario $a = 0.1, b = 0.70$, it still contains bias even when $n = 500$.

In the relative efficiency (RE) Table 4.2, the entries of the last column are included for comparison with ARE found in Chapter 2. We notice that for both MTM and MWM approaches, the RE remains practically unaffected by the sample size and attains its corresponding ARE level when $n \geq 100$.

Table 4.1: Logistic model–mean values of $\widehat{\mu}/\mu$ and $\widehat{\sigma}/\sigma$.

a	b	n=50				n=100			
		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$	
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM
0	0	1.00		0.98		1.00		0.99	
0.05	0.05	1.00	1.00	1.05	1.01	1.00	1.00	1.01	0.99
0.10	0.10	1.00	1.00	1.01	0.98	1.00	1.00	1.01	0.99
0.25	0.25	1.00	1.00	1.06	1.01	1.00	1.00	1.01	0.99
0.49	0.49	1.00	1.00	1.72	1.00	1.00	1.00	0.87	0.51
0.10	0.70	1.02	0.95	1.03	0.91	1.01	0.98	1.02	0.96
0.25	0.00	0.99	1.00	0.99	0.99	1.00	1.00	0.99	0.99
MLE		1.00		0.99		1.00		0.99	
a	b	n=250				n=500			
		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$	
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM
0	0	1.00		1.00		1.00		1.00	
0.05	0.05	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.25	0.25	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00
0.49	0.49	1.00	1.00	1.24	1.00	1.00	1.00	1.03	0.90
0.10	0.70	1.00	0.99	1.01	0.98	1.00	1.00	1.00	0.99
0.25	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MLE		1.00		1.00		1.00		1.00	

The entries are mean values based on 100,000 samples. For μ, all the standard errors are 0.000. For σ, most standard errors are ≤ 0.001 except for (i) $n = 50$: $a = b = 0.49$ (MTM = 0.005, MWM = 0.003); (ii) $n = 100$: $a = b = 0.49$ (MTM = 0.002); (iii) $n = 250$: $a = b = 0.49$ (MTM = 0.002, MWM = 0.002).

Table 4.2: Logistic model–RE: finite-sample efficiency of MTMs and MWMs relative to MLEs.

a	b	n=50				n=100					
		RE		SD		RE		SD			
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM		
0	0	0.91		0.003		0.90		0.003			
0.05	0.05	0.82	0.89	0.002	0.003	0.93	0.91	0.004	0.003		
0.10	0.10	0.86	0.88	0.002	0.002	0.87	0.87	0.004	0.004		
0.25	0.25	0.58	0.68	0.001	0.002	0.62	0.68	0.002	0.002		
0.49	0.49	0.06	0.11	0.000	0.000	0.08	0.10	0.000	0.000		
0.10	0.70	0.27	0.32	0.001	0.001	0.28	0.32	0.001	0.001		
0.25	0.00	0.71	0.79	0.002	0.002	0.70	0.78	0.003	0.003		
MLE		0.99		0.003		1.00		0.003			
a	b	n=250				n=500				ARE	
		RE		SD		RE		SD			
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM
0	0	0.90		0.002		0.89		0.003		0.893	
0.05	0.05	0.91	0.91	0.003	0.003	0.93	0.91	0.003	0.004	0.936	0.913
0.10	0.10	0.87	0.88	0.003	0.003	0.87	0.88	0.003	0.002	0.874	0.878
0.25	0.25	0.62	0.68	0.001	0.002	0.62	0.68	0.002	0.002	0.625	0.680
0.49	0.49	0.08	0.10	0.000	0.000	0.09	0.10	0.000	0.000	0.095	0.104
0.10	0.70	0.28	0.32	0.001	0.001	0.28	0.32	0.001	0.001	0.283	0.323
0.25	0.00	0.69	0.78	0.002	0.002	0.68	0.78	0.003	0.003	0.681	0.774
MLE		1.00		0.004		1.00		0.003		1	

The entries are relative efficiency (RE) and standard error (SD) values based on 100,000 samples.

4.2 Laplace Model

We continue our simulation study with the Laplace distribution LAP($\mu = 2, \sigma = 5$) using the following parameters:

- *Sample size*: $n = 50, 100, 250, 500$.
- *Estimators of* μ, σ:
  - MLE
  - Q (MTM or MWM) 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 $b = 0.00$.

Here, the MLE of Laplace distribution is derived from equation (2.22) in Section 2.3.4.

Simulation results are summarized in Tables 4.3 and 4.4. The Laplace distribution exhibits similar findings as the logistic distribution in Section 4.1.

Table 4.3: Laplace model–mean values of $\widehat{\mu}/\mu$ and $\widehat{\sigma}/\sigma$.

a	b	n=50				n=100			
		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$	
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM
0	0	1.00		0.98		1.00		0.99	
0.05	0.05	1.00	1.00	1.08	1.02	1.00	1.00	1.01	0.99
0.10	0.10	1.00	1.00	1.02	0.99	1.00	1.00	1.01	0.99
0.25	0.25	1.00	1.00	1.09	1.02	1.00	1.00	1.02	0.99
0.49	0.49	1.00	1.00	1.90	1.10	1.00	1.00	0.93	0.54
0.10	0.70	1.04	0.79	1.03	0.91	1.03	0.89	1.02	0.96
0.25	0.00	0.97	0.99	0.98	0.98	1.01	1.01	0.98	0.99
MLE		1.00		0.99		1.00		0.99	
a	b	n=250				n=500			
		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$	
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM
0	0	1.00		1.00		1.00		1.00	
0.05	0.05	1.00	1.00	1.02	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.25	0.25	1.00	1.00	1.02	1.01	1.00	1.00	1.00	1.00
0.49	0.49	1.00	1.00	1.30	1.04	1.00	1.00	1.06	0.93
0.10	0.70	1.01	0.96	1.01	0.98	1.01	0.98	1.00	0.99
0.25	0.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
MLE		1.00		1.00		1.00		1.00	

The entries are mean values based on 100,000 samples. For μ, most standard errors are $\leq$ 0.001 except for (i) $n = 50$: $a = 0.10, b = 0.70$ (MTM = 0.002, MWM = 0.002); $a = 0.25$, $b = 0.00$ (MTM=0.002) (ii) $n = 100$: $a = 0.10, b = 0.70$ (MTM=0.003, MWM=0.002). For σ, most standard errors are ≤ 0.001 except for (i) $n = 50$: $a = b = 0.49$ (MTM = 0.005, MWM = 0.003); (ii) $n = 100$: $a = b = 0.49$ (MTM = 0.003, MWM = 0.002).

Table 4.4: Laplace model–RE: finite-sample efficiency of MTMs and MWMs relative to MLEs.

a	b	n=50				n=100			
		RE		SD		RE		SD	
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM
0	0	0.65		0.002		0.64		0.002	
0.05	0.05	0.60	0.64	0.002	0.002	0.71	0.67	0.003	0.002
0.10	0.10	0.69	0.66	0.002	0.002	0.70	0.66	0.002	0.002
0.25	0.25	0.52	0.56	0.002	0.002	0.59	0.57	0.002	0.002
0.49	0.49	0.06	0.12	0.000	0.000	0.10	0.13	0.000	0.000
0.10	0.70	0.23	0.28	0.001	0.001	0.24	0.28	0.001	0.001
0.25	0.00	0.58	0.61	0.002	0.002	0.56	0.60	0.002	0.002
MLE		0.91		0.003		0.93		0.003	

a	b	n=250				n=500				ARE	
		RE		SD		RE		SD			
		MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM
0	0	0.63		0.001		0.63		0.002		0.633	
0.05	0.05	0.69	0.65	0.002	0.002	0.72	0.66	0.002	0.002	0.723	0.663
0.10	0.10	0.71	0.65	0.002	0.002	0.71	0.65	0.002	0.002	0.714	0.653
0.25	0.25	0.58	0.57	0.001	0.001	0.60	0.58	0.001	0.002	0.607	0.575
0.49	0.49	0.09	0.13	0.000	0.000	0.12	0.14	0.000	0.000	0.128	0.140
0.10	0.70	0.24	0.28	0.001	0.001	0.24	0.29	0.001	0.001	0.243	0.287
0.25	0.00	0.55	0.59	0.001	0.001	0.54	0.59	0.002	0.002	0.539	0.588
MLE		0.95		0.001		0.97		0.002		1	

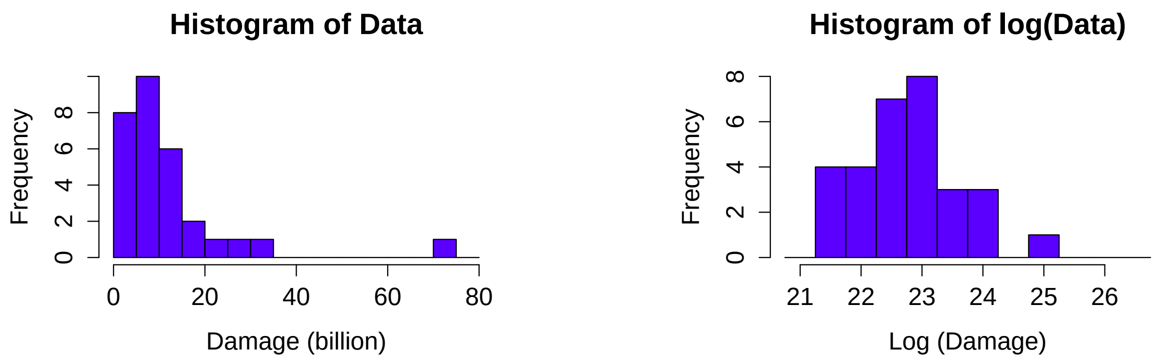
The entries are relative efficiency (RE) and standard error (SD) values based on 100,000 samples.

5 Real Data Illustrations

In this chapter we apply the MTM and MWM 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).

5.1 Model Fitting

The shape of the density curve (see Figure 5.1) is similar to many insurance loss distributions, for example, lognormal, log-logistic and log-Laplace distributions. We next discuss the robustness and fit of the MTM and MWM estimators in each model.

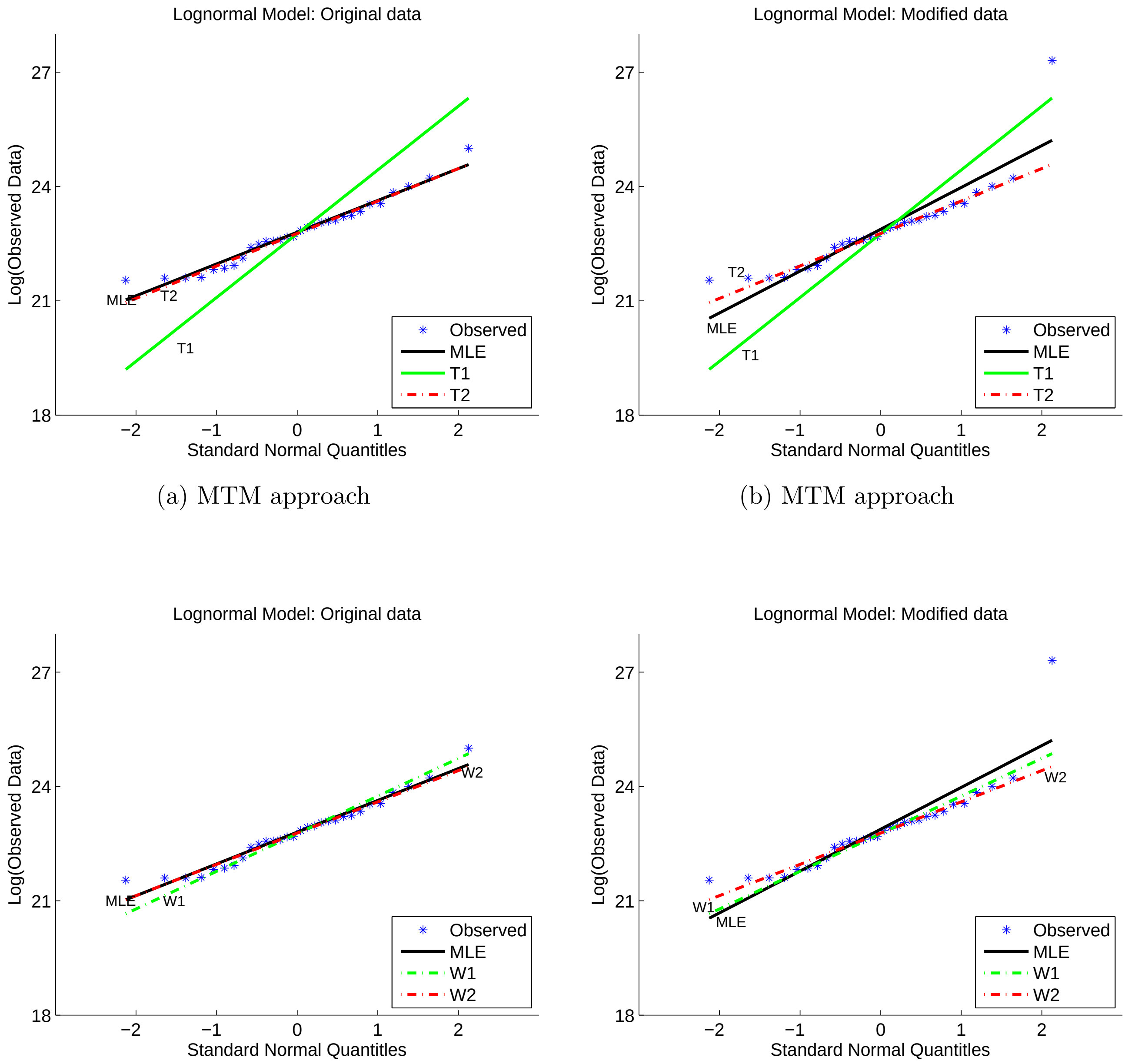


(a) Original Data (b) Log transformation of the Data

Figure 5.1: The histogram of the top 30 damaging hurricanes.

