## Robust measures of semiparametric models II: Moments

## **Tuban Lee**

This manuscript was compiled on June 14, 2023

A. Invariant Moments. All popular robust location estimators, such as the symmetric trimmed mean, symmetric Winsorized mean, Hodges-Lehmann estimator, Huber M-estimator, and median of means, are symmetric. As shown in RSSM I, a  $\gamma$ -weighted Hodges-Lehmann mean (WHLM<sub>k, $\epsilon,\gamma$ </sub>) can achieve consistency for the population mean in any  $\gamma$ -symmetric distribution with a finite mean. However, it falls considerably short of consistently handling other parametric distributions that are not  $\gamma$ -symmetric. Shifting from semiparametrics to parametrics, consider a robust estimator with a non-sampledependent breakdown point (defined in Subsection??) which is consistent simultaneously for both a semiparametric distribution and a parametric distribution that does not belong to that semiparametric distribution, it is named with the prefix 'invariant' followed by the name of the population parameter it is consistent with. Here, the recombined I-statistic is defined as

$$RI_{d,h_{\mathbf{k}},\mathbf{k}_{1},\mathbf{k}_{2},k_{1},k_{2},\epsilon_{1},\epsilon_{2},\gamma_{1},\gamma_{2},n,LU_{1},LU_{2}} := \lim_{c \to \infty} \left( \frac{\left(LU_{1}_{h_{\mathbf{k}},\mathbf{k}_{1},k_{1},\epsilon_{1},\gamma_{1},n} + c\right)^{d+1}}{\left(LU_{2}_{h_{\mathbf{k}},\mathbf{k}_{2},k_{2},\epsilon_{2},\gamma_{2},n} + c\right)^{d}} - c \right),$$

where d is the key factor for bias correction,  $LU_{h_{\mathbf{k}},\mathbf{k},k,\epsilon,\gamma,n}$  is the LU-statistic,  $\mathbf{k}$  is the degree of the U-statistic, k is the degree of the LL-statistic,  $\epsilon$  is the upper asymptotic breakdown point of the LU-statistic. It is assumed in this series that in the subscript of an estimator, if  $\mathbf{k}$ , k and  $\gamma$  are omitted,  $\mathbf{k}=1$ , k=1,  $\gamma=1$  are assumed, if just one  $\gamma$  is indicated,  $\gamma_1=\gamma_2$ , if n is omitted, only the asymptotic behavior is considered, in the absence of subscripts, no assumptions are made. The subsequent theorem shows the significance of a recombined n-statistic.

Theorem A.1. Definerecombined11 12  $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n,WL_1,WL_2}$  $RI_{d,h_{\mathbf{k}}=x,\mathbf{k}_{1}=1,\mathbf{k}_{2}=1,k_{1},k_{2},\epsilon_{1},\epsilon_{2},\gamma_{1},\gamma_{2},n,LU_{1}=WL_{1},LU_{2}=WL_{2}}$ 13  $\begin{array}{l} \text{suming finite means, } rm \\ d = \frac{\mu - WL_{1}k_{1}, \epsilon_{1}, \gamma_{1}}{WL_{1}k_{1}, \epsilon_{1}, \gamma_{1}} \frac{1}{WL_{2}k_{2}, \epsilon_{2}, \gamma_{2}}, k_{1}, k_{2}, \epsilon_{1}, \epsilon_{2}, \gamma_{1}, \gamma_{2}, W_{1}, k_{2}, k_{3}, k_{4}, k_{5}, k$ 14 15 where  $\mu$ ,  $WL_{1k_1,\epsilon_1,\gamma_1}$ , and  $WL_{2k_2,\epsilon_2,\gamma_2}$  are different location parameters from that location-scale distribution. If  $\gamma_1 = \gamma_2$ , WL = WHLM, rm is also consistent for any  $\gamma$ -symmetric distributions.

