Robust estimations from distribution structures: III. Invariant Moments

Tuobang Li

This manuscript was compiled on November 28, 2023

Descriptive statistics for parametric models are currently highly sensative to departures, gross errors, and/or random errors. Here, leveraging the structures of parametric distributions and their central moment kernel distributions, a class of estimators, consistent simultanously for both a semiparametric distribution and a distinct parametric distribution, is proposed. These efficient estimators are robust to both gross errors and departures from parametric assumptions, making them ideal for estimating the mean and central moments of common unimodal distributions. This article also illuminates the understanding of the common nature of probability distributions and the measures of them.

moments | invariant | unimodal | adaptive estimation | U-statistics

The potential biases of robust location estimators in estimating the population mean have been noticed for more than two centuries (1), with numerous significant attempts made to address them. In calculating a robust estimator, the procedure of identifying and downweighting extreme values inherently necessitates the formulation of distributional assumptions. Previously, it was demonstrated that, due to the presence of infinite-dimensional nuisance shape parameters, the semiparametric approach struggles to consistently address distributions with shapes more intricate than symmetry. Newcomb (1886) provided the first modern approach to robust parametric estimation by developing a class of estimators that gives "less weight to the more discordant observations" (2). In 1964, Huber (3) used the minimax procedure to obtain M-estimator for the contaminated normal distribution, which has played a pre-eminent role in the later development of robust statistics. However, as previously demonstrated, under growing asymmetric departures from normality, the bias of the Huber M-estimator (HM) increases rapidly. This is a common issue in parametric robust statistics. For example, He and Fung (1999) constructed (4) a robust M-estimator (HFM) for the two-parameter Weibull distribution, from which the mean and central moments can be calculated. Nonetheless, it is inadequate for other parametric distributions, e.g., the gamma, Perato, lognormal, and the generalized Gaussian distributions (SI Dataset S1). Another interesting approach is based on L-estimators, such as percentile estimators. For examples of percentile estimators for the Weibull distribution, the reader is referred to the works of Menon (1963) (5), Dubey (1967) (6), Marks (2005) (7), and Boudt, Caliskan, and Croux (2011) (8). At the outset of the study of percentile estimators, it was known that they arithmetically utilize the invariant structures of parametric distributions (5, 6). An estimator is classified as an I-statistic if it asymptotically satisfies $\mathrm{I}(\mathrm{LE}_1, \ldots, \mathrm{LE}_l) = (\theta_1, \ldots, \theta_q)$ for the distribution it is consistent, where LEs are calculated with the use of LU-statistics (defined in Subsection **??**), I is defined using arithmetic operations and constants but may also incorporate transcendental functions and quantile functions, and θs are the population parameters it estimates. In this article, two subclasses of I-statistics are introduced, recombined I-statistics and quantile I-statistics. Based on LU-statistics, I-statistics are naturally robust. Compared to probability density functions (pdfs) and cumulative distribution functions (cdfs), the quantile functions of many parametric distributions are more elegant. Since the expectation of an L-estimator can be expressed as an integral of the quantile function, I-statistics are often analytically obtainable. However, it is observed that even when the sample follows a gamma distribution, which belongs to the same larger family as the Weibull model, the generalized gamma distribution, a misassumption can still lead to substantial biases in Marks percentile estimator (MP) for the Weibull distribution (7) (SI Dataset S1).

On the other hand, while robust estimation of scale has also been intensively studied with established methods (9, 10), the development of robust measures of asymmetry and kurtosis lags behind, despite the availability of several approaches (11–15). The purpose of this paper is to demonstrate that, in light of previous works, the estimation of central moments can be transformed into a location estimation problem by using U-statistics, the central moment kernel distributions possess desirable properties, and by utilizing the invariant structures of unimodal distributions, a suite of robust estimators can be constructed whose biases are typically smaller than the variances (as seen in Table 1 for $n = 5184$).

A. Invariant Moments. Most 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 REDS I, a symmetric weighted Hodges-Lehmann mean ($\mathrm{SWHLM}_{k,\epsilon}$) can achieve consistency for the population mean in any symmetric distribution with a finite mean. However, it falls considerably short of consistently handling other parametric distributions that are not symmetric. Shifting from semiparametrics to parametrics, consider a robust estimator with a non-sample-dependent breakdown point (defined in

Significance Statement

Bias, variance, and contamination are the three main errors in statistics. Consistent robust estimation is unattainable without distributional 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

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

$$\mathrm{RI}_{d,h_{\mathbf{k}},\mathbf{k}_1,\mathbf{k}_2,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,n,LU_1,LU_2} :=$$

$$\lim_{c\to\infty}\left(\frac{\left(LU_{1h_{\mathbf{k}},\mathbf{k}_1,k_1,\epsilon_1,\gamma_1,n}+c\right)^{d+1}}{\left(LU_{2h_{\mathbf{k}},\mathbf{k}_2,k_2,\epsilon_2,\gamma_2,n}+c\right)^{d}}-c\right),$$

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, ϵ 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 γ 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 γ 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,* $rm_{d=\frac{\mu-WL_{1k_1,\epsilon_1,\gamma_1}}{WL_{1k_1,\epsilon_1,\gamma_1}-WL_{2k_2,\epsilon_2,\gamma_2}},k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,WL_1,WL_2}$ *is a consistent mean estimator for a location-scale distribution, where* μ, $WL_{1k_1,\epsilon_1,\gamma_1}$, *and* $WL_{2k_2,\epsilon_2,\gamma_2}$ *are different location parameters from that location-scale distribution. If* $\gamma_1=\gamma_2=1$, *WL = SWHLM, rm is also consistent for any symmetric distributions.*

