Robust and Efficient Fitting of Severity Models and the Method of Winsorized Moments

Qian Zhao[1]
Robert Morris University

Vytaras Brazauskas[2]
University of Wisconsin-Milwaukee

Jugal Ghorai[3]
University of Wisconsin-Milwaukee

To appear in *ASTIN Bulletin*
(*Submitted*: October 4, 2016 *Revised*: April 20, 2017; July 21, 2017 *Accepted*: August 2, 2017)

Abstract. Continuous parametric distributions are useful tools for modeling and pricing insurance risks, measuring income inequality in economics, investigating reliability of engineering systems, and in many other areas of application. In this paper, we propose and develop a new method for estimation of their parameters—the method of Winsorized moments (MWM)—which is conceptually similar to the method of trimmed moments (MTM) and thus is robust and computationally efficient. Both approaches yield explicit formulas of parameter estimators for log-location-scale families and their variants, which are commonly used to model claim severity. Large-sample properties of the new estimators are provided and corroborated through simulations. Their performance is also compared to that of MTM and the maximum likelihood estimators (MLE). In addition, 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 layer are computed when the lognormal and log-logistic models are fitted to the data using the MWM, MTM, and MLE methods.

Keywords. Claim Severity; Efficiency; Insurance Layer; Robustness; Winsorized Data.

---

[1] Qian Zhao, Ph.D., is an Assistant Professor in the Department of Mathematics, Robert Morris University, Moon Township, PA 15108, USA. *e-mail*: `zhao@rmu.edu` Much of this work was completed while the author was a Ph.D. student in the Department of Mathematical Sciences at the University of Wisconsin-Milwaukee.

[2] Corresponding author: Vytaras Brazauskas, Ph.D., ASA, is a Professor in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA. *e-mail*: `vytaras@uwm.edu`

[3] Jugal Ghorai, Ph.D., is a Professor in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA. *e-mail*: `jugal@uwm.edu`

1 Introduction

The insurance industry relies on models for claim frequency and severity, aggregate loss, payments, reserves, and other variables. In practice, all models are inevitably dependent upon simplifying assumptions, imperfect parameter estimates and other inputs, which create 'model risk'. Model risk can be defined as the risk that a company incurs because its models are misspecified or because some of the assumptions underlying those models are not met in practice (see McNeil *et al.*, 2005). Taking claim severity, for instance, we might work with a lognormal distribution to model losses whereas the true underlying distribution is heavy-tailed. Hence being aware of such a risk, practitioners spend significant amount of their time stress-testing assumptions, constructing several competing models for the same problem, and performing sensitivity studies.

The field of statistics that is concerned with *model mis-specification* and *data quality* (e.g., measurement errors, outliers, typos) is called robust statistics. In this area, the main focus is on parametric models, their fitting to the observed data, and identification of outliers. Fitting of the model is accomplished by employing procedures that are designed to have limited sensitivity to changes in the underlying assumptions as well as to "unexpected" data points (for illustrations, see Section 5). The procedures that possess such properties are called *robust*. This suggests that problems involving model risk are related to, and can be (in part) managed with, robust statistical methods.

Initial formal studies on robust statistical methods have appeared in the statistical literature in the mid-1960's (for the main trends, theory, and techniques of this field, see Huber and Ronchetti, 2009). In the financial literature, the concepts of model risk and robust statistics have been discussed by Cont (2006) and Dell'Aquila and Embrechts (2006). In economics, Hansen and Sargent (2008) provide an extensive account on robust macroeconomic models. In actuarial science, robustness studies have been carried out by Künsch (1992), Gisler and Reinhard (1993), Brazauskas and Serfling (2000, 2003), Marceau and Rioux (2001), Serfling (2002), Dornheim and Brazauskas (2007), and others. Further, for measuring insurance risks, contract specifications define data layers that play essential role in pricing and thus have to be taken into consideration when estimating parameters of the model. Such reasoning lead to introduction of the *method of trimmed moments* (MTM) – a parameter estimation method which is robust, computationally efficient, and can be easily adapted to insurance contract specifications (see Brazauskas *et al*., 2009, and Brazauskas, 2009).

The MTM procedure works like the standard method-of-moments but instead of moments uses trimmed moments (which are always finite). Fully worked out examples of such estimators are available for location-scale families, log-folded-normal, -Cauchy, and -Student's t distributions (with known degrees of freedom), as well as exponential, Pareto I, generalized Pareto, and gamma distributions (see Brazauskas and Kleefeld, 2009, 2011, and Kleefeld and Brazauskas, 2012). MTMs have also attracted attention from applied researchers: Horbenko *et al*. (2011), Opdyke and Cavallo (2012), and Chau (2013) have discussed its feasibility in operational risk modeling and Kim and Jeon (2013) have used this approach in credibility studies. Nonetheless, one often mentioned drawback of MTM is that it discards all information contained in outlying data points even though they might be legitimate observations from the actual assumed loss model.

To address this shortcoming and thus improve MTM's efficiency, in this article we introduce a new method which will be referred to as the *method of Winsorized moments* (MWM). In this method, instead of discarding the extreme observations we replace them by a few non-extreme sample order statistics. This approach is known as data Winsorization (hence the name of the method) and it works as described in the following example.

Example 1.1. Suppose we have a data set of ten ordered observations:

$$1,\ 2,\ 3,\ 10,\ 20,\ 30,\ 100,\ 200,\ 300,\ 1000.$$

Data Winsorization means 'pulling back' low and high extremes towards the middle, where the bulk of data reside. For example, the 20% left- and 10% right-Winsorized version of this data set is:

$$3,\ 3,\ 3,\ 10,\ 20,\ 30,\ 100,\ 200,\ 300,\ 300.$$

We see that 70% of data are original observations $(3, 10, 20, 30, 100, 200, 300)$ but their end-points (3 and 300) are used to replace 'extremes' $(1, 2$ and $1000)$. So the new data set is 20% left-Winsorized and 10% right-Winsorized. For practical examples involving data Winsorization in insurance ratemaking, see Dornheim and Brazauskas (2014, Section 2.2). □

Parameter estimators obtained this way retain all the desirable properties of MTMs—they are robust, computationally efficient, and responsive to insurance contract specifications—but typically are more efficient, although both approaches yield estimates that are still less efficient than those obtained by MLE which can be numerically unstable (for illustrations of this fact, see Section 5.1). Moreover, data Winsorization is a well-known and widely-accepted approach in other areas, not only insurance. For example, it is used for measuring income inequality in economics (see, e.g., Van Kerm, 2007) and for improving the reliability of engineering solutions (see, e.g., Ko and Lee, 1991).

The remainder of this paper is organized as follows. In Section 2, we present the MWM idea, along with the asymptotic properties of the MWM estimators. Examples of estimators for location-scale families and two loss severity models—lognormal and log-logistic—are provided in Section 3. In the simulation study of Section 4, the finite-sample performance of MWM estimators is compared to that of MTM and MLE, with the objective to see how large the sample size should be for the estimators to achieve asymptotic unbiasedness and reach their asymptotic efficiency levels. In Section 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 layer are computed when the lognormal and log-logistic models are fitted to the data using the MWM, MTM, and MLE methods. We conclude the paper with a brief summary of main findings in Section 6.

2 Method of Winsorized Moments

In this section we describe the MWM idea, along with the asymptotic properties of the obtained estimators, and conclude with several examples of MWM estimators for parametric models.

2.1 Definition

Suppose $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables, which follow a parametric distribution F with $k \geq 1$ 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 MWM estimators of $\theta_1, \ldots, \theta_k$ are derived in three steps:

1. Compute the sample Winsorized moments

$$\widehat{W}_{jn} = \frac{1}{n}\Big[m_n(j)h_j\big(X_{m_n(j)+1:n}\big) + \sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(X_{i:n}) + m_n^*(j)h_j\big(X_{n-m_n^*(j):n}\big)\Big] \tag{2.1}$$

   for $1 \le j \le k$. Here $m_n(j)$ and $m_n^*(j)$ are integers $0 \le m_n(j) < n - m_n^*(j) \le n$ 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. Also, h_j are specially chosen data transformations that usually represent moments or moment-like functions (e.g., $h_j(x) = x^j$, $h_j(x) = (\log x)^j$, or $h_j(x) = (1/x)^j$).

2. Derive the corresponding population Winsorized moments

$$W_j = W_j(\theta_1, \ldots, \theta_k) = a_j h_j(F^{-1}(a_j)) + \int_{a_j}^{1-b_j} h_j(F^{-1}(u))\, du + b_j h_j(F^{-1}(1-b_j)) \tag{2.2}$$

   for $1 \le j \le k$. Here $F^{-1}(u) = \inf\{x \in \mathbb{R} : u \le F(x)\}$ is the quantile function. (Notice that when $a_j = b_j = 0$, then $W_j = \mathbb{E}[h_j(X)]$ which, depending upon the distribution F, may be infinite. On the other hand, when $a_j > 0$ and $b_j > 0$, the Winsorized moment W_j is always finite.)

3. Match the population and sample Winsorized moments and solve the system of equations

$$\begin{cases} W_1(\theta_1, \ldots, \theta_k) & = & \widehat{W}_{1n}, \\ & \vdots & \\ W_k(\theta_1, \ldots, \theta_k) & = & \widehat{W}_{kn} \end{cases} \tag{2.3}$$

   with respect to $\theta_1, \ldots, \theta_k$. The obtained solutions, which we denote by $\widehat{\theta}_j = g_j(\widehat{W}_{1n}, \ldots, \widehat{W}_{kn})$, $1 \le j \le k$, are, by definition, the MWM estimators of $\theta_1, \ldots, \theta_k$. Notice that the functions g_j are such that $\theta_j = g_j(W_1, \ldots, W_k)$.