20 Proof. Finding d that make  $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,\mathrm{WL}_1,\mathrm{WL}_2}$  a consistent mean estimator is equivalent for finding the solution of  $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,\mathrm{WL}_1,\mathrm{WL}_2} = \mu$ . First consider the location-scale distribution. Since  $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,\mathrm{WL}_1,\mathrm{WL}_2} = \lim_{c \to \infty} \left( \frac{\left(\mathrm{WL}_{1k_1,\epsilon_1,\gamma_1} + c\right)^{d+1}}{\left(\mathrm{WL}_{2k_2,\epsilon_2,\gamma_2} + c\right)^d} - c \right) = (d+1)\,\mathrm{WL}_{1k_1,\epsilon_1,\gamma} - \frac{1}{2} \left( \frac{\mathrm{WL}_{2k_2,\epsilon_2,\gamma_2} + c}{2} \right)^d + \frac{1}{2} \left( \frac{\mathrm{WL}_{2k_2,\epsilon_2,\gamma_2} + c}{2} \right)^d +$ 

 $d\mathrm{WL}_{2k_2,\epsilon_2,\gamma} = \mu. \quad \mathrm{So}, \ d = \frac{\mu - \mathrm{WL}_{1k_1,\epsilon_1,\gamma_1}}{\mathrm{WL}_{1k_1,\epsilon_1,\gamma_1} - \mathrm{WL}_{2k_2,\epsilon_2,\gamma_2}}. \quad \mathrm{In} \quad \mathrm{RSSM} \, \mathrm{I}, \ \mathrm{it} \ \mathrm{was} \ \mathrm{established} \ \mathrm{that} \ \mathrm{any} \ \mathrm{WL}(k,\epsilon,\gamma) \ \mathrm{can} \ \mathrm{be} \ \mathrm{expressed} \ \mathrm{as} \ \lambda \mathrm{WL}_0(k,\epsilon,\gamma) + \mu \ \mathrm{for} \ \mathrm{a} \ \mathrm{location}\text{-scale} \ \mathrm{distribution} \ \mathrm{parameterized} \ \mathrm{by} \ \mathrm{a} \ \mathrm{location} \ \mathrm{parameter} \ \mu \ \mathrm{and} \ \mathrm{a} \ \mathrm{scale} \ \mathrm{parameter} \ \lambda, \ \mathrm{where} \ \mathrm{WL}_0(k,\epsilon,\gamma) \ \mathrm{is} \ \mathrm{a} \ \mathrm{function} \ \mathrm{of} \ Q_0(p), \ \mathrm{the} \ \mathrm{quantile} \ \mathrm{function} \ \mathrm{of} \ \mathrm{a} \ \mathrm{standard} \ \mathrm{distribution} \ \mathrm{without} \ \mathrm{any} \ \mathrm{shifts} \ \mathrm{or} \ \mathrm{scaling}, \ \mathrm{according} \ \mathrm{to} \ \mathrm{the} \ \mathrm{definition} \ \mathrm{of} \ \mathrm{the} \ \mathrm{weighted} \ L\text{-statistic.} \ \mathrm{The} \ \mathrm{simultaneous} \ \mathrm{cancellation} \ \mathrm{of} \ \mu \ \mathrm{and} \ \lambda \ \mathrm{in} \ \frac{(\lambda\mu_0+\mu)-(\lambda\mathrm{WL}_{10}(k_1,\epsilon_1,\gamma_1)+\mu)}{(\lambda\mathrm{WL}_{10}(k_1,\epsilon_1,\gamma_1)+\mu)-(\lambda\mathrm{WL}_{20}(k_2,\epsilon_2,\gamma_2)+\mu)} \ \mathrm{assures} \ \mathrm{that} \ \mathrm{the} \ d \ \mathrm{in} \ rm \ \mathrm{is} \ \mathrm{always} \ \mathrm{a} \ \mathrm{constant} \ \mathrm{for} \ \mathrm{a} \ \mathrm{location}\text{-scale} \ \mathrm{distribution}. \ \mathrm{The} \ \mathrm{proof} \ \mathrm{of} \ \mathrm{the} \ \mathrm{second} \ \mathrm{assertion} \ \mathrm{follows} \ \mathrm{directly} \ \mathrm{from} \ \mathrm{the} \ \mathrm{coincidence} \ \mathrm{property}. \ \mathrm{According} \ \mathrm{to} \ \mathrm{Theorem} \ \mathrm{18} \ \mathrm{in} \ \mathrm{RSSM} \ \mathrm{I}, \ \mathrm{for} \ \mathrm{any} \ \gamma\text{-symmetric} \ \mathrm{distribution} \ \mathrm{with} \ \mathrm{a} \ \mathrm{finite} \ \mathrm{mean}, \ \mathrm{WHLM}_{1k_1,\epsilon_1,\gamma} = \ \mathrm{WHLM}_{2k_2,\epsilon_2,\gamma} = \mu. \ \mathrm{Then} \ rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\gamma,\mathrm{WHLM}_1,\mathrm{WHLM}_2} = \ \mathrm{lim}_{c\to\infty} \left(\frac{(\mu+c)^{d+1}}{(\mu+c)^d} - c\right) = \mu. \ \mathrm{This} \ \mathrm{completes} \ \mathrm{the} \ \mathrm{demonstration}. \ \ \Box$ 