Proof. Finding d that make $rm_{d,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,\mathrm{WL}_1,\mathrm{WL}_2}$ a consistent mean estimator is equivalent to finding the solution of $rm_{d,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,\mathrm{WL}_1,\mathrm{WL}_2}=\mu$. First consider the location-scale distribution. Since $rm_{d,k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,\mathrm{WL}_1,\mathrm{WL}_2} = \lim_{c\to\infty}\left(\frac{\left(\mathrm{WL}_{1k_1,\epsilon_1,\gamma_1}+c\right)^{d+1}}{\left(\mathrm{WL}_{2k_2,\epsilon_2,\gamma_2}+c\right)^{d}}-c\right) = (d+1)\,\mathrm{WL}_{1k_1,\epsilon_1,\gamma}-d\mathrm{WL}_{2k_2,\epsilon_2,\gamma}=\mu$. So, $d=\frac{\mu-\mathrm{WL}_{1k_1,\epsilon_1,\gamma_1}}{\mathrm{WL}_{1k_1,\epsilon_1,\gamma_1}-\mathrm{WL}_{2k_2,\epsilon_2,\gamma_2}}$. In REDS I, it was established that any $\mathrm{WL}(k,\epsilon,\gamma)$ can be expressed as $\lambda\mathrm{WL}_0(k,\epsilon,\gamma)+\mu$ for a location-scale distribution parameterized by a location parameter μ and a scale parameter λ, where $\mathrm{WL}_0(k,\epsilon,\gamma)$ 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 μ and λ in $\frac{(\lambda\mu_0+\mu)-(\lambda\mathrm{WL}_{1_0}(k_1,\epsilon_1,\gamma_1)+\mu)}{(\lambda\mathrm{WL}_{1_0}(k_1,\epsilon_1,\gamma_1)+\mu)-(\lambda\mathrm{WL}_{2_0}(k_2,\epsilon_2,\gamma_2)+\mu)}$ 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 19 in REDS I, for any symmetric distribution with a finite mean, $\mathrm{SWHLM}_{1k_1}=\mathrm{SWHLM}_{2k_2}=\mu$. Then $rm_{d,k_1,k_2,\epsilon_1,\epsilon_2,\mathrm{SWHLM}_1,\mathrm{SWHLM}_2}=\lim_{c\to\infty}\left(\frac{(\mu+c)^{d+1}}{(\mu+c)^{d}}-c\right)=\mu$. This completes the demonstration. □

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, α 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, $d_{exp,rm}=\frac{\mu_{exp}-Q_{exp}(p_1)}{Q_{exp}(p_1)-Q_{exp}(p_2)}=\frac{\lambda-\ln\left(\frac{1}{1-p_1}\right)\lambda}{\ln\left(\frac{1}{1-p_1}\right)\lambda-\ln\left(\frac{1}{1-p_2}\right)\lambda}=-\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$. Since $\lim_{\alpha\to\infty}\frac{\frac{\alpha}{\alpha-1}-(1-p_1)^{-1/\alpha}}{(1-p_1)^{-1/\alpha}-(1-p_2)^{-1/\alpha}}=-\frac{\ln(1-p_1)+1}{\ln(1-p_1)-\ln(1-p_2)}$, $d_{Per,rm}$ 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, $rm_{d=\frac{\mu-\mathrm{SWHLM}_{1k_1,\epsilon_1}}{\mathrm{SWHLM}_{1k_1,\epsilon_1}-\mathrm{SWHLM}_{2k_2,\epsilon_2}},k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\mathrm{SWHLM}_1,\mathrm{SWHLM}_2}$ is consistent for at least one particular case, where μ, $\mathrm{SWHLM}_{1k_1,\epsilon_1}$, and $\mathrm{SWHLM}_{2k_2,\epsilon_2}$ are different location parameters from an exponential distribution. Let $\mathrm{SWHLM}_{1k_1,\epsilon_1,\gamma}=\mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}$, $\mathrm{SWHLM}_{2k_2,\epsilon_2,\gamma}=m$, 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 $rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ for distributions with skewness between those of the exponential and symmetric distributions are tiny (SI Dataset S1). $rm_{d\approx 0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ exhibits excellent performance for all these common unimodal distributions (SI Dataset S1).

The recombined mean is a 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. The most popular method for computing the sample quantile function was proposed by Hyndman and Fan in 1996 (16). Another 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 $\hat{F}_n(x) := \frac{1}{n}\left(\frac{x - X_{cf}}{X_{cf+1} - X_{cf}} + cf\right)$, based on Hyndman and Fan's definition, or $\hat{F}_n(x) := \frac{1}{n-1}\left(cf - 1 + \frac{x - X_{cf}}{X_{cf+1} - X_{cf}}\right)$, based on the latter definition, where $cf = \sum_{i=1}^{n} \mathbf{1}_{X_i \le x}$, $\mathbf{1}_A$ 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 μ and $WL_{k,\epsilon,\gamma}$ are location parameters from that location-scale distribution. If $WL = SWHLM$, qm is also consistent for any symmetric distributions.*

Proof. When $F(\mathrm{WL}_{k,\epsilon,\gamma}) \ge \frac{\gamma}{1+\gamma}$, 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 λ and μ for a location-scale distribution. Therefore d in qm is also invariably a constant. For the symmetric case, $F(\mathrm{SWHLM}_{k,\epsilon}) = F(\mu) = F(Q(\frac{1}{2})) = \frac{1}{2}$ is valid for any symmetric distribution with a finite second moment, as the same values correspond to same percentiles. Then, $qm_{d,k,\epsilon,\mathrm{SWHLM}} = F^{-1}\left(\left(F(\mathrm{SWHLM}_{k,\epsilon}) - \frac{1}{2}\right)d + F(\mu)\right) = F^{-1}(0 + F(\mu)) = \mu$. 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. □

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

$$d_{exp,qm} = \frac{\left(1-e^{-1}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)}{\left(1-e^{-\ln\left(\frac{1}{1-p_1}\right)}\right)-\left(1-e^{-\ln\left(\frac{1}{1-p_2}\right)}\right)} = \frac{1-\frac{1}{e}-p_1}{p_1-p_2}.$$