Note 2.1. The system of equations (2.3) can be written as $W_j(\theta_1, \ldots, \theta_k) - \widehat{W}_{jn} = 0$ for $j = 1, \ldots, k$, and thus the MWM estimator can be viewed as an M-estimator (Huber and Ronchetti, 2009). Alternatively, the proposed approach can also be interpreted as a special case of Generalized Method of Moments (Hansen, 1982). These perspectives might be very useful if one chooses to extend MWM to more general settings (e.g., generalized linear models, econometric models, multivariate distributions, Bayesian analysis), because the estimation methods obtained this way would possess favorable robustness properties and computational efficiency. □

Note 2.2. The procedure (2.1)–(2.3) is presented for general k and in theory it should always work. In practice, however, it can happen that equations (2.3) do not have a solution, or that they are difficult to solve even numerically when k is large due to stability problems (think about the regression models with a large number of covariates, for example). To alleviate the problem, the functions h_j have to be chosen thoughtfully, i.e., we want to select data transformations that linearize the quantile function (in terms of parameters) as much as possible and keep the moments as low as possible because those yield simpler equations. On the other hand, for many applications in actuarial science, risk management, economics, business and engineering the most common models fall within the case $k \leq 4$ (see Klugman *et al.*, 2012, Appendix A; Kleiber and Kotz, 2003; and Johnson *et al.*, 1995). Moreover, a number of those distributions belong to a general class of location-scale families or their variants, which implies that $k \leq 2$. In the latter instances, the procedure is indeed straightforward and, as will be shown in Section 3.1, yields explicit formulas for MWM estimators. Finally, for parametric distributions that cannot be fully transformed to a location-scale family or its variant (e.g., gamma, log-t, two-parameter Pareto, GPD, GB2), we would choose $h_1(t) = h_2(t) = t$ but $a_1 \neq a_2$ and $b_1 \neq b_2$. Such an approach was implemented for MTM (see Brazauskas and Kleefeld, 2009, and Kleefeld and Brazauskas, 2012) and could be utilized for MWM as well. $\square$

2.2 Asymptotic Properties

For any fixed j, the sample Winsorized moment in equation (2.1) can be written as

$$\widehat{W}_{jn} = \frac{1}{n}\sum_{i=1}^{n} K_j\left(\frac{i}{n+1}\right) h_j(X_{i:n}) + c_{jn}^{(1)} h_j\left(X_{\lfloor np_j^{(1)}\rfloor+1:n}\right) + c_{jn}^{(2)} h_j\left(X_{\lceil np_j^{(2)}\rceil:n}\right),$$

where $\lfloor\cdot\rfloor$ denotes "smallest integer part" and $\lceil\cdot\rceil$ denotes "greatest integer part",

$$K_j(x) = \mathbf{1}\left\{p_j^{(1)} \leq x \leq p_j^{(2)}\right\} = \begin{cases} 1, & \text{if } p_j^{(1)} \leq x \leq p_j^{(2)}; \\ 0, & \text{otherwise;} \end{cases} \tag{2.4}$$

with $p_j^{(1)} = a_j$ and $p_j^{(2)} = 1 - b_j$ and where a_j and b_j represent left and right Winsorizing proportions, respectively. Also, $\lim_{n\to\infty} c_{jn}^{(1)} = c_j^{(1)} = p_j^{(1)} = a_j$ and $\lim_{n\to\infty} c_{jn}^{(2)} = c_j^{(2)} = 1 - p_j^{(2)} = b_j$.

For h_j functions used in this paper, which will be continuously differentiable on $(-\infty, \infty)$ or $(0, \infty)$, it is fairly easy to verify the conditions of Corollary 3 in Chernoff *et al.* (1967). Hence, for any fixed

j, $1 \le j \le k$, the statistic $\widehat{W}_{jn}$ is asymptotically normal with the mean

$$W_j = \int_{a_j}^{1-b_j} h_j\left(F^{-1}(u)\right)\, du + c_j^{(1)} h_j\left(F^{-1}\left(p_j^{(1)}\right)\right) + c_j^{(2)} h_j\left(F^{-1}\left(p_j^{(2)}\right)\right) \tag{2.5}$$

and the variance

$$n^{-1}\sigma_j^2 = n^{-1}\int_0^1 \alpha_j^2(u)\, du\,,$$

where

$$\alpha_j(u) = \frac{1}{1-u}\int_u^1 K_j(w)H_j^{'}(w)(1-w)\, dw + \sum_{m=1}^{2} \mathbf{1}\left\{p_j^{(m)} \ge u\right\} c_j^{(m)}\left(1-p_j^{(m)}\right) H_j^{'}\left(p_j^{(m)}\right)$$

and $H_j = h_j \circ F^{-1}$. Following Serfling (1980, p. 20), this asymptotic normality statement can be written concisely as

$$\widehat{W}_{jn} \;\sim\; \mathcal{AN}\left(W_j,\, \frac{\sigma_j^2}{n}\right).$$

Next, let us define matrix $\mathbf{\Sigma} := \left[\sigma_{ij}^2\right]_{i,j=1}^{k}$ with the entries

$$\begin{aligned}
\sigma_{ij}^2 &= \int_0^1 \alpha_i(u)\alpha_j(u)\, du \\
&= \int_0^1 \left\{\frac{1}{1-u}\left[\int_u^1 K_i(w)H_i^{'}(w)(1-w)\, dw + \sum_{m=1}^{2} \mathbf{1}\left\{p_i^{(m)} \ge u\right\} c_i^{(m)}\left(1-p_i^{(m)}\right) H_i^{'}\left(p_i^{(m)}\right)\right]\right. \\
&\quad \left. \times \frac{1}{1-u}\left[\int_u^1 K_j(v)H_j^{'}(v)(1-v)\, dv + \sum_{m=1}^{2} \mathbf{1}\left\{p_j^{(m)} \ge u\right\} c_j^{(m)}\left(1-p_j^{(m)}\right) H_j^{'}\left(p_j^{(m)}\right)\right]\right\} du \\
&=: A_{i,j}^{(1)} + A_{i,j}^{(2)} + A_{i,j}^{(3)} + A_{i,j}^{(4)}\,,
\end{aligned} \tag{2.6}$$

where

$$A_{i,j}^{(1)} = \int_0^1 \frac{1}{(1-u)^2}\left[\int_u^1 K_i(w)H_i^{'}(w)(1-w)\, dw \times \int_u^1 K_j(v)H_j^{'}(v)(1-v)\, dv\right] du\,,$$

$$A_{i,j}^{(2)} = \int_0^1 \frac{1}{(1-u)^2}\left[\int_u^1 K_i(w)H_i^{'}(w)(1-w)\, dw \times \sum_{m=1}^{2} \mathbf{1}\left\{p_j^{(m)} \ge u\right\} c_j^{(m)}\left(1-p_j^{(m)}\right) H_j^{'}\left(p_j^{(m)}\right)\right] du\,,$$

$$A_{i,j}^{(3)} = \int_0^1 \frac{1}{(1-u)^2}\left[\int_u^1 K_j(v)H_j^{'}(v)(1-v)\, dv \times \sum_{m=1}^{2} \mathbf{1}\left\{p_i^{(m)} \ge u\right\} c_i^{(m)}\left(1-p_i^{(m)}\right) H_i^{'}\left(p_i^{(m)}\right)\right] du\,,$$

$$\begin{aligned}
A_{i,j}^{(4)} &= \int_0^1 \frac{1}{(1-u)^2}\left[\sum_{m=1}^{2} \mathbf{1}\left\{p_i^{(m)} \ge u\right\} c_i^{(m)}\left(1-p_i^{(m)}\right) H_i^{'}\left(p_i^{(m)}\right)\right. \\
&\quad \left. \times \sum_{m=1}^{2} \mathbf{1}\left\{p_j^{(m)} \ge u\right\} c_j^{(m)}\left(1-p_j^{(m)}\right) H_j^{'}\left(p_j^{(m)}\right)\right] du\,.
\end{aligned}$$

Now, using the notation introduced above, in the following theorem we establish asymptotic normality of sample Winsorized moments and the corresponding MWM estimators.

Theorem 2.1. *Let $X_{1:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of a random sample $X_1, \ldots, X_n$ from a continuous distribution F. Suppose functions h_j, $j = 1, \ldots, k$, are continuously differentiable on $(-\infty, \infty)$ or $(0, \infty)$. Then the following statements follow from Remark 9 in Chernoff et al. (1967):*

(i) $\left(\widehat{W}_{1n}, \ldots, \widehat{W}_{kn}\right) \sim \mathcal{AN}\left((W_1, \ldots, W_k), \frac{1}{n}\mathbf{\Sigma}\right)$,

(ii) $\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right) \sim \mathcal{AN}\left((\theta_1, \ldots, \theta_k), \frac{1}{n}\mathbf{D}\mathbf{\Sigma}\mathbf{D}'\right)$,

where $\mathbf{D} := \left[d_{ij}\right]_{i,j=1}^{k}$ is the Jacobian of the transformations $g_1, \ldots, g_k$ evaluated at $\left(W_1, \ldots, W_k\right)$ and $\mathbf{\Sigma}$ is the covariance-variance matrix introduced in equation (2.6).

From equations (2.4) and (2.5), we see that the entries of $\mathbf{\Sigma}$ actually depend on the proportions (a_i, b_i) and (a_j, b_j). In total, there are six possible arrangements of these proportions:

1. $0 \leq a_i \leq 1 - b_i < a_j \leq 1 - b_j \leq 1$,
2. $0 \leq a_i \leq a_j < 1 - b_i \leq 1 - b_j \leq 1$,
3. $0 \leq a_i \leq a_j < 1 - b_j \leq 1 - b_i \leq 1$,
4. $0 \leq a_j \leq 1 - b_j < a_i \leq 1 - b_i \leq 1$,
5. $0 \leq a_j \leq a_i < 1 - b_j \leq 1 - b_i \leq 1$,
6. $0 \leq a_j \leq a_i < 1 - b_i \leq 1 - b_j \leq 1$.

Each choice results in a different expression of σ_{ij}^2. To get a better sense of how the terms $A_{i,j}^{(1)}, \ldots, A_{i,j}^{(4)}$ look like, in Lemma A.1 (see Appendix) we provide their expressions for the case when the respective lower and upper proportions of two sample Winsorized moments are identical, i.e., $0 \leq a = a_i = a_j < 1 - b_i = 1 - b_j = 1 - b \leq 1$. In Section 3, we will use this selection of a's and b's and derive explicit formulas of MWM estimators for location and scale parameters for several choices of F and h_j.