25 26

29

30

31

32

33

34

36

38

40

41

42

43

46

47

48

49

51

52

53

54

55

 $Q_{Par}\left(p\right) = x_m(1-p)^{-\frac{1}{\alpha}}$ , where  $x_m$  is the minimum possible value that a random variable following the Pareto distribution can take, serving a scale parameter,  $\alpha$  is a shape parameter. The mean of the Pareto distribution is given by  $\frac{\alpha x_m}{\alpha-1}$ . As  $\mathrm{WL}(k,\epsilon,\gamma)$  can be expressed as a function of Q(p), one can set the two  $\mathrm{WL}_{k,\epsilon,\gamma}$ s in the d value of rm as two arbitrary quantiles  $Q_{Par}(p_1)$  and  $Q_{Par}(p_2)$ . For the Pareto distribution,  $d_{Per,rm} = \frac{\mu_{Per} - Q_{Par}(p_1)}{Q_{Par}(p_1) - Q_{Par}(p_2)} = \frac{\frac{\alpha x_m}{\alpha-1} - x_m(1-p_1)^{-\frac{1}{\alpha}}}{x_m(1-p_1)^{-\frac{1}{\alpha}} - x_m(1-p_2)^{-\frac{1}{\alpha}}}$ .  $x_m$  can be canceled out. Intriguingly, the quantile function of exponential distribution is  $Q_{exp}(p) = \ln\left(\frac{1}{1-p}\right)\lambda$ ,  $\lambda \geq 0$ .  $\mu_{exp} = \lambda$ . Then,  $d_{exp,rm} = \frac{\mu_{exp} - Q_{exp}(p_1)}{Q_{exp}(p_1) - Q_{exp}(p_2)} =$ 

For example, the Pareto distribution has a quantile function

 $\lambda \geq 0. \quad \mu_{exp} = \lambda. \quad \text{Then,} \quad d_{exp,rm} = \frac{\mu_{exp} - Q_{exp}(p_1)}{Q_{exp}(p_1) - Q_{exp}(p_2)} = \frac{\lambda - \ln\left(\frac{1}{1-p_1}\right)\lambda}{\ln\left(\frac{1}{1-p_1}\right)\lambda - \ln\left(\frac{1}{1-p_2}\right)\lambda} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}. \quad \text{Since}$   $\lim_{\alpha \to \infty} \frac{\frac{\alpha}{\alpha-1} - (1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha} - (1-p_2)^{-1/\alpha}} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)},$   $d_{Per,rm} \quad \text{approaches} \quad d_{exp,rm}, \quad \text{as} \quad \alpha \to \infty, \quad \text{regardless of the type of weighted $L$-statistic used.} \quad \text{That}$ 

## Significance Statement

Bias, variance, and contamination are the three main errors in statistics. Consistent robust estimation is unattainable without parametric assumptions. In this article, invariant moments are proposed as a means of achieving near-consistent and robust estimations of moments, even in scenarios where moderate violations of distributional assumptions occur, while the variances are sometimes smaller than those of the sample moments.

T.L. designed research, performed research, analyzed data, and wrote the paper.

The author declares no competing interest.

