Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 370–393
https://doi.org/10.1214/19-BJPS432

Measuring symmetry and asymmetry of multiplicative distortion measurement errors data

Jun Zhang[a], Yujie Gai[b], Xia Cui[c] and Gaorong Li[d]

[a]*Shenzhen University*
[b]*Central University of Finance and Economics*
[c]*Guangzhou University*
[d]*Beijing Normal University*

Abstract. This paper studies the measure of symmetry or asymmetry of a continuous variable under the multiplicative distortion measurement errors setting. The unobservable variable is distorted in a multiplicative fashion by an observed confounding variable. First, two direct plug-in estimation procedures are proposed, and the empirical likelihood based confidence intervals are constructed to measure the symmetry or asymmetry of the unobserved variable. Next, we propose four test statistics for testing whether the unobserved variable is symmetric or not. The asymptotic properties of the proposed estimators and test statistics are examined. We conduct Monte Carlo simulation experiments to examine the performance of the proposed estimators and test statistics. These methods are applied to analyze a real dataset for an illustration.

1 Introduction

Measurement errors may exist in many disciplines. At least this seems to be true if data are collected from medical research, health science and economics, due to improper instrument calibration or many other reasons. However, very seldom the exact characteristics of these errors are known. Therefore, the biassing effects on estimation and testing are mostly considered under the assumption of white noise added to the variable under discussion Fuller (1987). When the variables have been measured with errors, the presence of measurement errors causes biased and inconsistent parameter estimates and leads to erroneous conclusions to various degrees in practical analysis even in a large sample. Techniques for addressing measurement error problems can be classified along two dimensions. Different techniques are employed in errors-in-variables linear models and in errors-in-variables non-linear models. The attenuation bias in a simple measurement error linear regression is an underestimate of the coefficient (Fuller (1987)). Bias in nonlinear models is more complex than linear regression models (Carroll et al. (2006)). Due to the importance of the measurement error problems, some research on measurement error models has been carefully studied, see, for example, Liang and Ren (2005), Liang, Härdle and Carroll (1999), Li, Zhang and Feng (2016).

In this paper, we consider that a unobservable variable X is measured with multiplicative errors and involved by a confounding variable:

$$\widetilde{X} = X\psi(U), \tag{1.1}$$

where, X is the unobservable continuous variable of interest, $\widetilde{X}$ is the available observed variable, and $\psi(\cdot)$ is a contaminating unknown function of the observed confounding variable U, which is assumed to be independent of X. The multiplicative distortion function $\psi(U)$

Key words and phrases. Confounding variable, errors-in-variables, correlation coefficient, empirical likelihood, symmetry.

Received May 2018; accepted January 2019.

satisfies $E[\psi(U)] = 1$ for identifiability. The multiplicative distortion measurement errors model (1.1) is introduced by Şentürk and Müller (2005). The model (1.1) is different from Hwang (1986), who considered a multiplicative measurement error model where the observed value W is assumed by $W = XU$, and X and U are unobserved variables and asymmetric distribution such that $E(U) = 1$ and $\text{Var}(U) = \sigma^2$. In model (1.1), the confounding variable U is observed but distorting function $\psi(u)$ is unknown. For the regression models in the multiplicative fashion, Chen et al. (2010) proposed the least absolute relative error estimation method and Chen et al. (2016), Liu and Xia (2018) proposed the least product relative error for estimating parameters in the multiplicative regression models.

Recently, a number of authors have studied the multiplicative distortion measurement errors model (1.1) in various parametric or semi-parametric setting. See, for example, Şentürk and Müller (2006) considered the estimation for multiplicative distortion measurement errors linear regression models. Cui et al. (2009) studied the nonlinear multiplicative distortion measurement errors models. Li, Lin and Cui (2010) considered partial linear models, where the linear part is observed with multiplicative distortion measurement errors, and Li and Lu (2018) considered to use lasso-type penalty functions including lasso and adaptive lasso for simultaneous variable selection and parameter estimation. Delaigle, Hall and Zhou (2016) obtained a fundamental work of nonparametric estimation of a regression curve when the data are observed with multiplicative distortion. Recently, Zhao and Xie (2018) and Li et al. (2018) respectively, considered the adaptive model test and adaptive estimation for the nonparametric multiplicative distortion measurement error models. Toward this end, there are no systematic studies on measuring the symmetry or asymmetry of X under the distortion measurement errors model (1.1).

In this paper, we study the symmetry or asymmetry of a continuous variable X under the distortion measurement errors models (1.1). Let $f(x)$ and $F(x)$, $x \in \mathbb{R}$ be the probability density function and the distribution function, respectively. We say X is symmetric about γ if $F(\gamma - x) = 1 - F(\gamma + x)$ (or equivalently $f(\gamma - x) = f(\gamma + x)$) for every $x \in \mathbb{R}$. The concept of symmetry plays an important role in mathematics as well as in statistics. For the validity of signed rank procedures, symmetry is a crucial assumption (Kraft and van Eeden (1972)) for the Mann–Whitney type tests and also the Wilcoxon signed-rank test. It is known that the Wilcoxon signed-rank test is not robust against the assumption of symmetry, so it is necessary to check the assumption of symmetry before employing the Wilcoxon signed-rank procedure in practice. There are several tests available in the literature to assess the symmetry of an unknown density function $f(x)$ based on an *i.i.d.* random sample. See, for example, Hill and Rao (1977) considered Cramér–von Mises type statistics to test the symmetry, Butler (1969) in which a test statistic is proposed by using a sample version of $\sup_{x\le 0} |F(\gamma + x) + F(\gamma - x) - 1|$, Davis and Quade (1978) proposed U-statistics by using the triple sample for testing the symmetry. Recently, Patil, Patil and Bagkavos (2012) proposed measuring the symmetry and asymmetry by using the correlation coefficient between the density function and the distribution function, namely, $\rho_{(f,F)} = \frac{\text{Cov}(f(X),F(X))}{\sqrt{\sigma_f^2 \sigma_F^2}}$ when $0 < \sigma_f^2 < \infty$, where $\sigma_f^2 = \text{Var}(f(X))$ and $\sigma_F^2 = \text{Var}(F(X))$ $(= \frac{1}{12})$. Patil, Patil and Bagkavos (2012) showed that $\rho_{(f,F)}$ works well in capturing the visual impression of asymmetry in a given density curve. Some other test statistics are referred to in Gupta (1967), Csörgő and Heathcote (1987) and the references therein.

As the variable X is observed with multiplicative distortion measurement errors, the existing literature for measuring the symmetry or asymmetry of X cannot be directly used. There is no literature to study the symmetry or asymmetry of a continuous variable under the model setting (1.1). In this paper, we propose a general kth correlation coefficient $\rho_k = \frac{\text{Cov}(f^k(X),F(X))}{\sqrt{\sigma_{f^k}^2 \sigma_F^2}}$ with $\sigma_{f^k}^2 = \text{Var}(f^k(X))$ for some $k > 0$ as a measure of symmetry or asymmetry. Note that ρ_1 is the correlation coefficient $\rho_{(f,F)}$ proposed by Patil, Patil and Bagkavos (2012). Under multiplicative distortion measurement errors setting (1.1), we first propose two direct plug-in estimation procedures to calibrate unobserved X by using the estimation methods proposed in Cui et al. (2009) and Delaigle, Hall and Zhou (2016), and further use calibrated variable $\hat{X}$ to estimate ρ_k. Next, we construct empirical likelihood based confidence intervals of ρ_k, which can be used to judge the symmetry or asymmetry of X. We also consider a special case by using the symmetry or asymmetry of observed $\widetilde{X}$ to judge the symmetry or asymmetry of X. Finally, we propose four test statistics for testing whether X is symmetric or not. The asymptotic properties of the proposed estimators and test statistics are examined. We conduct Monte Carlo simulation experiments to examine the performance of the proposed estimators and test statistics.

The paper is organized as follows. In Section 2, we propose two direct plug-in estimation procedures and derive related asymptotic results. The empirical likelihood based confidence intervals are also investigated. In Section 3, we propose four test statistics for measuring the symmetry or asymmetry of the unobserved variable, and study the asymptotic properties of the test statistics. In Section 4, simulation studies are conducted to examine the performance of the proposed estimators and test statistics. In Section 5, the analysis of a real dataset is presented. In Section 6, some discussion of the proposed methods is presented. Technical proofs of theorems are provided in the on-line supplementary materials (Zhang et al. (2019)).

2 Direct plug-in estimation procedure

2.1 General setting

As the distorted $\widetilde{X}$ is available, we first calibrate unobservable X by using the observed *i.i.d.* sample $\{\widetilde{X}_i, U_i\}_{i=1}^n$. To ensure identifiability of model (1.1), it is assumed that

$$E\big[\psi(U)\big] = 1. \tag{2.1}$$

The identifiability condition (2.1) is introduced by Şentürk and Müller (2005), and it is analogous to the classical additive measurement error setting: $E(\theta) = 0$ for $W = Z + \theta$, where W is error-prone and Z is error-free. For the distorting function $\psi(u)$ and the expectation of X, it is commonly assumed in the literature as:

Assumption M1: $E(X) \neq 0$, the unknown smoothing distorting function $\psi(u)$ satisfies $\psi(u) \neq 0$ for all $u \in [\mathcal{U}_L, \mathcal{U}_R]$, $\mathcal{U}_L < \mathcal{U}_R$, where $[\mathcal{U}_L, \mathcal{U}_R]$ denotes the compact support of U.

Assumption M2: the unknown smoothing distorting function $\psi(u)$ satisfies $\psi(u) > 0$ for all $u \in [\mathcal{U}_L, \mathcal{U}_R]$.

By using two assumptions M1 and M2, we have two different estimation procedures of ρ_k.

Step 1.1. Using the identifiability condition (2.1) and assumption M1, we have $E(\widetilde{X}) = E(X)$ and $\psi(u) = E(\widetilde{X}|U = u)/E(\widetilde{X})$. Then, the local linear estimator $\hat{\psi}_1(u)$ of $\psi(u)$ is proposed as

$$\hat{\psi}_1(u) = \frac{S_{n2}(u)Q_{n0,\widetilde{X}}(u) - S_{n1}(u)Q_{n1,\widetilde{X}}(u)}{[S_{n2}(u)S_{n0}(u) - [S_{n1}(u)]^2]\overline{\widetilde{X}}}, \tag{2.2}$$

where $S_{n\omega}(u) = \frac{1}{nh_1}\sum_{i=1}^n (\frac{U_i - u}{h_1})^\omega K(\frac{U_i - u}{h_1})$ for $\omega = 0, 1, 2$, and $Q_{n\delta,\widetilde{X}}(u) = \frac{1}{nh_1} \times \sum_{i=1}^n (\frac{U_i - u}{h_1})^\delta K(\frac{U_i - u}{h_1})\widetilde{X}_i$ for $\delta = 0, 1$, and $\overline{\widetilde{X}} = \frac{1}{n}\sum_{i=1}^n \widetilde{X}_i$. Here $K(\cdot)$ denotes a kernel density function, and h_1 is a bandwidth.

Step 1.2. Using the identifiability condition (2.1) and assumption M2, we have $E(|\widetilde{X}|) = E(|X|)$ and $\psi(u) = E(|\widetilde{X}||U = u)/E(|\widetilde{X}|)$. Then, the local linear estimator $\hat{\psi}_2(u)$ is used to estimate $\psi(u)$ by

$$\hat{\psi}_2(u) = \frac{S_{n2}(u)V_{n0,|\widetilde{X}|}(u) - S_{n1}(u)V_{n1,|\widetilde{X}|}(u)}{[S_{n2}(u)S_{n0}(u) - [S_{n1}(u)]^2]\overline{|\widetilde{X}|}}, \tag{2.3}$$

where $V_{n\delta,|\widetilde{X}|}(u) = \frac{1}{nh_1}\sum_{i=1}^{n}(\frac{U_i-u}{h_1})^{\delta}K(\frac{U_i-u}{h_1})|\widetilde{X}_i|$ for $\delta = 0, 1$, and $\overline{|\widetilde{X}|} = \frac{1}{n}\sum_{i=1}^{n}|\widetilde{X}_i|$.

Step 2. Using (2.2) and (2.3), we obtain $\hat{X}_i^{[1]} = \widetilde{X}_i/\hat{\psi}_1(U_i)$ and $\hat{X}_i^{[2]} = \widetilde{X}_i/\hat{\psi}_2(U_i)$, $i = 1, \ldots, n$ and the estimators of $f(x)$ and $F(x)$ are constructed as

$$\begin{cases} \hat{f}_1(x) = \dfrac{1}{nh_2}\displaystyle\sum_{i=1}^{n} K\Big(\dfrac{\hat{X}_i^{[1]} - x}{h_2}\Big), \\ \hat{F}_1(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} I\{\hat{X}_i^{[1]} \le x\}, \end{cases} \tag{2.4}$$

$$\begin{cases} \hat{f}_2(x) = \dfrac{1}{nh_3}\displaystyle\sum_{i=1}^{n} K\Big(\dfrac{\hat{X}_i^{[2]} - x}{h_3}\Big), \\ \hat{F}_2(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} I\{\hat{X}_i^{[2]} \le x\}, \end{cases} \tag{2.5}$$

where h_2 and h_3 are two positive-valued bandwidths.

Step 3. Directly using $E[F(X)] = \frac{1}{2}$, $\sigma_F^2 = \frac{1}{12}$, two moment based estimators of ρ_k are proposed as, for $s = 1, 2$,

$$\hat{\rho}_k^{[s]} = \sqrt{12}\frac{\frac{1}{n}\sum_{i=1}^{n}\hat{f}_s^k(\hat{X}_i^{[s]})\hat{F}_s(\hat{X}_i^{[s]}) - \frac{1}{2n}\sum_{i=1}^{n}\hat{f}_s^k(\hat{X}_i^{[s]})}{\{\frac{1}{n}\sum_{i=1}^{n}[\hat{f}_s^k(\hat{X}_i^{[s]}) - \frac{1}{n}\sum_{i=1}^{n}\hat{f}_s^k(\hat{X}_i^{[s]})]^2\}^{1/2}}.$$

In the simulation study in Section 4, k will be considered as $k = 0.5$, $k = 1$, $k = 2$ and $k = 3$ for illustrations.

We first list some of the conditions needed for the proofs of our asymptotic results.

(A1) The density function $f_U(u)$ of the random variable U is bounded away from 0 and satisfies the Lipschitz condition of order 1 on $[\mathcal{U}_L, \mathcal{U}_R]$.
(A2) The distorting function $\psi(u)$ has three continuous derivatives on $[\mathcal{U}_L, \mathcal{U}_R]$.
(A3) The density functions $f(x)$ of X and $\tilde{f}(\tilde{x})$ of $\widetilde{X}$ have two continuous derivatives, satisfying $\int_{-\infty}^{\infty} f^{1+4k}(x)\,dx < \infty$, $\int_{-\infty}^{\infty}\tilde{f}^{1+4k}(\tilde{x})\,d\tilde{x} < \infty$, $k > 0$.
(A4) The kernel function $K(\cdot)$ is a univariate bounded, continuous and symmetric density function about zero, supported on $[-C, C]$, $C > 0$. The second derivative of $K(\cdot)$ is bounded on $[-C, C]$, satisfying a Lipschitz condition. Moreover, $\int_{-C}^{C} t^2K(t)\,dt \neq 0$ and $\int_{-C}^{C}|t|^jK(t)\,dt < \infty$ for $j = 1, 2, 3$.
(A5) As $n \to \infty$, $nh_s^4 \to 0$, $\frac{\log^2 n}{nh_s^2} \to 0$ for $s = 1, 2, 3, 4$.