Note 2.3. In the procedure (2.1)–(2.3), if one selects $a_j > 0$ and $b_j > 0$ ($0 < a_j + b_j < 1$), for $j = 1, 2$, then the resulting estimators will be resistant against outliers, i.e., they will be globally robust with the lower and upper breakdown points given by LBP $= \min\{a_1, a_2\}$ and UBP $= \min\{b_1, b_2\}$, respectively. The robustness of such estimators against extremely small or large outliers comes from the fact that in the computation of estimates the order statistics with the index less than $n \times$ LBP or higher than $n \times (1 -$ UBP$)$ are replaced with the ones having index $n \times$ LBP or $n \times (1 -$ UBP$)$, respectively. For more details on LBP and UBP for heavy-tailed asymmetric models, see Brazauskas and Serfling (2000) and Serfling (2002). □

Note 2.4. The primary competitor of MWM is the method of trimmed moments (MTM) introduced by Brazauskas *et al*. (2009). The MTM procedure is conceptually equivalent to (2.1)–(2.3), except that trimmed moments are used instead of Winsorized moments. Its asymptotic properties were derived using central limit theory for trimmed L-statistics (see Brazauskas *et al*., 2007), but they also follow directly from the results of this section. That is, in (2.4) we need to choose $\widetilde{K}_j(x) = K_j(x)/(1-a_j-b_j)$ with $p_j^{(1)} = a_j$ and $p_j^{(2)} = 1 - b_j$, and $\lim_{n\to\infty} c_{jn}^{(1)} = c_j^{(1)} = 0$ and $\lim_{n\to\infty} c_{jn}^{(2)} = c_j^{(2)} = 0$. Then, these choices imply that in expression (2.6) we have $A_{i,j}^{(2)} = A_{i,j}^{(3)} = A_{i,j}^{(4)} = 0$, and $\widetilde{A}_{i,j}^{(1)} = A_{i,j}^{(1)} \big/ \big[(1-a_i-b_i)(1-a_j-b_j)\big]$. Clearly, the statements of Theorem 2.1 remain valid for these new selections and hence MTMs. □

3 Examples

In this section, we derive MWM estimators of location and scale parameters for general (i.e., not necessarily symmetric) location-scale families, and obtain the entries of their asymptotic covariance-variance matrix. For specific numerical illustrations, we choose lognormal and log-logistic distributions and evaluate the asymptotic relative efficiency (ARE) of the MWM estimators with respect to the maximum likelihood estimator (MLE):

$$\text{ARE}\big(\text{MWM, MLE}\big) \;=\; \frac{\text{asymptotic variance of MLE estimator}}{\text{asymptotic variance of MWM estimator}}\,.$$

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 more details on these issues, the reader may be referred, for example, to Serfling (1980, Section 4.1).

3.1 Location-Scale Families

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

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

where location $-\infty < \mu < \infty$ and scale $\sigma > 0$ are unknown parameters, and F_0 is the standard (i.e., with $\mu = 0$ and $\sigma = 1$) parameter-free version of F. The corresponding quantile function is

$$F^{-1}(u) = \mu + \sigma F_0^{-1}(u), \qquad 0 < u < 1.$$

Since F has two unknown parameters, we employ two Winsorized moments. Choosing $h_1(t) = t$ and $h_2(t) = t^2$, and then following the procedure of Section 2.1, we have

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

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

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

$$\begin{aligned}
W_1 := W_1(\mu,\sigma) &= a\,F^{-1}(a) + \int_a^{1-b} F^{-1}(u)\,du + b\,F^{-1}(1-b) \\
&= a\left[\mu + \sigma\,F_0^{-1}(a)\right] + \int_a^{1-b}\left[\mu + \sigma\,F_0^{-1}(u)\right]du + b\left[\mu + \sigma\,F_0^{-1}(1-b)\right] \\
&= \mu + \sigma\left\{a\,F_0^{-1}(a) + \int_a^{1-b} F_0^{-1}(u)\,du + b\,F_0^{-1}(1-b)\right\} \\
&= \mu + \sigma\,c_1,
\end{aligned}$$

$$
\begin{aligned}
W_2 \;:=\; W_2(\mu,\sigma) &= a\left[F^{-1}(a)\right]^2 + \int_a^{1-b} [F^{-1}(u)]^2\,du + b\left[F^{-1}(1-b)\right]^2 \\
&= a\left[\mu + \sigma\,F_0^{-1}(a)\right]^2 + \int_a^{1-b} \left[\mu + \sigma\,F_0^{-1}(u)\right]^2 du + b\left[\mu + \sigma\,F_0^{-1}(1-b)\right]^2 \\
&= \mu^2 + 2\mu\sigma\left\{a\,F_0^{-1}(a) + \int_a^{1-b} F_0^{-1}(u)\,du + b\,F_0^{-1}(1-b)\right\} \\
&\quad + \sigma^2\left\{a\left[F_0^{-1}(a)\right]^2 + \int_a^{1-b} \left[F_0^{-1}(u)\right]^2 du + b\left[F_0^{-1}(1-b)\right]^2\right\} \\
&= \mu^2 + 2\mu\sigma\,c_1 + \sigma^2\,c_2,
\end{aligned}
$$

where $c_k \equiv c_k(F_0,a,b) := a\left[F_0^{-1}(a)\right]^k + \int_a^{1-b}\left[F_0^{-1}(u)\right]^k du + b\left[F_0^{-1}(1-b)\right]^k,\ k=1,2$, do not depend on any unknown parameters and can be easily evaluated using numerical methods.

Equating $\widehat{W}_{1n}$ to W_1 and $\widehat{W}_{2n}$ to W_2, and then solving the resulting system of equations with respect to μ and σ, we obtain the MWM estimators

$$
\begin{cases}
\widehat{\mu}_{\mathrm{MWM}} = \widehat{W}_{1n} - c_1\,\widehat{\sigma}_{\mathrm{MWM}} \;=:\; g_1\left(\widehat{W}_{1n}, \widehat{W}_{2n}\right); \\
\widehat{\sigma}_{\mathrm{MWM}} = \sqrt{\left(\widehat{W}_{2n} - \widehat{W}_{1n}^2\right)\cdot\left(c_2 - c_1^2\right)^{-1}} \;=:\; g_2\left(\widehat{W}_{1n}, \widehat{W}_{2n}\right).
\end{cases}
\tag{3.2}
$$

The entries of the covariance-variance matrix $\mathbf{\Sigma}$ are calculated using the formulas of Section 2.2. After lengthy but straightforward derivations, we can obtain quite simple expressions for $\sigma_{i,j}^2$:

$$
\begin{aligned}
\sigma_{11}^2 &= \sigma^2 C_1, \\
\sigma_{12}^2 &= \sigma_{21}^2 \;=\; 2\mu\sigma^2 C_1 + 2\sigma^3 C_2, \\
\sigma_{22}^2 &= 4\mu^2\sigma^2 C_1 + 8\mu\sigma^3 C_2 + 4\sigma^4 C_3,
\end{aligned}
$$

where the constants $C_k \equiv C_k(F_0,a,b)$ can be written in terms of the earlier introduced constants c_k and do not depend on any unknown parameters; their expressions are provided in the appendix.

For calculating the matrix $\mathbf{D}$, we differentiate the functions g_i in (3.2):

$$\begin{aligned}
d_{11} &= \frac{\partial g_1}{\partial \widehat{W}_{1n}}\bigg|_{(W_1,W_2)} = 1 - c_1\frac{\partial g_2}{\partial \widehat{W}_{1n}}\bigg|_{(W_1,W_2)} = \frac{c_1\mu + c_2\sigma}{\sigma(c_2 - c_1^2)},\\
d_{12} &= \frac{\partial g_1}{\partial \widehat{W}_{2n}}\bigg|_{(W_1,W_2)} = -c_1\frac{\partial g_2}{\partial \widehat{W}_{2n}}\bigg|_{(W_1,W_2)} = \frac{-0.5c_1}{\sigma(c_2 - c_1^2)},\\
d_{21} &= \frac{\partial g_2}{\partial \widehat{W}_{1n}}\bigg|_{(W_1,W_2)} = \frac{-\widehat{W}_{1n}}{\sqrt{(c_2-c_1^2)(\widehat{W}_{2n} - \widehat{W}_{1n}^2)}}\bigg|_{(W_1,W_2)} = \frac{-\mu - c_1\sigma}{\sigma(c_2 - c_1^2)},\\
d_{22} &= \frac{\partial g_2}{\partial \widehat{W}_{2n}}\bigg|_{(W_1,W_2)} = \frac{0.5}{\sqrt{(c_2-c_1^2)(\widehat{W}_{2n} - \widehat{W}_{1n}^2)}}\bigg|_{(W_1,W_2)} = \frac{0.5}{\sigma(c_2 - c_1^2)}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbf{D\Sigma D'} &= \begin{bmatrix} d_{11} & d_{12}\\ d_{21} & d_{22}\end{bmatrix}\begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2\\ \sigma_{21}^2 & \sigma_{22}^2\end{bmatrix}\begin{bmatrix} d_{11} & d_{21}\\ d_{12} & d_{22}\end{bmatrix}\\
&= \frac{\sigma^2}{(c_2-c_1^2)^2}\begin{bmatrix} C_1c_2^2 - 2c_1c_2C_2 + c_1^2C_3 & -C_1c_1c_2 + c_2C_2 + c_1^2C_2 - c_1C_3\\ -C_1c_1c_2 + c_2C_2 + c_1^2C_2 - c_1C_3 & C_1c_1^2 - 2c_1C_2 + C_3\end{bmatrix}. \qquad (3.3)
\end{aligned}$$

We summarize the above findings as a theorem stated below.

Theorem 3.1. *Suppose $X_1, \ldots, X_n$ are i.i.d. random variables from a continuous location-scale family with cdf F defined by equation (3.1). Let $\widehat{\mu}_{MWM}$ and $\widehat{\sigma}_{MWM}$ denote the MWM estimators of μ and σ, respectively. Then*

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

where the matrix $\mathbf{D\Sigma D'}$ is specified by equation (3.3).

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.

To get a better sense of how broad the results of Section 3.1 are and how versatile the MWM approach is, Table 3.1 lists selected location-scale distributions and their variants used for modeling insurance loss severity, along with key inputs for equations (3.2).

Table 3.1. Distributional characteristics, functions $h_k(t)$ and constants $c_k(F_0, a, b)$ in equations (3.2) for selected location-scale distributions and related loss models.

Distribution	Standard Quantile Function, F_0^{-1}	Functions $h_1(t)$	$h_2(t)$	$a = b = 0.05$ c_1	c_2	$a = b = 0.10$ c_1	c_2	$a = 0.25,\ b = 0.01$ c_1	c_2
Location-scale distributions $\left(F^{-1}(u) = \mu + \sigma F_0^{-1}(u)\right)$									
$Exp\,(0, \sigma)$	$-\log(1-u)$	t	–	0.9513	–	0.9054	–	1.0277	–
$N\,(\mu, \sigma)$	$\Phi^{-1}(u)$	t	t^2	0	0.8313	0	0.6787	0.1458	0.6315
$C\,(\mu, \sigma)$	$\tan(\pi(u-0.5))$	t	t^2	0	7.1058	0	3.0537	1.0594	20.0825
$L\,(\mu, \sigma)$	$-\log(1/u-1)$	t	t^2	0	2.4779	0	1.9312	0.2776	1.9272
$G\,(\mu, \sigma)$	$-\log(-\log(u))$	t	t^2	0.5397	1.5408	0.5070	1.2369	0.6859	1.7009
Log-location-scale distributions $\left(F^{-1}(u) = \exp\left\{\mu + \sigma F_0^{-1}(u)\right\}\right)$									
$PaI\,(\log\theta, 1/\alpha)$	$-\log(1-u)$	$\log t$	–	0.9513	–	0.9054	–	1.0277	–
$logN\,(\mu, \sigma)$	$\Phi^{-1}(u)$	$\log t$	$\log^2 t$	0	0.8313	0	0.6787	0.1458	0.6315
$logC\,(\mu, \sigma)$	$\tan(\pi(u-0.5))$	$\log t$	$\log^2 t$	0	7.1058	0	3.0537	1.0594	20.0825
$logL\,(\mu, \sigma)$	$-\log(1/u-1)$	$\log t$	$\log^2 t$	0	2.4779	0	1.9312	0.2776	1.9272
$W\,(\log\theta, 1/\tau)$	$\log(-\log(1-u))$	$\log t$	$\log^2 t$	-0.5397	1.5408	-0.5070	1.2369	-0.3108	0.7480
$iW\,(\log\theta, 1/\tau)$	$-\log(-\log(u))$	$\log t$	$\log^2 t$	0.5397	1.5408	0.5070	1.2369	0.6859	1.7009
Folded $\left(F^{-1}(u) = \sigma F_0^{-1}((u+1)/2)\right)$ and *log-folded distributions* $\left(F^{-1}(u) = \exp\left\{\sigma F_0^{-1}((u+1)/2)\right\}\right)$									
$FN\,(\sigma)$	$\Phi^{-1}((u+1)/2)$	t	–	0.7806	–	0.7624	–	0.8349	–
$FC\,(\sigma)$	$\tan(\pi(u/2))$	t	–	2.2576	–	1.8203	–	3.3340	–
$logFN\,(\sigma)$	$\Phi^{-1}((u+1)/2)$	$\log t$	–	0.7806	–	0.7624	–	0.8349	–
$logFC\,(\sigma)$	$\tan(\pi(u/2))$	$\log t$	–	2.2576	–	1.8203	–	3.3340	–