5.1.1 Lognormal Model

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. We now fit the lognormal distribution to the data using both the MTM and MWM approaches with two pairs of trimming and Winsorizing proportions a and b. For comparison, we also fit this model using the MLE estimator (which corresponds to $a = b = 0$).



(a) MTM approach

(b) MTM approach

(c) MWM approach

(d) MWM approach

Figure 5.2: Lognormal fits to the original (left panel) and modified (right panel) hurricane data.

The initial fits from the MTM and MWM approaches are illustrated in the upper and lower left panel of Figure 5.2, respectively, where T1 and W1 denote the MTM and MWM estimator with $a = b = \frac{14}{30}$, respectively. Likewise, T2 and W2 correspond to the case $a = b = \frac{1}{30}$. The parameter estimates, AREs and goodness-of-fit measurements (Fit) appear in Table 5.1. The goodness-of-fit is measured using the mean absolute deviation $(1/30)\sum_{i=1}^{30} |\log \widehat{F}^{-1}((j-0.5)/30) - \log X_{j:30}|$ between the log-fitted and log-observed data.

It is clear that the parameter estimates and model fits are strongly dependent on the trimming (Winsorizing) proportions a and b and thus the proportions should be chosen carefully. In particular, the T1 and W1 estimators are highly robust but very inefficient, with the ARE being only 13.9% for MTM and 15.5% for MWM. 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 (Winsorizing) of one observation at each tail leads to the T2 and W2 fits which are practically identical to the MLE fit.

Moreover, in terms of accuracy, the MWM approach has obvious advantage over MTM for high values of a and b (i.e., $a = b = \frac{14}{30}$): the fit error decreases from 0.660 to 0.140. Note also that it has a slightly better ARE. For small values of a and b (i.e., $a = b = \frac{1}{30}$), MWM has better ARE and comparable goodness-of-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 5.2. The new parameter estimates and goodness-of-fit measurements (marked with superscript $*$) are reported in Table 5.1. As we see, for both MTM and MWM approaches, the T1, T2 and W1, W2 parameter estimates are not affected by the data modification whereas the new MLE fit is significantly different.

Table 5.1: Statistical quantities and goodness-of-fit measurements for the fitted models.

Estimators	Lognormal				Log-logistic				Log-Laplace			
	ARE	$\widehat{\mu}$	$\widehat{\sigma}$	Fit	ARE	$\widehat{\mu}$	$\widehat{\sigma}$	Fit	ARE	$\widehat{\mu}$	$\widehat{\sigma}$	Fit
MLE	1.000	22.800	0.834	0.104	1.000	22.775	0.477	0.104	1.000	22.760	0.663	0.122
MLE modified	1.000	22.877	1.098	0.293	1.000	22.777	0.531	0.185	1.000	22.760	0.740	0.196
T1$(a=\frac{14}{30}, b=\frac{14}{30})$	0.139	22.760	1.673	0.660	0.179	22.760	1.408	0.767	0.233	22.760	2.045	1.336
T1$^*(a=\frac{14}{30}, b=\frac{14}{30})$	0.139	22.760	1.673	0.649	0.179	22.760	1.408	0.709	0.233	22.760	2.045	1.260
T2$(a=\frac{1}{30}, b=\frac{1}{30})$	0.910	22.766	0.852	0.101	0.949	22.766	0.497	0.101	0.718	22.766	0.687	0.123
T2$^*(a=\frac{1}{30}, b=\frac{1}{30})$	0.910	22.766	0.852	0.178	0.949	22.766	0.497	0.178	0.718	22.766	0.687	0.193
W1$(a=\frac{14}{30}, b=\frac{14}{30})$	0.155	22.760	0.988	0.140	0.199	22.760	0.619	0.191	0.250	22.760	1.198	0.515
W1$^*(a=\frac{14}{30}, b=\frac{14}{30})$	0.155	22.760	0.988	0.216	0.199	22.760	0.619	0.249	0.250	22.760	1.198	0.470
W2$(a=\frac{1}{30}, b=\frac{1}{30})$	0.942	22.776	0.820	0.104	0.919	22.776	0.470	0.106	0.663	22.776	0.629	0.128
W2$^*(a=\frac{1}{30}, b=\frac{1}{30})$	0.942	22.776	0.820	0.181	0.919	22.776	0.470	0.183	0.663	22.776	0.629	0.204

5.1.2 Log-Logistic Model

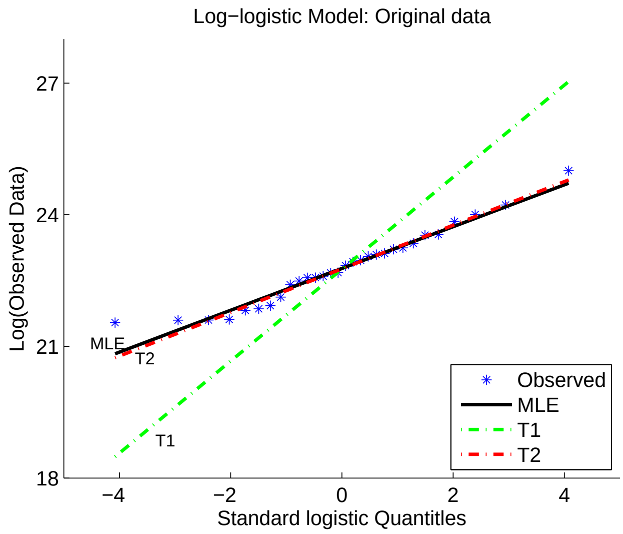
As an alternative to the lognormal model, in this section we repeat the model-fitting exercise with the log-logistic model. T1, T2 and W1, W2 are based on the same trimming (Winsorizing) proportions as before. The fitted models are shown in Figure 5.3 and the results are recorded in Table 5.1.

As in Section 2.3.3, the MLE here is found by using the Newton method. To guarantee convergence of the iteration, initial values $\widehat{\mu}_{\text{start}}$ and $\widehat{\sigma}_{\text{start}}$ are set as median and normalized *median absolute deviation* (MAD) about the median of log-transformed data $\mathbf{X} = (x_1, \ldots, x_{30})$ respectively, which are

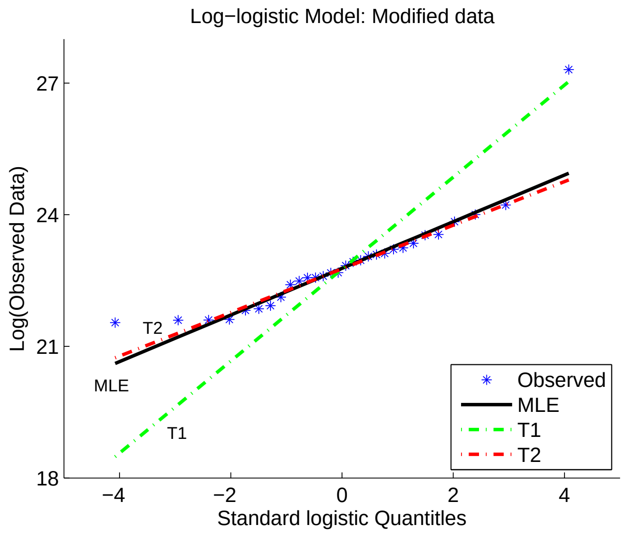
$$\begin{cases} \widehat{\mu}_0 = \text{Median}(\log(\mathbf{X})); \\ \\ \widehat{\sigma}_0 = \dfrac{\text{Median}|\log(\mathbf{X}) - \text{Median}(\log(\mathbf{X}))|}{0.6745}. \end{cases}$$

Here, the median and MAD are robust alternatives to the mean and standard deviation, respectively. Both of them are not influenced by the presence of a large outlier. Note that for non-robust choices of $\widehat{\mu}_{\text{start}}, \widehat{\sigma}_{\text{start}}$ the method *fails to converge.*

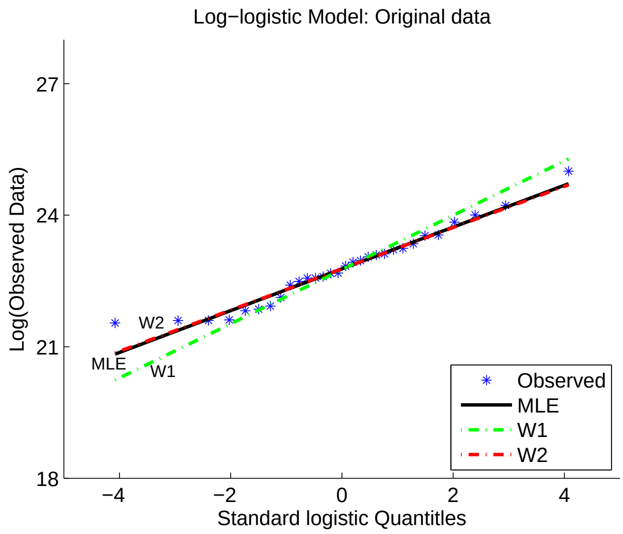
As we see in Figure 5.3 and Table 5.1, the T1, T2 and W1, W2 parameter estimates are not affected by the data modification while the new MLE fit is slightly different. That means, MLE in the log-logistic case, being the most efficient estimator, still has some degree of robustness. Moreover, both the ARE and goodness-of-fit show that MWM is more appropriate for $a = b = \frac{14}{30}$ (i.e., W1) while MTM performs a little better for $a = b = \frac{1}{30}$ (i.e., T2), which matches the findings from Section 3.3.3.



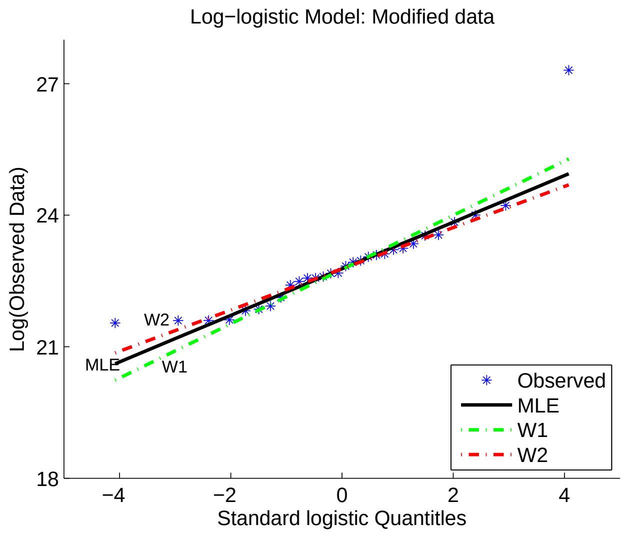
(a) MTM approach



(b) MTM approach



(c) MWM approach

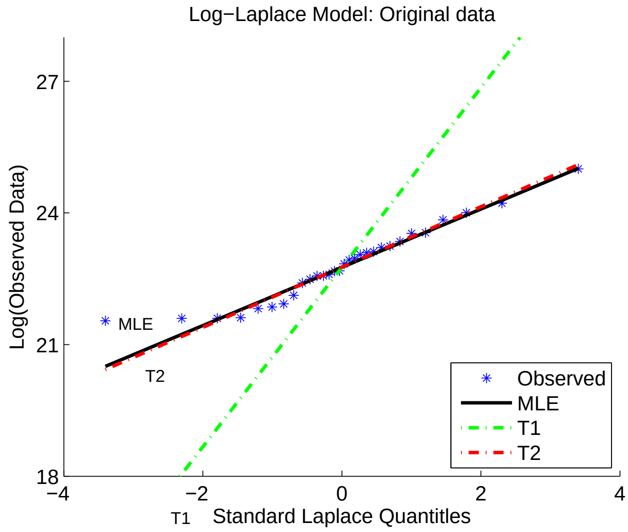


(d) MWM approach

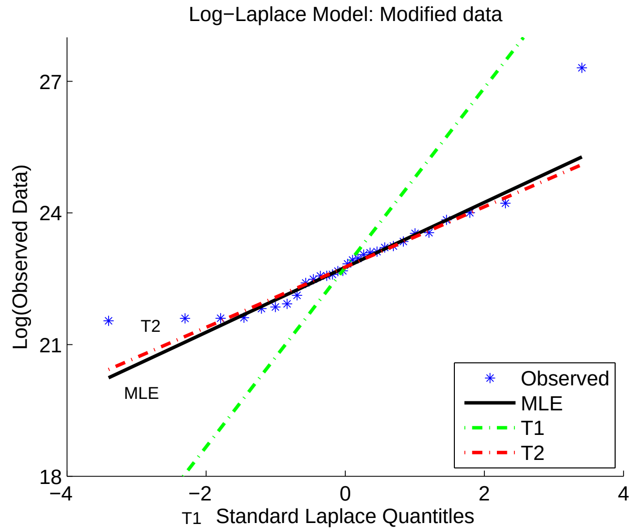
Figure 5.3: Log-logistic fits to the original (left panel) and modified (right panel) hurricane data.