So, for the Weibull, gamma, Pareto, lognormal and generalized Gaussian distribution, $qm_{d=\frac{F_{exp}(\mu)-F_{exp}(\mathrm{SWHLM}_{k,\epsilon})}{F_{exp}(\mathrm{SWHLM}_{k,\epsilon})-\frac{1}{2}},k,\epsilon,\mathrm{SWHLM}}$ is also consistent for at least one particular case, provided that μ and $\mathrm{SWHLM}_{k,\epsilon}$ are different location parameters from an exponential distribution and $F(\mu)$, $F(\mathrm{SWHLM}_{k,\epsilon})$ and $\frac{1}{2}$ are all within the range of $[\epsilon, 1-\epsilon]$. Also let $\mathrm{SWHLM}_{k,\epsilon,\gamma} = \mathrm{BM}_{\nu=3,\epsilon=\frac{1}{24}}$ 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 $rm_{d\approx0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ and $qm_{d\approx0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ 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 $rm_{d\approx0.103,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM},m}$ and $qm_{d\approx0.088,\nu=3,\epsilon=\frac{1}{24},\mathrm{BM}}$ 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 ϵ in $[0, u]$, but increasing in $[u, \frac{1}{1+\gamma}]$, since all quantile averages with breakdown points from ϵ 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 νth γ-orderliness on which the $\mathrm{WA}_{\epsilon,\gamma}$ is based, if $0 \le \gamma \le 1$, the mean-$\mathrm{WA}_{\epsilon,\gamma}$-$\gamma$-median inequality still holds. This is due to the violation of νth γ-orderliness being bounded, when $0 \le \gamma \le 1$, as shown in REDS 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 ϵ when the shape parameter $\alpha > \frac{1}{1-\ln(2)} \approx 3.259$ as shown in the SI Text of REDS 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 $\alpha \le 3.387$. Another key factor in determining the risk of violation of orderliness is the skewness of the distribution. In REDS 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 (12). When $\gamma = 1$, the over-corrections in rm and qm are dependent on the SWA_ϵ-median difference, which can be a reasonable measure of skewness after standardization (11, 13), 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 γ-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. $\psi_{\mathbf{k}}\left(x_1 = \lambda x_1 + \mu, \cdots, x_{\mathbf{k}} = \lambda x_{\mathbf{k}} + \mu\right) = \lambda^{\mathbf{k}} \psi_{\mathbf{k}}\left(x_1, \cdots, x_{\mathbf{k}}\right)$.

A direct result of Theorem A.3 is that, 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, $r\mathbf{k}m_{d=\frac{\mu_{\mathbf{k}}-\mathrm{WHL}\mathbf{k}m_{1k_1,\epsilon_1,\gamma_1}}{\mathrm{WHL}\mathbf{k}m_{1k_1,\epsilon_1,\gamma_1}-\mathrm{WHL}\mathbf{k}m_{2k_2,\epsilon_2,\gamma_2}},k_1,k_2,\epsilon=\min(\epsilon_1,\epsilon_2),\gamma_1,\gamma_2,\mathrm{WHL}\mathbf{k}m_1,\mathrm{WHL}\mathbf{k}m_2}$ is a consistent $\mathbf{k}$th central moment estimator for a location-scale distribution, where $\mu_{\mathbf{k}}$, $\mathrm{WHL}\mathbf{k}m_{1k_1,\epsilon_1,\gamma_1}$, and $\mathrm{WHL}\mathbf{k}m_{2k_2,\epsilon_2,\gamma_2}$ 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. According to Theorem ??, if the original distribution is unimodal, the central moment kernel distribution is always a heavy-tailed distribution, as the degree term amplifies its skewness and tailedness. From the better performance of the quantile mean in heavy-tailed distributions, the quantile $\mathbf{k}$th central moments are generally better than the recombined $\mathbf{k}$th central moments regarding asymptotic bias.

Finally, the recombined standardized $\mathbf{k}$th moment is defined to be

$$rs\mathbf{k}m_{\epsilon=\min(\epsilon_1,\epsilon_2),k_1,k_2,k_3,k_4,\gamma_1,\gamma_2,\gamma_3,\gamma_4,n,\mathrm{WHL}\mathbf{k}m_1,\mathrm{WHL}\mathbf{k}m_2,\mathrm{WHL}var_1,\mathrm{WHL}var_2} \coloneqq \frac{r\mathbf{k}m_{d,k_1,k_2,\epsilon_1,\gamma_1,\gamma_2,n,\mathrm{WHL}\mathbf{k}m_1,\mathrm{WHL}\mathbf{k}m_2}}{(rvar_{d,k_3,k_4,\epsilon_2,\gamma_3,\gamma_4,n,\mathrm{WHL}var_1,\mathrm{WHL}var_2})^{\mathbf{k}/2}}.$$

The quantile standardized $\mathbf{k}$th moment is defined similarly,

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

B. A shape-scale distribution as the consistent distribution. In the last section, the parametric robust estimation is limited to a location-scale distribution, with the location parameter often being omitted for simplicity. For improved fit to observed skewness or kurtosis, shape-scale distributions with shape parameter (α) and scale parameter (λ) are commonly utilized. Weibull, gamma, Pareto, lognormal, and generalized Gaussian distributions (when μ is a constant) are all shape-scale unimodal distributions. Furthermore, if either the shape parameter α or the skewness or kurtosis is constant, the shape-scale distribution is reduced to a location-scale distribution. Let $D(|skewness|, kurtosis, \mathbf{k}, etype, dtype, n) = d_{i\mathbf{k}m}$ denote the function to specify d values, where the first input is the absolute value of the skewness, the second input is the kurtosis, the third is the order of the central moment (if $\mathbf{k} = 1$, the mean), the fourth is the type of estimator, the fifth is the type of consistent distribution, and the sixth input is the sample size. For simplicity, the last three inputs will be omitted in the following discussion. Hold in awareness that since skewness and kurtosis are interrelated, specifying d values for a shape-scale distribution only requires either skewness or kurtosis, while the other may be also omitted. Since many common shape-scale distributions are always right-skewed (if not, only the right-skewed or left-skewed part is used for calibration, while the other part is omitted), the absolute value of the skewness should be the same as the skewness of these distributions. This setting also handles the left-skew scenario well.

For recombined moments up to the fourth ordinal, the object of using a shape-scale distribution as the consistent distribution is to find solutions for the system of equations
$$\begin{cases} rm\left(\mathrm{WHLM}, \gamma m, D(|rskew|, rkurt, 1)\right) = \mu \\ rvar\left(\mathrm{WHL}var, \gamma mvar, D(|rskew|, rkurt, 2)\right) = \mu_2 \\ rtm\left(\mathrm{WHL}tm, \gamma mtm, D(|rskew|, rkurt, 3)\right) = \mu_3 \\ rfm\left(\mathrm{WHL}fm, \gamma mfm, D(|rskew|, rkurt, 4)\right) = \mu_4 \\ rskew = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} \\ rkurt = \frac{\mu_4}{\mu_2^2} \end{cases},$$
where μ_2, μ_3 and μ_4 are the population second, third and fourth central moments. $|rskew|$ and $rkurt$ should be the invariant points of the functions $\varsigma(|rskew|) = \left|\frac{rtm(\mathrm{WHL}tm,\gamma mtm,D(|rskew|,3))}{rvar(\mathrm{WHL}var,\gamma mvar,D(|rskew|,2))^{\frac{3}{2}}}\right|$ and $\varkappa(rkurt) = \frac{rfm(\mathrm{WHL}fm,\gamma mfm,D(rkurt,4))}{rvar(\mathrm{WHL}var,\gamma mvar,D(rkurt,2))^2}$. Clearly, this is an overdetermined nonlinear system of equations, given that the skewness and kurtosis are interrelated for a shape-scale distribution. Since an overdetermined system constructed with random coefficients is almost always inconsistent, it is natural to optimize them separately using the fixed-point iteration (see Algorithm 1, only $rkurt$ is provided, others are the same).