Abbreviations: Exp = exponential; N = normal; C = Cauchy; L = logistic; G = Gumbel; W = Weibull; PaI = single-parameter Pareto (with θ known); FN = folded normal; FC = folded Cauchy; iW = inverse Weibull.

Note 3.1. Table 3.1 provides only a fraction of location-scale families and their transformations that are available in the literature. More examples can be found in Johnson *et al.* (1995) and Kleiber and Kotz (2003). Also, for more examples of folded distributions that have been used for modeling insurance data, see Nadarajah and Bakar (2015). In all these cases, the formulas for $(\widehat{\mu}_{\text{MWM}}, \widehat{\sigma}_{\text{MWM}})$ are given by (3.2); the only adjustment one has to make is to recompute constants c_k. □

3.2 Lognormal Model

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

$$\text{Lognormal}\,(\mu, \sigma): \quad F(x) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right), \qquad x > 0, \tag{3.4}$$

where log-location $-\infty < \mu < \infty$ and log-scale $\sigma > 0$ are unknown parameters with Φ denoting the standard normal cdf. Since the logarithmic transformation makes this distribution normal, which is a member of the location-scale family, results of Section 3.1 apply with two modifications: $h_1(t) = \log(t)$ and $h_2(t) = \log^2(t)$. Hence, the MWM estimators of μ and σ have the same closed form expression as estimators in (3.2), except that the sample Winsorized moments are now defined as

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

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

with $m_n(1)/n = m_n(2)/n \to a$ and $m_n^*(1)/n = m_n^*(2)/n \to b$ as $n \to \infty$. Note also that the above choice of functions h_1 and h_2 ensures that the formulas of c_1 and c_2 do not change when one computes the corresponding MWM estimators for location and scale parameters of the normal distribution (for which $h_1(t) = t$ and $h_2(t) = t^2$). That is,

$$c_k = c_k(\Phi, a, b) = a\left[\Phi^{-1}(a)\right]^k + \int_a^{1-b} \left[\Phi^{-1}(u)\right]^k du + b\left[\Phi^{-1}(1-b)\right]^k, \quad k = 1, 2.$$

Let us summarize this discussion as a corollary.

Corollary 3.1. *Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with cdf F defined by equation (3.4). Let $\widehat{\mu}_{MWM}$ and $\widehat{\sigma}_{MWM}$ denote the MWM estimators of μ and σ, respectively. Then it follows from Theorem 3.1 that*

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

where the matrix $\mathbf{D\Sigma D}'$ is specified by equation (3.3), but now with the standard normal cdf Φ instead of the therein used standardized location-scale distribution.

We next examine how much efficiency is lost due to using $\left(\widehat{\mu}_{\text{MWM}}, \widehat{\sigma}_{\text{MWM}}\right)$ instead of $\left(\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}\right)$. The following note provides key facts about the lognormal distribution MLEs.

Note 3.2. The MLE of lognormal distribution parameters has explicit form

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

It is well known (see, e.g., Serfling, 2002) that

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

Hence, it follows that

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

In addition, when $m_n(1) = m_n^*(1) = m_n(2) = m_n^*(2) = 0$, then the MWM estimators (3.2) become $(\widehat{\mu}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}})$. Also note that since $\mathbf{S} \to \mathbf{S_0}$ (element-wise convergence) when $a = b \to 0$, then the MLE asymptotic distribution follows from Corollary 3.1. □

Numerical values of the AREs are provided in Table 3.2 for chosen proportions a and b. Since the logarithmic transformation of the lognormal random variable makes the variable normal, which is symmetric, we can see similar performances of the MWM estimators when similar Winsorizing schemes are used. For example, the AREs are identical for the MWM estimators with reversed Winsorizing proportions: $(a, b) = (0.1, 0.25)$ has ARE = 0.701 and $(a, b) = (0.25, 0.1)$ also has ARE = 0.701. Further, 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 (see, e.g., Huber and Ronchetti, 2009). This can be explained by the fact that σ, unlike μ, is a parameter that measures data spread – thus spacings between a few observations in the middle are not necessarily representative of the dispersion for the entire sample. 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 = 0.914 (good); $a = b = 0.25$ has ARE = 0.571 (moderate); $a = b = 0.49$ has ARE = 0.081 (very poor). Finally, it is also of interest to compare the MWM approach with the MTM. Thus Table 3.2 has a second part that contains ARE entries for the MTM estimators, which are taken from Brazauskas *et al.* (2009). We clearly see that MWM uniformly outperforms MTM in terms of ARE, while still offering identical breakdown points (degrees of resistance against lower and upper outliers) and computational efficiency. The only point of overlap between the two methods is when $a = b = 0$, where they both become the MLE and thus have ARE = 1.

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

Method of Estimation	Proportion a	Proportion b 0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
MWM	0	1	0.957	0.916	0.874	0.791	0.581	0.379	0.214
	0.05	0.957	0.914	0.872	0.830	0.745	0.534	0.330	0.163
	0.10	0.916	0.872	0.829	0.786	0.701	0.489	0.284	0.109
	0.15	0.874	0.830	0.786	0.744	0.658	0.444	0.236	–
	0.25	0.791	0.745	0.701	0.658	0.571	0.354	0.126	–
	0.49	0.581	0.534	0.489	0.444	0.354	0.081	–	–
	0.70	0.379	0.330	0.284	0.236	0.126	–	–	–
	0.85	0.214	0.163	0.109	–	–	–	–	–
MTM	0	1	0.932	0.874	0.821	0.722	0.502	0.312	0.169
	0.05	0.932	0.872	0.820	0.771	0.678	0.470	0.286	0.142
	0.10	0.874	0.820	0.769	0.722	0.633	0.430	0.248	0.097
	0.15	0.821	0.771	0.722	0.676	0.590	0.390	0.208	–
	0.25	0.722	0.678	0.633	0.590	0.507	0.312	0.113	–
	0.49	0.502	0.470	0.430	0.390	0.312	0.074	–	–
	0.70	0.312	0.286	0.248	0.208	0.113	–	–	–
	0.85	0.169	0.142	0.097	–	–	–	–	–

3.3 Log-logistic Model

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

$$\text{Log-logistic}\,(\mu, \sigma): \quad F(x) = G_0\left(\frac{\log(x) - \mu}{\sigma}\right), \qquad x > 0, \tag{3.5}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$ are unknown parameters, and

$$G_0(y) = \frac{1}{1 + \exp\{-y\}}, \qquad -\infty < y < \infty, \tag{3.6}$$

denotes the standard logistic cdf. The quantile function corresponding to (3.5) is given by

$$F^{-1}(u) = \exp\left\{\mu + \sigma G_0^{-1}(u)\right\}, \qquad 0 < u < 1,$$

where $G_0^{-1}(u) = -\log(1/u - 1)$ is the quantile function of the standard logistic distribution. Applying all the steps of Section 3.1 with $h_1(t) = \log(t)$ and $h_2(t) = \log^2(t)$, the MWM estimators of μ and σ are obtained in the same form as in (3.2). Note that the sample Winsorized moments $\widehat{W}_{1n}$ and $\widehat{W}_{2n}$ are computed as in Section 3.2.

We next examine how much efficiency is lost due to using $(\widehat{\mu}_{\text{MWM}}, \widehat{\sigma}_{\text{MWM}})$ instead of $(\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}})$. The following note provides key facts about the log-logistic distribution MLEs.

Note 3.3. The log-likelihood of the log-logistic distribution is given by

$$\begin{aligned}\log \mathcal{L}\big(\mu, \sigma \,|\, X_1, \ldots, X_n\big) &= \sum_{i=1}^{n} \log\left(\frac{g_0\big((\log(X_i) - \mu)/\sigma\big)}{\sigma X_i}\right) \\ &= -\sum_{i=1}^{n} \frac{\log(X_i) - \mu}{\sigma} - n\log(\sigma) - \sum_{i=1}^{n} \log(X_i) - 2\sum_{i=1}^{n} \log\left(1 + \exp\left\{-\frac{\log(X_i) - \mu}{\sigma}\right\}\right),\end{aligned}$$

where g_0 denotes the density function corresponding to (3.6). One can clearly see from this log-likelihood expression that there is no closed-form solution for $(\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}})$; thus it has to be found using iterative numerical procedures. Further, it is known (see, e.g., deCani and Stine, 1986) that

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

Hence, it follows that

$$\text{ARE}\big((\widehat{\mu}_{\text{MWM}}, \widehat{\sigma}_{\text{MWM}}), (\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}})\big) = \big(\det(\mathbf{S_0})/\det(\mathbf{S})\big)^{1/2} = \sqrt{\frac{27}{(3+\pi^2)\det(\mathbf{S})}}.$$

Finally, unlike the lognormal model case, the MLE of log-logistic distribution parameters is not a special or limiting case of the MWM estimators. □

In Table 3.3 we provide numerical values of the AREs for selected proportions a and b. After the logarithmic transformation, the log-logistic model becomes logistic which is symmetric, and thus we again see 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 $= 0.801$. Further, when a and b increase, we observe a gradual decrease in efficiency, this pattern was also observed in Table 3.2. However, unlike the lognormal case, the ARE reaches its maximum at $a = b = 0.05$ (or in some small neighborhood of it), not at $a = b = 0$. Finally, comparison with the MTM estimators, shows that at the peak MTM reaches a higher ARE value than MWM, but that advantage is valid only for nearby located a and b values. Once we look outside the square $\{(a,b):\ 0 \le a \le 0.10,\ 0 \le b \le 0.10\}$, MWM uniformly improves MTM according to the ARE criterion.