5.1.3 Log-Laplace Model

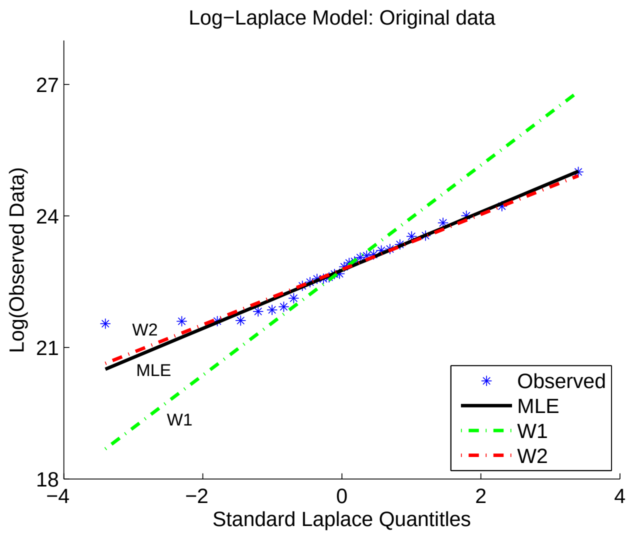
In this section we re-fit the hurricane data with the log-Laplace model as yet another alternative to the lognormal model. Figure 5.4 shows that, except for T1, all other estimators, including the MLE, fit the data well. This is true for both original and modified data. The success of MLE in this case is not surprising because $\widehat{\mu}_{\text{MLE}}$ is the median of data (thus the overall MLE inherits some robustness). The most notable advantage of MWM over MTM is the goodness-of-fit criterion for $a = b = \frac{14}{30}$. Also, we can see from Table 5.1 that all procedures – MTM, MWM, and MLE – yield nearly identical estimates of μ. For the scale parameter σ, however, there are some differences. Most importantly, MTM and MWM are completely unaffected by the data modification, whereas the MLE of σ is significantly influenced by the single outlier.



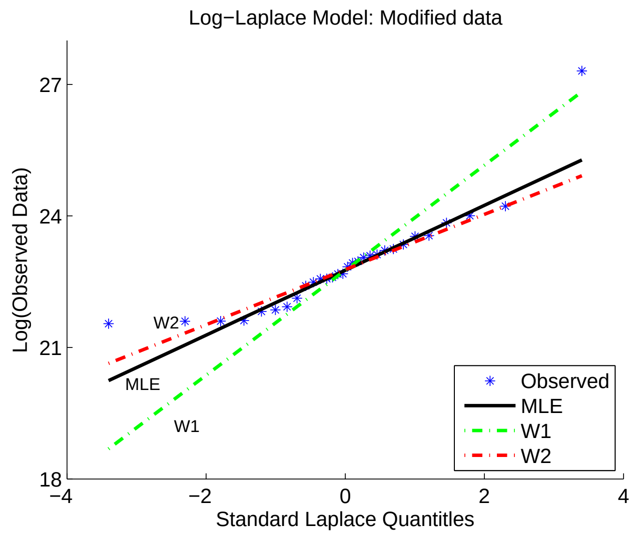
(a) MTM approach



(b) MTM approach



(c) MWM approach



(d) MWM approach

Figure 5.4: Log-Laplace fits to the original (left panel) and modified (right panel) hurricane data.

5.2 Actuarial Premiums

Consider now estimation of the severity component of the pure premium for an insurance benefit (Z) that equals to the amount by which a hurricane's damage (X) exceeds 5 (billion) with a maximum benefit of 20 (billion). That is,

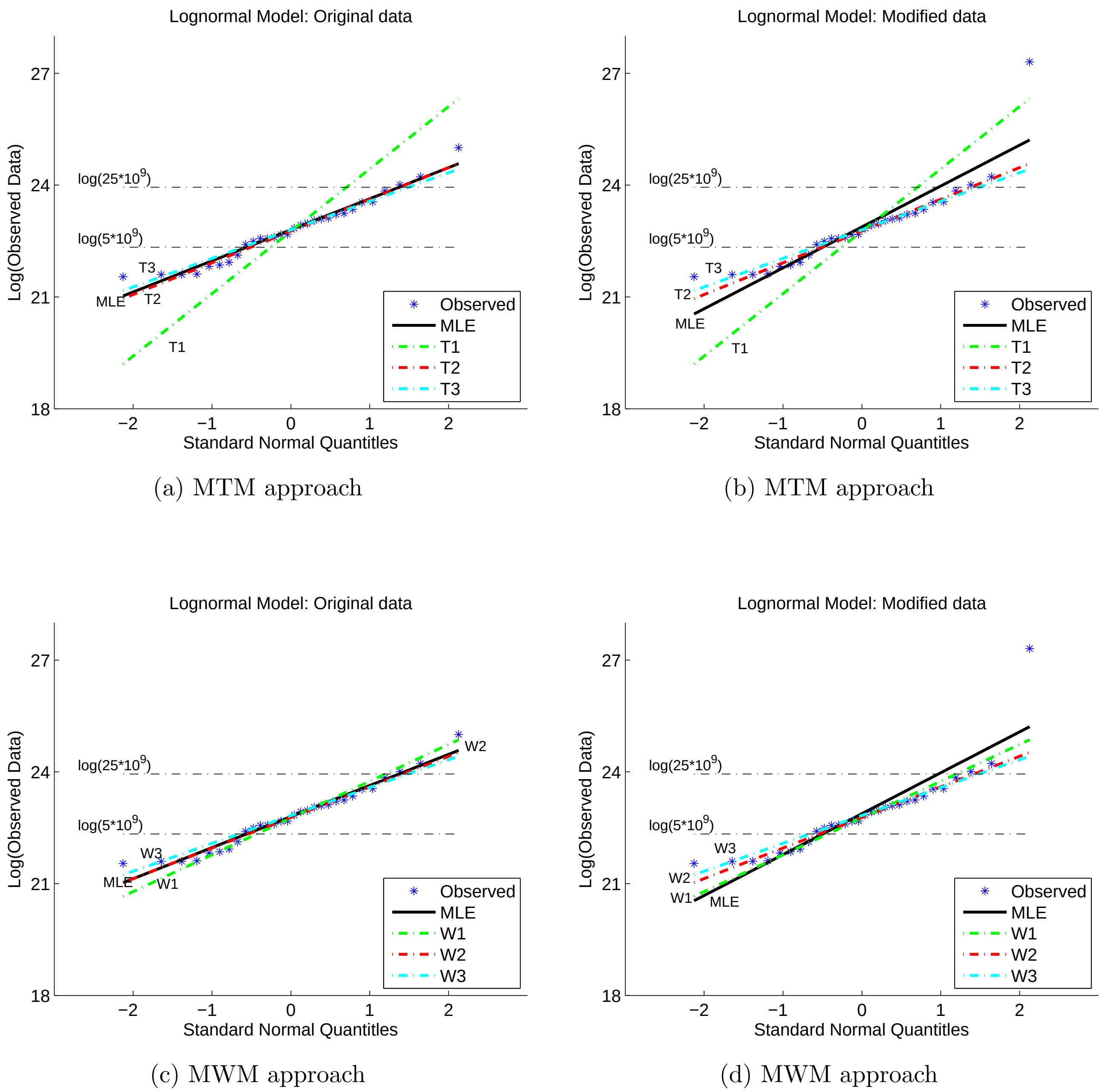
$$Z = \begin{cases} 0, & \text{if } X \leq 5; \\ X - 5, & \text{if } 5 < X \leq 25; \\ 20, & \text{if } X > 25. \end{cases} \tag{5.1}$$

If X follows the distribution function F, we seek

$$\begin{aligned} \mathit{\Pi}(F) &= \mathbb{E}[X \wedge 25] - \mathbb{E}[X \wedge 5] \\ &= \int_5^{25} (x - 5)\, dF(x) + 20(1 - F(25)) \\ &= 20 - \int_5^{25} F(x)\, dx. \end{aligned} \tag{5.2}$$

Since it is now most important that our fitted distribution captures the behavior of the underlying damage distribution between 5 and 25, the MTM and MWM estimators are most natural with the choices $a = \frac{8}{30}$ (which corresponds to the proportion of observations below 5) and $b = \frac{3}{30}$ (which corresponds to the proportion of observations above 25). We denote these MTM and MWM estimators by T3 and W3, respectively.

As we see from the left panels of Figures 5.5, 5.6 and 5.7, for each distribution, the overall T3 (W3) fits are very similar to those of T2 (W2) and MLE, but they yield a closer fit than the latter two procedures over the layer of interest, which is [5, 25], and they are unaffected by the large outlier (modification). Modified data fits are shown in the right panels of the figures.



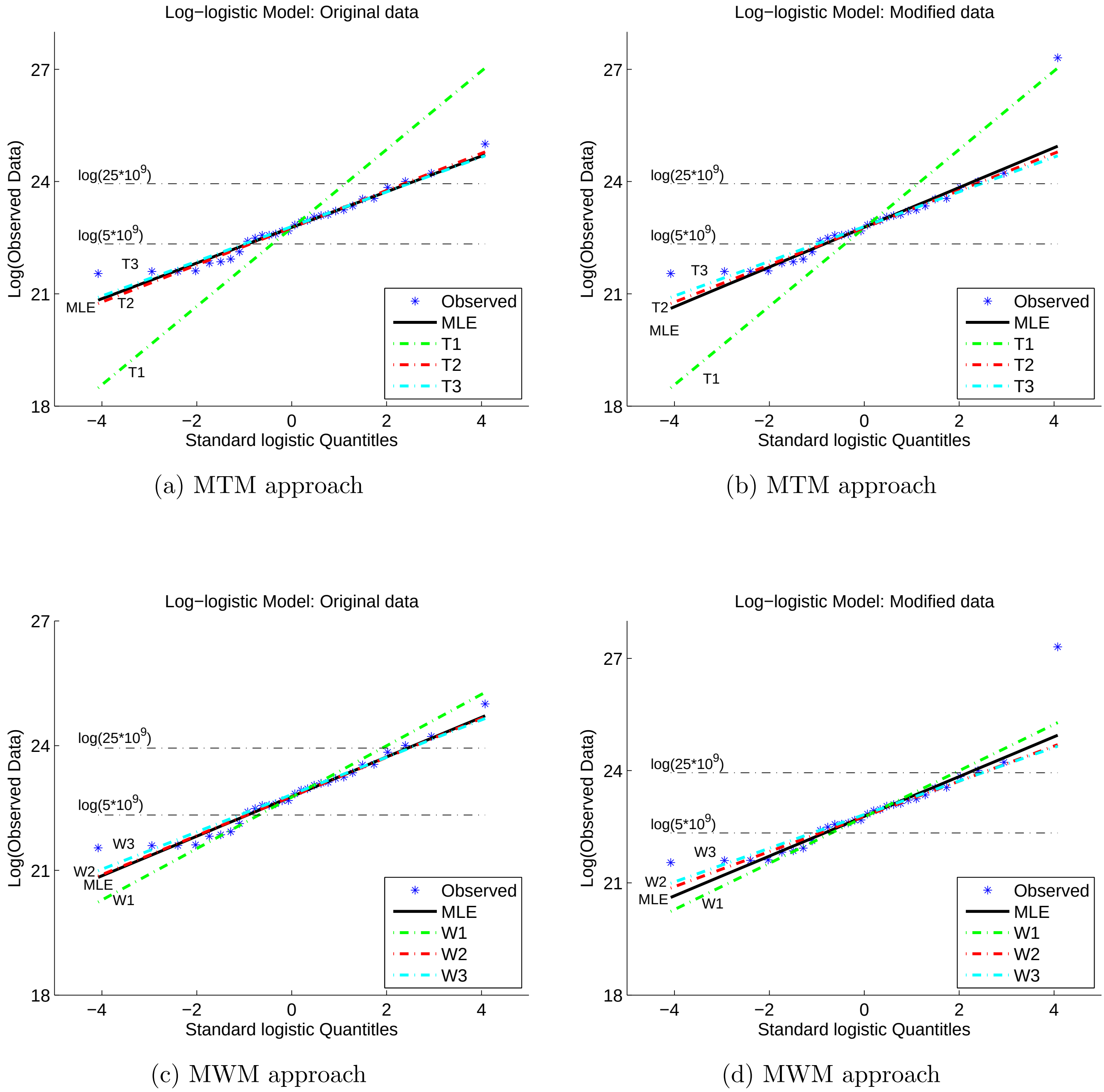
(a) MTM approach

(b) MTM approach

(c) MWM approach

(d) MWM approach

Figure 5.5: Lognormal fits to the original (left panel) and modified (right panel) hurricane data, with the insurance layer [5, 25].