Algorithm 1 *rkurt* for a shape-scale distribution

Input: D; WHLvar; WHLfm; $\gamma mvar$; γmfm; $maxit$; δ
Output: $rkurt_{i-1}$

  $i = 0$
  $rkurt_i \leftarrow \varkappa(kurtosis_{max})$ ▷ Using the maximum kurtosis available in D as an initial guess.
  repeat
    $i = i + 1$
    $rkurt_{i-1} \leftarrow rkurt_i$
    $rkurt_i \leftarrow \varkappa(rkurt_{i-1})$
  until $i > maxit$ or $|rkurt_i - rkurt_{i-1}| < \delta$ ▷ $maxit$ is the maximum number of iterations, δ is a small positive number.

The following theorem shows the validity of Algorithm 1.

Theorem B.1. *Assuming $\gamma = 1$ and $m\mathbf{k}ms$, where $2 \le \mathbf{k} \le 4$, are all equal to zero, $|rskew|$ and $rkurt$, defined as the largest attracting fixed points of the functions $\varsigma(|rskew|)$ and $\varkappa(rkurt)$, are consistent estimators of $\tilde{\mu}_3$ and $\tilde{\mu}_4$ for a shape-scale distribution whose $\mathbf{k}$th central moment kernel distributions are U-congruent, as long as they are within the domain of D, where $\tilde{\mu}_3$ and $\tilde{\mu}_4$ are the population skewness and kurtosis, respectively.*

Proof. Without loss of generality, only $rkurt$ is considered, while the logic for $|rskew|$ is the same. Additionally, the second central moments of the underlying sample distribution and consistent distribution are assumed to be 1, with other cases simply multiplying a constant factor according to Theorem A.3. From the definition of D, $\frac{\varkappa(rkurt_D)}{rkurt_D} = \frac{\frac{fm_D - \mathrm{SWHL}fm_D}{\mathrm{SWHL}fm_D - mfm_D}(\mathrm{SWHL}fm - mfm) + \mathrm{SWHL}fm}{rkurt_D\left(\frac{var_D - \mathrm{SWHL}var_D}{\mathrm{SWHL}var_D - mvar_D}(\mathrm{SWHL}var - mvar) + \mathrm{SWHL}var\right)^2}$, where the subscript D indicates that the estimates are from the central moment kernel distributions generated from the consistent distribution, while other estimates are from the underlying distribution of the sample.

Then, assuming the $m\mathbf{k}ms$ are all equal to zero and $var_D = 1$, $\frac{\varkappa(rkurt_D)}{rkurt_D} = \frac{\frac{fm_D - \mathrm{SWHL}fm_D}{\mathrm{SWHL}fm_D}(\mathrm{SWHL}fm) + \mathrm{SWHL}fm}{rkurt_D\left(\frac{\mathrm{SWHL}var}{\mathrm{SWHL}var_D}\right)^2} = \frac{\left(\frac{fm_D - \mathrm{SWHL}fm_D}{\mathrm{SWHL}fm_D} + 1\right)(\mathrm{SWHL}fm)}{fm_D\left(\frac{\mathrm{SWHL}var}{\mathrm{SWHL}var_D}\right)^2} = \frac{\mathrm{SWHL}fm\mathrm{SWHL}var_D^2}{\mathrm{SWHL}fm_D\mathrm{SWHL}var^2} = \frac{\frac{\mathrm{SWHL}fm}{\mathrm{SWHL}var^2}}{\frac{\mathrm{SWHL}fm_D}{\mathrm{SWHL}var_D{}^2}} = \frac{\mathrm{SWHL}kurt}{\mathrm{SWHL}kurt_D}$. Since SWHL$fm_D$ are from the same fourth central moment kernel distribution as $fm_D = rkurt_D var_D{}^2$, according to the definition of U-congruence, an increase in fm_D will also result in an increase in SWHLfm_D. Combining with Theorem A.3, SWHL$kurt$ is a measure of kurtosis that is invariant to location and scale, so $\lim_{rkurt_D \to \infty} \frac{\varkappa(rkurt_D)}{rkurt_D} < 1$. As a result, if there is at least one fixed point, let the largest one be fix_{max}, then it is attracting since $|\frac{\partial(\varkappa(rkurt_D))}{\partial(rkurt_D)}| < 1$ for all $rkurt_D \in [fix_{max}, kurtosis_{max}]$, where $kurtosis_{max}$ is the maximum kurtosis available in D.

□

As a result of Theorem B.1, assuming continuity, $m\mathbf{k}ms$ are all equal to zero, and U-congruence of the central moment kernel distributions, Algorithm 1 converges surely provided that a fixed point exists within the domain of D. At this stage, D can only be approximated through a Monte Carlo study. The continuity of D can be ensured by using linear interpolation. One common encountered problem is that the domain of D depends on both the consistent distribution and the Monte Carlo study, so the iteration may halt at the boundary if the fixed point is not within the domain. However, by setting a proper maximum number of iterations, the algorithm can return the optimal boundary value. For quantile moments, the logic is similar, if the percentiles do not exceed the breakdown point. If this is the case, consistent estimation is impossible, and the algorithm will stop due to the maximum number of iterations. The fixed point iteration is, in principle, similar to the iterative reweighing in Huber M-estimator, but an advantage of this algorithm is that it is solely related to the inputs in Algorithm 1 and is independent of the sample size. Since they are consistent for a shape-scale distribution, $|rskew|$ can specify d_{rm} and d_{tm}, $rkurt$ can specify d_{rvar} and d_{rfm}. Algorithm 1 enables the robust estimations of all four moments to reach a near-consistent level for common unimodal distributions (Table 1, SI Dataset S1), just using the Weibull distribution as the consistent distribution.