Condition (A1) ensures that the density function $f_U(u)$ is positive, which implies that the denominators involved in the nonparametric estimators are bounded away from zero in a large sample setting. Condition (A2) is a mild smoothness condition on the distorting function $\psi(u)$. Condition (A3) is the technique condition of density functions to ensure that the asymptotic variances in Theorem 2.1 and Theorem 2.3 are finite. Condition (A4) is the common condition for the kernel function $K(t)$. The Epanechnikov kernel complies with this condition. Condition (A5) is generally required for bandwidth h_s in nonparametric smoothing. Bandwidths h_s, $s = 1, 2, 3$ are used in Section 2.1 and bandwidth h_4 is used in Section 2.2.

In the following, we define $\overline{\hat{f}_s^k \hat{F}_s} = \frac{1}{n}\sum_{i=1}^n \hat{f}_s^k(\hat{X}_i^{[s]})\hat{F}_s(\hat{X}_i^{[s]})$, $\overline{\hat{f}_s^k} = \frac{1}{n}\sum_{i=1}^n \hat{f}_s^k(\hat{X}_i^{[s]})$ and $\hat{\sigma}_{s,f^k} = \{\frac{1}{n}\sum_{i=1}^n [\hat{f}_s^k(\hat{X}_i^{[s]}) - \overline{\hat{f}_s^k}]^2\}^{1/2}$, $E(f^k F) = E[f^k(X)F(X)]$, $E(f^k) = E[f^k(X)]$. Moreover, let $\mu_{f^d} = \int f^d(x)\,dx$ and

$$\delta_k(t, f, F) = \begin{pmatrix} (k+1)f^k(t)F(t) + \int_t^\infty f^{k+1}(x)\,dx \\ (k+1)f^k(t) \\ \frac{2k+1}{2\sigma_{f^k}} f^{2k}(t) + \frac{(k+1)\mu_{f^{k+1}}}{\sigma_{f^k}} f^k(t) \end{pmatrix}, \qquad \eta_k = \begin{pmatrix} E(f^k F) \\ E(f^k) \\ \sigma_{f^k} \end{pmatrix}.$$

Theorem 2.1. *Assume that conditions* (A1)–(A5) *hold*,

(a) *we have*

$$\sqrt{n}\big((\overline{\hat{f}_s^k \hat{F}_s} - E(f^k F)), (\overline{\hat{f}_s^k} - E(f^k)), (\hat{\sigma}_{s,f^k} - \sigma_{f^k})\big)^{\mathrm{T}} \xrightarrow{L} N(0, \Sigma_{k,s}),$$

where

$$\Sigma_{k,s} = k^2 \eta_k \eta_k^{\mathrm{T}} E(X^2)\operatorname{Var}(\psi(U))\left(\frac{I\{s=1\}}{[E(X)]^2} + \frac{I\{s=2\}}{[E(|X|)]^2}\right) + \operatorname{Cov}(\delta_k(X, f, F));$$

(b) *let* $\theta_k = \frac{1}{\sigma_{f^k}}(\sqrt{12}, -\frac{\sqrt{12}}{2}, -\rho_k)^{\mathrm{T}}$, *we have*

$$\sqrt{n}(\hat{\rho}_k^{[s]} - \rho_k) \xrightarrow{L} N(0, \theta_k^{\mathrm{T}} \Sigma_{k,s} \theta_k).$$

Remark 1. The term $k^2\eta_k\eta_k^{\mathrm{T}} E(X^2)\operatorname{Var}(\psi(U))(\frac{I\{s=1\}}{[E(X)]^2} + \frac{I\{s=2\}}{[E(|X|)]^2})$ in Theorem 2.1(a) is caused by the distortion function $\psi(U)$. One can see that, the estimator $\hat{\rho}_k^{[s]}$ performs efficiently only when $\operatorname{Var}(\psi(U)) = 0$ that is, $\psi(u) \equiv 1$. When $\operatorname{Var}(\psi(U)) > 0$, the increment of the asymptotic variance of $\hat{\rho}_k^{[s]}$ is caused by using the calibrated variables $\{\hat{X}_i^{[s]}, i = 1, \ldots, n\}$ instead of unobserved $\{X_i, i = 1, \ldots, n\}$, i.e., the effect of distorting function $\psi(U)$ exists.

Remark 2. It is known that $|E(X)| \le E(|X|)$, and $\frac{1}{[E(X)]^2} \ge \frac{1}{[E|X|]^2}$ when $E(X) \neq 0$. This inequality tells us the estimator $\hat{\rho}_k^{[2]}$ will perform better than $\hat{\rho}_k^{[1]}$ in an asymptotic way when $E(X) \neq 0$. Moreover, when X is a almost surely positive-valued random variable, that is, $P(X < 0) = 0$, we have $E(X) \neq 0$ and $E(|X|) = E(X)$. Thus, $\Sigma_{k,1} = \Sigma_{k,2}$ and the asymptotic variance of $\hat{\rho}_k^{[1]}$ is the same as $\hat{\rho}_k^{[2]}$. If $E(X) = 0$, we only use $\hat{\rho}_k^{[2]}$ because $\overline{\tilde{X}} = O_P(n^{-1/2})$ and the denominator in (2.2) will converge to zero in probability, and the performance of estimator $\hat{\rho}_k^{[1]}$ is unstable.

We now use empirical likelihood (EL) (Owen (1991)) method to construct confidence intervals of ρ_k. Empirical likelihood method avoids to estimate asymptotic covariances, improves the accuracy of coverage in a moderate sample setting, and also is easily implemented and automatically studentized. So, the EL method is widely applied in practice. See, for example, Liu and Xia (2018), Kiwitt and Neumeyer (2012). It is noted that the asymptotic results of Theorem 2.1 can be used to construct asymptotic confidence intervals when one gets an estimator of asymptotic variance, namely, $\hat{\theta}_{k,s}^{\mathrm{T}}\hat{\Sigma}_{k,s}\hat{\theta}_{k,s}$. In details, the confidence intervals for $\hat{\rho}_k^{[s]}$, $s = 1, 2$, got from normal approximation are, $I_{\alpha,\mathrm{NOR}}^{[s]} = \{\rho'_{k,s} : n(\hat{\rho}_k^{[s]} - \rho'_{k,s})^2/(\hat{\theta}_k^{\mathrm{T}} \hat{\Sigma}_{k,s} \hat{\theta}_{k,s}) \le c_\alpha\}$. Partlett and Patil (2017) claimed that a numerical method is need to evaluate the integrals in the asymptotic variance when $f(x)$ is symmetric for $\rho_1 = 0$. When $\rho_1 \neq 0$, the performance of estimator of $\hat{\theta}_{1,s}^{\mathrm{T}} \hat{\Sigma}_{1,s} \hat{\theta}_{1,s}$ may not be appropriate, as one need to estimate so many complicated terms in the asymptotic variance and their finite-sample behaviors may not perform well. Although the estimator can be shown to be consistent under some mild assumptions. In the following, we make statistical inference based on the EL principle for ρ_k.

The correlation coefficient ρ_k can be estimated through the estimating equation in the population level:

$$E\left[\left\{\sqrt{12}\left(F(X) - \frac{1}{2}\right) - \rho_k\left(\frac{f^k(X) - E[f^k(X)]}{\sigma_{f^k}}\right)\right\} \times \left(\frac{f^k(X) - E[f^k(X)]}{\sigma_{f^k}}\right)\right] = 0.$$

Motivated by this equation, the empirical log-likelihood ratio function is defined as

$$\hat{\ell}_n^{[s]}(\rho_k) = -2 \max\left\{\sum_{i=1}^n \log(np_i) : p_i \ge 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\Im}_{n,i}^{[s]}(\rho_k) = 0\right\},$$

where, for $i = 1, \ldots, n$, $s = 1, 2$,

$$\hat{\Im}_{n,i}^{[s]}(\rho_k) = \left\{\sqrt{12}\left(\hat{F}_s(\hat{X}_i^{[s]}) - \frac{1}{2}\right) - \rho_k\left(\frac{\hat{f}_s^k(\hat{X}_i^{[s]}) - \overline{\hat{f}_s^k}}{\hat{\sigma}_{s,f^k}}\right)\right\} \times \left(\frac{\hat{f}_s^k(\hat{X}_i^{[s]}) - \overline{\hat{f}_s^k}}{\hat{\sigma}_{s,f^k}}\right).$$

By the Lagrange multiplier method, we have $\hat{\ell}_n^{[s]}(\rho_k) = 2\sum_{i=1}^n \log\{1 + \hat{\lambda}_s \hat{\Im}_{n,i}^{[s]}(\rho_k)\}$, where $\hat{\lambda}_s$ is determined by the equation $\frac{1}{n}\sum_{i=1}^n \frac{\hat{\Im}_{n,i}^{[s]}(\rho_k)}{1+\hat{\lambda}_s \hat{\Im}_{n,i}^{[s]}(\rho_k)} = 0$.

Theorem 2.2. *Assume that conditions* (A1)–(A5) *hold,* $\hat{\ell}_n^{[s]}(\rho_k)$ *converges to a centered chi-squared distribution with degree of freedom one.*

From Theorem 2.2, an EL confidence interval for ρ_k is constructed as $I_{\rho_k} = \{\rho'_{k,s} : \hat{\ell}_n^{[s]}(\rho'_{k,s}) \le c_\kappa\}$, where c_κ denotes the κ quantile of the chi-squared distribution with degree of freedom one.

2.2 A special setting

Suppose that X is symmetric about γ, i.e., $F(x) = 1 - F(2\gamma - x)$, and $\psi(u) > 0$ for all $u \in [\mathcal{U}_L, \mathcal{U}_R]$ under assumption M2. Using the independence condition between U and X, it is seen that

$$\begin{aligned}
\widetilde{F}(x) &= EI\{\widetilde{X} \le x\} = E\big[EI\{X \le x/\psi(U)\}|U\big] = E\big[F\big(x/\psi(U)\big)|U\big] \\
&= E\big[1 - F\big(2\gamma - x/\psi(U)\big)|U\big] \\
&= E\big[1 - EI\{X \le 2\gamma - x/\psi(U)\}|U\big] \\
&= E\big[1 - EI\{\widetilde{X} \le 2\gamma\psi(U) - x\}|U\big].
\end{aligned} \tag{2.6}$$

From (2.6), if X is symmetric about zero ($\gamma = 0$), then, $\widetilde{F}(x) = 1 - \widetilde{F}(-x)$. This implies that $\widetilde{X}$ is also symmetric about zero, and the effect of multiplicative distortion for testing the symmetry of X about zero vanishes. In this case, testing the symmetry of X about zero is equivalent to testing the symmetry of $\widetilde{X}$ about zero. Similar to (2.4) and (2.5), we propose the estimators of the density function of $\widetilde{X}$, $\tilde{f}(\tilde{x})$ and distribution function of $\widetilde{X}$, $\widetilde{F}(\tilde{x})$ as

$$\hat{\tilde{f}}(x) = \frac{1}{nh_4}\sum_{i=1}^{n} K\left(\frac{\widetilde{X}_i - x}{h_4}\right), \qquad \hat{\widetilde{F}}(x) = \frac{1}{n}\sum_{i=1}^{n} I\{\widetilde{X}_i \le x\},$$

where h_4 is a positive-valued bandwidth.

Directly using $E[\widetilde{F}(\widetilde{X})] = \frac{1}{2}$, $\sigma_{\widetilde{F}}^2 = \operatorname{Var}(\widetilde{F}(\widetilde{X})) = \frac{1}{12}$, a moment based estimator of $\tilde{\rho}_k = \frac{\operatorname{Cov}(\tilde{f}^k(\widetilde{X}), \widetilde{F}(\widetilde{X}))}{\sqrt{\sigma_{\widetilde{F}}^2 \sigma_{\tilde{f}^k}^2}}$, $\sigma_{\tilde{f}^k}^2 = \operatorname{Var}(\tilde{f}(\widetilde{X}))$, is proposed as

$$\hat{\tilde{\rho}}_k = \sqrt{12}\frac{\frac{1}{n}\sum_{i=1}^{n} \hat{\tilde{f}}^k(\widetilde{X}_i)\hat{\widetilde{F}}(\widetilde{X}_i) - \frac{1}{2n}\sum_{i=1}^{n} \hat{\tilde{f}}^k(\widetilde{X}_i)}{\{\frac{1}{n}\sum_{i=1}^{n}[\hat{\tilde{f}}^k(\widetilde{X}_i) - \frac{1}{n}\sum_{i=1}^{n} \hat{\tilde{f}}^k(\widetilde{X}_i)]^2\}^{1/2}}.$$

Next, we define $\overline{\hat{\tilde{f}}^k \hat{\widetilde{F}}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\tilde{f}}^k(\widetilde{X}_i)\hat{\widetilde{F}}(\widetilde{X}_i)$, $\overline{\hat{\tilde{f}}^k} = \frac{1}{n}\sum_{i=1}^{n} \hat{\tilde{f}}^k(\widetilde{X}_i)$ and $\hat{\sigma}_{\tilde{f}^k} = \{\frac{1}{n} \times \sum_{i=1}^{n}[\hat{\tilde{f}}^k(\widetilde{X}_i) - \overline{\hat{\tilde{f}}^k}]^2\}^{1/2}$, $E(\tilde{f}^k\widetilde{F}) = E[\tilde{f}^k(\widetilde{X})\widetilde{F}(\widetilde{X})]$, $E(\tilde{f}^k) = E[\tilde{f}^k(\widetilde{X})]$. Moreover, let $\mu_{\tilde{f}^d} = \int \tilde{f}^d(\tilde{x})\,d\tilde{x}$ and

$$\delta_k(t, \tilde{f}, \widetilde{F}) = \begin{pmatrix} (k+1)\tilde{f}^k(t)\widetilde{F}(t) + \int_t^{\infty} \tilde{f}^{k+1}(\tilde{x})\,d\tilde{x} \\ (k+1)\tilde{f}^k(t) \\ \frac{2k+1}{2\sigma_{\tilde{f}^k}}\tilde{f}^{2k}(t) + \frac{(k+1)\mu_{\tilde{f}^{k+1}}}{\sigma_{\tilde{f}^k}}\tilde{f}^k(t) \end{pmatrix}.$$

Theorem 2.3. *Assume that conditions* (A1)–(A5) *hold*,

(a) *let* $\widetilde{\Sigma}_k = \operatorname{Cov}(\delta_k(\widetilde{X}, \tilde{f}, \widetilde{F}))$, *we have*

$$\sqrt{n}\big((\overline{\hat{\tilde{f}}^k\hat{\widetilde{F}}} - E(\tilde{f}^k\widetilde{F})), (\overline{\hat{\tilde{f}}^k} - E(\tilde{f}^k)), (\hat{\sigma}_{\tilde{f}^k} - \sigma_{\tilde{f}^k})\big)^{\mathrm{T}} \xrightarrow{L} N(0, \widetilde{\Sigma}_k);$$

(b) $\sqrt{n}(\hat{\tilde{\rho}}_k - \tilde{\rho}_k) \xrightarrow{L} N(0, \tilde{\theta}_k^{\mathrm{T}}\widetilde{\Sigma}_k\tilde{\theta}_k)$, *where* $\tilde{\theta}_k = \frac{1}{\sigma_{\tilde{f}^k}}(\sqrt{12}, -\frac{\sqrt{12}}{2}, -\tilde{\rho}_k)^{\mathrm{T}}$.