TABLE 3.3. Log-logistic model: $\text{ARE}((\widehat{\mu}_{\text{MWM}}, \widehat{\sigma}_{\text{MWM}}),\ (\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$ and $\text{ARE}((\widehat{\mu}_{\text{MTM}}, \widehat{\sigma}_{\text{MTM}}),\ (\widehat{\mu}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}))$ for selected values of a and b, with the boxed numbers highlighting the case $a = b$.

Method of Estimation	Proportion a	Proportion b: 0	0.05	0.10	0.15	0.25	0.49	0.70	0.85
MWM	0	0.893	0.896	0.873	0.843	0.774	0.571	0.358	0.187
	0.05	0.896	0.913	0.895	0.868	0.801	0.589	0.359	0.169
	0.10	0.873	0.895	0.878	0.852	0.783	0.564	0.323	0.118
	0.15	0.843	0.868	0.852	0.825	0.754	0.528	0.277	–
	0.25	0.774	0.801	0.783	0.754	0.680	0.439	0.153	–
	0.49	0.571	0.589	0.564	0.528	0.439	0.104	–	–
	0.70	0.358	0.359	0.323	0.277	0.153	–	–	–
	0.85	0.187	0.169	0.118	–	–	–	–	–
MTM	0	0.893	0.884	0.834	0.782	0.681	0.449	0.258	0.127
	0.05	0.884	0.936	0.903	0.861	0.768	0.529	0.311	0.146
	0.10	0.834	0.903	0.874	0.835	0.745	0.507	0.283	0.106
	0.15	0.782	0.861	0.835	0.797	0.709	0.473	0.245	–
	0.25	0.681	0.768	0.745	0.709	0.625	0.391	0.138	–
	0.49	0.449	0.529	0.507	0.473	0.391	0.095	–	–
	0.70	0.258	0.311	0.283	0.245	0.138	–	–	–
	0.85	0.127	0.146	0.106	–	–	–	–	–

4 Simulation Study

In this section, we supplement our theoretical results concerning the MWM and MTM estimators with their finite-sample performance evaluations. The objective is to compare the two methods and 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 (MSE). That is,

$$\text{RE}\big(\text{Q, MLE}\big) \;=\; \frac{\text{asymptotic generalized variance of MLE estimator}}{\text{small-sample generalized MSE of Q estimator}}\,,$$

where Q represents MWM, MTM, or MLE, the numerator is defined as in Section 3, and the denominator is the square root of the determinant of

$$\begin{bmatrix} \mathbb{E}\big[(\widehat{\theta}_1-\theta_1)^2\big] & \mathbb{E}\big[(\widehat{\theta}_1-\theta_1)(\widehat{\theta}_2-\theta_2)\big] \\ \mathbb{E}\big[(\widehat{\theta}_1-\theta_1)(\widehat{\theta}_2-\theta_2)\big] & \mathbb{E}\big[(\widehat{\theta}_2-\theta_2)^2\big] \end{bmatrix}.$$

From a specified distribution F (i.e., lognormal or log-logistic), we generate 100,000 samples of a specified length n using simulations. For each sample we estimate the parameters of F using various estimators (MLE, MWM, and MTM). This process results in 100,000 copies of estimates for each estimator, which are then used to compute the estimator's RE and its standard error. 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 Lognormal Model

We start the study with the lognormal distribution LN($\mu = 5, \sigma = 1$) using the following parameters:

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

We summarize the simulation results in Tables 4.1 and 4.2, where the first table represents analysis of the bias and the second one that of the relative efficiency. In both tables, the estimator with $a = b = 0$ corresponds to the MLE.

TABLE 4.1. Lognormal model, LN($\mu = 5, \sigma = 1$): Mean values of $\widehat{\mu}/\mu$ and $\widehat{\sigma}/\sigma$ for selected n and several MWM (denoted as 'W') and MTM (denoted as 'T') estimators. For $a = b = 0$, MWM and MTM correspond to the MLE.

Proportion		$n = 100$				$n = 250$				$n = 500$				$n \to \infty$			
a	b	$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$	
		W	T	W	T	W	T	W	T	W	T	W	T	W	T	W	T
0	0	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1	1
0.05	0.05	1.00	1.00	0.99	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1	1
0.10	0.10	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1	1
0.25	0.25	1.00	1.00	0.98	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1	1	1	1
0.49	0.49	1.00	1.00	0.51	0.87	1.00	1.00	1.00	1.24	1.00	1.00	0.90	1.03	1	1	1	1
0.10	0.70	0.98	1.01	0.96	1.01	0.99	1.00	0.98	1.01	1.00	1.00	0.99	1.00	1	1	1	1
0.25	0	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1	1

NOTE: The entries for $n < \infty$ are mean values based on 100,000 samples. Their standard errors are ≤ 0.001, except for the estimators with $a = b = 0.49$, for which the standard errors are ≤ 0.003.

It is easy to see from Table 4.1 that all MWM and MTM estimators successfully estimate the log-location parameter μ. Indeed, they practically become unbiased for samples of size $n \geq 100$. Estimation of σ, however, reveals a different story. Although most estimators have less than 1% relative bias for $n \geq 100$, the median-type estimators (i.e., when $a = b = 0.49$) perform erratically: they have the respective relative bias of -49% (MWM) and -13% (MTM) for $n = 100$, 0% and +24% for $n = 250$, -10% and +3% for $n = 500$. Further, if we look at Table 4.2, we see that the simulated RE's of these estimators for $n > 100$ (which are equal to 0.08 and 0.07) are almost identical to the corresponding ARE's (which are equal to 0.081 and 0.074). Asymptotically all the estimators under investigation are unbiased, but in finite size samples we obviously see that bias is possible. Nonetheless, the relatively large biases we observe for fixed n have little effect on the RE's as they converge to the corresponding ARE's for $n > 100$. Finally, the large bias and variance of the median-type estimators can be explained by the fact that too few of the original observations are used to estimate σ. This parameter, unlike μ, reflects the spread of data and in finite size samples cannot be accurately estimated using only a few spacings between observations in the middle.

TABLE 4.2. Lognormal model, LN($\mu = 5, \sigma = 1$): Finite-sample efficiencies of MWMs and MTMs relative to MLEs. The ratios W/T represent efficiency of MWM relative MTM. For $a = b = 0$, MWM and MTM correspond to the MLE.

Proportion		$n = 100$			$n = 250$			$n = 500$			$n \to \infty$		
a	b	MWM	MTM	W/T	MWM	MTM	W/T	MWM	MTM	W/T	MWM	MTM	W/T
0	0	1.00	1.00	1	1.00	1.00	1	1.00	1.00	1	1	1	1
0.05	0.05	0.91	0.87	1.05	0.91	0.87	1.05	0.91	0.87	1.05	0.914	0.872	1.05
0.10	0.10	0.82	0.77	1.06	0.82	0.77	1.06	0.83	0.77	1.08	0.829	0.769	1.08
0.25	0.25	0.57	0.50	1.14	0.57	0.50	1.14	0.57	0.51	1.12	0.571	0.507	1.13
0.49	0.49	0.08	0.06	1.33	0.08	0.07	1.14	0.08	0.07	1.14	0.081	0.074	1.09
0.10	0.70	0.28	0.25	1.12	0.28	0.25	1.12	0.28	0.25	1.12	0.284	0.248	1.15
0.25	0	0.79	0.72	1.10	0.79	0.72	1.10	0.79	0.72	1.10	0.791	0.722	1.10

NOTE: The entries for $n < \infty$ are mean values based on 100,000 samples. Their standard errors are ≤ 0.003.

4.2 Log-logistic Model

We continue our simulation study with the log-logistic distribution LL($\mu = 5, \sigma = 1$) using the following parameters:

- *Sample size*: $n = 100,\ 250,\ 500$.
- *Estimators of* μ, σ:
  - MLE.
  - MWM and MTM with: $a = b = 0$; $a = b = 0.05$; $a = b = 0.10$; $a = b = 0.25$; $a = b = 0.49$; $a = 0.10$ and $b = 0.70$; $a = 0.25$ and $b = 0$.

In this case, the MLE estimates have to be found numerically (see Note 3.3). For that we will use the Newton algorithm, with the tolerance limit for the error being 10^{-5}. To guarantee the convergence of Newton's iterations, we choose $(\widehat{\mu}_{\text{start}}, \widehat{\sigma}_{\text{start}}) = (4.9, 1.1)$ as initial values, which is sufficiently close to the root. We summarize the simulation results in Tables 4.3 and 4.4, where the first table represents analysis of the bias and the second one that of the relative efficiency.

TABLE 4.3. Log-logistic model, LL($\mu = 5, \sigma = 1$): Mean values of $\widehat{\mu}/\mu$ and $\widehat{\sigma}/\sigma$ for selected n, MLE, MWM (denoted as 'W') and MTM (denoted as 'T') estimators.

Proportion		$n = 100$				$n = 250$				$n = 500$				$n \to \infty$			
a	b	$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$		$\widehat{\mu}/\mu$		$\widehat{\sigma}/\sigma$	
		W	T	W	T	W	T	W	T	W	T	W	T	W	T	W	T
0	0	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1	1
0.05	0.05	1.00	1.00	0.99	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1	1	1	1
0.10	0.10	1.00	1.00	0.99	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1	1
0.25	0.25	1.00	1.00	0.99	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1	1	1	1
0.49	0.49	1.00	1.00	0.51	0.87	1.00	1.00	1.00	1.24	1.00	1.00	0.90	1.03	1	1	1	1
0.10	0.70	0.98	1.01	0.96	1.02	0.99	1.00	0.98	1.01	1.00	1.00	0.99	1.00	1	1	1	1
0.25	0	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1	1
MLE		1.00		0.99		1.00		1.00		1.00		1.00		1		1	

NOTE: The entries for $n < \infty$ are mean values based on 100,000 samples. Their standard errors are ≤ 0.001, except for the estimators with $a = b = 0.49$, for which the standard errors are ≤ 0.002.

Similar to the case of lognormal model, we observe in Table 4.3 that all estimators of μ and σ become practically unbiased for $n \geq 100$, with the exception of the $a = b = 0.49$ estimators for σ, which are again volatile. Fortunately, the relatively large bias in estimating σ has no influence on those estimators' RE (see Table 4.4) as it converges to its corresponding ARE level for $n \geq 100$. Note that finite-sample performance of all other estimators, including the MLE, is reliable and predictable.