C. Multiple consistent distributions.

D. Variance. As one of the fundamental theorems in statistics, the Central Limit Theorem declares that the standard deviation of the limiting form of the sampling distribution of the sample mean is $\frac{\sigma}{\sqrt{n}}$. The principle, asymptotic normality, was later applied to the sampling distributions of robust location estimators. Bickel and Lehmann, also in the landmark series (17, 18), argued that meaningful comparisons of the efficiencies of various kinds of location estimators can be accomplished by studying their standardized variances, asymptotic variances, and efficiency bounds. Standardized variance, $\frac{\mathrm{Var}(\hat{\theta})}{\theta^2}$, allows the use of simulation studies or empirical data to compare the variances of estimators of distinct parameters. However, a limitation of this approach is the inverse square dependence of the standardized variance on θ. If $\mathrm{Var}\left(\hat{\theta}_1\right) = \mathrm{Var}\left(\hat{\theta}_2\right)$, but θ_1 is close to zero and θ_2 is relatively large, their standardized variances will still differ dramatically. Here, the scaled standard error (SSE) is proposed as a method for estimating the variances of estimators measuring the same attribute, offering a standard error more comparable to that of the sample mean and much less influenced by the magnitude of θ.

Definition D.1 (Scaled standard error). Let $\mathcal{M}_{s_i s_j} \in \mathbb{R}^{i \times j}$ denote the sample-by-statistics matrix, i.e., the first column corresponds to $\widehat{\theta}$, which is the mean or a U-central moment measuring the same attribute of the distribution as the other columns, the second to the jth column correspond to $j-1$ statistics required to scale, $\widehat{\theta_{r_1}}$, $\widehat{\theta_{r_2}}$, …, $\widehat{\theta_{r_{j-1}}}$. Then, the scaling factor $\mathcal{S} = \left[1, \frac{\bar{\theta_{r_1}}}{\bar{\theta_m}}, \frac{\bar{\theta_{r_2}}}{\bar{\theta_m}}, \ldots, \frac{\bar{\theta_{r_{j-1}}}}{\bar{\theta_m}}\right]^T$ is a $j \times 1$ matrix, which $\bar{\theta}$ is the mean of the column of $\mathcal{M}_{s_i s_j}$. The normalized matrix is $\mathcal{M}_{s_i s_j}^N = \mathcal{M}_{s_i s_j}\mathcal{S}$. The SSEs are the unbiased standard deviations of the corresponding columns of $\mathcal{M}_{s_i s_j}^N$.

The U-central moment (the central moment estimated by using U-statistics) is essentially the mean of the central moment kernel distribution, so its standard error should be generally close to $\frac{\sigma_{\mathbf{k}m}}{\sqrt{n}}$, although not exactly since the kernel distribution is not i.i.d., where $\sigma_{\mathbf{k}m}$ is the asymptotic standard deviation of the central moment kernel distribution. If the statistics of interest coincide asymptotically, then the standard errors should still be used, e.g, for symmetric location estimators and odd ordinal central moments for the symmetric distributions, since the scaled standard error will be too sensitive to small changes when they are zero.

The SSEs of all robust estimators proposed here are often, although many exceptions exist, between those of the sample median and those of the sample mean or median central moments and U-central moments (SI Dataset S1). This is because similar monotonic relations between breakdown point and variance are also very common, e.g., Bickel and Lehmann (18) proved that a lower bound for the efficiency of TM_ϵ to sample mean is $(1-2\epsilon)^2$ and this monotonic bound holds true for any distribution. However, the direction of monotonicity differs for distributions with different kurtosis. Lehmann and Scheffé (1950, 1955) (19, 20) in their two early papers provided a way to construct a uniformly minimum-variance unbiased estimator (UMVUE). From that, the sample mean and unbiased sample second moment can be proven as the UMVUEs for the population mean and population second moment for the Gaussian distribution. While their performance for sub-Gaussian distributions is generally satisfied, they perform poorly when the distribution has a heavy tail and completely fail for distributions with infinite second moments. For sub-Gaussian distributions, the variance of a robust location estimator is generally monotonic increasing as its robustness increases, but for heavy-tailed distributions, the relation is reversed. So, unlike bias, the variance-optimal choice can be very different for distributions with different kurtosis.

Lai, Robbins, and Yu (1983) proposed an estimator that adaptively chooses the mean or median in a symmetric distribution and showed that the choice is typically as good as the better of the sample mean and median regarding variance (21). Another approach which can be dated back to Laplace (1812) (22) is using $w\bar{x}+(1-w)m_n$ as a location estimator and w is deduced to achieve optimal variance. Inspired by Lai et al's approach (21), in this study, for $rkurt$, there are 364 combinations based on 14 SWHLfms and 26 SWHLvars (SI Text). Each combination has a root mean square error (RMSE) for a single-parameter distribution, which can be inferred using a Monte Carlo study. For $qkurt$, there are another 364 combinations, but if the percentiles of quantile moments exceed the breakdown point, that combination is excluded. Then, the combination with the smallest RMSE, calibrated by a two-parameter distribution, is chosen. Similar to Subsection B, let $I(kurtosis, dtype, n) = ikurt_{\mathrm{WHL}fm,\mathrm{WHL}var}$ represent these relationships. In this article, the breakdown points of the SWHLMs in SWHL$\mathbf{k}m$ were adjusted to ensure the overall breakdown points were $\frac{1}{24}$, as detailed in Theorem **??**. There are two approaches to determine $ikurt$. The first one is computing all 364+364 $rkurt$ and $qkurt$, and then, since $\lim_{ikurt\to\infty}\frac{I(ikurt)}{ikurt}<1$, the same fix point iteration algorithm as Algorithm 1 can be used to choose the RMSE-optimum combination. The only difference is that unlike D, I is defined to be discontinuous but linear interpolation can also ensure continuity. The second approach is shown in SI Algorithm 2. The RMSEs of these $ikurt$ from the two approaches can be further determined by a Monte Carlo study. Algorithm 1 can also be used to determine the optimum choice among the two approaches. The 364+364 $rkurt$ and $qkurt$ can form a vector, $Vkurt$, where the $Q_{Vkurt}(\frac{1}{5})$ to $Q_{Vkurt}(\frac{4}{5})$ can be used to determine the d values of $r\mathbf{k}ms$ and $q\mathbf{k}ms$. The RMSEs of those $r\mathbf{k}ms$ and $q\mathbf{k}ms$ can also be estimated by a Monte Carlo study and the estimator with the smallest RMSE of each ordinal is named as $i\mathbf{k}m$. When $\mathbf{k}$ is even, the $ikurt$ determined by Ism (detailed in the SI Text) is used to determine $i\mathbf{k}m$. This approach yields results that are often nearly optimal (SI Dateset S1). The estimations of skewness and $i\mathbf{k}m$, when $\mathbf{k}$ is odd, follow the same logic.