The correlation coefficient $\tilde{\rho}_k$ can be estimated through the estimating equation in the population level:

$$E\left[\left\{\sqrt{12}\left(\widetilde{F}(\widetilde{X}) - \frac{1}{2}\right) - \tilde{\rho}_k\left(\frac{\tilde{f}^k(\widetilde{X}) - E[\tilde{f}^k(\widetilde{X})]}{\sigma_{\tilde{f}^k}}\right)\right\} \times \left(\frac{\tilde{f}^k(\widetilde{X}) - E[\tilde{f}^k(\widetilde{X})]}{\sigma_{\tilde{f}^k}}\right)\right] = 0.$$

The empirical log-likelihood ratio function is defined as

$$\hat{\tilde{\ell}}_n(\tilde{\rho}_k) = -2\max\left\{\sum_{i=1}^{n}\log(n\tilde{p}_i) : \tilde{p}_i \ge 0, \sum_{i=1}^{n}\tilde{p}_i = 1, \sum_{i=1}^{n}\tilde{p}_i\hat{\tilde{\Im}}_{n,i}(\tilde{\rho}_k) = 0\right\},$$

where, for $i = 1, \ldots, n$,

$$\hat{\tilde{\Im}}_{n,i}(\tilde{\rho}_k) = \left\{ \sqrt{12}\left(\hat{\tilde{F}}(\widetilde{X}_i) - \frac{1}{2} \right) - \tilde{\rho}_k \left(\frac{\hat{\tilde{f}}^k(\widetilde{X}_i) - \overline{\hat{\tilde{f}}^k}}{\hat{\sigma}_{\tilde{f}^k}} \right) \right\} \times \left(\frac{\hat{\tilde{f}}^k(\widetilde{X}_i) - \overline{\hat{\tilde{f}}^k}}{\hat{\sigma}_{\tilde{f}^k}} \right).$$

Using Lagrange multiplier, we have $\hat{\tilde{\ell}}_n(\tilde{\rho}_k) = 2\sum_{i=1}^n \log\{1 + \hat{\tilde{\lambda}}\hat{\tilde{\Im}}_{n,i}(\tilde{\rho}_k)\}$, where $\hat{\tilde{\lambda}}$ is determined by the equation $\frac{1}{n}\sum_{i=1}^n \frac{\hat{\tilde{\Im}}_{n,i}(\tilde{\rho}_k)}{1+\hat{\tilde{\lambda}}\hat{\tilde{\Im}}_{n,i}(\tilde{\rho}_k)} = 0$.

Theorem 2.4. *Assume that conditions* (A1)–(A5) *hold*, $\hat{\tilde{\ell}}_n(\tilde{\rho}_k)$ *converges to a centered chi-squared distribution with degree of freedom one.*

From Theorem 2.4, an EL confidence interval for $\tilde{\rho}_k$ is constructed as $\tilde{I}_{\tilde{\rho}_k} = \{\tilde{\rho}'_k : \hat{\tilde{\ell}}_n(\tilde{\rho}'_k) \le c_\kappa\}$, where c_κ denotes the κ quantile of the chi-squared distribution with degree of freedom one. From (2.6), the symmetry of $\widetilde{X}$ and X about γ is equivalent when $\psi(U) \equiv 1$. We can also conduct a hypothesis $\mathcal{H}_0$: $\psi(u) = 1$, $u \in [\mathcal{U}_L, \mathcal{U}_R]$ and $\mathcal{H}_1$: $\psi(u) \neq 1$, for some $u \in [\mathcal{U}_L, \mathcal{U}_R]$ at first. If the hypothesis $\mathcal{H}_0$ holds, one can directly use the distorted variable $\widetilde{X}$, estimator $\hat{\tilde{\rho}}_k$ and Theorem (2.4) to conclude whether $\widetilde{X}$ is symmetric or not. If $\widetilde{X}$ is symmetric, then X is symmetric because $\psi(U) \equiv 1$. Otherwise, if $\widetilde{X}$ is asymmetric, the symmetry of X can be determined through the proposed estimators $\hat{\rho}_k^{[l]}$, $l = 1, 2$, or the proposed test statistics presented in the following section. There are many literature to discuss whether a nonparametric function is a constant function for the hypothesis $\mathcal{H}_0$. One can use one of them to test $\mathcal{H}_0$. There is unnecessary to repeat these statistics and we omit them here.

3 Some tests of symmetry for distortion measurement errors

In parametric inference, there are many methods for measuring symmetry or asymmetry for a continuous random variable X. The historically known measures for detecting departures from symmetry include

$$S_P = \frac{E(X) - \vartheta}{\sigma}, \qquad S_G = \frac{E(X) - \vartheta}{E|X - \vartheta|},$$
$$S_M = \frac{E[(X - E(X))^3]}{\sigma^3}, \qquad S_{Q,\delta} = \frac{q_{1-\delta} + q_\delta - 2\vartheta}{q_{1-\delta} - q_\delta}, \qquad \delta \in (0, 0.5),$$

where ϑ stands for the median, σ^2 stands for the variance and q_δ stands for the δ-quantile for a continuous random variable X, respectively. There are many literature on the study of S_P, S_G, S_M and $S_{Q,\delta}$ when X is observed without measurement errors. For the multiplicative distortion measurement errors data considered in this paper, as $\{X_1, X_2, \ldots, X_n\}$ are distorted and unobservable, and only $\{(\widetilde{X}_i, U_i), i = 1, \ldots, n\}$ are available. In this section, we study how to estimate S_P, S_G, S_M and $S_{Q,\delta}$ under the multiplicative measurement errors setting (1.1) and propose some test statistics for measuring symmetry or asymmetry of X.

3.1 Skewness measures S_P and S_G

Pearson's skewness is defined as $S_P = \frac{E(X)-\vartheta}{\sigma}$, Hotelling and Solomons (1932) showed that $|S_P| \leq 1$. Suppose that $\{X_1, X_2, \dots, X_n\}$ are observable, Cabilio and Cabilio (1996) considered to use $\widehat{S}_P = \frac{\bar{X}-\mathrm{me}(X)}{s}$ for testing symmetry versus skewness, where $\bar{X}$ and $\mathrm{me}(X)$ are the sample mean and median respectively, and s^2 is the sample variance.

To estimate S_P in the multiplicative distortion measurement errors setting, we propose

$$\widehat{\widehat{S}}_{P,l} = \frac{\overline{\widetilde{X}} - \mathrm{me}(\hat{X}^{[l]})}{\hat{s}_l}, \quad l = 1, 2,$$

where, $\mathrm{me}(\hat{X}^{[l]})$ is the median of calibrated variables $\{\hat{X}_1^{[l]}, \dots, \hat{X}_n^{[l]}\}$, $\hat{s}_l^2$ is the estimator of σ^2, and $\hat{s}_l^2$ is defined as $\hat{s}_l^2 = \frac{1}{n}\sum_{i=1}^n \frac{(\widetilde{X}_i - \hat{E}(\widetilde{X}_i|U_i))^2}{\hat{\psi}_l^2(U_i)}$, $l = 1, 2$, here, $\hat{E}(\widetilde{X}_i|U_i) = \frac{S_{n2}(U_i)Q_{n0,\widetilde{X}}(U_i) - S_{n1}(U_i)Q_{n1,\widetilde{X}}(U_i)}{S_{n2}(U_i)S_{n0}(U_i) - [S_{n1}(U_i)]^2}$.

Theorem 3.1. *Assume that conditions* (A1)–(A5) *hold,* $f(\vartheta) > 0$, *we have*

$$\sqrt{n}(\widehat{\widehat{S}}_{P,l} - S_P) \xrightarrow{L} N(0, \sigma^2_{S_{P,l}}),$$

where

$$\sigma^2_{S_{P,1}} = \frac{1}{\sigma^2}\mathrm{Var}\left(\frac{\vartheta(2\widetilde{X} - X)}{E(X)} - \frac{I\{X \leq \vartheta\}}{f(\vartheta)} - \frac{E(X)-\vartheta}{2\sigma^2}[X - E(X)]^2\right),$$

$$\sigma^2_{S_{P,2}} = \frac{1}{\sigma^2}\mathrm{Var}\left(\widetilde{X} + \frac{2\vartheta - E(X)}{E(|X|)}|\widetilde{X}| - \frac{\vartheta}{E(|X|)}|X| - \frac{I\{X \leq \vartheta\}}{f(\vartheta)} - \frac{E(X)-\vartheta}{2\sigma^2}[X - E(X)]^2\right).$$

Remark 3. Note that if X is a positive random, we have $|X| = X$ and $E(|X|) = E(X)$, and the assumption M2 also entails that $|\widetilde{X}| = |X|\psi(U) = X\psi(U) = \widetilde{X}$. Then, the asymptotic variances $\sigma^2_{S_{P,1}}, \sigma^2_{S_{P,2}}$ satisfy $\sigma^2_{S_{P,1}} = \sigma^2_{S_{P,2}}$. If X is a negative random, we have $|X| = -X$ and $E(|X|) = -E(X)$, and the assumption M2 also entails that $|\widetilde{X}| = |X|\psi(U) = -X\psi(U) = -\widetilde{X}$. Then, the asymptotic variances $\sigma^2_{S_{P,1}}, \sigma^2_{S_{P,2}}$ also satisfy $\sigma^2_{S_{P,1}} = \sigma^2_{S_{P,2}}$.

If X is symmetric with mean $E(X)$ and median ϑ, it is easily seen that $E(X) = \vartheta$. Then, the asymptotic variances $\sigma^2_{S_{P,1}}, \sigma^2_{S_{P,2}}$ reduce to

$$\sigma^2_{S_{P,1}} = \left\{1 + \frac{4E(\widetilde{X}^2)}{\sigma^2}\frac{\mathrm{Var}(\psi(U))}{\mathrm{Var}(\psi(U))+1} + \frac{1}{4\sigma^2 f^2(\vartheta)} + \frac{E|X-\vartheta|}{\sigma^2 f(\vartheta)}\right\} \times I\{\vartheta \neq 0\},$$

$$\sigma^2_{S_{P,2}} = \frac{\mathrm{Var}(\widetilde{X})}{\sigma^2} + \frac{\vartheta^2 E(\widetilde{X}^2)}{[E|X|]^2\sigma^2}\frac{\mathrm{Var}(\psi(U))}{\mathrm{Var}(\psi(U))+1} + \frac{2\vartheta E(\widetilde{X}|\widetilde{X}|)}{E|X|\sigma^2}\frac{\mathrm{Var}(\psi(U))}{\mathrm{Var}(\psi(U))+1} + \frac{1}{4\sigma^2 f^2(\vartheta)} + \frac{E|X-\vartheta|}{\sigma^2 f(\vartheta)}.$$

Moreover, if $\vartheta = 0$, the asymptotic variance $\sigma^2_{S_{P,2}}$ further reduces to $\frac{\mathrm{Var}(\widetilde{X})}{\sigma^2} + \frac{1}{4\sigma^2 f^2(0)} + \frac{E|X|}{\sigma^2 f(0)}$.

Next, we propose consistent estimators of $\sigma^2_{S_{P,1}}$ and $\sigma^2_{S_{P,2}}$ when X is symmetric. A consistent estimator of $\sigma^2_{S_{P,1}}$ is proposed as

$$\hat{\sigma}^2_{S_{P,1}} = 1 + \frac{4\overline{\widetilde{X}^2}}{\hat{s}_1^2}\frac{\widehat{\operatorname{Var}}_1(\psi(U))}{\widehat{\operatorname{Var}}_1(\psi(U))+1} + \frac{1}{4\hat{s}_1^2\hat{f}_1^2(\operatorname{me}(\hat{X}^{[1]}))} + \frac{\hat{E}_1(|X-\vartheta|)}{\hat{s}_1^2\hat{f}_1(\operatorname{me}(\hat{X}^{[1]}))},$$

where $\overline{\widetilde{X}^2} = \frac{1}{n}\sum_{i=1}^n \widetilde{X}_i^2$, $\hat{f}_1(\operatorname{me}(\hat{X}^{[1]}))$ is defined in (2.4) by substituting x with $\operatorname{me}(\hat{X}^{[1]})$, $\widehat{\operatorname{Var}}_1(\psi(U))$ is defined as $\widehat{\operatorname{Var}}_1(\psi(U)) = \frac{1}{n}\sum_{i=1}^n \hat{\psi}_1^2(U_i) - \{\overline{\hat{\psi}_1(U)}\}^2$, $\overline{\hat{\psi}_1(U)} = \frac{1}{n} \times \sum_{i=1}^n \hat{\psi}_1(U_i)$, and $\hat{E}_1(|X-\vartheta|) = \frac{1}{n}\sum_{i=1}^n |\hat{X}_i^{[1]} - \operatorname{me}(\hat{X}^{[1]})|$.

A consistent estimator of $\sigma^2_{S_{P,2}}$ is proposed as

$$\begin{aligned}\hat{\sigma}^2_{S_{P,2}} = {} & \frac{\widehat{\operatorname{Var}}(\widetilde{X})}{\hat{s}_2^2} + \frac{\operatorname{me}^2(\hat{X}^{[2]})\overline{\widetilde{X}^2}}{(\overline{|\widetilde{X}|})^2\hat{s}_2^2}\frac{\widehat{\operatorname{Var}}_2(\psi(U))}{\widehat{\operatorname{Var}}_2(\psi(U))+1} \\ & + \frac{2\operatorname{me}(\hat{X}^{[2]})\overline{\widetilde{X}|\widetilde{X}|}}{\overline{|\widetilde{X}|}\hat{s}_2^2}\frac{\widehat{\operatorname{Var}}_2(\psi(U))}{\widehat{\operatorname{Var}}_2(\psi(U))+1} \\ & + \frac{1}{4\hat{s}_2^2\hat{f}_2^2(\operatorname{me}(\hat{X}^{[2]}))} + \frac{\hat{E}_2|X-\vartheta|}{\hat{s}_2^2\hat{f}_2(\operatorname{me}(\hat{X}^{[2]}))},\end{aligned}$$

where, $\widehat{\operatorname{Var}}(\widetilde{X}) = \frac{1}{n}\sum_{i=1}^n(\widetilde{X}_i - \overline{\widetilde{X}})^2$, $\overline{\widetilde{X}|\widetilde{X}|} = \frac{1}{n}\sum_{i=1}^n \widetilde{X}_i|\widetilde{X}_i|$, $\widehat{\operatorname{Var}}_2(\psi(U)) = \frac{1}{n} \times \sum_{i=1}^n \hat{\psi}_2^2(U_i) - \{\overline{\hat{\psi}_2(U)}\}^2$, $\overline{\hat{\psi}_2(U)} = \frac{1}{n}\sum_{i=1}^n \hat{\psi}_2(U_i)$, and also $\hat{E}_2(|X-\vartheta|) = \frac{1}{n}\sum_{i=1}^n |\hat{X}_i^{[2]} - \operatorname{me}(\hat{X}^{[2]})|$, moreover, $\hat{f}_2(\operatorname{me}(\hat{X}^{[2]}))$ is defined in (2.5) by substituting x with $\operatorname{me}(\hat{X}^{[2]})$. In particular, if $\vartheta = 0$, a consistent estimator of the asymptotic variance $\sigma^2_{S_{P,2}}$ is defined as $\hat{\sigma}^2_{0,S_{P,2}} = \frac{\widehat{\operatorname{Var}}(\widetilde{X})}{\hat{s}_2^2} + \frac{1}{4\hat{s}_2^2\hat{f}_2^2(0)} + \frac{\overline{|\widetilde{X}|}}{\hat{s}_2^2\hat{f}_2(0)}$.