TABLE 4.4. Log-logistic model, LL($\mu = 5, \sigma = 1$): Finite-sample efficiencies of MLE, MWMs and MTMs relative to the asymptotic variance of MLE. The ratios W/T represent efficiency of MWM relative MTM.

Proportion		$n = 100$			$n = 250$			$n = 500$			$n \to \infty$		
a	b	MWM	MTM	W/T	MWM	MTM	W/T	MWM	MTM	W/T	MWM	MTM	W/T
0	0	0.90	0.90	1	0.90	0.90	1	0.89	0.89	1	0.893	0.893	1
0.05	0.05	0.91	0.93	0.98	0.91	0.93	0.98	0.91	0.93	0.98	0.913	0.936	0.98
0.10	0.10	0.87	0.87	1.00	0.88	0.87	1.01	0.88	0.87	1.01	0.878	0.874	1.00
0.25	0.25	0.68	0.62	1.10	0.68	0.62	1.10	0.68	0.62	1.10	0.680	0.625	1.09
0.49	0.49	0.10	0.08	1.25	0.10	0.08	1.25	0.10	0.09	1.11	0.104	0.095	1.09
0.10	0.70	0.32	0.28	1.14	0.32	0.28	1.14	0.32	0.28	1.14	0.323	0.283	1.14
0.25	0	0.78	0.70	1.11	0.78	0.69	1.13	0.78	0.68	1.15	0.774	0.681	1.14
MLE			1.00			1.00			1.00			1	

NOTE: The entries for $n < \infty$ are mean values based on 100,000 samples. Their standard errors are ≤ 0.004.

4.3 Risk Measurement

To demonstrate what trade-offs the use of robust estimators entails in estimating distribution tails, in this section we perform an additional simulation study. Here, we parametrically estimate the 95% and 99% value-at-risk, VaR, measures of the lognormal model LN($\mu = 5, \sigma = 1$) using the MLE, MWM and MTM estimators. For the lognormal model, the VaR estimates are computed as follows:

$$\widehat{\mathrm{VaR}(\alpha)} = \widehat{F}^{-1}(\alpha) = \exp\{\widehat{\mu} + \widehat{\sigma}\Phi^{-1}(\alpha)\} \quad \text{with} \;\; \alpha = 0.95,\, 0.99.$$

For the choice of parameters $\mu = 5$ and $\sigma = 1$, the true values of these risk measures are:

$$\mathrm{VaR}(0.95) = \exp\{5 + \Phi^{-1}(0.95)\} = 768.93 \quad \text{and} \quad \mathrm{VaR}(0.99) = \exp\{5 + \Phi^{-1}(0.99)\} = 1519.30.$$

In Table 4.5, we provide the bias, standard deviation, and root-MSE of the parametric VaR estimators. We clearly see that, aside from the cases of extreme trimming/Winsorizing (e.g., $a = b = 0.49$ and $a = 0.10, b = 0.70$), the robust estimators perform quite well. For a fixed choice of a and b, MWM outperforms MTM, as was predicted by those estimators' large- and small-sample properties. The MLE (i.e., MWM and MTM with $a = b = 0$) performs best because it is an optimal method when data *exactly* follow the assumed lognormal model, which is the case in this study. However, if that assumption were violated, the MLE's performance would be degraded, and in those situations, robust procedures would be indispensable.

Table 4.5. Lognormal model, LN($\mu = 5, \sigma = 1$): Bias, standard deviation and root-MSE of the 95% and 99% value-at-risk measures estimated using MLE, MWMs, MTMs. For $a = b = 0$, MWM and MTM correspond to the MLE.

Proportion		VaR(0.95)						VaR(0.99)					
a	b	Bias		Std. Deviation		$\sqrt{\text{MSE}}$		Bias		Std. Deviation		$\sqrt{\text{MSE}}$	
		MWM	MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM	MTM	MWM	MTM
0	0	-0.9	-0.9	118.4	118.4	118.4	118.4	0.6	0.6	295.6	295.6	295.6	295.6
0.05	0.05	-0.8	12.4	124.9	131.0	124.9	131.6	1.9	39.6	315.9	335.9	315.9	338.2
0.10	0.10	-1.0	15.3	132.7	142.5	132.7	143.3	2.5	49.6	340.2	372.2	340.2	375.5
0.25	0.25	-2.8	28.8	170.7	195.4	170.7	197.5	5.1	98.8	457.9	542.4	457.9	551.3
0.49	0.49	-46.7	44k	8k	5m	8k	5m	3k	18m	0.3m	3b	0.3m	3b
0.10	0.70	10.1	171.5	431.8	645.3	432.0	667.7	94.2	567.8	1224	2045	1228	2122
0.25	0	-1.3	0.5	125.3	127.8	125.3	127.8	0.1	9.0	325.2	338.3	325.2	338.5

Note: The true values of the risk measures are: VaR(0.95) = 768.93, VaR(0.99) = 1519.30. The results are based on 100,000 simulated samples of size 100. For $a = b = 0.49$, k stands for $\times 10^3$, m for $\times 10^6$, b for $\times 10^9$.

5 Real Data Illustrations

In this section, we apply the MWM, MTM and MLE to analyze the normalized damage amounts from the 30 most damaging hurricanes in the United States from 1925 to 1995, as recorded by Pielke and Landsea (1998). The damages were normalized to 1995 dollars by inflation, personal property increases, and coastal county population changes. Our goal is to investigate what effect initial assumptions and parameter estimation methods have on model fit (Section 5.1) and how that impacts insurance contract pricing (Section 5.2).

5.1 Model Fitting

As can be seen from Figure 5.1, the shape of the histogram of the top 30 damaging hurricanes is similar to that of many insurance loss distributions—it is right-skewed and heavy-tailed. That is, relatively small losses are most frequent, but as the size of loss increases their frequency declines; and there is also one outlier ($\sim$ \$72 billion). A histogram of log-transformed damages (see right-hand panel of Figure 5.1) shows that a roughly bell-shaped density curve will provide a satisfactory, though not perfect, overall fit to the data. Therefore, we will fit lognormal and log-logistic models to this data set using MLE, MWM, and MTM with several choices of proportions $a = a_1 = a_2$ and $b = b_1 = b_2$.

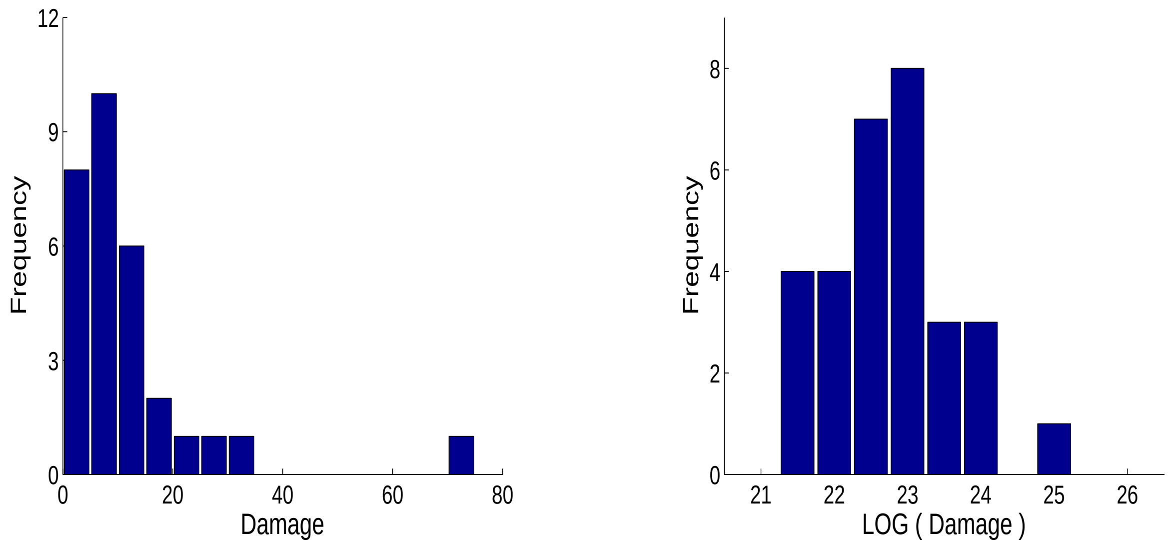

Figure 5.1. The histograms of the top 30 damaging hurricanes.
Left-hand panel: Original data (in billions). Right-hand panel: Log-transformed data.