Due to combinatorial explosion, the bootstrap (23), introduced by Efron in 1979, is indispensable for computing invariant central moments in practice. In 1981, Bickel and Freedman (24) showed that the bootstrap is asymptotically valid to approximate the original distribution in a wide range of situations, including U-statistics. The limit laws of bootstrapped trimmed U-statistics were proven by Helmers, Janssen, and Veraverbeke (1990) (25). In REDS I, the advantages of quasi-bootstrap were discussed (26–28). By using quasi-sampling, the impact of the number of repetitions of the bootstrap, or bootstrap size, on variance is very small (SI Dataset S1). An estimator based on the quasi-bootstrap approach can be seen as a complex deterministic estimator that is not only computationally efficient but also statistical efficient. The only drawback of quasi-bootstrap compared to non-bootstrap is that a small bootstrap size can produce additional finite sample bias (SI Text). In general, the variances of invariant central moments are much smaller than those of corresponding unbiased sample central moments (deduced by Cramér (29, 30)), except that of the corresponding second central moment (Table 1).

Discussion

Moments, including raw moments, central moments, and standardized moments, are the most common parameters that describe probability distributions. Central moments are preferred over raw moments because they are invariant to translation. In 1947, Hsu and Robbins proved that the arithmetic mean converges completely to the population mean provided the second moment is finite (31). The strong law of large numbers (proven by Kolmogorov in 1933) (32) implies that the $\mathbf{k}$th sample central moment is asymptotically unbiased. Recently, fascinating statistical phenomena regarding Taylor's law for distributions with infinite moments have been discovered by Drton and Xiao (2016) (33), Pillai and Meng (2016) (34), Cohen, Davis, and Samorodnitsky (2020) (35), and Brown, Cohen, Tang, and Yam (2021) (36). Lindquist and Rachev (2021) raised a critical question in their inspiring comment to Brown et al's paper (36): "What are the proper measures for the location, spread, asymmetry, and dependence (association) for random samples with infinite mean?" (37). From a different perspective, this question closely aligns with the essence of Bickel and Lehmann's open question in 1979 (10). They suggested using median, interquartile range, and medcouple (38) as the robust versions of the first three moments. While answering this question is not the focus of this paper, it is almost certain that the estimators proposed in this series will have a place. Since the efficiency of an L-statistic to the sample mean is generally monotonic with respect to the breakdown point (18), and the estimation of central moments can be transformed into the location estimation of the central moment kernel distribution, similar monotonic relations can be expected. In the case of a distribution with an infinite mean, non-robust estimators will not converge and will not provide valid estimates since their variances will be infinitely large.

Table 1. Evaluation of invariant moments for five common unimodal distributions in comparison with current popular methods

Errors	$\bar{x}$	TM	H-L	SM	HM	WM	SQM	BM	MoM	MoRM	mHLM	$rm_{exp,\text{BM}}$	$qm_{exp,\text{BM}}$
WASAB	0.000	0.107	0.088	0.078	0.078	0.066	0.048	0.048	0.034	0.035	0.034	0.002	0.003
WRMSE	0.014	0.111	0.092	0.083	0.083	0.070	0.053	0.053	0.041	0.041	0.038	0.017	0.018
WASB$_{n=5184}$	0.000	0.108	0.089	0.078	0.079	0.066	0.048	0.048	0.034	0.036	0.033	0.002	0.003
WSE $\vee$ WSSE	0.014	0.014	0.014	0.015	0.014	0.014	0.014	0.015	0.017	0.014	0.014	0.017	0.017

Errors		HFM$_\mu$	MP$_\mu$	rm	qm	im	var	var_{bs}	Tsd^2	HFM$_{\mu_2}$	MP$_{\mu_2}$	$rvar$	$qvar$	$ivar$
WASAB		0.037	0.043	0.001	0.002	0.001	0.000	0.000	0.200	0.027	0.042	0.005	0.018	0.003
WRMSE		0.049	0.055	0.015	0.015	0.014	0.017	0.017	0.198	0.042	0.062	0.019	0.026	0.019
WASB$_{n=5184}$		0.038	0.043	0.001	0.002	0.001	0.000	0.001	0.198	0.027	0.043	0.005	0.018	0.003
WSE $\vee$ WSSE		0.018	0.021	0.015	0.015	0.014	0.017	0.017	0.015	0.024	0.032	0.018	0.017	0.018

Errors	tm	tm_{bs}	HFM$_{\mu_3}$	MP$_{\mu_3}$	rtm	qtm	itm	fm	fm_{bs}	HFM$_{\mu_4}$	MP$_{\mu_4}$	rfm	qfm	ifm
WASAB	0.000	0.000	0.052	0.059	0.006	0.083	0.034	0.000	0.000	0.037	0.046	0.024	0.038	0.011
WRMSE	0.019	0.018	0.063	0.074	0.018	0.083	0.044	0.026	0.023	0.049	0.062	0.037	0.043	0.029
WASB$_{n=5184}$	0.001	0.003	0.052	0.059	0.007	0.082	0.038	0.001	0.009	0.037	0.047	0.024	0.036	0.013
WSE $\vee$ WSSE	0.019	0.018	0.021	0.091	0.015	0.012	0.017	0.024	0.021	0.020	0.027	0.021	0.020	0.022

The first table presents the use of the exponential distribution as the consistent distribution for five common unimodal distributions: Weibull, gamma, Pareto, lognormal, and generalized Gaussian distributions. Popular robust mean estimators discussed in REDS 1 were used as comparisons. The breakdown points of mean estimators in the first table, besides H-L estimator and Huber M-estimator, are all $\frac{1}{8}$. The second and third tables present the use of the Weibull distribution as the consistent distribution not plus/plus using the lognormal distribution for the odd ordinal moments optimization and the generalized Gaussian distribution for the even ordinal moments optimization. SQM is the robust mean estimator used in recombined/quantile moments. Unbiased sample central moments (var, tm, fm), U-central moments with quasi-bootstrap (var_{bs}, tm_{bs}, fm_{bs}), and other estimators were used as comparisons. The generalized Gaussian distribution was excluded for He and Fung M-Estimator and Marks percentile estimator, since the logarithmic function does not produce results for negative inputs. The breakdown points of estimators in the second and third table, besides M-estimators and percentile estimator, are all $\frac{1}{24}$. The tables include the average standardized asymptotic bias (ASAB, as $n \to \infty$), root mean square error (RMSE, at $n = 5184$), average standardized bias (ASB, at $n = 5184$) and variance (SE $\vee$ SSE, at $n = 5184$) of these estimators, all reported in the units of the standard deviations of the distribution or corresponding kernel distributions. W means that the results were weighted by the number of Google Scholar search results on May 30, 2022 (including synonyms). The calibrations of d values and the computations of ASAB, ASB, and SSE were described in Subsection D, **??** and SI Methods. Detailed results and related codes are available in SI Dataset S1 and GitHub.