Theorem 3.2. *Under the conditions of Theorem 3.1, if X is symmetric and satisfies $\vartheta = E(X)$, we have*:

(i) *If $\vartheta \neq 0$, $\sqrt{n}\hat{S}_{P,1}/\hat{\sigma}_{S_{P,1}} \xrightarrow{L} N(0,1)$.*
(ii) $\sqrt{n}\hat{S}_{P,2}/\hat{\sigma}_{S_{P,2}} \xrightarrow{L} N(0,1)$.
(iii) *In particular, if $\vartheta = 0$, $\sqrt{n}\hat{S}_{P,2}/\hat{\sigma}_{0,S_{P,2}} \xrightarrow{L} N(0,1)$.*

Using Theorem 3.2, for a sample of size n, the proposed α-level asymptotic free test of $\mathcal{H}_0^*$: $f(x+\vartheta) = f(\vartheta - x)$ when $\vartheta \neq 0$ is $|\sqrt{n}\hat{\widetilde{S}}_{P,l}/\hat{\sigma}_{S_{P,l}}| \geq z_{1-\alpha/2}$, $l = 1, 2$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution, i.e., $\Phi(z_{1-\alpha/2}) = 1-\alpha/2$, $\alpha \in (0,1)$. If $\vartheta = 0$, we can use $|\sqrt{n}\hat{\widetilde{S}}_{P,2}/\hat{\sigma}_{S_{P,2}}| \geq z_{1-\alpha/2}$ or $|\sqrt{n}\hat{\widetilde{S}}_{P,2}/\hat{\sigma}_{0,S_{P,2}}| \geq z_{1-\alpha/2}$ to test the hypothesis $\mathcal{H}_0^*$.

The measure S_G is known as Groeneveld & Meeden's measure of skewness (Groeneveld and Meeden (1984)). An estimator of S_G is proposed as

$$\hat{\widetilde{S}}_{G,l} = \frac{\overline{\widetilde{X}} - \operatorname{me}(\hat{X}^{[l]})}{\hat{E}_l|X-\vartheta|}, \quad l = 1, 2.$$

We have the following asymptotic result.

Theorem 3.3. *Under the condition of Theorem 3.1, we have*

$$\sqrt{n}(\hat{\widetilde{S}}_{G,l} - S_G) \xrightarrow{L} N(0, \sigma^2_{S_{G,l}}),$$

where

$$\sigma^2_{S_{G,1}} = \frac{1}{[E|X-\vartheta|]^2}\mathrm{Var}\biggl(\frac{2\vartheta}{E(X)}\widetilde{X} - \frac{2\vartheta - E(X)}{E(X)}X - \frac{I\{X \le \vartheta\}}{f(\vartheta)} - \frac{E(X)-\vartheta}{E|X-\vartheta|}|X-\vartheta|\biggr),$$

$$\sigma^2_{S_{G,2}} = \frac{1}{[E|X-\vartheta|]^2}\mathrm{Var}\biggl(\widetilde{X} + \frac{2\vartheta - E(X)}{E(|X|)}(|\widetilde{X}| - |X|) - \frac{I\{X \le \vartheta\}}{f(\vartheta)} - \frac{E(X)-\vartheta}{E|X-\vartheta|}|X-\vartheta|\biggr).$$

Remark 4. Note that if X is a positive random, we have $|X| = X$ and $E(|X|) = E(X)$, and the assumption M2 also entails that $|\widetilde{X}| = |X|\psi(U) = X\psi(U) = \widetilde{X}$. Then, the asymptotic variances $\sigma^2_{S_{G,1}}$, $\sigma^2_{S_{G,2}}$ also satisfy $\sigma^2_{S_{G,1}} = \sigma^2_{S_{G,2}}$. If X is a negative random, we have $|X| = -X$ and $E(|X|) = -E(X)$, and the assumption M2 also entails that $|\widetilde{X}| = |X|\psi(U) = -X\psi(U) = -\widetilde{X}$. Then, the asymptotic variances $\sigma^2_{S_{G,1}}$, $\sigma^2_{S_{G,2}}$ still satisfy $\sigma^2_{S_{G,1}} = \sigma^2_{S_{G,2}}$.

When X is symmetric about ϑ and $E(X) = \vartheta$, if $\vartheta \neq 0$, the asymptotic variance $\sigma^2_{S_{G,1}}$ reduces to $\sigma^2_{S_{G,1}} = \frac{\sigma^2}{[E|X-\vartheta|]^2}\sigma^2_{S_{P,1}}$, then a consistent estimator of $\sigma^2_{S_{G,1}}$ is obtained as $\hat{\sigma}^2_{S_{G,1}} = \frac{\hat{s}_1^2}{[\hat{E}_1|X-\vartheta|]^2}\hat{\sigma}^2_{S_{P,1}}$. For the estimator $\widehat{\widehat{S}}_{G,2}$, the asymptotic variance $\sigma^2_{S_{G,2}}$ also reduces to $\sigma^2_{S_{G,2}} = \frac{\sigma^2}{[E|X-\vartheta|]^2}\sigma^2_{S_{P,2}}$, and then a consistent estimator of $\sigma^2_{S_{G,2}}$ is obtained as $\hat{\sigma}^2_{S_{G,2}} = \frac{\hat{s}_2^2}{[\hat{E}_2|X-\vartheta|]^2}\hat{\sigma}^2_{S_{P,2}}$. Thus, the proposed α-level asymptotic free test of $\mathcal{H}_0^*: f(x+\vartheta) = f(\vartheta - x)$ is $|\sqrt{n}\widehat{\widehat{S}}_{G,l}/\hat{\sigma}_{S_{G,l}}| \ge z_{1-\alpha/2}$, $l = 1, 2$. In particular, if $\vartheta = 0$, a consistent estimator of the asymptotic variance $\sigma^2_{S_{G,2}}$ is defined as $\hat{\sigma}^2_{0,S_{G,2}} = \frac{\hat{s}_2^2}{[\hat{E}_2|X-\vartheta|]^2}\hat{\sigma}^2_{0,S_{P,2}}$. We have the following theorem.

Theorem 3.4. *Under the conditions of Theorem* 3.1, *if X is symmetric and satisfies $\vartheta = E(X)$, we have*:

(i) *If* $\vartheta \neq 0$, $\sqrt{n}\hat{S}_{G,1}/\hat{\sigma}_{S_{G,1}} \xrightarrow{L} N(0, 1)$.

(ii) $\sqrt{n}\hat{S}_{G,2}/\hat{\sigma}_{S_{G,2}} \xrightarrow{L} N(0, 1)$.

(iii) *In particular, if* $\vartheta = 0$, $\sqrt{n}\hat{S}_{G,2}/\hat{\sigma}_{0,S_{G,2}} \xrightarrow{L} N(0, 1)$.

3.2 The measure S_M

An asymptotic free test based on the sample skewness $\widehat{S}_M = \frac{n^{-1}\sum_{i=1}^n (X_i - \bar{X})^3}{[n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2]^{3/2}}$ is proposed by Gupta (1967) when the *i.i.d.* sample $\{X_1, \ldots, X_n\}$ are available and have finite population moments up to the sixth order. For the multiplicative distortion measurement errors considered in this paper, we proposed an estimator of skewness as

$$\widehat{\widehat{S}}_{M,l} = \frac{n^{-1}\sum_{i=1}^n (\widetilde{X}_i - \hat{E}(\widetilde{X}_i|U_i))^3}{[n^{-1}\sum_{i=1}^n \hat{\psi}_l^3(U_i)]\hat{s}_l^3}.$$

Theorem 3.5. *Assume that conditions* (A1)–(A5) *hold*, $E(X^6) < \infty$, *we have*

$$\sqrt{n}(\widehat{\widehat{S}}_{M,l} - S_M) \xrightarrow{L} N(0, \sigma^2_{S_M}), \quad l = 1, 2,$$

where,

$$\sigma^2_{S_M} = \mathrm{Var}\Big(\frac{\psi^3(U)(X-E(X))^3}{E[\psi^3(U)]\sigma^3} - \frac{S_M}{E[\psi^3(U)]}\psi^3(U) - \frac{3S_M}{2\sigma^2}(X-E(X))^2\Big).$$

Remark 5. The asymptotic variance $\sigma^2_{S_M}$ is expressed as $\sigma^2_{S_M} = \frac{\delta_\psi(6)\mu_X(6)}{\delta^2_\psi(3)\sigma^6} - S^2_M\frac{\delta_\psi(6)}{\delta^2_\psi(3)} + S^2_M\frac{9\mu_X(4)}{4\sigma^4} - 3S_M\frac{\mu_X(5)}{\sigma^5} + \frac{3}{4}S^2_M$, where $\delta_\psi(d) = E[\psi^d(U)]$ and $\mu_X(d) = E[(X-E(X))^d]$. When X is symmetric with $E(X)$ and $E(X^6) < \infty$, we have $S_M = 0$ and the asymptotic variance $\sigma^2_{S_M}$ reduces to $\sigma^2_{S_M} = \frac{\delta_\psi(6)\mu_X(6)}{\delta^2_\psi(3)\sigma^6}$. It is known that when X is symmetric and observed without measurement errors, the asymptotic variance of $\widehat{S}_M$ is shown to be $\frac{\mu_X(6)}{\sigma^6}$ (Gupta (1967)). The Cauchy-inequality entails that $\delta_\psi(6) \geq \delta^2_\psi(3)$, and then $\sigma^2_{S_M} \geq \frac{\mu_X(6)}{\sigma^6}$ when X is symmetric. This implies the multiplicative distortion function $\psi(U)$ increases the classical asymptotic variance of sample skewness.

Note that $E\{[\widetilde{X} - E(\widetilde{X}|U)]^6\} = \delta_\psi(6)\mu_X(6)$. When X is symmetric about $E(X)$ and $E(X) \neq 0$, a consistent estimator of $\sigma^2_{S_M}$ is proposed as $\hat{\sigma}^2_{S_{M,l}} = \frac{n^{-1}\sum_{i=1}^n(\widetilde{X}_i - \hat{E}(\widetilde{X}_i|U_i))^6}{\{n^{-1}\sum_{i=1}^n\hat{\psi}^3_l(U_i)\}^2\hat{s}^6_l}$, $l = 1, 2$. Thus, a α-level asymptotic free test of $\mathcal{H}^*_0 : f(x+\vartheta) = f(\vartheta - x)$ is $|\sqrt{n}\widehat{S}_{M,l}/\hat{\sigma}_{S_{M,l}}| \geq z_{1-\alpha/2}$. If $E(X) = 0$, the α-level asymptotic free test is $|\sqrt{n}\widehat{\widehat{S}}_{M,2}/\hat{\sigma}_{S_{M,2}}| \geq z_{1-\alpha/2}$.

3.3 The measure $S_{Q,\delta}$

The Galton's measure of skewness (also known as Bowley's measure of skewness) is known as $S_{Q,0.25}$ when $\delta = 0.25$, and the Kelley's measure of skewness is known as $S_{Q,0.1}$ when $\delta = 0.1$. For the multiplicative distortion measurement errors, we propose estimators of $S_{Q,\delta}$ as

$$\widehat{\widehat{S}}^{[l]}_{Q,\delta} = \frac{\widehat{\widehat{q}}^{[l]}_{1-\delta} + \widehat{\widehat{q}}^{[l]}_{\delta} - 2\,\mathrm{me}(\hat{X}^{[l]})}{\widehat{\widehat{q}}^{[l]}_{1-\delta} - \widehat{\widehat{q}}^{[l]}_{\delta}}, \quad \delta \in (0, 0.5), l = 1, 2,$$

where $\widehat{\widehat{q}}^{[l]}_{\delta}$ is the δ-quantile of $\{\hat{X}^{[l]}_1, \ldots, \hat{X}^{[l]}_n\}$, $l = 1, 2$.

Theorem 3.6. *Assume that conditions* (A1)–(A5) *hold, if* $f(\vartheta) > 0$, $f(q_\delta) > 0$ *and* $f(q_{1-\delta}) > 0$, *we have*

$$\sqrt{n}(\widehat{\widehat{S}}^{[l]}_{Q,\delta} - S_{Q,\delta}) \xrightarrow{L} N(0, \sigma^2_{S_{Q,\delta}}), \quad l = 1, 2,$$

where, let $\omega_\delta = \frac{q_{1-\delta}-\vartheta}{q_{1-\delta}-q_\delta}$, *and*

$$\begin{aligned}\sigma^2_{S_{Q,\delta}} = &\frac{4\delta(1-\delta)}{(q_{1-\delta}-q_\delta)^2}\Big[\frac{\omega^2_\delta}{f^2(q_\delta)} + \frac{(1-\omega_\delta)^2}{f^2(q_{1-\delta})}\Big] + \frac{1}{(q_{1-\delta}-q_\delta)^2 f^2(\vartheta)} \\ &+ \frac{8\delta^2\omega_\delta(1-\omega_\delta)}{(q_{1-\delta}-q_\delta)^2 f(q_\delta)f(q_{1-\delta})} - \frac{4\delta\omega_\delta}{(q_{1-\delta}-q_\delta)^2 f(q_\delta)f(\vartheta)} \\ &- \frac{4\delta(1-\omega_\delta)}{(q_{1-\delta}-q_\delta)^2 f(q_{1-\delta})f(\vartheta)}.\end{aligned}$$

Remark 6. Note that estimators $\widehat{\widehat{S}}^{[l]}_{Q,\delta}$ of skewness are *efficient*. In other words, the proposed estimator $\widehat{\widehat{S}}^{[l]}_{Q,\delta}$ eliminate the effect caused by the additive adjusted covariate $\psi(U)$, that is, the effect of additive errors vanishes.