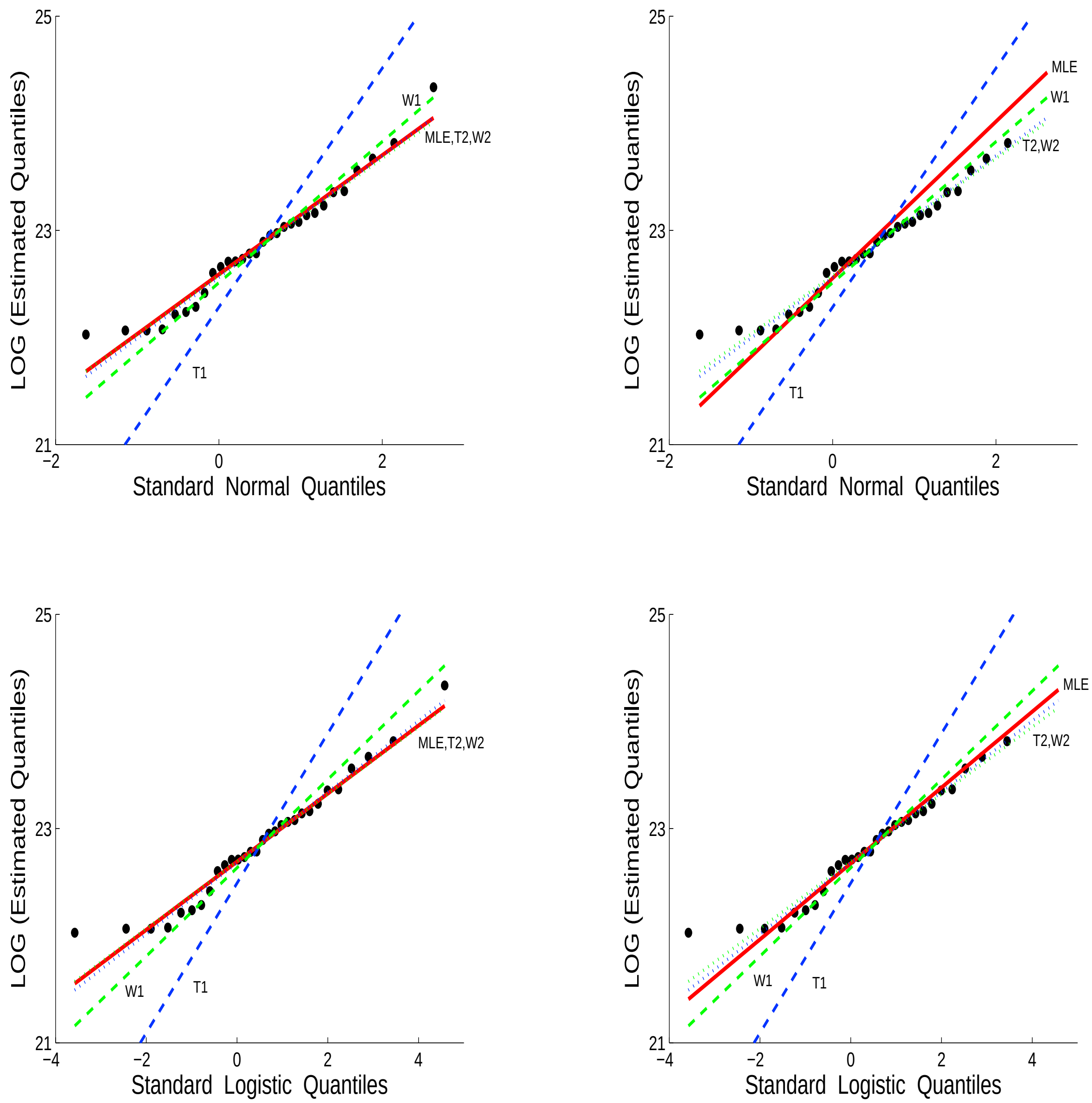

FIGURE 5.2. Model fits to the original (left-hand panels) and modified (right-hand panels) hurricane data. Top row: Lognormal models. Bottom row: Log-logistic models. The lines represent the relationship between standard quantiles and the log-quantiles of the estimated distribution. The labels for the lines show which method was used to estimate model parameters. The empirical quantiles are marked by '•'.

The initial fits are illustrated in the left-hand panels of Figure 5.2, where W1 and T1 denote the MWM and MTM estimators, respectively, with $a = b = 14/30$ (highly robust but inefficient estimators), and W2 and T2 correspond to the case $a = b = 1/30$ (minimally robust but highly efficient estimators). The parameter estimates and goodness-of-fit measurements appear in Table 5.1, where the fit is measured using the mean absolute deviation between the log-parametrically-fitted and log-empirically-evaluated quantiles,

$$(1/30)\sum_{j=1}^{30}\left|\log \widehat{F}^{-1}((j-0.5)/30)-\log X_{j:30}\right|. \tag{5.1}$$

(In Table 5.2, the same measure is used, except that it is restricted to the contract specifications of the example in Section 5.2.) In addition, to see the benefits of robust fitting, we have slightly modified the original data set by replacing the largest observation 72.303×10^9 with 723.03×10^9 and then re-computed the estimates and goodness-of-fit measures. The outcomes of this exercise are illustrated in the right-hand panels of Figure 5.2 and Table 5.1.

TABLE 5.1. Parameter estimates and goodness-of-fit measurements (FIT) of the lognormal and log-logistic models for the original and modified hurricane data.

Estimator	Proportion		Lognormal Model			Log-logistic Model		
	a	b	$\widehat{\mu}$	$\widehat{\sigma}$	FIT	$\widehat{\mu}$	$\widehat{\sigma}$	FIT
MLE	–	–	22.800	0.834	0.104	22.775	0.477	0.104
MLE (modified)	–	–	22.877	1.098	0.293	22.777	0.531	0.185
W1	14/30	14/30	22.760	0.988	0.140	22.760	0.619	0.191
W1 (modified)	14/30	14/30	22.760	0.988	0.216	22.760	0.619	0.249
T1	14/30	14/30	22.760	1.673	0.660	22.760	1.048	0.767
T1 (modified)	14/30	14/30	22.760	1.673	0.649	22.760	1.048	0.709
W2	1/30	1/30	22.776	0.820	0.104	22.776	0.470	0.106
W2 (modified)	1/30	1/30	22.776	0.820	0.181	22.776	0.470	0.183
T2	1/30	1/30	22.766	0.852	0.101	22.766	0.497	0.101
T2 (modified)	1/30	1/30	22.766	0.852	0.178	22.766	0.497	0.178

Several conclusions emerge from this analysis. First, the robust MWM and MTM estimates are not affected at all by the data modification whereas the MLE fit is substantially different from the original one at the lognormal model and changes slightly at the log-logistic model. Second, as mentioned in Note 3.2, the log-logistic MLE has to be found numerically, and when applied to the hurricane data, the Newton algorithm *fails to converge*. Hence, to guarantee convergence of the iteration, we used robust estimates of μ and σ as starting values for the algorithm, which inadvertantly improved stability of the MLE procedure at the log-logistic model. Results at the lognormal model are more indicative of the non-robust nature of MLE. Third, as was noticed in Section 3.2, highly robust MWM and MTM estimators ($a = b = 14/30$) produce nearly identical estimates of μ at both models, but their inefficiency for estimating σ makes its estimates volatile. This divergence vanishes when we choose more efficient MWM and MTM estimators ($a = b = 1/30$). Fourth, in terms of goodness-of-fit, major differences between MWM and MTM emerge only for $a = b = 14/30$, all other cases produce nearly identical model fits, which are not much different from those of MLE (for original data). We interpret this occurrence as a coincidence that was observed for one data set.

5.2 Actuarial Premiums

Let us consider estimation of the loss 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 dollars with a maximum benefit of 20 billion dollars. That is,

$$Z = \begin{cases} 0, & \text{if } X \leq x_1; \\ X - x_1, & \text{if } x_1 < X \leq x_2; \\ x_2 - x_1, & \text{if } X > x_2, \end{cases} \tag{5.2}$$

and, if X follows the distribution function F, we seek

$$\mathit{\Pi}[F] \;=\; \mathbb{E}[Z] \;=\; \int_{x_1}^{x_2} (x - x_1)\, dF(x) + (x_2 - x_1)[1 - F(x_2)], \tag{5.3}$$

where $x_1 = 5 \times 10^9$ and $x_2 = 25 \times 10^9$.

Since it is now most important that our fitted distribution captures the behavior of the underlying damage distribution between x_1 and x_2, the MWM and MTM estimators are most natural with the choices $a = 8/30$ (which corresponds to the proportion of observations below x_1) and $b = 3/30$ (which corresponds to the proportion of observations above x_2). We denote these MWM and MTM estimators as W3 and T3, respectively.

As can be seen from Table 5.2, for each model, the overall W3 and T3 fits are very similar to those of W2, T2 and MLE, but they yield a closer fit than those three procedures over the layer of interest, which is $[x_1; x_2]$. Further, in Table 5.2 we also provide the actuarial premiums calculated using equation (5.3) for each fitted model and compare them with the empirical premium $\Pi[\widehat{F}_n]$, where $\widehat{F}_n$ denotes the empirical distribution function. In addition, Table 5.2 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 (5.3) together with the MWM, MTM and MLE asymptotic distributions, which have been discussed in Sections 3.2 and 3.3. For constructing the empirical interval, we use the classical central limit theorem and have that

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

where $V[F]$ is derived from equations (5.2) and (5.3). That is,

$$V[F] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \int_{x_1}^{x_2} (x - x_1)^2 \, dF(x) + (x_2 - x_1)^2[1 - F(x_2)] - \big(\Pi[F]\big)^2,$$

which is estimated by replacing F with $\widehat{F}_n$.

TABLE 5.2. Parameter estimates, goodness-of-fit measurements (rFIT, defined by (5.1) but restricted to the data in $[x_1;\, x_2]$), and actuarial premiums for the layer $[x_1;\, x_2]$ with 95% confidence intervals in parentheses.

Estimator	Proportion		Lognormal Model				Log-logistic Model			
	a	b	$\widehat{\mu}$	$\widehat{\sigma}$	rFIT	PREMIUM	$\widehat{\mu}$	$\widehat{\sigma}$	rFIT	PREMIUM
MLE	–	–	22.80	0.83	0.054	5.60 (3.37; 7.84)	22.78	0.48	0.045	5.29 (3.02; 7.60)
W1	14/30	14/30	22.76	0.99	0.105	5.86 (0.86; 10.86)	22.76	0.62	0.117	5.96 (1.23; 10.69)
T1	14/30	14/30	22.76	1.67	0.412	7.34 (2.55; 12.13)	22.76	1.05	0.433	7.37 (2.71; 12.03)
W2	1/30	1/30	22.78	0.82	0.050	5.38 (3.17; 7.60)	22.78	0.47	0.044	5.26 (2.98; 7.54)
T2	1/30	1/30	22.77	0.85	0.057	5.44 (3.17; 7.70)	22.77	0.50	0.050	5.36 (3.06; 7.65)
W3	8/30	3/30	22.83	0.75	0.046	5.49 (3.26; 7.72)	22.83	0.45	0.041	5.46 (3.19; 7.74)
T3	8/30	3/30	22.80	0.77	0.042	5.34 (3.07; 7.61)	22.80	0.46	0.040	5.37 (3.11; 7.63)

NOTE: The empirical point and interval estimates of the pure premium are 5.42 and (3.11; 7.72), respectively. Point and interval estimates of the actuarial premiums are measured in billions.

As Table 5.2 suggests, the MWM and MTM estimators with appropriate proportions a and b (i.e., the estimators W2, T2 and W3, T3) lead to premium estimates that are closer to the empirical estimate than those obtained with highly robust but inefficient (i.e., W1, T1) or highly efficient but non-robust estimators (i.e., MLE). The best fits over the restricted range, which are almost identical, are achieved by the estimators that are constructed to be closest to the data over the interval of interest (i.e., W3, T3). In this example, we again observe substantial differences between MWM and

MTM fits only for $a = b = 14/30$. Due to similar theoretical properties and design, the other MWM and MTM estimators have practically the same rFIT values for this data set. In addition, note the remarkable stability of point and interval estimates of the pure premium based on W3 and T3 when the distributional assumption is changed from lognormal to log-logistic. In both cases, the point estimate changes only about 0.5%. Also, the main advantage of parametric procedures (MWM, MTM, and MLE) over the empirical approach is that in general they produce shorter confidence intervals for the measures of interest, though the advantage is minimal in the current example. In summary, the illustration we have provided in this section exemplifies the idea that the MWM and MTM estimators are an appropriate choice for various model-fitting situations including those when a close fit in one or both tails of the distribution is not required.

6 Concluding Remarks

In this paper, we have introduced and developed a new method for estimating the parameters of continuous distributions: the method of Winsorized moments (MWM). The method utilizes the underlying principle of the classical method-of-moments and its actions on data are easily understood, which is a most appealing feature for practitioners. We have described the asymptotic properties of the MWM estimators, provided examples of estimators for location-scale families and several insurance loss models—lognormal and log-logistic—and compared MWM with its main competitor, the method of trimmed moments (MTM).

