# Robust measures of semiparametric models II: Moments

**Tuban Lee**

This manuscript was compiled on June 16, 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 ($\text{WHLM}_{k,\epsilon,\gamma}$) 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-sample-dependent 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

$$\text{RI}_{d,h_{\mathbf{k}},\mathbf{k}_1,\mathbf{k}_2,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,n,LU_1,LU_2} := \lim_{c\to\infty}\left(\frac{\left(LU_{1h_{\mathbf{k}},\mathbf{k}_1,k_1,\epsilon_1,\gamma_1,n}+c\right)^{d+1}}{\left(LU_{2h_{\mathbf{k}},\mathbf{k}_2,k_2,\epsilon_2,\gamma_2,n}+c\right)^{d}}-c\right),$$

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 $\mathbf{k}$ is indicated, $\mathbf{k}_1=\mathbf{k}_2$, 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 $I$-statistic.

**Theorem A.1.** *Define the recombined mean as* $rm_{d,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,n,WL_1,WL_2} := RI_{d,h_{\mathbf{k}}=x,\mathbf{k}_1=1,\mathbf{k}_2=1,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,n,LU_1=WL_1,LU_2=WL_2}$. *Assuming finite means,* $rm_{d=\frac{\mu-WL_{1k_1,\epsilon_1,\gamma_1}}{WL_{1k_1,\epsilon_1,\gamma_1}-WL_{2k_2,\epsilon_2,\gamma_2}},k_1,k_2,\epsilon=\min(\epsilon_1,}$ *is a consistent mean estimator for a location-scale distribution, 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.*

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

For example, the Pareto distribution has a quantile function $Q_{Par}(p) = x_m(1-p)^{-\frac{1}{\alpha}}$, 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 $\text{WL}(k,\epsilon,\gamma)$ can be expressed as a function of $Q(p)$, one can set the two $\text{WL}_{k,\epsilon,\gamma}$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, $d_{exp,rm} = \frac{\mu_{exp}-Q_{exp}(p_1)}{Q_{exp}(p_1)-Q_{exp}(p_2)} = \frac{\lambda-\ln\left(\frac{1}{1-p_1}\right)\lambda}{\ln\left(\frac{1}{1-p_1}\right)\lambda-\ln\left(\frac{1}{1-p_2}\right)\lambda} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$. Since $\lim_{\alpha\to\infty}\frac{\frac{\alpha}{\alpha-1}-(1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha}-(1-p_2)^{-1/\alpha}} = -\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$,


## 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.

[1] To whom correspondence should be addressed. E-mail: tl@biomathematics.org

$d_{Per,rm}$ approaches $d_{exp,rm}$, as $\alpha \to \infty$, regardless of the type of weighted $L$-statistic used. That means, for the Weibull, gamma, Pareto, lognormal and generalized Gaussian distribution, $rm_{d=\frac{\mu-\mathrm{WHLM1}_{k_1,\epsilon_1,\gamma}}{\mathrm{WHLM1}_{k_1,\epsilon_1,\gamma}-\mathrm{WHLM2}_{k_2,\epsilon_2,\gamma}},k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma,\mathrm{WHLM}_1,\mathrm{WHLM}_2}$ is consistent for at least one particular case, where $\mu$, $\mathrm{WHLM1}_{k_1,\epsilon_1,\gamma}$, and $\mathrm{WHLM2}_{k_2,\epsilon_2,\gamma}$ are different location parameters from an exponential distribution. Let $\mathrm{WHLM1}_{k_1,\epsilon_1,\gamma} = \mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}$, $\mathrm{WHLM2}_{k_2,\epsilon_2,\gamma} = m$, then $\mu = \lambda$, $m = Q\left(\frac{1}{2}\right) = \ln 2\lambda$, $\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}} = \lambda\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)$, the detailed formula is given in the SI Text. So, $d = \frac{\mu-\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}}{\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}-m} = \frac{\lambda-\lambda\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)}{\lambda\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)-\ln 2\lambda} = -\frac{\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)}{1-\ln(2)+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)} \approx 0.103$. The biases of $rm_{d\approx0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ for distributions with skewness between those of the exponential and symmetric distributions are tiny (SI Dataset S1). $rm_{d\approx0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ 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 $\mathrm{QI}_{d,h_{\mathbf{k}},\mathbf{k},k,\epsilon,\gamma,,n,LU} := \begin{cases} \hat{Q}_{n,h_{\mathbf{k}}}\left(\left(\hat{F}_{n,h_{\mathbf{k}}}(LU)-\frac{\gamma}{1+\gamma}\right)d+\hat{F}_{n,h_{\mathbf{k}}}(LU)\right) & \hat{F}_{n,h_{\mathbf{k}}}(LU) \geq \frac{\gamma}{1+\gamma} \\ \hat{Q}_{n,h_{\mathbf{k}}}\left(\hat{F}_{n,h_{\mathbf{k}}}(LU)-\left(\frac{\gamma}{1+\gamma}-\hat{F}_{n,h_{\mathbf{k}}}(LU)\right)d\right) & \hat{F}_{n,h_{\mathbf{k}}}(LU) < \frac{\gamma}{1+\gamma} \end{cases}$ where $LU$ is $LU_{\mathbf{k},k,\epsilon,\gamma,n}$, $\hat{F}_{n,h_{\mathbf{k}}}(x)$ is the empirical cumulative distribution function of the $h_{\mathbf{k}}$ kernel distribution, $\hat{Q}_{n,h_{\mathbf{k}}}$ is 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 $\hat{F}_n(\mathrm{WL}_{k,\epsilon,\gamma,n}) \geq \frac{\gamma}{1+\gamma}$ 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 way, as shown in the subsequent proof.