When the density function of X is symmetric with $E(X) \neq 0$, we have $\omega_\delta = \frac{1}{2}$, $f(q_\delta) = f(q_{1-\delta})$, the asymptotic variance $\sigma^2_{S_{Q,\delta}}$ reduces to $\sigma^2_{S_{Q,\delta}} = \frac{1}{(q_{1-\delta}-q_\delta)^2}[\frac{2\delta}{f^2(q_\delta)} + \frac{1}{f^2(\vartheta)} - \frac{4\delta}{f(q_\delta)f(\vartheta)}]$, and a consistent estimator of $\sigma^2_{S_{Q,\delta}}$ is proposed as

$$(\hat{\sigma}^{[l]}_{S_{Q,\delta}})^2 = \frac{1}{(\widehat{\widetilde{q}}^{[l]}_{1-\delta} - \widehat{\widetilde{q}}^{[l]}_{\delta})^2}\Big[\frac{2\delta}{\hat{f}_l^2(\widehat{\widetilde{q}}^{[l]}_{\delta})} + \frac{1}{\hat{f}_l^2(\mathrm{me}(\hat{X}^{[l]}))} - \frac{4\delta}{\hat{f}_l(\widehat{\widetilde{q}}^{[l]}_{\delta})\hat{f}_l(\mathrm{me}(\hat{X}^{[l]}))}\Big].$$

The α-level asymptotic free test of $\mathcal{H}_0^*$: $f(x+\vartheta) = f(\vartheta - x)$, can be constructed as $|\sqrt{n}\widehat{\widetilde{S}}^{[l]}_{Q,\delta}/\hat{\sigma}^{[l]}_{S_{Q,\delta}}| \geq z_{1-\alpha/2}$, where $\hat{\sigma}^{[l]}_{S_{Q,\delta}} = \sqrt{(\hat{\sigma}^{[l]}_{S_{Q,\delta}})^2}$, $l = 1, 2$. If $E(X) = 0$, the α-level asymptotic free test is $|\sqrt{n}\widehat{\widetilde{S}}^{[2]}_{Q,\delta}/\hat{\sigma}^{[2]}_{S_{Q,\delta}}| \geq z_{1-\alpha/2}$.

4 Simulation studies

In this section, we present numerical results to evaluate the performance of the proposed estimators and test statistics. In the following simulation and real data analysis, the Epanechnikov kernel $K(t) = 0.75(1-t^2)^+$ is used. To select bandwidth h_s, $s = 1, \dots, 4$, a undersmoothing bandwidth for h_s is needed due to condition (A5). An ad-hoc but reasonable choice is $O(n^{-1/5}) \times n^{-2/15} = O(n^{-1/3})$. Then, we use the rule of thumb and the bandwidth h_s's are chosen as $h_1 = \hat{\sigma}_U n^{-1/3}$, $h_2 = \hat{\sigma}_{\hat{X}^{[1]}} n^{-1/3}$, $h_3 = \hat{\sigma}_{\hat{X}^{[2]}} n^{-1/3}$ and $h_4 = \hat{\sigma}_{\widetilde{X}} n^{-1/3}$, where $\hat{\sigma}_U$, $\hat{\sigma}_{\hat{X}^{[1]}}$, $\hat{\sigma}_{\hat{X}^{[2]}}$ and $\hat{\sigma}_{\widetilde{X}}$ are the sample standard deviation of $\{U_i, \hat{X}^{[1]}_i, \hat{X}^{[2]}_i, \widetilde{X}_i\}_{i=1}^n$, respectively.

Example 1. In this example, we generate 2000 realization and the sample size is chosen as $n = 100$, 300 and 500. The variable U is generated from a uniform distribution $U(0, 1)$ and multiplicative distortion function is chosen as $\psi(U) = 1 - 0.5\cos(2\pi U)$.

In Table 1, we consider that the variable $X \sim N(-1, 1)$. In this case, $E(X) = -1 \neq 0$, the estimators $\hat{\rho}^{[s]}_k$, $s = 1, 2$ works. We investigate the performances of the benchmark estimator $\hat{\rho}^B_k$ (using the true covariate X in the simulation), the proposed estimator $\hat{\rho}^{[s]}_k$, $s = 1, 2$, and the naive estimator $\hat{\tilde{\rho}}_k$ for $k = 0.5, 1, 2, 3$. We report the means, standard errors, and the 95% confidence intervals by using the proposed empirical likelihood method, associated with the lower and upper bounds, the average lengthes and the coverage probabilities. In Table 1, we see that when the sample size n is 300 or 500, the performances of $\hat{\rho}^B_{0.5}$, $\hat{\rho}^{[s]}_{0.5}$, $s = 1, 2$ are better than $\hat{\rho}^{[s]}_k$ and $\hat{\rho}^B_k$, $k = 1, 2, 3$ because the values of mean of $\hat{\rho}^{[s]}_{0.5}$ and $\hat{\rho}^B_{0.5}$ are close to zero and the values of MSE are smaller. We find that the performances of $\hat{\rho}^{[s]}_k$, $s = 1, 2$ are almost the same as $\hat{\rho}^B_k$ when the sample size n gets larger. We see that the empirical likelihood confidence intervals of $\hat{\rho}^{[s]}_{0.5}$, $s = 1, 2$ and $\hat{\rho}^B_{0.5}$ show satisfactory performances, and the coverage probabilities of $\hat{\rho}^{[s]}_{0.5}$, $\hat{\rho}^B_{0.5}$ are much better than $\hat{\rho}^{[s]}_k$ and $\hat{\rho}^B_k$, $k = 1, 2, 3$, which indicate that a larger value of k may cause lower coverage probabilities with shorter average lengthes but has much smaller values of MSE. When the sample size n increases, the performances of $\hat{\rho}^{[s]}_{0.5}$, $s = 1, 2$ and $\hat{\rho}^B_{0.5}$ become better both in terms of average lengthes of the confidence intervals and the coverage probabilities. Generally, when the sample size n gets larger, such as 300 or 500, the values of MSE of $\hat{\rho}^{[2]}_{0.5}$ are all smaller than those of $\hat{\rho}^{[1]}_{0.5}$. This is not surprised and coincided with the Remark 2 in Theorem 2.1, the estimator $\hat{\rho}^{[2]}_{0.5}$ performs better than $\hat{\rho}^{[1]}_{0.5}$ in estimation because $|E(X)| = 1$, $E(|X|) = 1.166$ and $\frac{1}{[E(X)]^2} > \frac{1}{[E|X|]^2}$ holds true. Besides, the naive estimator $\hat{\tilde{\rho}}_k$ fails to recover the symmetry of the underlying unobserved variable X. The values of mean of $\hat{\tilde{\rho}}_k$ all depart from zero, and the confidence

Table 1 *The means (MEAN), standard errors (SE), mean squared errors (MSE) and the* 95% *confidence intervals for* $\hat{\rho}_k^B$, $\hat{\rho}_k^{[1]}$, $\hat{\rho}_k^{[2]}$ *and* $\hat{\hat{\rho}}_k$. *"Lower" stands for the lower bound, "Upper" stands for upper bound, "AL" stands for average length, "P" stands for the coverage probabilities when* $X \sim N(-1, 1)$. *MSE is in the scale of* 10^{-3}

			MEAN	SE	MSE	Lower	Upper	AL	P
$n = 100$	$k = 0.5$	$\hat{\rho}_k^B$	0.0795	0.1456	27.5361	−0.2577	0.2572	0.5105	94.6%
		$\hat{\rho}_k^{[1]}$	0.0850	0.1344	25.3149	−0.2486	0.2672	0.5158	95.3%
		$\hat{\rho}_k^{[2]}$	0.0670	0.1452	25.5766	−0.2713	0.2433	0.5146	94.2%
		$\hat{\hat{\rho}}_k$	0.3165	0.1285	20.8157	−0.0276	0.4664	0.4940	95.1%
	$k = 1$	$\hat{\rho}_k^B$	0.0473	0.1539	25.9349	−0.2485	0.2469	0.4954	91.6%
		$\hat{\rho}_k^{[1]}$	0.0545	0.1421	23.1571	−0.2387	0.2559	0.4947	94.2%
		$\hat{\rho}_k^{[2]}$	0.0324	0.1525	24.3222	−0.2632	0.2309	0.4942	92.6%
		$\hat{\hat{\rho}}_k$	0.2948	0.1326	18.5536	−0.0061	0.4609	0.4671	92.7%
	$k = 2$	$\hat{\rho}_k^B$	0.0297	0.1649	28.0782	−0.2306	0.2272	0.4579	87.8%
		$\hat{\rho}_k^{[1]}$	0.0363	0.1525	24.5767	−0.2201	0.2346	0.4548	89.3%
		$\hat{\rho}_k^{[2]}$	0.0124	0.1620	26.3909	−0.2465	0.2096	0.4562	87.6%
		$\hat{\hat{\rho}}_k$	0.2730	0.1378	19.0801	0.0122	0.4402	0.4280	89.2%
	$k = 3$	$\hat{\rho}_k^B$	0.0209	0.1649	27.6467	−0.2126	0.2165	0.4292	80.1%
		$\hat{\rho}_k^{[1]}$	0.0272	0.1538	24.4122	−0.1998	0.2251	0.4249	83.0%
		$\hat{\rho}_k^{[2]}$	0.0061	0.1602	25.6939	−0.2233	0.2032	0.4265	80.7%
		$\hat{\hat{\rho}}_k$	0.2483	0.1361	18.5477	0.0292	0.4266	0.3974	84.0%
$n = 300$	$k = 0.5$	$\hat{\rho}_k^B$	0.0252	0.0801	7.0508	−0.1579	0.1583	0.3162	94.4%
		$\hat{\rho}_k^{[1]}$	0.0414	0.0779	7.7852	−0.1424	0.1732	0.3156	95.5%
		$\hat{\rho}_k^{[2]}$	0.0323	0.0796	7.3831	−0.1517	0.1643	0.3160	94.5%
		$\hat{\hat{\rho}}_k$	0.2690	0.0715	5.4458	0.0885	0.3892	0.3006	97.5%
	$k = 1$	$\hat{\rho}_k^B$	0.0151	0.0871	7.8229	−0.1491	0.1499	0.2991	90.2%
		$\hat{\rho}_k^{[1]}$	0.0299	0.0844	8.0284	−0.1348	0.1636	0.2984	90.6%
		$\hat{\rho}_k^{[2]}$	0.0200	0.0859	7.7906	−0.1446	0.1539	0.2985	90.4%
		$\hat{\hat{\rho}}_k$	0.2696	0.0759	5.7961	0.1107	0.3885	0.2778	93.2%
	$k = 2$	$\hat{\rho}_k^B$	0.0094	0.0974	9.5877	−0.1353	0.1371	0.2724	82.0%
		$\hat{\rho}_k^{[1]}$	0.0222	0.0943	9.4000	−0.1222	0.1488	0.2711	82.6%
		$\hat{\rho}_k^{[2]}$	0.0121	0.0955	9.2653	−0.1327	0.1391	0.2718	82.0%
		$\hat{\hat{\rho}}_k$	0.2612	0.0820	6.7278	0.1231	0.3720	0.2488	87.2%
	$k = 3$	$\hat{\rho}_k^B$	0.0080	0.1056	11.2267	−0.1304	0.1235	0.2539	75.8%
		$\hat{\rho}_k^{[1]}$	0.0189	0.1035	11.0796	−0.1172	0.1347	0.2520	77.3%
		$\hat{\rho}_k^{[2]}$	0.0095	0.1052	11.1605	−0.1266	0.1260	0.2526	77.8%
		$\hat{\hat{\rho}}_k$	0.2486	0.0836	7.0114	0.1261	0.3571	0.2310	84.9%

intervals exclude zero when the sample size n is larger or equal to 300. This indicates that the distorting function $\psi(U)$ ruins the symmetry of the unobserved variable X, and we could not ignore the multiplicative effect caused by the confounding variable U.

In Table 2, we consider that the variable $X \sim N(0, 1)$. In this case, $E(X) = 0$, the estimator $\hat{\rho}_k^{[1]}$ fails but the estimator $\hat{\rho}_k^{[2]}$ works. We investigate the performance of $\hat{\rho}_k^B$, $\hat{\rho}_k^{[2]}$ and $\hat{\hat{\rho}}_k$ for $k = 0.5, 1, 2, 3$. Note that X is symmetric about zero, then the naive estimator $\hat{\hat{\rho}}_k$ works

Table 1 (*Continued*)

			MEAN	SE	MSE	Lower	Upper	AL	P
$n = 500$	$k = 0.5$	$\hat{\rho}_k^B$	0.0132	0.0630	4.1481	−0.1251	0.1242	0.2494	94.5%
		$\hat{\rho}_k^{[1]}$	0.0304	0.0632	4.9321	−0.1070	0.1423	0.2493	95.5%
		$\hat{\rho}_k^{[2]}$	0.0246	0.0635	4.6453	−0.1137	0.1356	0.2493	95.3%
		$\hat{\hat{\rho}}_k$	0.2618	0.0557	3.2251	0.1254	0.3625	0.2371	96.5%
	$k = 1$	$\hat{\rho}_k^B$	0.0073	0.0690	4.8211	−0.1183	0.1174	0.2357	91.5%
		$\hat{\rho}_k^{[1]}$	0.0227	0.0689	5.2672	−0.1017	0.1336	0.2354	92.2%
		$\hat{\rho}_k^{[2]}$	0.0168	0.0693	5.0872	−0.1085	0.1269	0.2354	91.8%
		$\hat{\hat{\rho}}_k$	0.2674	0.0591	3.5070	0.1464	0.3645	0.2181	93.8%
	$k = 2$	$\hat{\rho}_k^B$	0.0037	0.0780	6.1007	−0.1080	0.1060	0.2140	83.8%
		$\hat{\rho}_k^{[1]}$	0.0167	0.0773	6.2578	−0.0932	0.1200	0.2133	83.7%
		$\hat{\rho}_k^{[2]}$	0.0108	0.0778	6.1736	−0.0998	0.1137	0.2136	84.4%
		$\hat{\hat{\rho}}_k$	0.2618	0.0640	4.0967	0.1561	0.3508	0.1946	86.3%
	$k = 3$	$\hat{\rho}_k^B$	0.0027	0.0874	7.6495	−0.1011	0.0978	0.1989	76.9%
		$\hat{\rho}_k^{[1]}$	0.0151	0.0841	7.3065	−0.0883	0.1092	0.1976	74.8%
		$\hat{\rho}_k^{[2]}$	0.0092	0.0855	7.4030	−0.0946	0.1034	0.1980	74.1%
		$\hat{\hat{\rho}}_k$	0.2494	0.0661	4.3832	0.1526	0.3323	0.1796	80.1%

for testing the symmetry of X about zero. In Table 2, we see that when the sample size n is 300 or 500, the performances of $\hat{\rho}_{0.5}^B$, $\hat{\rho}_{0.5}^{[2]}$ and $\hat{\hat{\rho}}_{0.5}$ are better than $\hat{\rho}_k^B$, $\hat{\rho}_k^{[2]}$ and $\hat{\hat{\rho}}_k$, $k = 1, 2, 3$, because the values of mean of $\hat{\rho}_{0.5}^{[2]}$ and $\hat{\rho}_{0.5}^B$ are close to zero and the values of MSE are smaller, and coverage probabilities of empirical likelihood confidence intervals of $\hat{\rho}_{0.5}^B$, $\hat{\rho}_{0.5}^{[2]}$ and $\hat{\hat{\rho}}_k$ show satisfactory performances and are closer to 95%. Similar to the simulation results reported in Table 1, a larger value of k causes lower coverage probabilities with shorter average lengthes even it results in smaller values of MSE. When the sample size n increases, the performances of $\hat{\rho}_{0.5}^{[2]}$ are close to the benchmark estimator $\hat{\rho}_{0.5}^B$ both in terms of the estimation and confidence intervals.