<sup>&</sup>lt;sup>1</sup>To whom correspondence should be addressed. E-mail: tl@biomathematics.org

for the Weibull. gamma. Pareto. logmeans. 57 generalized distribution, normal and Gaussian 58  $n_{d=\frac{\mu-\text{WHLM}_{1}_{k_{1},\epsilon_{1},\gamma}}{\text{WHLM}_{1}_{k_{1},\epsilon_{1},\gamma}-\text{WHLM}_{2}_{k_{2},\epsilon_{2},\gamma}},k_{1},k_{2},\epsilon_{1},\epsilon_{2},\gamma,\text{WHLM}_{1},\text{WHLM}_{2}}$  consistent for at least one particular case, where and WHLM<sub> $2k_2,\epsilon_2,\gamma$ </sub>  $WHLM_{1k_1,\epsilon_1,\gamma}$ , aredifferent location parameters from an exponential dis-62 Let WHLM<sub>1 $k_1,\epsilon_1,\gamma$ </sub>  $BM_{\nu=3,\epsilon=\frac{1}{24}},$ 63 WHLM<sub> $2k_2,\epsilon_2,\gamma$ </sub> = m, then  $\mu = \lambda$ ,  $m = Q(\frac{1}{2}) = \ln 2\lambda$ , 64  $BM_{\nu=3,\epsilon=\frac{1}{24}} = \lambda \left( 1 + \ln \left( \frac{26068394603446272 \sqrt[6]{\frac{7}{247}} \sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}} \right) \right)$ the detailed formula is given in the SI Text. So, d = $\lambda - \lambda \left(1 + \ln\left(\frac{26068394603446272 \sqrt[6]{\frac{7}{247}} \sqrt[3]{11}}{391^{5/6} 101898752449325\sqrt{5}}\right)\right)$  $\lambda \left(1 + \ln\left(\frac{26068394603446272\sqrt{\frac{6}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right) - \ln 2\lambda$  $26068394603446272 \sqrt[6]{\frac{7}{247}} \sqrt[3]{11}$  $391^{5/6}101898752449325\sqrt{5}$  $-\approx 0.103$ . The biases of  $1 - \ln(2) + \ln\left(\frac{26068394603446272\sqrt{6}\sqrt{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)$ 

 $rm_{d\approx 0.103, \nu=3, \epsilon_1=\frac{1}{24}, \epsilon_2=\frac{1}{2}, \mathrm{BM}, m}$  for distributions with skewness between those of the exponential and symmetric distributions are tiny (SI Dataset S1).  $rm_{d\approx 0.103, \nu=3, \epsilon_1=\frac{1}{24}, \epsilon_2=\frac{1}{2}, {\rm BM}, m}$ exhibits excellent performance for all these common unimodal distributions (SI Dataset S1).

The recombined mean is an recombined I-statistic. Consider an I-statistic whose LEs are percentiles of a distribution obtained by plugging LU-statistics into a cumulative distribution function, I is defined with arithmetic operations, constants and quantile functions, such an estimator is classified as a quantile I-statistic. One version of the quantile I-statistic can be defined as  $\mathrm{QI}_{d,h_{\mathbf{k}},\mathbf{k}_1,\mathbf{k}_2,k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n,LU_1,LU_2}\coloneqq$ 

$$\begin{cases} \hat{Q}_{n,h_{\mathbf{k}}}\left(\left(\hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) - \frac{\gamma}{1+\gamma}\right)d + \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right)\right) & \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) \geq \\ \hat{Q}_{n,h_{\mathbf{k}}}\left(\hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) - \left(\frac{\gamma}{1+\gamma} - \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right)\right)d\right) & \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) < \\ \text{where } LU \text{ is } LU_{\mathbf{k},k,\epsilon,\gamma,n}, \, \hat{F}_{n,h_{\mathbf{k}}}\left(x\right) \text{ is the empirical cumulative distribution function of the } h_{\mathbf{k}} \text{ kernel distribution, } \hat{Q}_{n,h_{\mathbf{k}}} \text{ is } \end{cases}$$

the quantile function of the  $h_{\mathbf{k}}$  kernel distribution.