**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 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.*

*Proof.* When $F(\mathrm{WL}_{k,\epsilon,\gamma}) \geq \frac{\gamma}{1+\gamma}$, the solution of $\left(F(\mathrm{WL}_{k,\epsilon,\gamma})-\frac{\gamma}{1+\gamma}\right)d + F(\mathrm{WL}_{k,\epsilon,\gamma}) = F(\mu)$ is $d = \frac{F(\mu)-F(\mathrm{WL}_{k,\epsilon,\gamma})}{F(\mathrm{WL}_{k,\epsilon,\gamma})-\frac{\gamma}{1+\gamma}}$. 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 $F(\mathrm{WL}(k,\epsilon,\gamma);\lambda,\mu) = F(\frac{\lambda\mathrm{WL}_0(k,\epsilon,\gamma)+\mu-\mu}{\lambda};1,0) = F(\mathrm{WL}_0(k,\epsilon,\gamma);1,0)$. 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, $F(\mathrm{WHLM}_{k,\epsilon,\gamma}) = F(\mu) = F(Q(\frac{\gamma}{1+\gamma})) = \frac{\gamma}{1+\gamma}$ is valid for any $\gamma$-symmetric distribution with a finite second moment, as the same values correspond to same percentiles. Then, $qm_{d,k,\epsilon,\gamma,\mathrm{WHLM}} = F^{-1}\left(\left(F(\mathrm{WHLM}_{k,\epsilon,\gamma})-\frac{\gamma}{1+\gamma}\right)d+F(\mu)\right) = F^{-1}(0+F(\mu)) = \mu$. 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 $p_2$, $d_{Par,qm} = \frac{1-\left(\frac{x_m}{\frac{\alpha x_m}{\alpha-1}}\right)^{\alpha}-\left(1-\left(\frac{x_m}{x_m(1-p_1)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)}{\left(1-\left(\frac{x_m}{x_m(1-p_1)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)-\left(1-\left(\frac{x_m}{x_m(1-p_2)^{-\frac{1}{\alpha}}}\right)^{\alpha}\right)} = \frac{1-\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}-p_1}{p_1-p_2}$. The $d$ value in $qm$ for the exponential distribution is always identical to $d_{Par,qm}$ as $\alpha\to\infty$, since $\lim_{\alpha\to\infty}\left(\frac{\alpha-1}{\alpha}\right)^{\alpha} = \frac{1}{e}$ and the cdf of the exponential distribution is $F_{exp}(x) = 1-e^{-\lambda^{-1}x}$, then $d_{exp,qm} = \frac{(1-e^{-1})-\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)}{\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_2}\right)}\right)} = \frac{1-\frac{1}{e}-p_1}{p_1-p_2}$. So, for the Weibull, gamma, Pareto, lognormal and generalized Gaussian distribution, $qm_{d=\frac{F_{exp}(\mu)-F_{exp}(\mathrm{WHLM}_{k,\epsilon,\gamma})}{F_{exp}(\mathrm{WHLM}_{k,\epsilon,\gamma})-\frac{\gamma}{1+\gamma}},k,\epsilon,\gamma,\mathrm{WHLM}}$ is also consistent for at least one particular case, provided that $\mu$ and $\mathrm{WHLM}_{k,\epsilon,\gamma}$ are different location parameters from an exponential distribution and $F(\mu)$, $F(\mathrm{WHLM}_{k,\epsilon,\gamma})$ and $\frac{\gamma}{1+\gamma}$ are all within the range of $[\gamma\epsilon, 1-\epsilon]$. Also let $\mathrm{WHLM}_{k,\epsilon,\gamma} = \mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}$ and $\mu=\lambda$, then $d = \frac{F_{exp}(\mu)-F_{exp}(\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}})}{F_{exp}(\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}})-\frac{1}{2}} = \frac{-e^{-1}+e^{-\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)}}{\frac{1}{2}-e^{-\left(1+\ln\left(\frac{26068394603446272\sqrt[6]{\frac{7}{247}}\sqrt[3]{11}}{391^{5/6}101898752449325\sqrt{5}}\right)\right)}} = \frac{\frac{101898752449325\sqrt{5}\sqrt[6]{\frac{247}{7}}391^{5/6}}{26068394603446272\sqrt[3]{11}e}-\frac{1}{e}}{\frac{1}{2}-\frac{101898752449325\sqrt{5}\sqrt[6]{\frac{247}{7}}391^{5/6}}{26068394603446272\sqrt[3]{11}e}} \approx 0.088$. $F_{exp}(\mu)$, $F_{exp}(\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}})$ 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 $rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ and $qm_{d\approx 0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ 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 $rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ 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.