In Table 3, we consider that the variable $X \sim \chi^2(5) - 5$. In this case, $E(X) = 0$ but X is asymmetric. The estimator $\hat{\rho}_k^{[1]}$ also fails and the estimator $\hat{\rho}_k^{[2]}$ works. We investigate the performances of $\hat{\rho}_k^B$, $\hat{\rho}_k^{[2]}$ and $\hat{\hat{\rho}}_k$ for $k = 0.5, 1, 2, 3$. In Table 3, we see that when the sample size n is 300 or 500, the performances of $\hat{\rho}_{0.5}^B$ and $\hat{\rho}_{0.5}^{[2]}$ are better than $\hat{\rho}_k^B$ and $\hat{\rho}_k^{[2]}$, $k = 1, 2, 3$, because those have larger values of mean and lower values of coverage probabilities of empirical likelihood confidence intervals. Moreover, $\hat{\rho}_{0.5}^B$ and $\hat{\rho}_{0.5}^{[2]}$ show better performances and are closer to 95%. Similar to the simulation results reported in Table 1 and Table 2, a larger value of k leads to lower coverage probabilities although the sample size n increases. Note that X is asymmetric, then the naive estimator $\hat{\hat{\rho}}_k$ fails and has lager bias compared with $\hat{\rho}_k^B$ and $\hat{\rho}_k^{[2]}$. It is seen that the right confidence intervals of $\hat{\rho}_k^B$ and $\hat{\rho}_k^{[2]}$ exclude the left confidence intervals of $\hat{\hat{\rho}}_k$ when the sample size n is 300 and 500. This implies that the bias of $\hat{\hat{\rho}}_k$ is non-ignorable, and the distorting function $\psi(U)$ also ruins the asymmetry of the unobserved variable X.

Example 2. In this example, we generate 2000 realization and the sample size is chosen as $n = 100$, 300 and 500. The variable X is designed as $X \sim N(-2, 1)$, $X \sim \chi_5^2 - 1$ (a

Table 2 *The means (MEAN), standard errors (SE), mean squared errors (MSE) and the 95% confidence intervals for $\hat{\rho}_k^B$, $\hat{\rho}_k^{[1]}$, $\hat{\rho}_k^{[2]}$ and $\hat{\hat{\rho}}_k$. "Lower" stands for the lower bound, "Upper" stands for upper bound, "AL" stands for average length, "P" stands for the coverage probabilities when $X \sim N(0,1)$. MSE is in the scale of 10^{-3}*

			MEAN	SE	MSE	Lower	Upper	AL	P
$n=100$	$k=0.5$	$\hat{\rho}_k^B$	0.0807	0.1419	26.6534	−0.2577	0.2594	0.5173	95.9%
		$\hat{\rho}_k^{[2]}$	0.0797	0.1470	27.9862	−0.2578	0.2604	0.5183	93.9%
		$\hat{\hat{\rho}}_k$	0.0658	0.1241	19.7466	−0.2524	0.2570	0.5094	96.0%
	$k=1$	$\hat{\rho}_k^B$	0.0491	0.1513	25.3132	−0.2505	0.2473	0.4979	93.1%
		$\hat{\rho}_k^{[2]}$	0.0500	0.1543	26.3073	−0.2508	0.2489	0.4998	92.0%
		$\hat{\hat{\rho}}_k$	0.0417	0.1296	18.5500	−0.2439	0.2416	0.4855	96.0%
	$k=2$	$\hat{\rho}_k^B$	0.0328	0.1610	27.0065	−0.2320	0.2286	0.4609	87.0%
		$\hat{\rho}_k^{[2]}$	0.0338	0.1653	28.4590	−0.2325	0.2306	0.4632	85.6%
		$\hat{\hat{\rho}}_k$	0.0285	0.1368	19.5254	−0.2236	0.2208	0.4444	91.6%
	$k=3$	$\hat{\rho}_k^B$	0.0268	0.1664	28.4159	−0.2437	0.2450	0.4887	85.8%
		$\hat{\rho}_k^{[2]}$	0.0278	0.1717	30.2500	−0.2458	0.2457	0.4915	83.9%
		$\hat{\hat{\rho}}_k$	0.0234	0.1401	20.1806	−0.2369	0.2352	0.4722	91.2%
$n=300$	$k=0.5$	$\hat{\rho}_k^B$	0.0250	0.0832	7.5550	−0.1591	0.1566	0.3157	95.3%
		$\hat{\rho}_k^{[2]}$	0.0245	0.0834	7.5687	−0.1601	0.1562	0.3164	95.7%
		$\hat{\hat{\rho}}_k$	0.0190	0.0706	5.3600	−0.1519	0.1551	0.3071	97.2%
	$k=1$	$\hat{\rho}_k^B$	0.0157	0.0895	8.2707	−0.1502	0.1489	0.2992	91.9%
		$\hat{\rho}_k^{[2]}$	0.0157	0.0897	8.2952	−0.1512	0.1484	0.2996	91.4%
		$\hat{\hat{\rho}}_k$	0.0143	0.0741	5.7005	−0.1420	0.1460	0.2880	94.2%
	$k=2$	$\hat{\rho}_k^B$	0.0107	0.1001	10.1389	−0.1369	0.1357	0.2726	85.5%
		$\hat{\rho}_k^{[2]}$	0.0111	0.1006	10.2472	−0.1379	0.1352	0.2732	84.9%
		$\hat{\hat{\rho}}_k$	0.0104	0.0816	6.7688	−0.1274	0.1317	0.2592	88.4%
	$k=3$	$\hat{\rho}_k^B$	0.0091	0.1071	11.5617	−0.1429	0.1454	0.2884	83.9%
		$\hat{\rho}_k^{[2]}$	0.0096	0.1078	11.7203	−0.1437	0.1454	0.2892	83.2%
		$\hat{\hat{\rho}}_k$	0.0088	0.0861	7.4886	−0.1371	0.1352	0.2723	89.9%
$n=500$	$k=0.5$	$\hat{\rho}_k^B$	0.0168	0.0638	4.3616	−0.1249	0.1244	0.2493	95.3%
		$\hat{\rho}_k^{[2]}$	0.0164	0.0642	4.3990	−0.1253	0.1242	0.2496	95.7%
		$\hat{\hat{\rho}}_k$	0.0144	0.0545	3.1793	−0.1211	0.1206	0.2418	97.9%
	$k=1$	$\hat{\rho}_k^B$	0.0098	0.0717	5.2461	−0.1187	0.1172	0.2360	91.9%
		$\hat{\rho}_k^{[2]}$	0.0091	0.0721	5.2939	−0.1191	0.1172	0.2364	91.9%
		$\hat{\hat{\rho}}_k$	0.0076	0.0604	3.7145	−0.1140	0.1119	0.2259	95.5%
	$k=2$	$\hat{\rho}_k^B$	0.0062	0.0813	6.6469	−0.1086	0.1051	0.2138	83.3%
		$\hat{\rho}_k^{[2]}$	0.0056	0.0818	6.7307	−0.1088	0.1052	0.2141	83.4%
		$\hat{\hat{\rho}}_k$	0.0049	0.0666	4.4595	−0.1034	0.0993	0.2028	89.6%
	$k=3$	$\hat{\rho}_k^B$	0.0044	0.0878	7.7259	−0.1172	0.1091	0.2264	82.7%
		$\hat{\rho}_k^{[2]}$	0.0040	0.0883	7.8232	−0.1176	0.1092	0.2269	81.7%
		$\hat{\hat{\rho}}_k$	0.0038	0.0704	4.9710	−0.1111	0.1015	0.2126	87.4%

Table 3 *The means (MEAN), standard errors (SE), mean squared errors (MSE) and the 95% confidence intervals for $\hat{\rho}_k^B$, $\hat{\rho}_k^{[1]}$, $\hat{\rho}_k^{[2]}$ and $\hat{\hat{\rho}}_k$. "Lower" stands for the lower bound, "Upper" stands for upper bound, "AL" stands for average length, "P" stands for the coverage probabilities when $X \sim \chi^2(5) - 5$. MSE is in the scale of 10^{-3}*

			MEAN	SE	MSE	Lower	Upper	AL	P
$n = 100$	$k = 0.5$	$\hat{\rho}_k^B$	−0.5117	0.1006	22.5682	−0.7046	−0.4486	0.2560	95.3%
		$\hat{\rho}_k^{[2]}$	−0.4344	0.1196	49.9786	−0.6734	−0.3006	0.3727	93.6%
		$\hat{\hat{\rho}}_k$	−0.2669	0.1098	16.3878	−0.5186	−0.0897	0.4288	95.4%
	$k = 1$	$\hat{\rho}_k^B$	−0.5258	0.1096	20.9914	−0.6937	−0.4136	0.2801	93.3%
		$\hat{\rho}_k^{[2]}$	−0.4619	0.1230	40.2815	−0.6589	−0.2910	0.3678	92.7%
		$\hat{\hat{\rho}}_k$	−0.2692	0.1183	15.7163	−0.5081	−0.0807	0.4274	92.7%
	$k = 2$	$\hat{\rho}_k^B$	−0.4977	0.1210	22.8350	−0.6720	−0.3538	0.3181	87.4%
		$\hat{\rho}_k^{[2]}$	−0.4512	0.1261	34.6609	−0.6403	−0.2787	0.3616	85.9%
		$\hat{\hat{\rho}}_k$	−0.2480	0.1243	16.3481	−0.4740	−0.0655	0.4084	89.4%
	$k = 3$	$\hat{\rho}_k^B$	−0.4639	0.1257	24.1410	−0.6442	−0.3130	0.3311	85.2%
		$\hat{\rho}_k^{[2]}$	−0.4281	0.1273	32.3629	−0.6153	−0.2608	0.3544	83.7%
		$\hat{\hat{\rho}}_k$	−0.2297	0.1259	16.6221	−0.6228	−0.2717	0.3511	88.8%
$n = 300$	$k = 0.5$	$\hat{\rho}_k^B$	−0.5896	0.0571	4.3919	−0.6824	−0.5328	0.1495	95.7%
		$\hat{\rho}_k^{[2]}$	−0.5398	0.0680	11.5940	−0.6516	−0.4432	0.2083	95.1%
		$\hat{\hat{\rho}}_k$	−0.3167	0.0587	7.7006	−0.4513	−0.1973	0.2539	97.4%
	$k = 1$	$\hat{\rho}_k^B$	−0.5897	0.0640	5.0508	−0.6803	−0.5170	0.1632	91.7%
		$\hat{\rho}_k^{[2]}$	−0.5470	0.0692	10.1868	−0.6516	−0.4502	0.2014	90.8%
		$\hat{\hat{\rho}}_k$	−0.2296	0.0669	4.6038	−0.4346	−0.1822	0.2524	94.6%
	$k = 2$	$\hat{\rho}_k^B$	−0.5553	0.0735	6.4882	−0.6535	−0.4684	0.1851	85.9%
		$\hat{\rho}_k^{[2]}$	−0.5231	0.0748	9.8226	−0.6278	−0.4261	0.2017	84.6%
		$\hat{\hat{\rho}}_k$	−0.2689	0.0731	5.4241	−0.3963	−0.1590	0.2372	90.4%
	$k = 3$	$\hat{\rho}_k^B$	−0.5202	0.0788	7.4378	−0.6212	−0.4303	0.1909	83.8%
		$\hat{\rho}_k^{[2]}$	−0.4941	0.0786	9.9009	−0.5980	−0.3990	0.1990	83.1%
		$\hat{\hat{\rho}}_k$	−0.2482	0.0758	5.8392	−0.3681	−0.1460	0.2221	87.4%
$n = 500$	$k = 0.5$	$\hat{\rho}_k^B$	−0.6067	0.0445	2.2341	−0.6784	−0.5590	0.1194	95.1%
		$\hat{\rho}_k^{[2]}$	−0.5656	0.0533	6.1683	−0.6518	−0.4940	0.1577	95.2%
		$\hat{\hat{\rho}}_k$	−0.3263	0.0467	2.2229	−0.4293	−0.2296	0.1996	97.5%
	$k = 1$	$\hat{\rho}_k^B$	−0.6051	0.0498	2.7245	−0.6731	−0.5431	0.1299	91.7%
		$\hat{\rho}_k^{[2]}$	−0.5696	0.0533	5.4387	−0.6484	−0.4931	0.1552	91.9%
		$\hat{\hat{\rho}}_k$	−0.3074	0.0515	2.6722	−0.4086	−0.2109	0.1976	94.0%
	$k = 2$	$\hat{\rho}_k^B$	−0.5705	0.0579	3.6740	−0.6457	−0.4997	0.1460	83.8%
		$\hat{\rho}_k^{[2]}$	−0.5434	0.0583	5.4013	−0.6238	−0.4674	0.1564	83.1%
		$\hat{\hat{\rho}}_k$	−0.2752	0.0574	3.3002	−0.3700	−0.1846	0.1854	90.9%
	$k = 3$	$\hat{\rho}_k^B$	−0.5357	0.0628	4.3289	−0.6137	−0.4639	0.1498	82.1%
		$\hat{\rho}_k^{[2]}$	−0.5134	0.0619	5.5886	−0.5939	−0.4392	0.1547	81.0%
		$\hat{\hat{\rho}}_k$	−0.2544	0.0602	3.6343	−0.3437	−0.1700	0.1736	84.9%

centered chi-squared distribution with degree of freedom 5 and mean 4). The multiplicative distortion function is designed as $\psi(U) = 1 - 0.5\cos(2\pi U)$, and U is generated from a uniform distribution $U(0, 1)$.