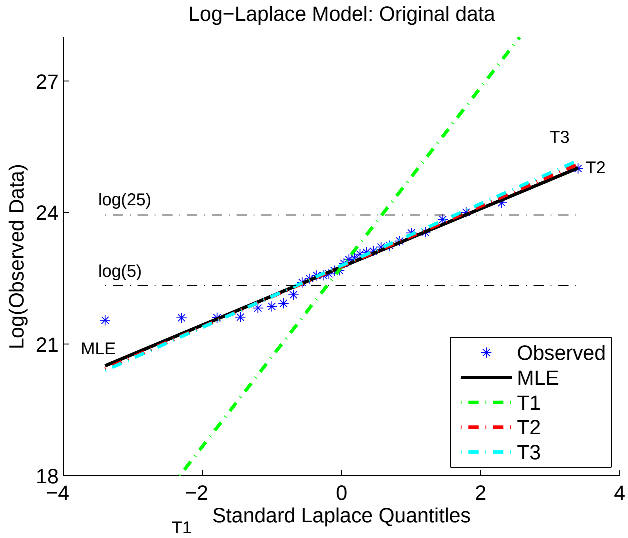
(a) MTM approach

(b) MTM approach

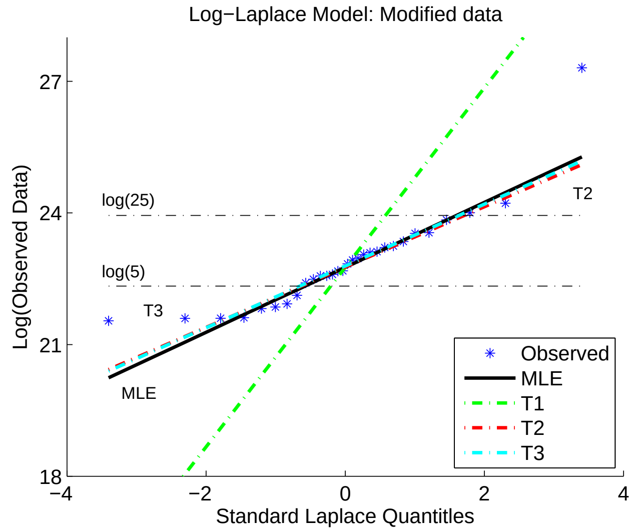
(c) MWM approach

(d) MWM approach

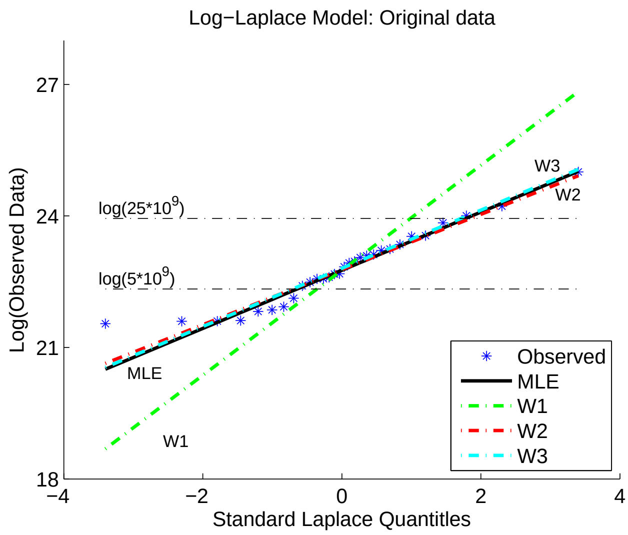
Figure 5.6: Log-logistic fits to the original (left panel) and modified (right panel) hurricane data, with the insurance layer [5, 25].



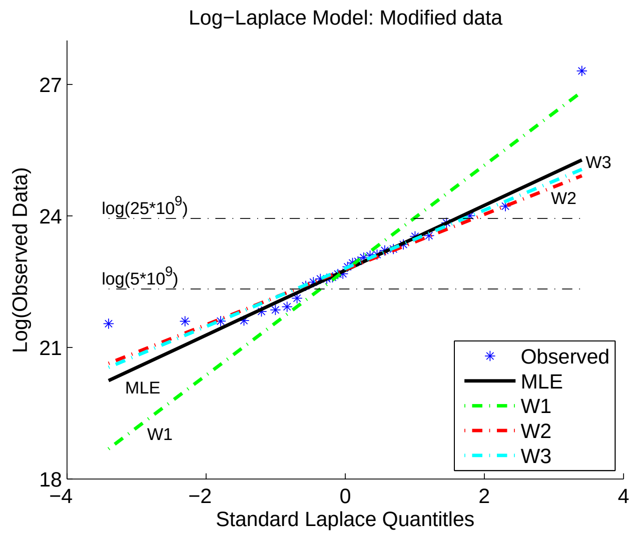
(a) MTM approach



(b) MTM approach



(c) MWM approach



(d) MWM approach

Figure 5.7: Log-Laplace fits to the original (left panel) and modified (right panel) hurricane data, with the insurance layer [5, 25].

In Table 5.2 we also provide the actuarial premiums (Pm.) calculated using equation (5.2) 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 5.2 also contains 95% confidence intervals (CIs) for the premium $\Pi(F)$. For parametric CIs, we use the delta method applied to the transformation of parameter estimators given by equation (5.2) together with the MTM, MWM and MLE asymptotic distributions. The

Table 5.2: 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].

Estimators	Lognormal			Log-logistic			Log-Laplace		
	RFit	Pm.	CIs	RFit	Pm.	CIs	RFit	Pm.	CIs
MLE	0.054	5.604	(3.368, 7.839)	0.045	5.290	(3.022, 7.599)	0.058	5.065	(3.257, 6.874)
MLE modified	0.202	6.896	(4.377, 9.416)	0.059	5.622	(3.256, 7.988)	0.048	5.338	(3.419, 7.257)
T1($a=\frac{14}{30}$,$b=\frac{14}{30}$)	0.412	7.342	(2.552, 12.133)	0.433	7.372	(2.711, 12.034)	0.619	7.675	(3.771, 11.579)
T1*($a=\frac{14}{30}$,$b=\frac{14}{30}$)	0.412	7.342	(2.552, 12.133)	0.433	7.372	(2.711, 12.034)	0.619	7.675	(3.771, 11.579)
T2($a=\frac{1}{30}$, $b=\frac{1}{30}$)	0.057	5.436	(3.168, 7.704)	0.050	5.356	(3.059, 7.652)	0.049	5.196	(2.864, 7.528)
T2*($a=\frac{1}{30}$, $b=\frac{1}{30}$)	0.057	5.436	(3.168, 7.704)	0.050	5.356	(3.059, 7.652)	0.049	5.196	(2.864, 7.528)
T3($a=\frac{8}{30}$, $b=\frac{3}{30}$)	0.042	5.335	(3.065, 7.605)	0.040	5.370	(3.106, 7.633)	0.041	5.421	(3.123, 7.718)
T3*($a=\frac{8}{30}$, $b=\frac{3}{30}$)	0.042	5.335	(3.065, 7.605)	0.040	5.370	(3.106, 7.633)	0.041	5.421	(3.123, 7.718)
W1($a=\frac{14}{30}$,$b=\frac{14}{30}$)	0.105	5.859	(0.857, 10.861)	0.117	5.959	(1.226, 10.692)	0.223	6.532	(2.536, 10.528)
W1*($a=\frac{14}{30}$,$b=\frac{14}{30}$)	0.105	5.859	(0.857, 10.861)	0.117	5.959	(1.226, 10.692)	0.223	6.532	(2.536, 10.528)
W2($a=\frac{1}{30}$, $b=\frac{1}{30}$)	0.050	5.384	(3.165, 7.603)	0.044	5.258	(2.976, 7.540)	0.060	5.046	(2.707, 7.385)
W2*($a=\frac{1}{30}$, $b=\frac{1}{30}$)	0.050	5.384	(3.165, 7.603)	0.044	5.258	(2.976, 7.540)	0.060	5.046	(2.707, 7.385)
W3($a=\frac{8}{30}$, $b=\frac{3}{30}$)	0.046	5.486	(3.257, 7.715)	0.041	5.464	(3.189, 7.739)	0.044	5.369	(2.979, 7.760)
W3*($a=\frac{8}{30}$, $b=\frac{3}{30}$)	0.046	5.486	(3.257, 7.715)	0.041	5.464	(3.189, 7.739)	0.044	5.369	(2.979, 7.760)
Empirical	-	5.416	(3.111, 7.722)	-	5.416	(3.111, 7.722)	-	5.416	(3.111, 7.722)

RFit: Goodness-of-fit measurements (restricted to the data in [5, 25]);
Pm.: Actuarial Premiums; CIs: Confidence Intervals.

resulting confidence intervals are constructed as follows:

$$\widehat{\Pi(F)} \pm z_{\frac{\alpha}{2}} \sqrt{\mathrm{Var}[\widehat{\Pi(F)}]},$$

where $\widehat{\mathrm{Var}[\Pi(F)]} = (\nabla h)' \, \Sigma_1 \, (\nabla h)$, $\nabla h = (\frac{\partial \Pi}{\partial \mu}, \frac{\partial \Pi}{\partial \sigma})'|_{(\widehat{\mu}, \widehat{\sigma})}$, $\Sigma_1 = \frac{\widehat{\sigma}^2}{n}\mathbf{S}$ or $\frac{\widehat{\sigma}^2}{n}\tilde{\mathbf{S}}$, and the level of CI is $1 - \alpha = 0.95$.

For constructing the empirical interval, we use the classical central limit theorem and have that

$$\Pi(\widehat{F_n}) \sim \mathcal{AN}\big(\Pi(F), n^{-1}V(F)\big),$$

where $V(F)$ is derived from equations (5.1) and (5.2), that is

$$\begin{aligned} V(F) &= \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 \\ &= \int_5^{25} (x-5)^2 dF(x) + 400(1 - F(25)) - (\Pi(F))^2. \end{aligned}$$

Replacing 72.303 by 723.03 affects the MLE. This is evident from the results in Table 5.2. The magnitude of premium shift under the MLE approach is model dependent. The single outlier inflates the premium by 23% at the lognormal model, 6% at the log-logistic, and 5% at the log-Laplace distribution. MTM and MWM, on the other hand, maintain the stability of restricted fit (RFit), premium estimate, and confidence interval. Moreover, we observe that the MTM and MWM procedures with appropriate cutting proportions (i.e., the T2, T3 and W2, W3) lead to premium estimates that are closer to the empirical estimate than those obtained with over-cutting (i.e., T1 and W1) or under-cutting (i.e., MLE). Also, the main advantage of parametric procedures (MTM, MWM and MLE) over the empirical approach is that in general they produce shorter confidence intervals for the measures of interest.

In summary, the illustration we have provided in this section exemplifies the idea that the MTM and MWM are appropriate choices for various model-fitting situations including those when a close fit in one or both tails of the distribution is not required.

6 Conclusion and Discussion

6.1 Summary

In this dissertation, we have presented two methods for estimating the parameters of claim severity distributions: the method of trimmed moments (MTM) and the method of Winsorized moments (MWM). For MTM, we closely followed the description presented by Brazauskas et al. (2009), described the asymptotic properties of the MTM estimators and provided examples of estimators for location-scale families and several loss models– exponential, logistic, Laplace and their exponential transformations. The main contribution of the dissertation is introduction and development of the MWM procedure. MWM is similar to MTM but now data trimming is replaced with data Winsorizing. Asymptotic properties of the new procedure have been rigorously studied and small-sample properties have been investigated using Monte Carlo Simulations.

Further, as demonstrated theoretically and via simulation, both methods can achieve various degree of robustness, which allows a controlled and desired balance between robustness and efficiency. Simulation study supplements our theoretical results and points out that MTM and MWM estimators are not only reliable, but also very computationally efficient estimators. A general asymptotic theory of new estimators is derived, which can be used to construct confidence intervals and to test hypotheses as well.

Finally, the effect of model choice and parameter estimation method on risk pricing is illustrated using actual data that represent hurricane damages in the United States from 1925 to 1995. In particular, the estimated pure premiums for an insurance contract are computed when the log-logistic, log-Laplace, and lognormal models are fitted to the data using the MTM, MWM and MLE methods. The real-data study reveals that calculating the premiums for the layers of insurance coverage is a task for which MTM and MWM are particularly natural.

6.2 Future Work

The findings of this dissertation suggest several directions for future research. In this section, we raise questions and discuss possible lines of attack for the future research problems.

6.2.1 MTM or MWM?

The first question to be asked is which method – MTM or MWM – is better? And the answer is "it depends". First of all, both methods are equally straightforward computationally. Secondly, using the same a_j and b_j proportions in both methods, we can achieve identical robustness properties for MTM and MWM estimators. Thirdly, in terms of ARE, the Winsorized estimators perform better when a_j and b_j are extreme (or large), but for smaller proportions there is no consistent winner as the outcome depends on the underlying distribution. This question needs more work and better insight on what characteristics of the model are the most influential drivers of the ARE of MTM and MWM.

6.2.2 Optimal Choice of h_j, a_j and b_j

The second question to be asked is how to choose functions h_j and proportions a_j and b_j? For the MTM procedure, this question was studied by Brazauskas (2009). It was noted that the choices $h_j(t) = t^j, j = 1, \ldots, k$, work well for the location-scale families and $h_j(t) = (\log(t))^j, j = 1, \ldots, k$, for log-location-scale families. One could guess that for the inverse type distributions $h_j(t) = t^{-j}, j = 1, \ldots, k$, might be a natural choice. Due to the conceptual similarities of MTM and MWM, we expect that similar choices of functions h_j should work for the MWM procedure. The question about proportions a_j and b_j is more difficult, and the answers proposed by Brazauskas (2009) rely on diagnostic plots and other qualitative assessments. To give more rigorous answers, one needs to introduce new criteria to complement the ARE versus robustness studies. At this moment it is not clear what criteria we should use, but it would be interesting to explore this question in depth.

6.2.3 Actuarial Applications

The real-data example of Chapter 5 shows that MTM and MWM are natural estimators for pricing insurance layers. This line of work could be future explored in the context of reinsurance where contracts have similar loss control schemes. In addition, one could investigate a blended version of MTM and MWM because that would mimic the joint effect of deductible and policy limit. Taking a step further, we can supplement deductible and policy limit with the co-insurance factor and construct an estimation method that is equivalent to insurance payment variables. Also, such methods may provide a useful alternative in modeling operational risk data, where data truncation from below creates well-known computational issues for the MLE based procedures.

References

[1] Bickle, P.J., 1965. On some robust estimators of location. *The Annals of Mathematical Statistics*, **36**, 847–858.

[2] Brazauskas, V., 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.

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

[4] Brazauskas, V., 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.

[5] Brazauskas, V., 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.

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

[7] Brazauskas, V., 2009. Robust and efficient fitting of loss models: diagnostic tools and insights. *North American Actuarial Journal*, **13**(3), 356–369.

[8] Brazauskas, V., Jones, B.L., Zitikis, R., 2009. Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference*, **139**(6), 2028–2043.

[9] Chernoff, H., Gastwirth, J.L., Jones, M.V., 1967. Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *The Annals of Mathematical Statistics*, **38**(1), 52–72.