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-$\mathrm{WA}_{\epsilon,\gamma}$-$\gamma$-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 $\mathrm{WA}_{\epsilon,\gamma}$, 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 $\mathrm{WA}_{\epsilon,\gamma}$ is based, if $0 \leq \gamma \leq 1$, the mean-$\mathrm{WA}_{\epsilon,\gamma}$-$\gamma$-median inequality still holds. This is due to the violation of $\nu$th $\gamma$-orderliness being bounded, when $0 \leq \gamma \leq 1$, 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-$\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}$-median inequality retains valid when $\alpha \leq 3.387$. Another key factor in determining the risk of violation of orderliness 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, only occurs as the distribution nears symmetry (1). When $\gamma = 1$, the over-corrections in $rm$ and $qm$ are dependent on the $\mathrm{SWA}_\epsilon$-median difference, which can be a reasonable measure of skewness after standardization (2, 3), implying that the over-correction is often tiny with moderate $d$. This qualitative analysis suggests the general reliability of $rm$ and $qm$ based on the mean-$\mathrm{WA}_{\epsilon,\gamma}$-$\gamma$-median inequality, especially for unimodal distributions with finite second moments when $0 \leq \gamma \leq 1$. Extending this rationale to other weighted $L$-statistics is possible, since the $\gamma$-$U$-orderliness can also be bounded with certain assumptions, as discussed previously.

Another crucial property of the central moment kernel distribution, location invariant, is introduced in the next theorem. The proof is provided in the SI Text.

**Theorem A.3.** $\psi_{\mathbf{k}}(x_1 = \lambda x_1 + \mu, \cdots, x_{\mathbf{k}} = \lambda x_{\mathbf{k}} + \mu) = \lambda^{\mathbf{k}} \psi_{\mathbf{k}}(x_1, \cdots, x_{\mathbf{k}})$.

A direct result of Theorem A.3 is that, $\mathrm{WHL}\mathbf{k}m$ after standardization is invariant to location and scale. So, the weighted H-L standardized $\mathbf{k}$th moment is defined to be

$$\mathrm{WHL}s\mathbf{k}m_{\epsilon=\min(\epsilon_1,\epsilon_2),k_1,k_2,\gamma_1,\gamma_2,n} \coloneqq \frac{\mathrm{WHL}\mathbf{k}m_{k_1,\epsilon_1,\gamma_1,n}}{(\mathrm{WHL}var_{k_2,\epsilon_2,\gamma_2,n})^{\mathbf{k}/2}}.$$