In Table 4, we investigate the performance of the benchmark estimators (using the true covariate X in the simulation)

$$\widehat{S}_P = \frac{\bar{X} - \text{me}(X)}{[n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2]^{1/2}}, \qquad \widehat{S}_G = \frac{\bar{X} - \text{me}(X)}{n^{-1}\sum_{i=1}^n |X_i - \text{me}(X)|},$$

$$\widehat{S}_M = \frac{n^{-1}\sum_{i=1}^n (X_i - \bar{X})^3}{[n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2]^{3/2}}, \qquad \widehat{S}_{Q,0.25} = \frac{\hat{q}_{0.75} + \hat{q}_{0.25} - 2\,\text{me}(X)}{\hat{q}_{0.75} - \hat{q}_{0.25}},$$

the distorted estimators

$$\widetilde{S}_P = \frac{\overline{\widetilde{X}} - \text{me}(\widetilde{X})}{[n^{-1}\sum_{i=1}^n (\widetilde{X}_i - \overline{\widetilde{X}})^2]^{1/2}}, \qquad \widetilde{S}_G = \frac{\overline{\widetilde{X}} - \text{me}(\widetilde{X})}{n^{-1}\sum_{i=1}^n |\widetilde{X}_i - \text{me}(\widetilde{X})|},$$

$$\widetilde{S}_M = \frac{n^{-1}\sum_{i=1}^n (\widetilde{X}_i - \overline{\widetilde{X}})^3}{[n^{-1}\sum_{i=1}^n (\widetilde{X}_i - \overline{\widetilde{X}})^2]^{3/2}}, \qquad \widetilde{S}_{Q,0.25} = \frac{\tilde{q}_{0.75} + \tilde{q}_{0.25} - 2\,\text{me}(\widetilde{X})}{\tilde{q}_{0.75} - \tilde{q}_{0.25}}$$

and the proposed estimators $\widehat{\widetilde{S}}_P$, $\widehat{\widetilde{S}}_G$, $\widehat{\widetilde{S}}_M$ and $\widehat{\widetilde{S}}_{Q,0.25}$. Here, $\text{me}(X)$ and $\hat{q}_\delta$ are the median and δ-quantile of $\{X_1, \ldots, X_n\}$, and $\text{me}(\widetilde{X})$ and $\tilde{q}_\delta$ are the median and δ-quantile of $\{\widetilde{X}_1, \ldots, \widetilde{X}_n\}$. Note that the true value of $(S_P, S_G, S_M, S_{Q,0.25})$ are $(0, 0, 0, 0)$ for normal distribution $N(-2, 1)$ and $(0.2051, 0.2718, 1.2649, 0.1512)$ for chi-squared distribution $\chi_5^2 - 1$. In Table 4, it is seen that the distorted estimators $\widetilde{S}_P$, $\widetilde{S}_G$, $\widetilde{S}_M$ and $\widetilde{S}_{Q,0.25}$ have non-ignorable bias, which results in large values of MSE compared with benchmark estimators and proposed estimators. It is easily seen that values of MSE for the distorted estimators generally do not decrease as the sample size n increase, which again implies that the distorted estimators produce non-ignorable bias. Meanwhile, the proposed estimators performs as well as the benchmark estimators, and the values of MSE become smaller as the sample size n becomes larger. It is also seen that the estimator $\widehat{\widetilde{S}}_{M,l}$ performs not well compared with the other three estimators, and the values of MSE are much larger especial for the chi-squared distribution.

In Table 5 and Table 6, we report the power functions $\widehat{\widetilde{S}}_{P,l}$, $\widehat{\widetilde{S}}_{G,l}$, $\widehat{\widetilde{S}}_{M,l}$ and $\widehat{\widetilde{S}}^{[l]}_{Q,0.25}$, $l = 1, 2$ based on 2000 realization. Here, we consider the data generating process (DGP) given as $X \sim (1 - b) * N(-0.5, 1) + b * (-\chi_1^2)$, $b = 0.2, \ldots, 1$. It is easily seen that the density of X is asymmetric when $b \neq 0$. The sample size n is chosen as $n = 100$, $n = 300$ and $n = 500$. We find that the power functions of $\hat{\rho}^{[2]}_{0.5}$, $\widehat{\widetilde{S}}_{P,2}$ and $\widehat{\widetilde{S}}_{G,2}$ are generally better than $\hat{\rho}^{[1]}_{0.5}$ $\widehat{\widetilde{S}}_{P,1}$ and $\widehat{\widetilde{S}}_{G,1}$ in this example. As the sample size n increases to 500, the power functions of these five estimators increase to one rapidly. It is seen that $\hat{\rho}^{[l]}_{0.5}$ works the best for detecting the asymmetry of the underlying density function and is the most powerful than other four test statistics, and $\widehat{\widetilde{S}}^{[l]}_{\delta,0.25}$ is the worst when the sample size n is 100 and gets better when the sample size n is 300 and 500.

5 A real data analysis

In this section, we analyze the baseline data collected from studies A and B of the Modification of Diet in Renal Disease (MDRD) Study (Rosman et al. (1984)). There are 827 samples in this dataset. The main goal of the original study was to demonstrate that dietary protein restriction can slow down the decline of the glomerular filtration rate (GFR). Here, we investigate the symmetry of the unobserved baseline glomerular filtration rate (GFR) and

Table 4 *The means (M), standard errors (SE), mean squared errors (MSE) of proposed estimators (P_l) $\widehat{\widehat{S}}_{P,l}$, $\widehat{\widehat{S}}_{G,l}$, $\widehat{\widehat{S}}_{M,l}$, $\widehat{\widehat{S}}^{[l]}_{Q,0.25}$, $l = 1, 2$, and benchmark estimators (B) and distorted estimators (D). MSE is in the scale of* 10^{-3}

			S_P			S_G			S_M			$S_{Q,0.25}$		
	n		M	SE	MSE	M	SE	MSE	M	SE	MSE	M	SE	MSE
$N(-2, 1)$	100	P_1	−0.0185	0.0723	5.5792	−0.0237	0.0920	9.0390	0.0010	0.2898	83.9628	−0.0025	0.1322	17.4985
		P_2	−0.0181	0.0728	5.6391	−0.0231	0.0925	9.0994	0.0001	0.2913	84.8528	−0.0031	0.1329	17.6846
		B	0.0027	0.0727	5.3009	0.0035	0.0917	8.4198	0.0059	0.2302	53.0160	0.0040	0.1347	18.1572
		D	−0.1711	0.0700	34.1925	−0.2207	0.0916	57.1264	−0.6781	0.2517	523.1803	−0.1678	0.1330	45.8475
	300	P_1	−0.0211	0.0428	2.2848	−0.0268	0.0541	3.6532	0.0004	0.1797	32.3091	−0.0036	0.0766	5.8798
		P_2	−0.0209	0.0428	2.2772	−0.0265	0.0541	3.6369	−0.0001	0.1798	32.3255	−0.0033	0.0770	5.9403
		B	0.0025	0.0426	1.8259	0.0032	0.0535	2.8780	0.0024	0.1391	19.3506	0.0042	0.0764	5.8583
		D	−0.1760	0.0397	32.5562	−0.2258	0.0519	53.7080	−0.7122	0.1547	531.2605	−0.1755	0.0746	36.3640
	500	P_1	−0.0219	0.0327	1.5546	−0.0276	0.04213	2.4714	−0.0023	0.1363	18.5970	−0.0043	0.0584	3.4329
		P_2	−0.0218	0.0329	1.5624	−0.0275	0.0415	2.4824	−0.0027	0.1365	18.6381	−0.0042	0.0584	3.4303
		B	0.0011	0.0338	1.1443	0.0014	0.0424	1.8009	−0.0019	0.1056	11.1632	0.0023	0.0597	3.5782
		D	−0.1750	0.0322	31.6685	−0.2241	0.0420	52.0180	−0.7143	0.1142	523.3961	−0.1720	0.0592	33.0892
$\chi^2_5 - 1$	100	P_1	0.1946	0.0719	5.2834	0.2620	0.0945	9.0257	0.9945	0.6087	443.5006	0.1406	0.1323	17.6125
		P_2	0.1954	0.0715	5.2147	0.2629	0.0944	8.9957	1.0009	0.6089	440.3624	0.1405	0.1320	17.5490
		B	0.2017	0.0671	4.5243	0.2692	0.0925	8.5681	1.1492	0.3978	171.6255	0.1455	0.1330	17.7233
		D	0.2584	0.0587	6.2945	0.3638	0.0879	16.1982	1.5643	0.4893	328.9876	0.2040	0.1302	19.7599
	300	P_1	0.2084	0.0411	1.7024	0.2778	0.0559	3.1625	1.1553	0.3980	170.3443	0.1463	0.0778	6.0774
		P_2	0.2086	0.0409	1.6896	0.2780	0.0558	3.1586	1.1578	0.3967	168.8061	0.1462	0.0777	6.0668
		B	0.2026	0.0396	1.5793	0.2692	0.0548	3.0140	1.2203	0.2634	71.3390	0.1467	0.0777	6.0565
		D	0.2638	0.0353	4.7044	0.3709	0.0533	12.6789	1.6768	0.3384	284.1171	0.2174	0.0774	10.3940
	500	P_1	0.2121	0.0306	0.9889	0.2821	0.0419	1.8635	1.2089	0.3465	123.1745	0.1500	0.0605	3.6704
		P_2	0.2121	0.0306	0.9898	0.2821	0.0419	1.8646	1.2101	0.3460	122.6922	0.1499	0.0605	3.6650
		B	0.2036	0.0297	0.8891	0.2703	0.0410	1.6836	1.2394	0.2113	45.3061	0.1486	0.0598	3.5921
		D	0.2639	0.0268	4.1858	0.3711	0.0402	11.4966	1.7155	0.2981	291.9543	0.2158	0.0609	7.9007

Table 5 *The simulation results for power calculations in Example 2 for $l = 1$*

b	$\hat{\rho}^{[1]}_{0.5}$	$\widehat{\widehat{S}}_{P,1}$	$\widehat{\widehat{S}}_{G,1}$	$\widehat{\widehat{S}}_{M,1}$	$\widehat{\widehat{S}}^{[1]}_{Q,0.25}$
			$n = 100$		
$b = 0.2$	0.0585	0.0090	0.0090	0.0345	0.0570
$b = 0.4$	0.3015	0.0640	0.0640	0.1565	0.0905
$b = 0.6$	0.9430	0.5625	0.5625	0.6855	0.1845
$b = 0.8$	1.0000	0.9355	0.9355	0.8075	0.4700
$b = 1.0$	1.0000	0.9540	0.9540	0.8230	0.5510
			$n = 300$		
$b = 0.2$	0.1170	0.0100	0.0100	0.0365	0.0665
$b = 0.4$	0.7575	0.3050	0.3050	0.7130	0.0980
$b = 0.6$	0.9990	0.9875	0.9875	0.9580	0.3730
$b = 0.8$	1.0000	0.9935	0.9935	0.9725	0.8390
$b = 1.0$	1.0000	0.9940	0.9940	0.9740	0.8815
			$n = 500$		
$b = 0.2$	0.1315	0.0140	0.0140	0.0540	0.0710
$b = 0.4$	0.9390	0.5690	0.5690	0.9305	0.1150
$b = 0.6$	1.0000	0.9990	0.9990	0.9905	0.5765
$b = 0.8$	1.0000	1.0000	1.0000	0.9910	0.9420
$b = 1.0$	1.0000	1.0000	1.0000	0.9920	0.9700

Table 6 *The simulation results for power calculations in Example 2 for $l = 2$*

b	$\hat{\rho}^{[2]}_{0.5}$	$\widehat{\widehat{S}}_{P,2}$	$\widehat{\widehat{S}}_{G,2}$	$\widehat{\widehat{S}}_{M,2}$	$\widehat{\widehat{S}}^{[2]}_{Q,0.25}$
			$n = 100$		
$b = 0.2$	0.0775	0.0620	0.0620	0.0245	0.0645
$b = 0.4$	0.3165	0.1570	0.1570	0.1565	0.0825
$b = 0.6$	0.9515	0.7455	0.7455	0.6855	0.1820
$b = 0.8$	1.0000	0.9985	0.9985	0.8075	0.4540
$b = 1.0$	1.0000	1.0000	1.0000	0.8230	0.5510
			$n = 300$		
$b = 0.2$	0.1105	0.0665	0.0665	0.0365	0.0665
$b = 0.4$	0.7705	0.5165	0.5165	0.7130	0.0920
$b = 0.6$	0.9995	0.9985	0.9985	0.9580	0.3645
$b = 0.8$	1.0000	1.0000	1.0000	0.9725	0.8275
$b = 1.0$	1.0000	1.0000	1.0000	0.9740	0.8815
			$n = 500$		
$b = 0.2$	0.1400	0.0770	0.0775	0.0540	0.0690
$b = 0.4$	0.9325	0.7770	0.7770	0.9305	0.1020
$b = 0.6$	1.0000	1.0000	1.0000	0.9905	0.5600
$b = 0.8$	1.0000	1.0000	1.0000	0.9910	0.9425
$b = 1.0$	1.0000	1.0000	1.0000	0.9920	0.9700

unobserved serum creatinine (SCr) data as an illustration of our method. Assume that the distorted GFR is $\widetilde{X}_1$ and the distorted SCr is $\widetilde{X}_2$. Suggested by Cui et al. (2009), the confounding variable U for this data is taken to be the body surface area (BSA), which is defined as $\mathrm{BSA}(m^2) = 0.007184 * \mathrm{Kg}^{0.425} * \mathrm{Cm}^{0.725}$.

We first present the patterns of $\hat{\psi}_{1,\mathrm{GFR}}(u)$ and $\hat{\psi}_{1,\mathrm{SCr}}(u)$ by using (2.2) in Figure 1 under the confounding variable-BSA. Figure 1 indicates that underlying distorting functions $\psi_{\mathrm{GFR}}(u)$ and $\psi_{\mathrm{SCr}}(u)$ are not a constant function, suggesting that the confounding variable-

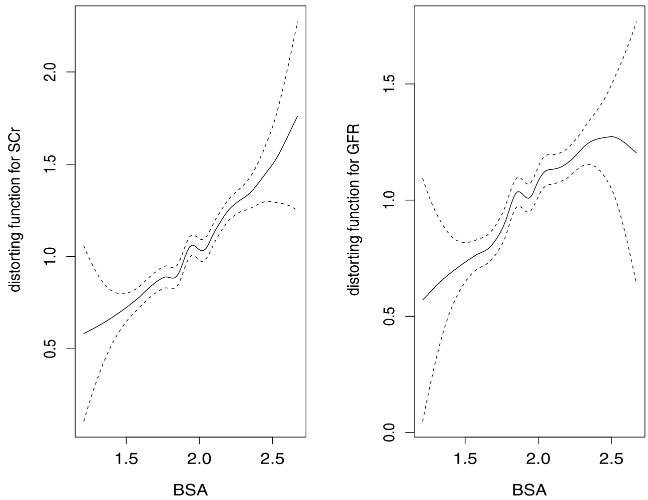

Figure 1 *The estimated curve of distorting functions* $\psi_{\mathrm{SCr}}(u)$ *and* $\psi_{\mathrm{GFR}}(u)$, *associated* 95% *pointwise confidence intervals* (*dotted lines*).