[10] Cont, R., 2006. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, **16**(3), 519–547.

[11] deCani, J.S., Stine, R.A., 1986. A note on the information matrix for a logistic distribution. *The American Statistician*, **40**(3), 220–222.

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

[13] Hampel, F.R., 1968. *Contributions to theTtheory of Robust Estimation.* Ph.D. dissertation, University of California, Berkeley.

[14] Hampel, F.R., 1971. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, **42**, 1887–1896.

[15] Hampel, F.R., 1974. Beyond location parameters: Robust concepts and methods (with discussion). *Bulletin of the International Statistical Institute*, **46**(1), 375–391.

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

[17] Hansen, L.P., Sargent, T.J., 2008. *Robustness.* Princeton University Press.

[18] Huber, P.J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**(1), 73–101.

[19] Huber, P.J., 1972. Robust statistics: a review. *The Annals of Mathematical Statistics*, **43**, 1041–1067.

[20] Huber, P.J., Ronchetti, E.M., 2009. *Robust Statistics*, 2nd edition. Wiley, Hoboken.

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

[22] Klugman, S.A., Panjer, H.H.,Willmot, G.E., 2008. *Loss Models: From Data to Decisions*, 3rd edition. Wiley, New York.

[23] Kotz, S., Kozubowski, T.J., Podgórski, K., 2001. *The Laplace Distribution and Generalizations–A Revisit with New Applications.* Birkhäuser.

[24] Marceau, E., Rioux, J., 2001. On robustness in risk theory. *Insurance: Mathematics and Economics*, **29**, 167–185.

[25] Maronna, R.A., Martin, R.D., Yohai, V.J., 2006. *Robust statistics: Theory and Methods.* Wiley, Chichester.

[26] McNeil, A.J., Frey, R., Embrechts P., 2005. *Quantitative Risk Management Concepts, Techniques and Tools.* Princeton University Press.

[27] Norton, R.M., 1984. The double exponential distribution: using calculus to find a maximum likelihood estimator. *The American Statistician*, **38**(2), 135–136.

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

[29] Rustum, R., Adeloye, A.J., 2007. Replacing outliers and missing values from activated sludge data using Kohonen Self Organizing Map. *Journal of Environmental Engineering*, **133**(9), 909–916.

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

[31] 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.

Appendix

Appendix A: Covariance-Variance Matrix: General Cases

For convenience, we repeat the following notation:

$$A_i = a_i(1-a_i)H_i^{'}(a_i), \quad B_i = b_i^2 H_i^{'}(1-b_i),$$

$$A_j = a_j(1-a_j)H_j^{'}(a_j), \quad B_j = b_j^2 H_j^{'}(1-b_j),$$

$$W(w) = H_i^{'}(w)(1-w), \quad V(v) = H_j^{'}(v)(1-v),$$

$$W_*(w) = \frac{W(w)}{(1-u)^2}, \quad V_*(v) = \frac{V(v)}{(1-u)^2}.$$

Case 2: $0 \leq a_i \leq a_j < 1-b_i \leq 1-b_j \leq 1$ $\quad 0 \quad a_i \quad a_j \quad 1-b_i \quad 1-b_j \quad 1$

The second term is:

$$= \int_0^{a_i} \frac{A_j + B_j}{(1-u)^2} \int_{a_i}^{1-b_i} W(w)\, dw\, du + \int_{a_i}^{a_j} \frac{A_j + B_j}{(1-u)^2} \int_{u}^{1-b_i} W(w)\, dw\, du$$

$$+ \int_{a_j}^{1-b_i} \frac{B_j}{(1-u)^2} \int_{u}^{1-b_i} W(w)\, dw\, du$$

$$= \frac{a_i}{1-a_i}\big[A_j + B_j\big] \int_{a_i}^{1-b_i} W(w)\, dw + \big[A_j + B_j\big] \int_{a_i}^{a_j} \int_{u}^{1-b_i} W_*(w)\, dw\, du$$

$$+ \big[B_j\big] \int_{a_j}^{1-b_i} \int_{u}^{1-b_i} W_*(w)\, dw\, du.$$

The third term is:

$$= \int_0^{a_i} \frac{A_i + B_i}{(1-u)^2} \int_{a_j}^{1-b_j} V(v)\, dv\, du + \int_{a_i}^{a_j} \frac{B_i}{(1-u)^2} \int_{a_j}^{1-b_j} V(v)\, dv\, du$$

$$+ \int_{a_j}^{1-b_i} \frac{B_i}{(1-u)^2} \int_{u}^{1-b_j} V(v)\, dv\, du$$

$$= \left[\frac{a_i}{1-a_i}A_i + \frac{a_j}{1-a_j}B_i\right]\int_{a_j}^{1-b_j} V(v)\,dv + \left[B_i\right]\int_{a_j}^{1-b_i}\int_u^{1-b_j} V_*(v)\,dv\,du.$$

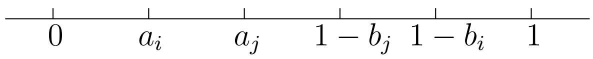
The fourth term is:

$$= \left[A_i + B_i\right]\left[A_j + B_j\right]\int_0^{a_i}\frac{1}{(1-u)^2}du + \left[B_i\right]\left[A_j + B_j\right]\int_{a_i}^{a_j}\frac{1}{(1-u)^2}du$$

$$+ \left[B_i\right]\left[B_j\right]\int_{a_j}^{1-b_i}\frac{1}{(1-u)^2}du$$

$$= \frac{a_i}{1-a_i}\left[A_i + B_i\right]\left[A_j + B_j\right] + \frac{a_j - a_i}{(1-a_i)(1-a_j)}\left[B_i\right]\left[A_j + B_j\right] + \frac{1-a_j-b_i}{(1-a_j)b_i}\left[B_i\right]\left[B_j\right].$$

Case 3: $0 \le a_i \le a_j < 1-b_j \le 1-b_i \le 1$



The second term is:

$$= \int_0^{a_i}\frac{A_j + B_j}{(1-u)^2}\int_{a_i}^{1-b_i} W(w)\,dw\,du + \int_{a_i}^{a_j}\frac{A_j + B_j}{(1-u)^2}\int_u^{1-b_i} W(w)\,dw\,du$$

$$+ \int_{a_j}^{1-b_j}\frac{B_j}{(1-u)^2}\int_u^{1-b_i} W(w)\,dw\,du$$

$$= \frac{a_i}{1-a_i}\left[A_j + B_j\right]\int_{a_i}^{1-b_i} W(w)\,dw + \left[A_j + B_j\right]\int_{a_i}^{a_j}\int_u^{1-b_i} W_*(w)\,dw\,du$$

$$+ \left[B_j\right]\int_{a_j}^{1-b_j}\int_u^{1-b_i} W_*(w)\,dw\,du.$$

The third term is:

$$= \int_0^{a_i}\frac{A_i + B_i}{(1-u)^2}\int_{a_j}^{1-b_j} V(v)\,dv\,du + \int_{a_i}^{a_j}\frac{B_i}{(1-u)^2}\int_{a_j}^{1-b_j} V(v)\,dv\,du$$

$$+ \int_{a_j}^{1-b_j}\frac{B_i}{(1-u)^2}\int_u^{1-b_j} V(v)\,dv\,du$$

$$= \left[\frac{a_i}{1-a_i}A_i + \frac{a_j}{1-a_j}B_i\right]\int_{a_j}^{1-b_j} V(v)\,dv + \left[B_i\right]\int_{a_j}^{1-b_j}\int_u^{1-b_j} V_*(v)\,dv\,du.$$

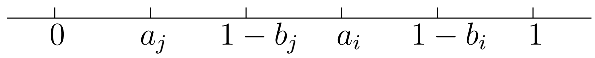
The fourth term is:

$$= [A_i + B_i][A_j + B_j] \int_0^{a_i} \frac{1}{(1-u)^2} du + [B_i][A_j + B_j] \int_{a_i}^{a_j} \frac{1}{(1-u)^2} du$$

$$+ [B_i][B_j] \int_{a_j}^{1-b_j} \frac{1}{(1-u)^2} du$$

$$= \frac{a_i}{1-a_i}[A_i + B_i][A_j + B_j] + \frac{a_j - a_i}{(1-a_i)(1-a_j)}[B_i][A_j + B_j] + \frac{1 - a_j - b_j}{(1-a_j)b_j}[B_i][B_j].$$

Case 4: $0 \le a_j \le 1 - b_j < a_i \le 1 - b_i \le 1$



The second term is:

$$= \int_0^{a_j} \frac{A_j + B_j}{(1-u)^2} \int_{a_i}^{1-b_i} W(w)\, dw\, du + \int_{a_j}^{1-b_j} \frac{B_j}{(1-u)^2} \int_{a_i}^{1-b_i} W(w)\, dw\, du$$

$$= \left[\frac{a_j}{1-a_j} A_j + \frac{1-b_j}{b_j} B_j\right] \int_{a_i}^{1-b_i} W(w)\, dw.$$

The third term is:

$$= \int_0^{a_j} \frac{A_i + B_i}{(1-u)^2} \int_{a_j}^{1-b_j} V(v)\, dv\, du + \int_{a_j}^{1-b_j} \frac{A_i + B_i}{(1-u)^2} \int_u^{1-b_j} V(v)\, dv\, du$$

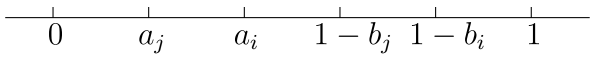
$$= \frac{a_j}{1-a_j}[A_i + B_i] \int_{a_j}^{1-b_j} V(v)\, dv + [A_i + B_i] \int_{a_j}^{1-b_j} \int_u^{1-b_j} V_*(v)\, dv\, du.$$

The four term is:

$$= [A_i + B_i][A_j + B_j] \int_0^{a_j} \frac{1}{(1-u)^2}\, du + [A_i + B_i][B_j] \int_{a_j}^{1-b_j} \frac{1}{(1-u)^2}\, du$$

$$= \frac{a_j}{1-a_j}[A_i + B_i][A_j + B_j] + \frac{1 - a_j - b_j}{(1-a_j)b_j}[A_i + B_i][B_j].$$

Case 5: $0 \le a_j \le a_i < 1 - b_j \le 1 - b_i \le 1$



The second term is:

$$
\begin{aligned}
&= \int_0^{a_j} \frac{A_j + B_j}{(1-u)^2} \int_{a_i}^{1-b_i} W(w)\, dw\, du + \int_{a_j}^{a_i} \frac{B_j}{(1-u)^2} \int_{a_i}^{1-b_i} W(w)\, dw\, du \\
&\quad + \int_{a_i}^{1-b_j} \frac{B_j}{(1-u)^2} \int_u^{1-b_i} W(w)\, dw\, du \\
&= \left[\frac{a_j}{1-a_j} A_j + \frac{a_i}{1-a_i} B_j\right] \int_{a_i}^{1-b_i} W(w)\, dw + \left[B_j\right] \int_{a_i}^{1-b_j} \int_u^{1-b_i} W_*(w)\, dw\, du.
\end{aligned}
$$

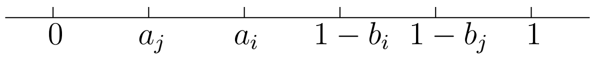
The third term is:

$$
\begin{aligned}
&= \int_0^{a_j} \frac{A_i + B_i}{(1-u)^2} \int_{a_j}^{1-b_j} V(v)\, dv\, du + \int_{a_j}^{a_i} \frac{A_i + B_i}{(1-u)^2} \int_u^{1-b_j} V(v)\, dv\, du \\
&\quad + \int_{a_i}^{1-b_j} \frac{B_i}{(1-u)^2} \int_u^{1-b_j} V(v)\, dv\, du \\
&= \frac{a_j}{1-a_j} \left[A_i + B_i\right] \int_{a_j}^{1-b_j} V(v)\, dv + \left[A_i + B_i\right] \int_{a_j}^{a_i} \int_u^{1-b_j} V_*(v)\, dv\, du \\
&\quad + \left[B_i\right] \int_{a_i}^{1-b_j} \int_u^{1-b_j} V_*(v)\, dv\, du.
\end{aligned}
$$

The fourth term is:

$$
\begin{aligned}
&= \left[A_i + B_i\right]\left[A_j + B_j\right] \int_0^{a_j} \frac{1}{(1-u)^2} du + \left[A_i + B_i\right]\left[B_j\right] \int_{a_j}^{a_i} \frac{1}{(1-u)^2} du \\
&\quad + \left[B_i\right]\left[B_j\right] \int_{a_i}^{1-b_j} \frac{1}{(1-u)^2} du \\
&= \frac{a_j}{1-a_j} \left[A_i + B_i\right]\left[A_j + B_j\right] + \frac{a_i - a_j}{(1-a_i)(1-a_j)} \left[A_i + B_i\right]\left[B_j\right] + \frac{1 - a_i - b_j}{(1-a_i) b_j} \left[B_i\right]\left[B_j\right].
\end{aligned}
$$

Case 6: $0 \le a_j \le a_i < 1 - b_i \le 1 - b_j \le 1$



The second term is:

$$= \int_0^{a_j} \frac{A_j + B_j}{(1-u)^2} \int_{a_i}^{1-b_i} W(w)\, dw\, du + \int_{a_j}^{a_i} \frac{B_j}{(1-u)^2} \int_{a_i}^{1-b_i} W(w)\, dw\, du$$

$$+ \int_{a_i}^{1-b_i} \frac{B_j}{(1-u)^2} \int_u^{1-b_i} W(w)\, dw\, du$$

$$= \left[\frac{a_i}{1-a_j} A_j + \frac{a_i}{1-a_i} B_j \right] \int_{a_i}^{1-b_i} W(w)\, dw + \big[B_j \big] \int_{a_i}^{1-b_i} \int_u^{1-b_i} W_*(w)\, dw\, du.$$