Similarly, the quantile mean can be defined as  $qm_{d,k,\epsilon,\gamma,n,\mathrm{WL}} := \mathrm{QI}_{d,h_{\mathbf{k}}=x,\mathbf{k}=1,k,\epsilon,\gamma,n,LU=\mathrm{WL}}$ . Moreover, in extreme right-skewed heavy-tailed distributions, if the calculated percentile exceeds  $1 - \epsilon$ , it will be adjusted to  $1 - \epsilon$ . In a left-skewed distribution, if the obtained percentile is smaller than  $\gamma \epsilon$ , it will also be adjusted to  $\gamma \epsilon$ . Without loss of generality, in the following discussion, only the case where  $\hat{F}_n\left(\mathrm{WL}_{k,\epsilon,\gamma,n}\right) \geq \frac{\gamma}{1+\gamma}$  is considered. A widely used method for calculating the sample quantile function involves employing linear interpolation of modes corresponding to the order statistics of the uniform distribution on the interval [0, 1], i.e.,  $\hat{Q}_n(p) = X_{\lfloor h \rfloor} + (h - \lfloor h \rfloor) \left( X_{\lceil h \rceil} - X_{\lfloor h \rfloor} \right), \ h = (n-1) p + 1.$ To minimize the finite sample bias, here, the inverse function of  $\hat{Q}_n$  is deduced as  $\hat{F}_n(x) := \frac{1}{n-1} \left( cf - 1 + \frac{x - X_{cf}}{X_{cf+1} - X_{cf}} \right)$ , where  $cf = \sum_{i=1}^{n} \mathbf{1}_{X_i \leq x}$ ,  $\mathbf{1}_A$  is the indicator of event A. The quantile mean uses the location-scale invariant in a different

Theorem A.2.  $qm_{d=\frac{F(\mu)-F(WL_{k,\epsilon,\gamma})}{F(WL_{k,\epsilon,\gamma})-\frac{\gamma}{1+\gamma},k,\epsilon,\gamma,\mathrm{WL}}}$  is a consistent

way, as shown in the subsequent proof.

mean estimator for a location-scale distribution provided that the means are finite and  $F(\mu)$ ,  $F(WL_{k,\epsilon,\gamma})$  and  $\frac{\gamma}{1+\gamma}$  are all within the range of  $[\gamma \epsilon, 1 - \epsilon]$ , where  $\mu$  and  $WL_{k,\epsilon,\gamma}$  are location parameters from that location-scale distribution. If WL = WHLM, qm is also consistent for any  $\gamma$ -symmetric distributions.

108

112

113

114

115

116

117

119

120

121

122

123

124

125

126

127

128

129

131 132

134

138

139

141

142

143

*Proof.* When  $F(WL_{k,\epsilon,\gamma}) \geq \frac{\gamma}{1+\gamma}$ , the solution of  $\left(F(WL_{k,\epsilon,\gamma}) - \frac{\gamma}{1+\gamma}\right)d + F(WL_{k,\epsilon,\gamma}) = F(\mu)$  is  $d = \frac{F(\mu) - F(WL_{k,\epsilon,\gamma})}{F(WL_{k,\epsilon,\gamma}) - \frac{\gamma}{1+\gamma}}$ . The d value for the case where  $F(\mathrm{WL}_{k,\epsilon,\gamma,n}) < \frac{\gamma}{1+\gamma}$  is the same. The definitions of the location and scale parameters are such that they must satisfy  $F(x; \lambda, \mu) = F(\frac{x-\mu}{\lambda}; 1, 0)$ , then  $F(\text{WL}(k, \epsilon, \gamma); \lambda, \mu) =$  $F(\frac{\lambda WL_0(k,\epsilon,\gamma)+\mu-\mu}{\lambda};1,0) = F(WL_0(k,\epsilon,\gamma);1,0).$  It follows that the percentile of any weighted L-statistic is free of  $\lambda$  and  $\mu$  for a location-scale distribution. Therefore d in qm is also invariably a constant. For the  $\gamma$ -symmetric case,  $F(WHLM_{k,\epsilon,\gamma}) = F(\mu) = F(Q(\frac{\gamma}{1+\gamma})) = \frac{\gamma}{1+\gamma}$ is valid for any  $\gamma$ -symmetric distribution with a finite second moment, as the same values correspond to same percentiles. Then,  $qm_{d,k,\epsilon,\gamma,WHLM}$  $F^{-1}\left(\left(F\left(WHLM_{k,\epsilon,\gamma}\right)-\frac{\gamma}{1+\gamma}\right)d+F\left(\mu\right)\right)$  $F^{-1}(0+F(\mu)) = \mu$ . To avoid inconsistency due to post-adjustment,  $F(\mu)$ ,  $F(\mathrm{WL}_{k,\epsilon,\gamma})$  and  $\frac{\gamma}{1+\gamma}$  must reside within the range of  $[\gamma \epsilon, 1 - \epsilon]$ . All results are now proven.  $\square$ 