Table 7 *The estimation and p-values of estimators* $\widehat{\widetilde{S}}_{P,1}$, $\widehat{\widetilde{S}}_{G,1}$, $\widehat{\widetilde{S}}_{M,1}$ *and* $\widehat{\widetilde{S}}^{[1]}_{Q,0.25}$

	$\widehat{\widetilde{S}}_{P,1}$		$\widehat{\widetilde{S}}_{G,1}$		$\widehat{\widetilde{S}}_{M,1}$		$\widehat{\widetilde{S}}^{[1]}_{Q,0.25}$	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
SCr	0.0719	0.4319	0.0856	0.4319	0.1619	0.0434	0.0827	0.0860
GFR	0.2733	0.0002	0.3618	0.0002	1.1476	2.0812×10^{-8}	0.2733	4.0810×10^{-7}

BSA definitely makes effect of GFR and SCr in this data. From the original data and the estimated curves of $\hat{\psi}_{1,\mathrm{GFR}}(u)$ and $\hat{\psi}_{1,\mathrm{SCr}}(u)$ in Figure 1, there is no negative values of the distorted GFR (i.e., $|\widetilde{X}_1| = \widetilde{X}_1$) and the distorted SCr ($|\widetilde{X}_2| = \widetilde{X}_2$), so the estimators $(\hat{\psi}_{1,\mathrm{GFR}}(u), \hat{\psi}_{1,\mathrm{SCr}}(u))$ and $(\hat{\psi}_{2,\mathrm{GFR}}(u), \hat{\psi}_{2,\mathrm{SCr}}(u))$ (by using (2.3) with a common bandwidth) satisfy $\hat{\psi}_{1,\mathrm{GFR}}(u) = \hat{\psi}_{2,\mathrm{GFR}}(u)$ and $\hat{\psi}_{1,\mathrm{SCr}}(u) = \hat{\psi}_{2,\mathrm{SCr}}(u)$ for this dataset. Figure 1 implies the values of $\hat{\psi}_{1,\mathrm{GFR}}(u)$ and $\hat{\psi}_{1,\mathrm{SCr}}(u)$ should be positive. Because all the values of $(\widetilde{X}_1, \widetilde{X}_2)$, $\hat{X}_1^{[1]}$, $\hat{X}_2^{[1]}$ and $\hat{\psi}_{1,\mathrm{GFR}}(u)$, $\hat{\psi}_{1,\mathrm{SCr}}(u)$ are all positive in this dataset, so we have $\{\hat{X}^{[1]}_{s,i} = \hat{X}^{[2]}_{s,i}\}_{i=1}^{827}$, $s = 1, 2$.

We now use $\hat{\rho}^{[1]}_{0.5}$, $\widehat{\widetilde{S}}_{P,1}$, $\widehat{\widetilde{S}}_{G,1}$, $\widehat{\widetilde{S}}_{M,1}$ and $\widehat{\widetilde{S}}^{[1]}_{Q,0.25}$ to investigate the symmetry of the unobserved GFR and SCr. The 95% confidence intervals of $\hat{\rho}^{[1]}_{0.5}$ are (−0.4534, −0.3034) for SCr, and (−0.7583, −0.6583) for GFR. Both two intervals exclude zero and indicate that SCr and GFR are asymmetric. We also present the values of estimators $\widehat{\widetilde{S}}_{P,1}$, $\widehat{\widetilde{S}}_{G,1}$, $\widehat{\widetilde{S}}_{M,1}$ and $\widehat{\widetilde{S}}^{[1]}_{Q,0.25}$ and associated p-values in Table 7. The plots of the histogram and density function estimate of original variables $\widetilde{X}_s$ and estimated variables $\hat{X}^{[1]}_s$, $s = 1, 2$ are presented in Figure 2. The p-values for GFR and figure in Figure 2 showed that the unobserved GFR is asymmetric. While, for SCr, the distorted $\widetilde{X}_2$ in Figure 2 implies asymmetry, but the estimated $\hat{X}^{[1]}_2$ shows a slightly symmetric and also investigated by statistics $\widehat{\widetilde{S}}_{P,1}$, $\widehat{\widetilde{S}}_{G,1}$ and $\widehat{\widetilde{S}}^{[1]}_{Q,0.25}$. The statistic $\widehat{\widetilde{S}}_{M,1}$ and $\hat{\rho}^{[1]}_{0.5}$ show that SCr should be asymmetric. Together with Figure 1 and the performances of statistic $\widehat{\widetilde{S}}_{M,1}$ and $\hat{\rho}^{[1]}_{0.5}$ presented in Table 7, we prefer to the conclusion that SCr is asymmetric for this dataset. In Figure 1, when the value of BSA is less than two, the values

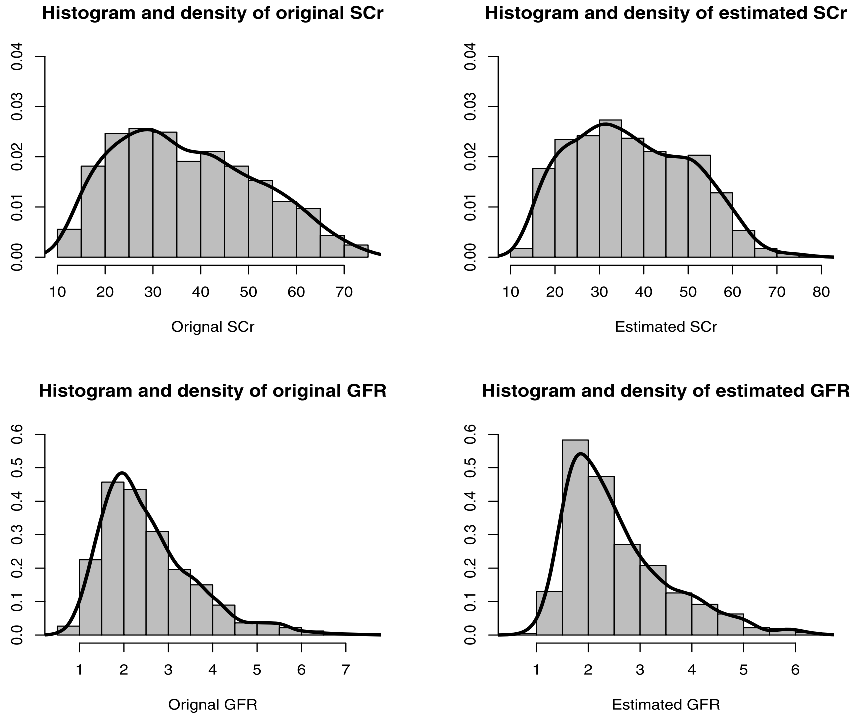

Figure 2 *The histograms and density curve estimates for estimated variable $\hat{X}$.*

of distorting functions $\hat{\psi}_{1,\text{GFR}}(u)$ and $\hat{\psi}_{1,\text{SCr}}(u)$ are less than one. This makes the observed values GFR ($\widetilde{X}_1$) and SCr ($\widetilde{X}_2$) have smaller values. Similarly, when the value of BSA is larger than two, the distorting functions $\hat{\psi}_{1,\text{GFR}}(u)$ and $\hat{\psi}_{1,\text{SCr}}(u)$ make the observed values GFR ($\widetilde{X}_1$) and SCr ($\widetilde{X}_2$) have larger values. Together with Figure 2, the distortion functions $\hat{\psi}_{1,\text{GFR}}(u)$ and $\hat{\psi}_{1,\text{SCr}}(u)$ make the unobserved GFR and SCr more dispersive. Figure 2 implies the Rayleigh distribution may be fit for the estimated GFR ($\hat{X}_1^{[1]}$) and the estimated SCr ($\hat{X}_2^{[1]}$) for employing in the field of Nutrition for linking dietary nutrient levels and human responses, moreover, the parameter in Rayleigh distribution may be used to calculate nutrient response relationship with estimated GFR and estimated SCr.

6 Discussions and further research

In this article, we start research of how to estimate and test the symmetry of a continuous variable under the multiplicative distortion measurement errors setting, and the associated asymptotic results are also investigated. Due to the importance of the symmetry in statistics, there are huge amount of papers on how to measure and test the symmetry; hence it is impossible for us to transform all the existing methods of the multiplicative distortion measurement errors setting. Instead of attempting to cover as many papers as we could, we intend to study relatively important methods in statistics literature on the hypothesis testing of the symmetry. Testing the symmetry for the model error under the multiplicative distortion measurement errors setting, such as parametric regression models and semi-parametric regression models, can be considered in the future work. The research is ongoing.

Acknowledgments

The authors thank the Editor, the associate editor and the referee for their constructive suggestions that helped them to improve the early manuscript. Xia Cui is a College Talent Cultivated by "Thousand-Hundred-Ten" Program of Guangdong Province and her research was supported by the National Natural Science Foundation of China (Grant No. 11471086), the Humans and Social Science Research Team of Guangzhou University (Grant No. 201503XSTD) and the Training Program for Excellent Young College Teachers of Guangdong Province (Grant No. Yq201404). Gaorong Li's research is supported by the National Natural Science Foundation of China (Grant Nos. 11871001 and 11471029) and Beijing Natural Science Foundation (Grant No. 1182003).

Supplementary Material

Supplement to "Measuring symmetry and asymmetry of multiplicative distortion measurement errors data" (DOI: 10.1214/19-BJPS432SUPP; .pdf). Supplementary materials: Complete proof of Theorems in this paper.

References

Butler, C. C. (1969). A test for symmetry using the sample distribution function. *The Annals of Mathematical Statistics* **40**, 2209–2210.

Cabilio, J. and Cabilio, P. (1996). A simple test of symmetry about an unknown median. *Canadian Journal of Statistics* **24**, 349–361.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Nonlinear Measurement Error Models: A Modern Perspective*, 2nd ed. New York: Chapman and Hall. MR2243417 https://doi.org/10.1201/9781420010138

Chen, K., Guo, S., Lin, Y. and Ying, Z. (2010). Least absolute relative error estimation. *Journal of the American Statistical Association* **105**, 1104–1112. MR2752606 https://doi.org/10.1198/jasa.2010.tm09307

Chen, K., Lin, Y., Wang, Z. and Ying, Z. (2016). Least product relative error estimation. *Journal of Multivariate Analysis* **144**, 91–98. MR3434942 https://doi.org/10.1016/j.jmva.2015.10.017

Csörgő, S. and Heathcote, C. R. (1987). Testing for symmetry. *Biometrika* **74**, 177–184. MR0885930 https://doi.org/10.2307/2336032

Cui, X., Guo, W., Lin, L. and Zhu, L. (2009). Covariate-adjusted nonlinear regression. *The Annals of Statistics* **37**, 1839–1870. MR2533473 https://doi.org/10.1214/08-AOS627

Davis, C. E. and Quade, D. (1978). U-statistics for skewness or symmetry. *Communication in Statistics: Theory and Methods* **7**, 413–418.

Delaigle, A., Hall, P. and Zhou, W.-X. (2016). Nonparametric covariate-adjusted regression. *The Annals of Statistics* **44**, 2190–2220. MR3546448 https://doi.org/10.1214/16-AOS1442

Fuller, W. A. (1987). *Measurement Error Models*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: John Wiley & Sons Inc. MR0898653 https://doi.org/10.1002/9780470316665

Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)* **33**, 391–399.

Gupta, M. K. (1967). An asymptotically nonparametric test of symmetry. *The Annals of Mathematical Statistics* **38**, 849–866. MR0208761 https://doi.org/10.1214/aoms/1177698879

Hill, D. L. and Rao, P. V. (1977). Tests of symmetry based on Cramér–von Mises statistics. *Biometrika* **64**, 489–494. MR0518893 https://doi.org/10.1093/biomet/64.3.489

Hotelling, H. and Solomons, L. M. (1932). The limits of a measure of skewness. *The Annals of Mathematical Statistics* **3**, 141–142.

Hwang, J. T. (1986). Multiplicative errors-in-variables models with applications to recent data released by the U.S. Department of Energy. *Journal of the American Statistical Association* **81**, 680–688. MR0860499

Kiwitt, S. and Neumeyer, N. (2012). Estimating the conditional error distribution in non-parametric regression. *Scandinavian Journal of Statistics* **39**, 259–281. MR2927025 https://doi.org/10.1111/j.1467-9469.2011.00763.x

Kraft, C. H. and van Eeden, C. (1972). Linearized rank estimates and signed-rank estimates for the general linear hypothesis. *The Annals of Mathematical Statistics* **43**, 42–57. MR0301857 https://doi.org/10.1214/aoms/1177692699

Li, F., Lin, L. and Cui, X. (2010). Covariate-adjusted partially linear regression models. *Communications in Statistics: Theory and Methods* **39**, 1054–1074. MR2745361 https://doi.org/10.1080/03610920902846539

Li, F., Lin, L., Lu, Y. and Feng, S. (2018). An adaptive estimation for covariate-adjusted nonparametric regression model. *Statistical Papers*. Available at https://doi.org/10.1007/s00362-019-01084-0. https://doi.org/10.1007/s00362-019-01084-0

Li, F. and Lu, Y. (2018). Lasso-type estimation for covariate-adjusted linear model. *Journal of Applied Statistics* **45**, 26–42. MR3736856 https://doi.org/10.1080/02664763.2016.1267121

Li, G., Zhang, J. and Feng, S. (2016). *Modern Measurement Error Models*. Beijing: Science Press.

Liang, H., Härdle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **27**, 1519–1535. MR1742498 https://doi.org/10.1214/aos/1017939140

Liang, H. and Ren, H. (2005). Generalized partially linear measurement error models. *Journal of Computational and Graphical Statistics* **14**, 237–250. MR2137900 https://doi.org/10.1198/106186005X37481

Liu, H. and Xia, X. (2018). Estimation and empirical likelihood for single-index multiplicative models. *Journal of Statistical Planning and Inference* **193**, 70–88. MR3713464 https://doi.org/10.1016/j.jspi.2017.08.003

Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics* **19**, 1725–1747. MR1135146 https://doi.org/10.1214/aos/1176348368

Partlett, C. and Patil, P. (2017). Measuring asymmetry and testing symmetry. *Annals of the Institute of Statistical Mathematics* **69**, 429–460. MR3611527 https://doi.org/10.1007/s10463-015-0547-4

Patil, P. N., Patil, P. P. and Bagkavos, D. (2012). A measure of asymmetry. *Statistical Papers* **53**, 971–985. MR2992928 https://doi.org/10.1007/s00362-011-0401-6

Rosman, J. B., ter Wee, P. M., Meijer, S., Piers-Becht, T. P. M., Sluiter, W. J. and Donker, A. J. M. (1984). Prospective randomized trial of early protein restriction in chronic renal failure. *The Lancet* **2**, 1291–1296.

Şentürk, D. and Müller, H.-G. (2005). Covariate adjusted correlation analysis via varying coefficient models. *Scandinavian Journal of Statistics: Theory and Applications* **32**, 365–383. MR2204625 https://doi.org/10.1111/j.1467-9469.2005.00450.x

Şentürk, D. and Müller, H.-G. (2006). Inference for covariate adjusted regression via varying coefficient models. *The Annals of Statistics* **34**, 654–679. MR2281880 https://doi.org/10.1214/009053606000000083

Zhang, J., Gai, J., Cui, X. and Li, G. (2019). Supplement to "Measuring symmetry and asymmetry of multiplicative distortion measurement errors data." https://doi.org/10.1214/19-BJPS432SUPP.

Zhao, J. and Xie, C. (2018). A nonparametric test for covariate-adjusted models. *Statistics & Probability Letters* **133**, 65–70. MR3732354 https://doi.org/10.1016/j.spl.2017.10.004

J. Zhang
College of Mathematics and Statistics
Institute of Statistical Sciences
Shenzhen University
518060, Shenzhen
China

Y. Gai
School of Statistics and Mathematics
Central University of Finance and Economics
100081, Beijing
China
E-mail: yujie_gai@126.com

X. Cui
School of Economics and Statistics
Guangzhou University
510006, Guangzhou
China

G. Li
School of Statistics
Beijing Normal University
100875, Beijing
China