The third term is:

$$= \int_0^{a_j} \frac{A_i + B_i}{(1-u)^2} \int_{a_j}^{1-b_j} V(v)\, dv\, du + \int_{a_j}^{a_i} \frac{A_i + B_i}{(1-u)^2} \int_u^{1-b_j} V(v)\, dv\, du$$

$$+ \int_{a_i}^{1-b_i} \frac{B_i}{(1-u)^2} \int_u^{1-b_j} V(v)\, dv\, du$$

$$= \frac{a_j}{1-a_j} \big[A_i + B_i \big] \int_{a_j}^{1-b_j} V(v)\, dv + \big[A_i + B_i \big] \int_{a_j}^{a_i} \int_u^{1-b_j} V_*(v)\, dv\, du$$

$$+ \big[B_i \big] \int_{a_i}^{1-b_i} \int_u^{1-b_j} V_*(v)\, dv\, du.$$

The fourth term is:

$$= \big[A_i + B_i \big] \big[A_j + B_j \big] \int_0^{a_j} \frac{1}{(1-u)^2} du + \big[A_i + B_i \big] \big[B_j \big] \int_{a_j}^{a_i} \frac{1}{(1-u)^2} du$$

$$+ \big[B_i \big] \big[B_j \big] \int_{a_i}^{1-b_i} \frac{1}{(1-u)^2} du$$

$$= \frac{a_j}{1-a_j} \big[A_i + B_i \big] \big[A_j + B_j \big] + \frac{a_i - a_j}{(1-a_i)(1-a_j)} \big[A_i + B_i \big] \big[B_j \big] + \frac{1 - a_i - b_i}{(1-a_i) b_i} \big[B_i \big] \big[B_j \big].$$

Appendix B: Covariance-Variance Matrix: Location–Scale Families

We begin by defining the terms $\tilde{c}^*, D^*, E^*, G^*$, which only depend on the parameter-free function F_0, and the Winsorizing fractions a and/or b.

$$\tilde{c}_1^* = \int_a^{1-b} \int_a^{1-b} \{\min(w,v) - wv\}\, dF_0^{-1}(v)\, dF_0^{-1}(w);$$

$$\tilde{c}_2^* = \int_a^{1-b} \int_a^{1-b} \{\min(w,v) - wv\}\, F_0^{-1}(v)\, dF_0^{-1}(v)\, dF_0^{-1}(w);$$

$$\tilde{c}_3^* = \int_a^{1-b} \int_a^{1-b} \{\min(w,v) - wv\}\, F_0^{-1}(v) F_0^{-1}(w)\, dF_0^{-1}(v)\, dF_0^{-1}(w);$$

$$a(1-a)H_1^{'}(a) = a(1-a)\sigma\big[F_0^{-1}(a)\big]^{'} = \sigma\, a(1-a)\big[F_0^{-1}(a)\big]^{'} = \sigma D_{1a}^*;$$

$$b^2 H_1^{'}(1-b) = b^2\sigma\big[F_0^{-1}(1-b)\big]^{'} = \sigma\, b^2\big[F_0^{-1}(1-b)\big]^{'} = \sigma D_{1b}^*;$$

$$\begin{aligned} a(1-a)H_2^{'}(a) &= a(1-a)\Big\{2\mu\sigma\big[F_0^{-1}(a)\big]^{'} + 2\sigma^2\, F_0^{-1}(a)\big[F_0^{-1}(a)\big]^{'}\Big\} \\ &= 2\mu\sigma\, a(1-a)\big[F_0^{-1}(a)\big]^{'} + 2\sigma^2 a(1-a) F_0^{-1}(a)\big[F_0^{-1}(a)\big]^{'} \\ &= 2\mu\sigma\, D_{1a}^* + 2\sigma^2\, D_{2a}^*; \end{aligned}$$

$$\begin{aligned} b^2 H_2^{'}(1-b) &= b^2\Big\{2\mu\sigma\big[F_0^{-1}(1-b)\big]^{'} + 2\sigma^2 F_0^{-1}(1-b)\big[F_0^{-1}(1-b)\big]^{'}\Big\} \\ &= 2\mu\sigma\, b^2\big[F_0^{-1}(1-b)\big]^{'} + 2\sigma^2\, b^2 F_0^{-1}(1-b)\big[F_0^{-1}(1-b)\big]^{'} \\ &= 2\mu\sigma\, D_{1b}^* + 2\sigma^2\, D_{2b}^*; \end{aligned}$$

$$\int_a^{1-b} H_1^{'}(w)(1-w)\, dw = \sigma \int_a^{1-b} (1-w)\, dF_0^{-1}(w) = \sigma\, E_1^*;$$

$$\begin{aligned} \int_a^{1-b} H_2^{'}(v)(1-v)\, dv &= 2\mu\sigma \int_a^{1-b} (1-v)\, dF_0^{-1}(v) + 2\sigma^2 \int_a^{1-b} (1-v)\, F_0^{-1}(v)\, dF_0^{-1}(v) \\ &= 2\mu\sigma\, E_1^* + 2\sigma^2\, E_2^*; \end{aligned}$$

$$\int_a^{1-b}\int_u^{1-b}\frac{H_1'(w)(1-w)}{(1-u)^2}\,dw\,du=\sigma\int_a^{1-b}\int_u^{1-b}\frac{1-w}{(1-u)^2}\,dF_0^{-1}(w)\,du=\sigma\,G_1^*;$$

$$\int_a^{1-b}\int_u^{1-b}\frac{H_2'(v)(1-v)}{(1-u)^2}\,dw\,du=2\mu\sigma\int_a^{1-b}\int_u^{1-b}\frac{1-v}{(1-u)^2}\,dF_0^{-1}(v)\,du$$

$$+\,2\sigma^2\int_a^{1-b}\int_u^{1-b}\frac{1-v}{(1-u)^2}\,F_0^{-1}(v)\,dF_0^{-1}(v)\,du$$

$$=2\mu\sigma\,G_1^*+2\sigma^2\,G_2^*.$$

Then the entries of the covariance-variance matrix are:

$$\sigma_{11}^2=\sigma^2\tilde{c}_1^*+2\left[\frac{a}{1-a}(\sigma\,D_{1a}^*+\sigma\,D_{1b}^*)\sigma\,E_1^*+\sigma\,D_{1b}^*\,\sigma\,G_1^*\right]+\frac{a}{1-a}\big(\sigma\,D_{1a}^*+\sigma\,D_{1b}^*\big)^2+\frac{1-a-b}{(1-a)b}\,(\sigma\,D_{1b}^*)^2$$

$$=\sigma^2\left[\tilde{c}_1^*+\frac{2a}{1-a}(D_{1a}^*+D_{1b}^*)E_1^*+2D_{1b}^*\,G_1^*+\frac{a}{1-a}(D_{1a}^*+D_{1b}^*)^2+\frac{1-a-b}{(1-a)b}(D_{1b}^*)^2\right]$$

$$=\sigma^2\,l_1^*;$$

$$\sigma_{12}^2=\sigma_{21}^2=2\mu\sigma^2\,\tilde{c}_1^*+2\sigma^3\,\tilde{c}_2^*+\frac{a}{1-a}\big(2\mu\sigma\,D_{1a}^*+2\sigma^2\,D_{2a}^*+2\mu\sigma\,D_{1b}^*+2\sigma^2\,D_{2b}^*\big)\sigma\,E_1^*+\big(2\mu\sigma\,D_{1b}^*$$

$$+\,2\sigma^2\,D_{2b}^*\big)\sigma\,G_1^*+\frac{a}{1-a}\big(\sigma\,D_{1a}^*+\sigma\,D_{1b}^*\big)\big(2\mu\sigma\,E_1^*+2\sigma^2\,E_2^*\big)+\sigma\,D_{1b}^*\,\big(2\mu\sigma\,G_1^*+2\sigma^2\,G_2^*\big)$$

$$+\,\frac{a}{1-a}\big(\sigma\,D_{1a}^*+\sigma\,D_{1b}^*\big)\big(2\mu\sigma\,D_{1a}^*+2\sigma^2\,D_{2a}^*+2\mu\sigma\,D_{1b}^*+2\sigma^2\,D_{2b}^*\big)$$

$$+\,\frac{1-a-b}{(1-a)b}\,\big(\sigma\,D_{1b}^*\big)\big(2\mu\sigma\,D_{1b}^*+2\sigma^2\,D_{2b}^*\big)$$

$$=2\mu\sigma^2\left[\tilde{c}_1^*+\frac{2a}{1-a}(D_{1a}^*+D_{1b}^*)E_1^*+2D_{1b}^*\,G_1^*+\frac{a}{1-a}(D_{1a}^*+D_{1b}^*)^2+\frac{1-a-b}{(1-a)b}(D_{1b}^*)^2\right]$$

$$+\,2\sigma^3\Big\{\tilde{c}_2^*+\frac{a}{1-a}\big[\big(D_{1a}^*+D_{1b}^*\big)E_2^*+\big(D_{2a}^*+D_{2b}^*\big)E_1^*\big]+\big(D_{1b}^*\,G_2^*+D_{2b}^*\,G_1^*\big)$$

$$+\frac{a}{1-a}\big(D^*_{1a}+D^*_{1b}\big)\big(D^*_{2a}+D^*_{2b}\big)+\frac{1-a-b}{(1-a)b}D^*_{1b}D^*_{2b}\Big\}$$

$$=2\mu\sigma^2\, l^*_1+2\sigma^3\, l^*_2;$$

$$\sigma^2_{22}=4\mu^2\sigma^2\,\tilde{c}^*_1+8\mu\sigma^3\,\tilde{c}^*_2+4\sigma^4\,\tilde{c}^*_3+\frac{2a}{1-a}\big(2\mu\sigma\, D^*_{1a}+2\sigma^2\, D^*_{2a}+2\mu\sigma\, D^*_{1b}+2\sigma^2\, D^*_{2b}\big)\big(2\mu\sigma\, E^*_1+2\sigma^2\, E^*_2\big)$$

$$+2\big(2\mu\sigma\, D^*_{1b}+2\sigma^2\, D^*_{2b}\big)\big(2\mu\sigma\, G^*_1+2\sigma^2\, G^*_2\big)+\frac{a}{1-a}\big(2\mu\sigma\, D^*_{1a}+2\sigma^2\, D^*_{2a}+2\mu\sigma\, D^*_{1b}+2\sigma^2\, D^*_{2b}\big)^2$$

$$+\frac{1-a-b}{(1-a)b}\,\big(2\mu\sigma\, D^*_{1b}+2\sigma^2\, D^*_{2b}\big)^2$$

$$=4\mu^2\sigma^2\Bigg[\tilde{c}^*_1+\frac{2a}{1-a}(D^*_{1a}+D^*_{1b})E^*_1+2D^*_{1b}\,G^*_1+\frac{a}{1-a}(D^*_{1a}+D^*_{1b})^2+\frac{1-a-b}{(1-a)b}(D^*_{1b})^2\Bigg]$$

$$+8\mu\sigma^3\Bigg\{\tilde{c}^*_2+\frac{a}{1-a}\big[\big(D^*_{1a}+D^*_{1b}\big)E^*_2+\big(D^*_{2a}+D^*_{2b}\big)E^*_1\big]+\big(D^*_{1b}\,G^*_2+D^*_{2b}\,G^*_1\big)$$

$$+\frac{a}{1-a}\big(D^*_{1a}+D^*_{1b}\big)\big(D^*_{2a}+D^*_{2b}\big)+\frac{1-a-b}{(1-a)b}D^*_{1b}D^*_{2b}\Bigg\}+4\sigma^4\Bigg[\tilde{c}^*_3+\frac{2a}{1-a}(D^*_{2a}+D^*_{2b})E^*_2$$

$$+2D^*_{2b}\,G^*_2+\frac{a}{1-a}(D^*_{2a}+D^*_{2b})^2+\frac{1-a-b}{(1-a)b}(D^*_{2b})^2\Bigg]$$

$$=4\mu^2\sigma^2\, l^*_1+8\mu\sigma^3\, l^*_2+4\sigma^4\, l^*_3.$$

Appendix C: Computer Code: ARE for Exponential Distribution

```
# Pareto ARE— MWM approach .R file
Calculate ARE for one−parameter distribution
library(cubature)
f <− function(x) { (min(x[1],x[2])−x[1]*x[2])/(1−x[1])/(1−x[2]) }
g <− function(x) { log(1−x) } # Io(a,1−b)
# "x" is vector J((a,1−b),(a,1−b))
a<−c(0.0000001,0.05,0.1,0.15,0.25,0.49,0.7,0.85)
b<−c(0.0000001,0.05,0.1,0.15,0.25,0.49,0.7,0.85)
ARE <− array(0, dim=c(length(a),length(b)))
mu <− array(0, dim=c(length(a),length(b)))
sigma2 <− array(0, dim=c(length(a),length(b)))
for (i in 1:length(a))
{
for (j in 1:length(b))
{
if( a[i]+b[j]>=1)break
I0<− integrate(g,lower=c(a[i]), upper = c(1−b[j]))
J0<−adaptIntegrate(f, lowerLimit = c(a[i], a[i]),upperLimit=c(1−b[j], 1−b[j]))
mu[i,j]=I0[[1]]+a[i]*log(1−a[i])+b[j]*log(b[j])
sigma2[i,j]=J0[[1]]+2*(1−a[i]−b[j])/(1−a[i])*(a[i]^2+a[i]*b[j]−b[j])−
2*b[j]*log(b[j]/(1−a[i]))+a[i]^3/(1−a[i])+2*a[i]^2*b[j]^2/(1−a[i])/b[j]
+b[j]*(1−b[j])
ARE[i,j]<−mu[i,j]^2/sigma2[i,j]
}
}
ARE
```