The cdf of the Pareto distribution is  $F_{Par}(x)$  $1 - \left(\frac{x_m}{x}\right)^{\alpha}$ . So, set the d value in qm with two arbitrary percentiles  $p_1$  and  $p_2$ ,  $d_{Par,qm}$ 

$$\frac{1 - \left(\frac{x_m}{\frac{\alpha x_m}{\alpha - 1}}\right)^{\alpha} - \left(1 - \left(\frac{x_m}{x_m(1 - p_1)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)}{\left(1 - \left(\frac{x_m}{x_m(1 - p_1)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right) - \left(1 - \left(\frac{x_m}{x_m(1 - p_2)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)} = 130$$

The d value in qm for the exponential  $\begin{cases} \hat{Q}_{n,h_{\mathbf{k}}}\left(\left(\hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) - \frac{\gamma}{1+\gamma}\right)d + \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right)\right) & \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) \geq \frac{\gamma}{1+\text{distribution}} & \text{is always identical to } d_{Par,qm} & \text{as } \alpha \to \infty, \\ \hat{Q}_{n,h_{\mathbf{k}}}\left(\hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) - \left(\frac{\gamma}{1+\gamma} - \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right)\right)d\right) & \hat{F}_{n,h_{\mathbf{k}}}\left(LU\right) < \frac{\gamma}{1+\gamma} & \text{ince } \lim_{\alpha \to \infty} \left(\frac{\alpha-1}{\alpha}\right)^{\alpha} = \frac{1}{e} & \text{and the cdf of the exponential} \end{cases}$ distribution is  $F_{exp}(x) = 1 - e^{-\lambda^{-1}x}$ , then  $d_{exp,qm} =$ 

$$\frac{\left(1-e^{-1}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)}{\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_2}\right)}\right)} = \frac{1-\frac{1}{e}-p_1}{p_1-p_2}. \text{ So, for the}$$
Weibull, gamma, Pareto, lognormal and generalized Gaus-

Weibull, gamma, Pareto, lognormal and generalized Gaussian distribution,  $qm_{d=\frac{F_{exp}(\mu)-F_{exp}(\mathbf{WHLM}_{k,\epsilon,\gamma})}{F_{exp}(\mathbf{WHLM}_{k,\epsilon,\gamma})-\frac{\gamma}{1+\gamma}},k,\epsilon,\gamma,\mathbf{WHLM}}$ 

is also consistent for at least one particular case, provided that  $\mu$  and WHLM<sub>k, $\epsilon$ , $\gamma$ </sub> are different location parameters from an exponential distribution and  $F(\mu)$ ,  $F(\text{WHLM}_{k,\epsilon,\gamma})$  and  $\frac{\gamma}{1+\gamma}$  are all within the range of  $[\gamma \epsilon, 1 - \epsilon]$ . Also let  $WHLM_{k,\epsilon,\gamma} = BM_{\nu=3,\epsilon=\frac{1}{24}}$  and  $\mu = \lambda$ , then  $d = \frac{F_{exp}(\mu) - F_{exp}(BM_{\nu=3,\epsilon=\frac{1}{24}})}{F_{exp}(BM_{\nu=3,\epsilon=\frac{1}{24}}) - \frac{1}{2}} =$ 

and 
$$\mu = \lambda$$
, then  $d = \frac{F_{exp}(\mu) - F_{exp}(\text{BM}_{\nu=3,\epsilon=\frac{1}{24}})}{F_{exp}(\text{BM}_{\nu=3,\epsilon=\frac{1}{24}}) - \frac{1}{2}} =$ 

$$-e^{-1} + e^{-\left(1 + \ln\left(\frac{26068394603446272\sqrt{\frac{6}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)} = -\left(1 + \ln\left(\frac{26068394603446272\sqrt{\frac{6}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)$$

$$\frac{\frac{101898752449325\sqrt{5}\sqrt{6}\sqrt{\frac{247}{4}}391^{5/6}}{26068394603446272\sqrt[3]{11e}} - \frac{1}{e}}{\frac{1}{2} - \frac{101898752449325\sqrt{5}\sqrt{6}\sqrt{\frac{247}{4}}391^{5/6}}{26068394603446272\sqrt[3]{11e}}} \approx 0.088. \quad F_{exp}(\mu), \quad 1.088$$