Therefore, the desired measures should be as robust as possible. Clearly now, if one wants to preserve the original relationship between each moment while ensuring maximum robustness, the natural choices are median, median variance, and median skewness. Similar to the robust version of L-moment (39) being trimmed L-moment (15), mean and central moments now also have their standard most robust version based on the complete congruence of the underlying distribution.

More generally, statistics, encompassing the collection, analysis, interpretation, and presentation of data, has evolved over time, with various approaches emerging to meet challenges in practice. Among these approaches, the use of probability models and measures of random variables for data analysis is often considered the core of statistics. While the early development of statistics was focused on parametric methods, there were two main approaches to point estimation. The Gauss–Markov theorem (1, 40) states the principle of minimum variance unbiased estimation which was further enriched by Neyman (1934) (41), Rao (1945) (42), Blackwell (1947) (43), and Lehmann and Scheffé (1950, 1955) (19, 20). Maximum likelihood was first introduced by Fisher in 1922 (44) in a multinomial model and later generalized by Cramér (1946), Hájek (1970), and Le Cam (1972) (29, 45, 46). In 1939, Wald (47) combined these two principles and suggested the use of minimax estimates, which involve choosing an estimator that minimizes the maximum possible loss. Following Huber's seminal work (3), M-statistics have dominated the field of parametric robust statistics for over half a century. Nonparametric methods, e.g., the Kolmogorov–Smirnov test, Mann-Whitney-Wilcoxon Test, and Hoeffding's independence test, emerged as popular alternatives to parametric methods in 1950s, as they do not make specific assumptions about the underlying distribution of the data. In 1963, Hodges and Lehmann proposed a class of robust location estimators based on the confidence bounds of rank tests (48). In REDS I, when compared to other semiparametric mean estimators with the same breakdown point, the H-L estimator was shown to be the bias-optimal choice, which aligns Devroye, and Lerasle, Lugosi, and Oliveira's conclusion that the median of means is near-optimal in terms of concentration bounds (49) as discussed. The formal study of semiparametric models was initiated by Stein (50) in 1956. Bickel, in 1982, simplified the general heuristic necessary condition proposed by Stein (50) and derived sufficient conditions for this type of problem, adaptive estimation (51). These conditions were subsequently applied to the construction of adaptive estimates (51). It has become increasingly apparent that, in robust statistics, many estimators previously called "nonparametric" are essentially semiparametric as they are partly, though not fully, characterized by some interpretable Euclidean parameters. This approach is particularly useful in situations where the data do not conform to a simple parametric distribution but still have some structure that can be exploited. In 1984, Bickel addressed the challenge of robustly estimating the parameters of a linear model while acknowledging the possibility that the model may be invalid but still within the confines of a larger model (52). He showed by carefully designing the estimators, the biases can be very small. The paradigm shift here opens up the possibility that by defining a large semiparametric model and constructing estimators simultaneously for two or more very different semiparametric/parametric models within the large semiparametric model, then even for a parametric model belongs to the large semiparametric model but not to the semiparametric/parametric models used for calibration, the performance of these estimators might still be near-optimal due to the common nature shared by the models used by the estimators. Maybe it can be named as comparametrics. Closely related topics are "mixture model" and "constraint defined model," which were generalized in Bickel, Klaassen, Ritov, and Wellner's classic semiparametric textbook (1993) (53) and the method of sieves, introduced by Grenander in 1981 (54). As the building blocks of statistics, invariant moments can reduce the overall errors of statistical results across studies and thus can enhance the replicability of the whole community (55, 56).

Methods

Methods of generating the Table 1 are summarized below, with details in the SI Text. The d values for the invariant moments of the Weibull distribution were approximated using a Monte Carlo study, with the formulae presented in Theorem A.1 and A.2. The computation of I functions is summarized in Subsection D and further explained in the SI Text. The computation of ASABs and ASBs is described in Subsection ??. The SEs and SSEs were computed by approximating the sampling distribution using 1000 pseudorandom samples for $n = 5184$ and 50 pseudorandom samples for $n = 2654208$. The impact of the bootstrap size, ranging from $n = 2.7 \times 10^2$ to $n = 2.765 \times 10^4$, on the variance of invariant moments and U-central moments was studied using the SEs and SSEs methods described above. A brute force approach was used to estimate the maximum biases of the robust estimators discussed for the five unimodal distributions. The validity of this approach is discussed in the SI Text.

Data and Software Availability. Data for Table 1 are given in SI Dataset S1-S4. All codes have been deposited in GitHub.