Consider two continuous distributions belonging to the same location–scale family, according to Theorem A.3, their corresponding $\mathbf{k}$th central moment kernel distributions only differ in scaling. Define the recombined $\mathbf{k}$th central moment as $r\mathbf{k}m_{d,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,n,\mathrm{WHL}\mathbf{k}m_1,\mathrm{WHL}\mathbf{k}m_2} \coloneqq \mathrm{RI}_{d,h_{\mathbf{k}}=\psi_{\mathbf{k}},\mathbf{k}_1=\mathbf{k},\mathbf{k}_2=\mathbf{k},k_1,k_2,\epsilon_1,\epsilon_2,\gamma_1,\gamma_2,n,LU_1=\mathrm{WHL}\mathbf{k}m_1,LU_2=\mathrm{WHL}\mathbf{k}m_2}$. Then, assuming finite $\mathbf{k}$th central moment and applying the same logic as in Theorem A.1, $r\mathbf{k}m_{d=\frac{\mu_{\mathbf{k}}-\mathrm{WHL}\mathbf{k}m_{1\,k_1,\epsilon_1,\gamma_1}}{\mathrm{WHL}\mathbf{k}m_{1\,k_1,\epsilon_1,\gamma_1}-\mathrm{WHL}\mathbf{k}m_{2\,k_2,\epsilon_2,\gamma_2}},k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,\mathrm{WHL}\mathbf{k}m_1,\mathrm{WHL}\mathbf{k}m_2}$ is a consistent $\mathbf{k}$th central moment estimator for a location-scale distribution, where $\mu_{\mathbf{k}}$, $\mathrm{WHL}\mathbf{k}m_{1\,k_1,\epsilon_1,\gamma_1}$, and $\mathrm{WHL}\mathbf{k}m_{2\,k_2,\epsilon_2,\gamma_2}$ are different $\mathbf{k}$th central moment parameters from that location-scale distribution. Similarly, the quantile will not change after scaling. The quantile $\mathbf{k}$th central moment is thus defined as

$$q\mathbf{k}m_{d,k,\epsilon,\gamma,n,\mathrm{WHL}\mathbf{k}m} \coloneqq \mathrm{QI}_{d,h_{\mathbf{k}}=\psi_{\mathbf{k}},\mathbf{k}=\mathbf{k},k,\epsilon,\gamma,n,LU=\mathrm{WHL}\mathbf{k}m}.$$

$q\mathbf{k}m_{d=\frac{F_{\psi_{\mathbf{k}}}(\mu_{\mathbf{k}})-F_{\psi_{\mathbf{k}}}(\mathrm{WHL}\mathbf{k}m_{k,\epsilon,\gamma})}{F_{\psi_{\mathbf{k}}}(\mathrm{WHL}\mathbf{k}m_{k,\epsilon,\gamma})-\frac{\gamma}{1+\gamma}},k,\epsilon,\gamma,\mathrm{WHL}\mathbf{k}m}$ is also a consistent $\mathbf{k}$th central moment estimator for a location-scale distribution provided that the $\mathbf{k}$th central moment is finite and $F_{\psi_{\mathbf{k}}}(\mu_{\mathbf{k}})$, $F_{\psi_{\mathbf{k}}}(\mathrm{WHL}\mathbf{k}m_{k,\epsilon,\gamma})$ and $\frac{\gamma}{1+\gamma}$ are all within the range of $[\gamma\epsilon, 1-\epsilon]$, where $\mu_{\mathbf{k}}$ and $\mathrm{WHL}\mathbf{k}m_{k,\epsilon,\gamma}$ are different $\mathbf{k}$th central moment parameters from that location-scale distribution.

So, the quantile standardized $\mathbf{k}$th moment is defined to be

$$qs\mathbf{k}m_{\epsilon=\min(\epsilon_1,\epsilon_2),k_1,k_2,\gamma_1,\gamma_2,n,\mathrm{WHL}\mathbf{k}m,\mathrm{WHL}var} \coloneqq \frac{q\mathbf{k}m_{d,k_1,\epsilon_1,\gamma_1,n,\mathrm{WHL}\mathbf{k}m}}{(qvar_{d,k_2,\epsilon_2,\gamma_2,n,\mathrm{WHL}var})^{\mathbf{k}/2}}$$

The recombined standardized $\mathbf{k}$th moment $(rs\mathbf{k}m_{\epsilon=\min(\epsilon_1,\epsilon_2),k_1,k_2,\gamma_1,\gamma_2,n,\mathrm{WHL}\mathbf{k}m_1,\mathrm{WHL}\mathbf{k}m_2,\mathrm{WHL}var_1,\mathrm{WHL}var_2})$ is defined similarly and not repeated here. From the better performance of the quantile mean in heavy-tailed distributions, the quantile $\mathbf{k}$th central moments are generally better than recombined $\mathbf{k}$th central moments regarding asymptotic bias.

## B. A shape-scale distribution as the consistent distribution.


1. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
2. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
3. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).