 $F_{exp}(\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}})$  and  $\frac{1}{2}$ are all within the range of  $[\frac{1}{24},\frac{23}{24}].~qm_{d\approx0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$  works better in the fat-tail scenarios (SI Dataset S1). Theorem A.1 and A.2 show that  $rm_{d\approx 0.103, \nu=3, \epsilon_1=\frac{1}{24}, \epsilon_2=\frac{1}{2}, \text{BM}, m}$  and  $qm_{d\approx 0.088, \nu=3, \epsilon=\frac{1}{24}, \text{BM}}$ 

2 |

70

72

73

75

76

77

80

82

83

84

85

86

87

88

89

91

92

93

94

95

96

97

98

100

101

102

103

104

are both consistent mean estimators for any symmetric distribution and the exponential distribution with finite second moments. It's obvious that the asymptotic breakdown points of  $rm_{d\approx 0.103, \nu=3, \epsilon_1=\frac{1}{24}, \epsilon_2=\frac{1}{2}, \mathrm{BM}, m}$  and  $qm_{d\approx 0.088, \nu=3, \epsilon=\frac{1}{24}, \mathrm{BM}}$  are both  $\frac{1}{24}$ . Therefore they are all invariant means.

150

151

152

153

154

155

156

157

158

159

160

161 162

163

164

165

167

168

169

170

171

172

173

175

176

177

180

181

182

183

184

To study the impact of the choice of WLs in rm and qm, it is constructive to recall that a weighted L-statistic is a combination of order statistics. While using a less-biased weighted L-statistic can generally enhance performance (SI Dataset S1), there is a greater risk of violation in the semiparametric framework. However, the mean-WA  $_{\epsilon,\gamma}$ - $\gamma$ -median inequality is robust to slight fluctuations of the QA function of the underlying distribution. Suppose for a right-skewed distribution, the QA function is generally decreasing with respect to  $\epsilon$  in [0, u], but increasing in  $[u, \frac{1}{1+\gamma}]$ , since all quantile averages with breakdown points from  $\epsilon$  to  $\frac{1}{1+\gamma}$  will be included in the computation of  $\text{WA}_{\epsilon,\gamma}$ , as long as  $\frac{1}{1+\gamma} - u \ll \frac{1}{1+\gamma} - \gamma \epsilon$ , and other portions of the QA function satisfy the inequality constraints that define the  $\nu$ th  $\gamma$ -orderliness on which the WA<sub> $\epsilon,\gamma$ </sub> is based, if  $0 \le \gamma \le 1$ , the mean-WA<sub> $\epsilon, \gamma$ </sub>- $\gamma$ -median inequality still holds. This is due to the violation of  $\nu$ th  $\gamma$ -orderliness being bounded, when  $0 \le \gamma \le 1$ , as shown in RMSM I and therefore cannot be extreme for unimodal distributions with finite second moments. For instance, the SQA function of the Weibull distribution is non-monotonic with respect to  $\epsilon$  when the shape parameter  $\alpha>\frac{1}{1-\ln(2)}\approx$  3.259 as shown in the SI Text of RMSM I, the violation of the second and third orderliness starts near this parameter as well, yet the mean-BM  $_{\nu=3,\epsilon=\frac{1}{24}}\text{-median}$  in equality retains valid when  $\alpha \leq 3.387$ . Another key factor in determining the risk of violation is the skewness of the distribution. In RSM I, it was demonstrated that in a family of distributions differing by a skewness-increasing transformation in van Zwet's sense, the violation of orderliness, if it happens, often only occurs as the distribution nears symmetry (1)

1. WR van Zwet, Convex Transformations of Random Variables: Nebst Stellingen. (1964).