Appendix D: Computer Code: ARE for Laplace Distribution

```
Calculate ARE— MWM approach for two−parameter distributions
%Laplace ARE— MWM , Matlab File
%###############################################################################
% For Logistic , we change the functions f, g, h, A, C and D, they depend on
the distribution. We also need to change Fisher matrix, others could keep same.
%###############################################################################
f1=@(x,y) (0.5+min(x,y)−(x+0.5).∗(y+0.5))./(1−2.∗abs(x))./(1−2.∗abs(y));
f2=@(x,y) −(0.5+min(x,y)−(x+0.5).∗(y+0.5))./(1−2.∗abs(x))./(1−2.∗abs(y))
.∗sign(x).∗log(1−2.∗abs(x));
f3=@(x,y) (0.5+min(x,y)−(x+0.5).∗(y+0.5))./(1−2.∗abs(x))./(1−2.∗abs(y))
.∗sign(x).∗log(1−2.∗abs(x)).∗sign(y).∗log(1−2.∗abs(y));
g1=@(x,y) −sign(x).∗log(1−2.∗abs(x));
g2=@(x,y) (sign(x).∗log(1−2.∗abs(x))).^2;
g3=@(x,y) 2.∗(1−x)./(1−2.∗abs(x−0.5));
g4=@(x,y) −2.∗(1−x)./(1−2.∗abs(x−0.5)).∗sign(x−0.5).∗log(1−2.∗abs(x−0.5));
h1=@(x,y) (x>y).∗2.∗(1−x)./(1−2.∗abs(x−0.5))./(1−y).^2;
h2=@(x,y) −2.∗(x>y).∗(1−x)./(1−2.∗abs(x−0.5)).∗sign(x−0.5).∗log(1−2.∗abs(x−0.5))
./(1−y).^2;
a=[0.000001,0.05,0.1,0.15,0.25,0.49,0.7,0.85];
b=[0.000001,0.05,0.1,0.15,0.25,0.49,0.7,0.85];
Mu_1=zeros(length(a),length(b));
Mu_2=zeros(length(a),length(b));
ARE=zeros(length(a),length(b));
Theta=zeros(length(a),length(b));
Sigma=zeros(length(a),length(b));
I=[1, 0; 0, 1]; %Fisher 1/sigma^2∗[1, 0; 0, 1]
%##########################################################################
For Logistic I=[1/3, 0; 0, (3+pi^2)/9]; %Fisher 1/sigma^2∗[1/3, 0; 0; (3+pi^2)/9]
%##########################################################################
Fisher=inv(I);
for i=1:length(a)
for j=1:length(b)
if a(i)+b(j)>=1,
break;
```

```
end
A1=4.*dblquad(f1,a(i)-0.5,1-b(j)-0.5,a(i)-0.5,1-b(j)-0.5);
A2=4.*dblquad(f2,a(i)-0.5,1-b(j)-0.5,a(i)-0.5,1-b(j)-0.5);
A3=4.*dblquad(f3,a(i)-0.5,1-b(j)-0.5,a(i)-0.5,1-b(j)-0.5);
C1=quad(g1,a(i)-0.5,1-b(j)-0.5)-a(i).*sign(a(i)-0.5).*log(1-2.*abs(a(i)-0.5))
-b(j).*sign(1-b(j)-0.5).*log(1-2.*abs(1-b(j)-0.5));
C2=quad(g2,a(i)-0.5,1-b(j)-0.5)+a(i).*(sign(a(i)-0.5).*log(1-2.*abs(a(i)-0.5))).^2
+b(j).*(sign(1-b(j)-0.5).*log(1-2.*abs(1-b(j)-0.5))).^2;
D_1a=2.*a(i).*(1-a(i))./(1-2.*abs(a(i)-0.5));
D_1b=2.*(b(j)).^2./(1-2.*abs(1-b(j)-0.5));
D_2a=-2.*a(i).*(1-a(i))./(1-2.*abs(a(i)-0.5)).*sign(a(i)-0.5).*log(1-2.*abs(a(i)-0.5));
D_2b=-2.*(b(j)).^2./(1-2.*abs(1-b(j)-0.5)).*sign(1-b(j)-0.5).*log(1-2.*abs(1-b(j)-0.5));
E1=quad(g3,a(i),1-b(j));
E2=quad(g4,a(i),1-b(j));
F1=dblquad(h1,a(i),1-b(j),a(i),1-b(j));
F2=dblquad(h2,a(i),1-b(j),a(i),1-b(j));
C1_star=A1+2.*a(i)./(1-a(i)).*(D_1a+D_1b).*E1+2.*D_1b.*F1+a(i)./(1-a(i))
.*(D_1a+D_1b).^2+(1-a(i)-b(j))./(1-a(i))./b(j).*(D_1b).^2;
C2_star=A2+a(i)./(1-a(i)).*((D_1a+D_1b).*E2+(D_2a+D_2b).*E1)+(D_1b.*F2
+D_2b.*F1)+a(i)./(1-a(i)).*(D_1a+D_1b).*(D_2a+D_2b)+(1-a(i)-b(j))./(1-a(i))
./b(j).*D_1b.*D_2b;
C3_star=A3+2.*a(i)./(1-a(i)).*(D_2a+D_2b).*E2+2.*D_2b.*F2+a(i)./(1-a(i))
.*(D_2a+D_2b).^2+(1-a(i)-b(j))./(1-a(i))./b(j).*(D_2b).^2;
s1(i,j)=C1_star.*C2.^2-2.*C1.*C2.*C2_star+C1.^2.*C3_star;
s2(i,j)=-C1_star.*C1.*C2+C2.*C2_star+C1.^2.*C2_star-C1.*C3_star;
s3=s2;
s4(i,j)=C1_star.*C1.^2-2.*C1.*C2_star+C3_star;
D(i,j)=(C2-C1.^2).^2;
S=1./D(i,j).*[s1(i,j),s2(i,j);s3(i,j),s4(i,j)];
ARE(i,j)=(det(Fisher)/det(S)).^0.5;
end
end
ARE
```

Appendix E: Computer Code: Simulation Study

```
% Logistic simulation- for both M T M and MWM Matlab file
%#######################################################################
For Laplace simulation, we change g, C_Win, they depend on the distribution,
and we also change the fisher matrix
%#######################################################################
clc;clear;
x0=1; Theta0=5; Sigma0=2;
g1=@(x,y) -log(1./x-1) ;
g2=@(x,y) (log(1./x-1)).^2;
N_repeat=10; N_sample=10000; N_size=50;
a=[0.0000001,0.05,0.1,0.25,0.49,0.1,0.25];
b=[0.0000001,0.05,0.1,0.25,0.49,0.7,0.0000001];
Mean_Sigma=zeros(N_repeat,length(a)); Mean_Theta=zeros(N_repeat,length(a));
Mean_RE=zeros(N_repeat,length(a));
Mean_Sigma_Win=zeros(N_repeat,length(a)); Mean_Theta_Win=zeros(N_repeat,length(a));
Mean_RE_Win=zeros(N_repeat,length(a));
RE1=zeros(N_repeat,length(a)); RE_Win=zeros(N_repeat,length(a));
I=1./Sigma0.^2.*[1/3, 0; 0, (3+pi^2)/9]; %Fisher 1/sigma^2*[1/3, 0; 0; (3+pi^2)/9]
Fisher=inv(I)./N_size; %
for N=1:N_repeat
N
Mu_1=zeros(length(a),N_sample); Mu_2=zeros(length(a),N_sample);
Mu_1_Win=zeros(length(a),N_sample); Mu_2_Win=zeros(length(a),N_sample);
Theta_MTM=zeros(length(a),N_sample); Sigma_MTM=zeros(length(a),N_sample);
Theta_MTM_Win=zeros(length(a),N_sample); Sigma_MTM_Win=zeros(length(a),N_sample);
for j=1:N_sample
U=rand(N_size,1);
x=Theta0-Sigma0*log(1./U-1);
#######################################################################
Laplace x=Theta0-Sigma0*sign(U-0.5).*log(1-2.*abs(U-0.5));
#######################################################################
X=sort(x);
for i=1:length(a)
mu_1=0;
```

```
mu_2=0;
C1=1./(1-a(i)-b(i)).*quad(g1,a(i),1-b(i));
C2=1./(1-a(i)-b(i)).*quad(g2,a(i),1-b(i));
C1_Win=C1.*(1-a(i)-b(i))-a(i).*log(1./a(i)-1)
-b(i).*log(1./(1-b(i))-1);
C2_Win=C2.*(1-a(i)-b(i))+a(i).*(log(1./a(i)-1).^2)
+b(i).*(log(1./(1-b(i))-1).^2);
for n=floor(a(i).*N_size)+1:1:ceil(N_size-b(i).*N_size)
mu_1=mu_1+X(n);
mu_2=mu_2+(X(n)).^2;
end
Mu_1(i,j)=mu_1./(ceil(N_size-b(i).*N_size)-floor(a(i).*N_size));
Mu_2(i,j)=mu_2./(ceil(N_size-b(i).*N_size)-floor(a(i).*N_size));
Mu_1_Win(i,j)=1./N_size.*(mu_1+X(floor(a(i).*N_size)+1)
.*floor(a(i).*N_size)+X(ceil(N_size-b(i).*N_size)).*floor(b(i).*N_size));
Mu_2_Win(i,j)=1./N_size.*(mu_2+X(floor(a(i).*N_size)+1).^2.*floor(a(i)
.*N_size)+X(ceil(N_size-b(i).*N_size)).^2.*floor(b(i).*N_size));
Sigma_MTM(i,j)=sqrt((Mu_2(i,j)-Mu_1(i,j).^2)./(C2-C1.^2));
Theta_MTM(i,j)=Mu_1(i,j)-C1.*Sigma_MTM(i,j);
Sigma_MTM_Win(i,j)=sqrt((Mu_2_Win(i,j)-Mu_1_Win(i,j).^2)./(C2_Win-C1_Win.^2));
Theta_MTM_Win(i,j)=(Mu_1_Win(i,j)-C1_Win.*Sigma_MTM_Win(i,j));
end
end
MSE_Theta=zeros(length(a),1); MSE_Sigma=zeros(length(a),1);
COV_T_S=zeros(length(a),1); MSE_Theta_Win=zeros(length(a),1);
MSE_Sigma_Win=zeros(length(a),1); COV_T_S_Win=zeros(length(a),1);
for i=1:length(a)
for j=1:N_sample
MSE_Sigma(i)=MSE_Sigma(i)+(Sigma_MTM(i,j)-Sigma0).^2;
%Mean squared error MTM
MSE_Theta(i)=MSE_Theta(i)+(Theta_MTM(i,j)-Theta0).^2;
COV_T_S(i)=COV_T_S(i)+(Theta_MTM(i,j)-Theta0).*(Sigma_MTM(i,j)-Sigma0);
MSE_Sigma_Win(i)=MSE_Sigma_Win(i)+(Sigma_MTM_Win(i,j)-Sigma0).^2;
%Mean squared error MTM
MSE_Theta_Win(i)=MSE_Theta_Win(i)+(Theta_MTM_Win(i,j)-Theta0).^2;
```

```
COV_T_S_Win(i)=COV_T_S_Win(i)+(Theta_MTM_Win(i,j)
-Theta0).*(Sigma_MTM_Win(i,j)-Sigma0);
end
S_MTM=MSE_Theta(i).*MSE_Sigma(i)-COV_T_S(i).^2;
% determinant of covariance matrix_MTM
S_MTM=vpa(S_MTM,30);
S_MTM_Win=MSE_Theta_Win(i).*MSE_Sigma_Win(i)-COV_T_S_Win(i).^2;
% determinant of covariance matrix_MTM
S_MTM_Win=vpa(S_MTM_Win,30);
RE1(N,i)=(N_sample.^2.*det(Fisher)./S_MTM).^0.5 ; % RE
RE_Win(N,i)=(N_sample.^2.*det(Fisher)./S_MTM_Win).^0.5 ; % RE
end
Mean_Sigma(N,:)= mean(Sigma_MTM')./Sigma0;
Mean_Theta(N,:)= mean(Theta_MTM')./Theta0;
Mean_Sigma_Win(N,:)= mean(Sigma_MTM_Win')./Sigma0;
Mean_Theta_Win(N,:)= mean(Theta_MTM_Win')./Theta0;
end

MEAN_final_Theta=mean(Mean_Theta) %mean of the 10 repeats
MEAN_final_Sigma=mean(Mean_Sigma)
Var_Theta=std(Mean_Theta)./sqrt(N_repeat)
Var_Sigma=std(Mean_Sigma)./sqrt(N_repeat)
RE1_final=mean(RE1)
Var_R1=std(RE1)./sqrt(N_repeat)

MEAN_final_Theta_Win=mean(Mean_Theta_Win)
MEAN_final_Sigma_Win=mean(Mean_Sigma_Win)
Var_Theta_Win=std(Mean_Theta_Win)./sqrt(N_repeat)
Var_Sigma_Win=std(Mean_Sigma_Win)./sqrt(N_repeat)
RE_Win_final=mean(RE_Win)
Var_R_Win=std(RE_Win)./sqrt(N_repeat)
```