1. CF Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae.* (Henricus Dieterich), (1823).
2. S Newcomb, A generalized theory of the combination of observations so as to obtain the best result. *Am. journal Math.* **8**, 343–366 (1886).
3. PJ Huber, Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964).
4. X He, WK Fung, Method of medians for lifetime data with weibull models. *Stat. medicine* **18**, 1993–2009 (1999).
5. M Menon, Estimation of the shape and scale parameters of the weibull distribution. *Technometrics* **5**, 175–182 (1963).
6. SD Dubey, Some percentile estimators for weibull parameters. *Technometrics* **9**, 119–129 (1967).
7. NB Marks, Estimation of weibull parameters from common percentiles. *J. applied Stat.* **32**, 17–24 (2005).
8. K Boudt, D Caliskan, C Croux, Robust explicit estimators of weibull parameters. *Metrika* **73**, 187–209 (2011).
9. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models. iii. dispersion in *Selected works of EL Lehmann.* (Springer), pp. 499–518 (2012).
10. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models iv. spread in *Selected Works of EL Lehmann.* (Springer), pp. 519–526 (2012).
11. AL Bowley, *Elements of statistics.* (King) No. 8, (1926).
12. WR van Zwet, *Convex Transformations of Random Variables: Nebst Stellingen.* (1964).
13. RA Groeneveld, G Meeden, Measuring skewness and kurtosis. *J. Royal Stat. Soc. Ser. D (The Stat.* **33**, 391–399 (1984).
14. J SAW, Moments of sample moments of censored samples from a normal population. *Biometrika* **45**, 211–221 (1958).
15. EA Elamir, AH Seheult, Trimmed l-moments. *Comput. Stat. & Data Analysis* **43**, 299–314 (2003).
16. RJ Hyndman, Y Fan, Sample quantiles in statistical packages. *The Am. Stat.* **50**, 361–365 (1996).
17. P Bickel, E Lehmann, Descriptive statistics for nonparametric models i. introduction in *Selected Works of EL Lehmann.* (Springer), pp. 465–471 (2012).
18. PJ Bickel, EL Lehmann, Descriptive statistics for nonparametric models ii. location in *selected works of EL Lehmann.* (Springer), pp. 473–497 (2012).
19. EL Lehmann, H Scheffé, Completeness, similar regions, and unbiased estimation-part i in *Selected works of EL Lehmann.* (Springer), pp. 233–268 (2011).
20. EL Lehmann, H Scheffé, *Completeness, similar regions, and unbiased estimation—part II.* (Springer), (2012).
21. T Lai, H Robbins, K Yu, Adaptive choice of mean or median in estimating the center of a symmetric distribution. *Proc. Natl. Acad. Sci.* **80**, 5803–5806 (1983).
22. PS Laplace, *Theorie analytique des probabilities.* (1812).
23. B Efron, Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 1–26 (1979).
24. PJ Bickel, DA Freedman, Some asymptotic theory for the bootstrap. *The annals statistics* **9**, 1196–1217 (1981).
25. R Helmers, P Janssen, N Veraverbeke, *Bootstrapping U-quantiles.* (CWI. Department of Operations Research, Statistics, and System Theory [BS]), (1990).
26. RD Richtmyer, A non-random sampling method, based on congruences, for" monte carlo" problems, (New York Univ., New York. Atomic Energy Commission Computing and Applied . . .), Technical report (1958).
27. IM Sobol', On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **7**, 784–802 (1967).
28. KA Do, P Hall, Quasi-random resampling for the bootstrap. *Stat. Comput.* **1**, 13–22 (1991).
29. H Cramér, *Mathematical methods of statistics.* (Princeton university press) Vol. 43, (1999).
30. I Gerlovina, AE Hubbard, Computer algebra and algorithms for unbiased moment estimation of arbitrary order. *Cogent mathematics & statistics* **6**, 1701917 (2019).
31. PL Hsu, H Robbins, Complete convergence and the law of large numbers. *Proc. national academy sciences* **33**, 25–31 (1947).
32. A Kolmogorov, Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.* **4**, 83–91 (1933).
33. M Drton, H Xiao, Wald tests of singular hypotheses. *Bernoulli* **22**, 38–59 (2016).
34. NS Pillai, XL Meng, An unexpected encounter with cauchy and lévy. *The Annals Stat.* **44**, 2089–2097 (2016).
35. JE Cohen, RA Davis, G Samorodnitsky, Heavy-tailed distributions, correlations, kurtosis and taylor's law of fluctuation scaling. *Proc. Royal Soc. A* **476**, 20200610 (2020).
36. M Brown, JE Cohen, CF Tang, SCP Yam, Taylor's law of fluctuation scaling for semivariances and higher moments of heavy-tailed data. *Proc. Natl. Acad. Sci.* **118**, e2108031118 (2021).
37. WB Lindquist, ST Rachev, Taylor's law and heavy-tailed distributions. *Proc. Natl. Acad. Sci.* **118**, e2118893118 (2021).
38. G Brys, M Hubert, A Struyf, A robust measure of skewness. *J. Comput. Graph. Stat.* **13**, 996–1017 (2004).
39. JR Hosking, L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Royal Stat. Soc. Ser. B (Methodological)* **52**, 105–124 (1990).
40. AA Markov, *Wahrscheinlichkeitsrechnung.* (Teubner), (1912).
41. J Neyman, On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Royal Stat. Soc.* **97**, 558–606 (1934).
42. C Radhakrishna Rao, Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–91 (1945).
43. D Blackwell, Conditional expectation and unbiased sequential estimation. *The Annals Math. Stat.* pp. 105–110 (1947).
44. RA Fisher, On the mathematical foundations of theoretical statistics. *Philos. transactions Royal Soc. London. Ser. A, containing papers a mathematical or physical character* **222**, 309–368 (1922).
45. L LeCam, On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *The Annals Math. Stat.* **41**, 802–828 (1970).
46. J Hájek, Local asymptotic minimax and admissibility in estimation in *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability.* Vol. 1, pp. 175–194 (1972).
47. A Wald, Contributions to the theory of statistical estimation and testing hypotheses. *The Annals Math. Stat.* **10**, 299–326 (1939).
48. J Hodges Jr, E Lehmann, Estimates of location based on rank tests. *The Annals Math. Stat.* **34**, 598–611 (1963).
49. L Devroye, M Lerasle, G Lugosi, RI Oliveira, Sub-gaussian mean estimators. *The Annals Stat.* **44**, 2695–2725 (2016).
50. CM Stein, Efficient nonparametric testing and estimation in *Proceedings of the third Berkeley symposium on mathematical statistics and probability.* Vol. 1, pp. 187–195 (1956).
51. PJ Bickel, On adaptive estimation. *The Annals Stat.* **10**, 647–671 (1982).
52. P Bickel, Parametric robustness: small biases can be worthwhile. *The Annals Stat.* **12**, 864–879 (1984).
53. P Bickel, CA Klaassen, Y Ritov, JA Wellner, *Efficient and adaptive estimation for semiparametric models.* (Springer) Vol. 4, (1993).
54. U Grenander, *Abstract Inference.* (1981).
55. JT Leek, RD Peng, Reproducible research can still be wrong: adopting a prevention approach. *Proc. Natl. Acad. Sci.* **112**, 1645–1646 (2015).
56. E National Academies of Sciences, et al., *Reproducibility and Replicability in Science.* (National Academies Press), (2019).