Further, as was demonstrated theoretically and via simulations, both methods are equally straightforward computationally and possess identical (global) robustness properties if the same proportions a and b are used for MWM and MTM estimators. In terms of efficiency, the Winsorized estimators outperform MTMs when a and b are large, but for smaller proportions (e.g., inside the square $\{(a, b)\colon\ 0 \leq a \leq 0.10,\ 0 \leq b \leq 0.10\}$, where typically the highest point of efficieny is reached) there is no consistent winner as the outcome depends on the underlying distribution. For additional perspectives on the finite-sample performance of these estimators, see Zhao, Brazauskas, Ghorai (2017).

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 lognormal and log-logistic models are fitted to the data using the MWM, MTM, and MLE methods. The real-data study reveals that calculating the premiums for the layers of insurance coverage is a task for which MWM and MTM are particularly natural.

Acknowledgements

The authors are very appreciative of valuable insights and useful comments provided by two anonymous referees, which helped to substantially improve the paper.

References

[1] Brazauskas, V. (2009). Robust and efficient fitting of loss models: diagnostic tools and insights. *North American Actuarial Journal*, **13**(3), 1–14.

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

[3] Brazauskas, V., Jones, B., and 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.

[4] Brazauskas, V. and Kleefeld, A. (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics*, **45**(3), 424–435.

[5] Brazauskas, V. and Kleefeld, A. (2011). Folded and log-folded-t distributions as models for insurance loss data. *Scandinavian Actuarial Journal*, **2011**(1), 59–79.

[6] Brazauskas, V. and Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution (with discussion). *North American Actuarial Journal*, **4**(4), 12–27. Discussion: **5**(3), 123–126. Reply: **5**(3), 126–128.

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

[8] Chau, J. (2013). *Robust Estimation In Operational Risk Modeling*. M.S. Thesis, Utrecht University.

[9] Chernoff, H., Gastwirth, J.L., and 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. and Stine, R.A. (1986). A note on the information matrix for a logistic distribution. *The American Statistician*, **40**(3), 220–222.

[12] Dell'Aquila, R. and Embrechts, P. (2006). Extremes and robustness: a contradiction? *Financial Markets and Portfolio Management*, **20**(1), 103–118.

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

[14] Dornheim, H. and Brazauskas, V. (2014). Case studies using credibility and corrected adaptively truncated likelihood methods. *Variance*, **7**(2), 168–192.

[15] Gisler, A. and Reinhard, P. (1993). Robust credibility. *ASTIN Bulletin*, **23**(1), 118–143.

[16] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(4), 1029–1054.

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

[18] Horbenko, N., Ruckdeschel, P., and Bae, T. (2011). Robust estimation of operational risk. *Journal of Operational Risk*, **6**(2), 3–30.

[19] Huber, P.J. and Ronchetti, E.M. (2009). *Robust Statistics*, 2nd edition. Wiley, Hoboken.

[20] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2, 2nd edition. Wiley, New York.

[21] Kim, J.H.T. and Jeon, Y. (2013). Credibility theory based on trimming. *Insurance: Mathematics and Economics*, **53**(1), 36–47.

[22] Kleefeld, A. and Brazauskas, V. (2012). A statistical application of the quantile mechanics approach: MTM estimators for the parameters of t and gamma distributions. *European Journal of Applied Mathematics*, **23**(5), 593–610.

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

[24] Klugman, S.A., Panjer, H.H., and Willmot, G.E. (2012). *Loss Models: From Data to Decisions*, 4th edition. Wiley, New York.

[25] Ko, S-J. and Lee, Y.H. (1991). Theoretical analysis of Winsorizing smoothers and their applications to image processing. *Proceedings of 1991 International Conference on Acoustics, Speech, and Signal Processing*, 3001–3004. IEEE.

[26] Künsch, H.R. (1992). Robust methods for credibility. *ASTIN Bulletin*, **22**(1), 33–49.

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

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

[29] Nadarajah, S. and Bakar, S.A.A. (2015). New folded models for the log-transformed Norwegian fire claim data. *Communications in Statistics: Theory and Methods*, **44**(20), 4408–4440.

[30] Opdyke, J.D. and Cavallo, A. (2012). Estimating operational risk capital: the challenges of truncation, the hazards of MLE, and the promise of robust statistics. *Journal of Operational Risk*, **7**(3), 3–90.

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

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

[33] 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.

[34] Van Kerm, P. (2007). Extreme incomes and the estimation of poverty and inequality indicators from EU-SILC. *IRISS Working Paper 2007-01*, An Integrated Research Infrastructure in the Socio-Economic Sciences, 1–51.

[35] Zhao, Q., Brazauskas, V., and Ghorai, J. (2017). Small-sample performance of the MTM and MWM estimators for the parameters of log-location-scale families. *Submitted for publication*.

Appendix

Lemma A.1. *If the Winsorizing proportions satisfy* $0 \le a = a_i = a_j < 1 - b_i = 1 - b_j = 1 - b \le 1$, *then the entries of* $\mathbf{\Sigma}$ *in (2.6) are found by adding the following four terms:*

$$
\begin{aligned}
\widehat{A}_{i,j}^{(1)} &= \int_a^{1-b}\int_a^{1-b} H_i^{'}(w)H_j^{'}(v)\big[\min(v,w) - v\,w\big]\,dv\,dw \\
&= aH_i(a)H_j(a) + bH_i(1-b)H_j(1-b) - \Delta_i\Delta_j + \int_a^{1-b} H_i(w)H_j(w)\,dw\,,
\end{aligned}
$$

$$
\begin{aligned}
\widehat{A}_{i,j}^{(2)} &= \frac{a}{1-a}\big[a(1-a)H_j^{'}(a) + b^2H_j^{'}(1-b)\big]\int_a^{1-b} H_i^{'}(w)(1-w)\,dw \\
&\quad + b^2H_j^{'}(1-b)\int_a^{1-b}\int_u^{1-b} \frac{H_i^{'}(w)(1-w)}{(1-u)^2}\,dw\,du \\
&= \Delta_i\big[a^2H_j^{'}(a) - b^2H_j^{'}(1-b)\big] + b^2H_i(1-b)H_j^{'}(1-b) - a^2H_i(a)H_j^{'}(a)\,,
\end{aligned}
$$

$$
\begin{aligned}
\widehat{A}_{i,j}^{(3)} &= \frac{a}{1-a}\big[a(1-a)H_i^{'}(a) + b^2H_i^{'}(1-b)\big]\int_a^{1-b} H_j^{'}(v)(1-v)\,dv \\
&\quad + b^2H_i^{'}(1-b)\int_a^{1-b}\int_u^{1-b} \frac{H_j^{'}(v)(1-v)}{(1-u)^2}\,dv\,du \\
&= \Delta_j\big[a^2H_i^{'}(a) - b^2H_i^{'}(1-b)\big] + b^2H_j(1-b)H_i^{'}(1-b) - a^2H_j(a)H_i^{'}(a) \;=\; \widehat{A}_{j,i}^{(2)}\,,
\end{aligned}
$$

$$
\begin{aligned}
\widehat{A}_{i,j}^{(4)} &= \frac{a}{1-a}\big[a(1-a)H_i^{'}(a) + b^2H_i^{'}(1-b)\big]\big[a(1-a)H_j^{'}(a) + b^2H_j^{'}(1-b)\big] \\
&\quad + \frac{1-a-b}{(1-a)b}\big[b^2H_i^{'}(1-b)b^2H_j^{'}(1-b)\big] \\
&= a^3(1-a)H_i^{'}(a)H_j^{'}(a) + b^3(1-b)H_i^{'}(1-b)H_j^{'}(1-b) \\
&\quad + a^2b^2\big[H_i^{'}(a)H_j^{'}(1-b) + H_j^{'}(a)H_i^{'}(1-b)\big]\,,
\end{aligned}
$$

where $\Delta_k \equiv \Delta_k(F,a,b) := aH_k(a) + \int_a^{1-b} H_k(v)\,dv + bH_k(1-b), \quad k = i,j.$

Proof: All four terms are derived by applying the integration by parts formula. □

In Section 3.1, the entries of the covariance-variance matrix $\mathbf{\Sigma}$ were expressed in terms of the constants $C_k \equiv C_k(F_0,\, a,\, b)$ and then noted that the latter ones can in turn be expressed in terms of the constants $c_k \equiv c_k(F_0, a, b) := a\left[F_0^{-1}(a)\right]^k + \int_a^{1-b}\left[F_0^{-1}(u)\right]^k du + b\left[F_0^{-1}(1-b)\right]^k$. These expressions are as follows:

$$\begin{aligned}
C_1 &= c_2 - c_1^2 - a\,\frac{\partial\left(c_2 - c_1^2\right)}{\partial a} - b\,\frac{\partial\left(c_2 - c_1^2\right)}{\partial b} \\
&\quad +\ a(1-a)\left(\frac{\partial c_1}{\partial a}\right)^2 + b(1-b)\left(\frac{\partial c_1}{\partial b}\right)^2 - 2ab\,\frac{\partial c_1}{\partial a}\frac{\partial c_1}{\partial b}\,,
\end{aligned}$$

$$\begin{aligned}
2\,C_2 &= c_3 - c_1c_2 - a\,\frac{\partial\left(c_3 - c_1c_2\right)}{\partial a} - b\,\frac{\partial\left(c_3 - c_1c_2\right)}{\partial b} \\
&\quad +\ a(1-a)\frac{\partial c_1}{\partial a}\frac{\partial c_2}{\partial a} + b(1-b)\frac{\partial c_1}{\partial b}\frac{\partial c_2}{\partial b} - ab\left(\frac{\partial c_1}{\partial a}\frac{\partial c_2}{\partial b} + \frac{\partial c_1}{\partial b}\frac{\partial c_2}{\partial a}\right),
\end{aligned}$$

$$\begin{aligned}
4\,C_3 &= c_4 - c_2^2 - a\,\frac{\partial\left(c_4 - c_2^2\right)}{\partial a} - b\,\frac{\partial\left(c_4 - c_2^2\right)}{\partial b} \\
&\quad +\ a(1-a)\left(\frac{\partial c_2}{\partial a}\right)^2 + b(1-b)\left(\frac{\partial c_2}{\partial b}\right)^2 - 2ab\,\frac{\partial c_2}{\partial a}\frac{\partial c_2}{\partial b}\,.
\end{aligned}$$

Consequently, the entries σ_{ij}^2 are as follows:

$$\begin{aligned}
\sigma_{11}^2 &= \sigma^2 C_1\,, \\
\sigma_{12}^2 &= \sigma_{21}^2 \ = \ 2\mu\sigma^2 C_1 + 2\sigma^3 C_2\,, \\
\sigma_{22}^2 &= 4\mu^2\sigma^2 C_1 + 8\mu\sigma^3 C_2 + 4\sigma^4 C_3\,.
\end{aligned}$$