Appendix F: Computer Code: Real Data Illustrations

```
% Winsorized lognormal-real data model .Matlab file
%#########################################################################
For Logistic and Laplace Real Data Illustrations , we change f, g, h, C ,D,
they depend on the distribution , others keep same.
%#########################################################################
clc;clear;
clc; clear;
Data0=[72.3030,33.0940,26.6190,22.6020,16.8640,16.6290,13.7950,12.4340,12.0480,10.9650,
10.7050,10.2320,9.3800,9.0660,8.3080,7.0690,7.0390,6.5360,6.3130,6.2930,5.8380,5.3680,
4.0560,3.3380,3.1080,3.0000,2.4350,2.3990,2.3960,2.266];
Data=sort(Data0); N=length(Data);
x0=0; X=log(Data-x0);

f1=@(x,y) (min(x,y)-x.*y).*exp(.5.*(norminv(x,0,1)).^2+.5.*(norminv(y,0,1)).^2);
f2=@(x,y) (min(x,y)-x.*y).*exp(.5.*(norminv(x,0,1)).^2
+.5.*(norminv(y,0,1)).^2).*norminv(x,0,1);
f3=@(x,y) (min(x,y)-x.*y).*exp(.5.*(norminv(x,0,1)).^2
+.5.*(norminv(y,0,1)).^2).*norminv(x,0,1).*norminv(y,0,1);
g1=@(x,y) norminv(x,0,1);
g2=@(x,y) (norminv(x,0,1)).^2;
g3=@(x,y) (1-x).*exp(0.5.*(norminv(x,0,1)).^2);
g4=@(x,y) (1-x).*norminv(x,0,1).*exp(0.5.*(norminv(x,0,1)).^2);
h1=@(x,y) (x>y).*(1-x).*exp(0.5.*(norminv(x,0,1)).^2)./(1-y).^2;
h2=@(x,y) (x>y).*(1-x).*norminv(x,0,1).*exp(0.5.*(norminv(x,0,1)).^2)./(1-y).^2;
a=[0.0000001,14/30, 1/30, 8/30];
b=[0.0000001,14/30, 1/30, 3/30];
A=length(a);
Data_Sim=zeros(A,N);
ARE=zeros(A,1);
x0=0;
Mu_1=zeros(A,1); Mu_2=zeros(A,1);
Theta=zeros(A,1); Sigma=zeros(A,1);
Fit=zeros(A,1); Fit_Restrit=zeros(A,1);
F_25=zeros(A,1);
```

```
Premium=zeros(A,1);
G_prime_Theta=zeros(A,1); G_prime_Sigma=zeros(A,1);
CI_Low=zeros(A,1); CI_Up=zeros(A,1);

for i=1:A
A1=2*pi.*dblquad(f1,a(i),1-b(i),a(i),1-b(i));
A2=2*pi.*dblquad(f2,a(i),1-b(i),a(i),1-b(i));
A3=2*pi.*dblquad(f3,a(i),1-b(i),a(i),1-b(i));
C1=quad(g1,a(i),1-b(i))+a(i).*norminv(a(i))+b(i).*norminv(1-b(i));
C2=quad(g2,a(i),1-b(i))+a(i).*(norminv(a(i))).^2+b(i).*(norminv(1-b(i))).^2;
D_1a=sqrt(2*pi).*a(i).*(1-a(i)).*exp(0.5.*(norminv(a(i))).^2);
D_1b=sqrt(2*pi).*(b(i)).^2.*exp(0.5.*(norminv(1-b(i))).^2);
D_2a=sqrt(2*pi).*a(i).*(1-a(i)).*norminv(a(i)).*exp(0.5.*(norminv(a(i))).^2);
D_2b=sqrt(2*pi).*(b(i)).^2.*norminv(1-b(i)).*exp(0.5.*(norminv(1-b(i))).^2);
E1=sqrt(2*pi).*quad(g3,a(i),1-b(i));
E2=sqrt(2*pi).*quad(g4,a(i),1-b(i));
F1=sqrt(2*pi).*dblquad(h1,a(i),1-b(i),a(i),1-b(i));
F2=sqrt(2*pi).*dblquad(h2,a(i),1-b(i),a(i),1-b(i));
C1_star=A1+2.*a(i)./(1-a(i)).*(D_1a+D_1b).*E1+2.*D_1b.*F1
+a(i)./(1-a(i)).*(D_1a+D_1b).^2+(1-a(i)-b(i))./(1-a(i))./b(i).*(D_1b).^2;
C2_star=A2+a(i)./(1-a(i)).*((D_1a+D_1b).*E2+(D_2a+D_2b).*E1)
+(D_1b.*F2+D_2b.*F1)+a(i)./(1-a(i)).*(D_1a+D_1b).*(D_2a+D_2b)
+(1-a(i)-b(i))./(1-a(i))./b(i).*D_1b.*D_2b;
C3_star=A3+2.*a(i)./(1-a(i)).*(D_2a+D_2b).*E2+2.*D_2b.*F2
+a(i)./(1-a(i)).*(D_2a+D_2b).^2+(1-a(i)-b(i))./(1-a(i))./b(i).*(D_2b).^2;

s1(i)=C1_star.*C2.^2-2.*C1.*C2.*C2_star+C1.^2.*C3_star;
s2(i)=-C1_star.*C1.*C2+C2.*C2_star+C1.^2.*C2_star-C1.*C3_star;
s3(i)=s2(i);
s4(i)=C1_star.*C1.^2-2.*C1.*C2_star+C3_star; % asymptotic variance matrix
D(i)=(C2-C1.^2).^2;
S=1./D(i).*[s1(i),s2(i);s3(i),s4(i)];
ARE(i)=(0.5./det(S)).^0.5; % ARE
mu_1=0; mu_2=0;
for n=floor(a(i).*N)+1:1:ceil(N-b(i).*N) % Mu
```

```
mu_1=mu_1+log(Data(n)-x0);
mu_2=mu_2+log(Data(n)-x0).^2;
end
Mu_1(i)=1./N.*(mu_1+X(floor(a(i).*N)+1).*floor(a(i).*N)+X(ceil(N-b(i).*N))
.*floor(b(i).*N));
Mu_2(i)=1./N.*(mu_2+X(floor(a(i).*N)+1).^2.*floor(a(i).*N)
+X(ceil(N-b(i).*N)).^2.*floor(b(i).*N));
Sigma(i)=sqrt((Mu_2(i)-Mu_1(i).^2)./(C2-C1.^2));
% Parameter estimation--Sigma and Theta
Theta(i)=Mu_1(i)-C1.*Sigma(i)

h11=@(x,y) (x-5).*lognpdf(x,Theta(i),Sigma(i));
F_25(i)=normcdf((log(25)-Theta(i))./Sigma(i));
Premium(i)= quad(h11,5,25)+20.*(1-F_25(i)); %Premium
%-----------original form
%h12=@(x,y) (x-5).*lognpdf(x,Theta(i),Sigma(i)).*((log(x)-Theta(i))./Sigma(i).^2);
%G_prime_Theta(i)= quad(h12,5,25)+20./Sigma(i).*normpdf((log(25)-Theta(i))./Sigma(i));
%gradient of Theta
%h13=@(x,y)-(x-5)./Sigma(i).*lognpdf(x,Theta(i),Sigma(i)).*(1-(log(x)-Theta(i)).^2
./Sigma(i).^2);
%G_prime_Sigma(i)= quad(h13,5,25)+20./Sigma(i).^2.*normpdf((log(25)
-Theta(i))./Sigma(i)).*(log(25)-Theta(i)); % gradient of Sigma
%-----------smiplified form
h14=@(x,y) normpdf((log(x)-Theta(i))./Sigma(i),0,1);
G_prime_Theta(i)= quad(h14,5,25)./Sigma(i); % gradient of Theta
h15=@(x,y) normpdf((log(x)-Theta(i))./Sigma(i),0,1).*(log(x)-Theta(i));
G_prime_Sigma(i)= quad(h15,5,25)./Sigma(i).^2; % gradient of Sigma

Cov=Sigma(i).^2.*S./N ; % covariance matrix
Var_F=[G_prime_Theta(i),G_prime_Sigma(i)]*[Cov]*[G_prime_Theta(i);G_prime_Sigma(i)];
% Delta method, variance
CI_Low(i)=Premium(i)-1.96*sqrt(Var_F); % confidence interval
CI_Up(i)=Premium(i)+1.96*sqrt(Var_F);
for j=1:N
XX(j)=norminv((j-0.5)./N);
```

```
Data_Sim(i,j)=Theta(i)+Sigma(i).*XX(j);
Fit(i)=Fit(i)+1./N.*abs(Data_Sim(i,j)-log(Data(j)));
end
for k=9:27
Fit_Restrit(i)=Fit_Restrit(i)+1/19.*abs(Data_Sim(i,k)-log(Data(k)));
end
end

Sigma
Theta
Fit
Fit_Restrit
ARE
Premium
CI_Low
CI_Up

figure(2)
scatter(XX,log(Data),'*'); hold on;
line(XX,Data_Sim(1,:),'Color','k','LineWidth',2); hold on;
line(XX,Data_Sim(2,:),'Color','g','LineStyle','-.','LineWidth',2); hold on;
line(XX,Data_Sim(3,:),'Color','r','LineStyle','-.','LineWidth',2); hold on;
line(XX,Data_Sim(4,:),'Color','c','LineStyle','-.','LineWidth',2); hold on;
Line1=log(5).*ones(length(XX),1);
line(XX,Line1,'Color','k','LineStyle','-.')
text(-2,1.9,'log(5)'); hold on;
Line2=log(25).*ones(length(XX),1);
line(XX,Line2,'Color','k','LineStyle','-.')
text(-2,3.5,'log(25)'); hold on;
legend('Observed','MLE','T1','T2','T3',4)
gtext('MLE')
gtext('T1')
gtext('T2')
gtext('T3')
title('Original data-Winsorized');
```

```
xlabel('Standard Normal Quantitles');
ylabel('Log(Observed Data)');

C=0; C1=0; Cu=0; Vu=0;
for n=1:length(Data) % Emperical distribution
if Data(n)<=25
C=C+1;
if 5<=Data(n)
Cu=Cu+(Data(n)-5)./30;
Vu=Vu+(Data(n)-5).^2./30;
C1=C1+1;
end
end
end
Premium_Empirical=Cu+20*(1-C/30)
Var_Empirical=(Vu+400*(1-C/30)-(Premium_Empirical).^2)/30;
Empirical_CI_Low=Premium_Empirical-1.96*sqrt(Var__Empirical)
Empirical_CI_Up=Premium_Empirical+1.96*sqrt(Var__Empirical)
```

CURRICULUM VITAE

Qian Zhao

Place of birth: Dalian, Liaoning, China

Education

- PhD., University of Wisconsin-Milwaukee, May 2017

  Major: Mathematics (concentrations in Statistics and Actuarial Science)

  Minor: Economics

- M.S., China University of Geosciences-Beijing, June 2011

  Major: Music Education

- B.S., China University of Geosciences-Beijing, July 2008

  Major: Applied Mathematics

Dissertation Title: Robust and Computationally Efficient Methods for Fitting Loss Models and Pricing Insurance Risks. Advisors: Professors V. Brazauskas and J. Ghorai.

Actuatial Education

Society of Actuaries, Schaumburg, Illinois

Professional Examinations Passed: P/1, FM/2, MFE/3, C/4, MLC.

VEE Credits Earned: Applied Statistics, Corporate Finance, and Economics.

Experience

- Graduate Teaching Assistant, University of Wisconsin-Milwaukee, Wisconsin, 2011-2017.

  Full instructor responsibility, teaching in areas of SOA/CAS Exam FM/2 Preparation Course, Elementary Statistical Analysis, Intermediate Algebra and Beginning Algebra.

- Data Scientist Intern, Biogen, Cambridge, Massachusetts, 2015.

- Building Information Model Estimating Development Intern, William Ryan Homes, Brookfield, Wisconsin, 2014.

Professional Society Memberships

American Statistical Association (ASA), 2016-present.

American Mathematical Society (AMS), 2011-present.

Publicaitons

- Zhao, Q., Dunn, K.J., Liu, X.W., 2011. A simulation study of formation permeability as a function of methane hydrate concentration. *Applied Geophysics*, 24(1): 58-162.

- Huang, G.D., Zhao, R.J., Zhao, Q., 2007. The improved Gray-Markov model for forecasting house price in Beijing. *Chinese Journal of Engineering Mathematics*, 8(2): 101-109.

Working Papers

- Zhao, Q., Brazauskas, V., Ghorai, J., 2017. Robust and efficient fitting of severity models and the method of Winsorized moments. *Submitted for publication.*
- Zhao, Q., Brazauskas, V., Ghorai, J., 2017. Small-sample performance of robust estimators for the parameters of log-location-scale families. *Submitted for publication.*

Professional Activiteis and Presentations

"Robust and Computationally Efficient Methods for Fitting Log-Location-Scale Distributions, with Applications to Risk Pricing". *51st Actuarial Research Conference*, University of Minnesota & University of St.Thomas, Twin Cities, Minnesota, 2016.

Referee for:

Communications in Statistics–Theory and Methods.

Journal of Statistical Distributions and Applications.

Honors and Awards

- Research Excellence Award, *Department of Mathematics Sciences*, University of Wisconsin-Milwaukee, Wisconsin, 2016-2017.
- Actuarial Research Conference Travel Grant, *Society of Actuaries, Schaumburg*, Illinois, 2016.
- Graduate Student Travel Award, *University of Wisconsin-Milwaukee*, Wisconsin, 2016.
- Chancellor's Graduate Student Award, *University of Wisconsin-Milwaukee*, Wisconsin, 2011-2016.
- First-place (Beijing), China Undergraduate Mathematical Contest in Modeling, *China Society for Industrial and Applied Mathematics*, 2007.

Technical Skills

Analytics: Statistical Analysis, Machine Learning, Data Mining, Multivariate Modeling.

Computer Programming: R, Matlab, Minitab, LaTex, SQL, JMP, SAS, C++, Python.