Measuring Symmetry and Asymmetry of Multiplicative Distortion Measurement Errors Data

Jun Zhang\textsuperscript{a}, Yujie Gai\textsuperscript{b}, Xia Cui\textsuperscript{c} and Gaorong Li\textsuperscript{d}

\textsuperscript{a}Shenzhen University
\textsuperscript{b}Central University of Finance and Economics
\textsuperscript{c}Guangzhou University
\textsuperscript{d}Beijing University of Technology

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. Firstly, 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 biasing 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

\textit{MSC 2010 subject classifications:} Primary 62G05; 62G08; 62G20

\textit{Keywords and phrases.} Confounding variable, Errors-in-variables, Correlation coefficient, Empirical likelihood, Symmetry.
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 et al. (1999); Li et al. (2016).

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

$$\tilde{X} = X\psi(U),$$

where, $X$ is the unobservable continuous variable of interest, $\tilde{X}$ 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 $\psi(U)$ 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 et al. (2010) considered partial linear models, where the linear part is observed with multiplicative distortion measurement errors, and Li and Lu (2017) considered to use lasso-type penalty functions including lasso and adaptive lasso for simultaneous variable selection and parameter estimation. Delaigle et al. (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.
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 \( \gamma \) 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 \( \sup_{x \leq 0} |F(\gamma + x) + F(\gamma - x) - 1| \), Davis and Quade (1978) proposed \( U \)-statistics by using the triple sample for testing the symmetry. Recently, Patil et al. (2012) proposed measuring the symmetry and asymmetry by using the correlation coefficient between the density function and the distribution function, namely, \( \rho(f,F) = \frac{\text{Cov}(f(X),F(X))}{\sqrt{\sigma^2_f \sigma^2_F}} \) when \( 0 < \sigma^2_f < \infty \), where \( \sigma^2_f = \text{Var}(f(X)) \) and \( \sigma^2_F = \text{Var}(F(X))(=\frac{1}{12}) \). Patil et al. (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); Csörgö 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 \( k \)-th correlation coefficient \( \rho_k = \frac{\text{Cov}(f^k(X),F(X))}{\sqrt{\sigma^2_{f^k} \sigma^2_F}} \) with \( \sigma^2_{f^k} = \text{Var}(f^k(X)) \) for some \( k > 0 \) as a measure of symmetry or asymmetry. Note that \( \rho_1 \) is the correlation coefficient \( \rho(f,F) \) proposed by Patil et al. (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 et al. (2016), and further use calibrated variable \( \hat{X} \) to estimate \( \rho_k \). Next, we construct empirical likelihood based confidence intervals of \( \rho_k \), 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 $\tilde{X}$ 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 $X$, 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.

2 Direct plug-in estimation procedure

2.1 General setting

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

$$E[\psi(U)] = 1. \quad (2.1)$$

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 $u \in [U_L, U_R]$, $U_L < U_R$, where $[U_L, U_R]$ denotes the compact support of $U$.

Assumption M2: the unknown smoothing distorting function $\psi(u)$ satisfies $\psi(u) > 0$ for all $u \in [U_L, U_R]$.

By using two assumptions M1 and M2, we have two different estimation procedures of $\rho_k$. 

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

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

(2.2)

where $S_{n\omega}(u) = \frac{1}{nh_1}\sum_{i=1}^{n} (\frac{U_i - u}{h_1})^\omega K\left(\frac{U_i - u}{h_1}\right)$ for $\omega = 0, 1, 2$, and $Q_{n\delta,\bar{X}}(u) = \frac{1}{nh_1}\sum_{i=1}^{n} \left(\frac{U_i - u}{h_1}\right)^\delta K\left(\frac{U_i - u}{h_1}\right) \bar{X}_i$ for $\delta = 0, 1$, and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i$. 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 $E(|\bar{X}|) = E(|X|)$ and $\psi(u) = E(|\bar{X}||U = u)/E(|\bar{X}|)$. Then, the local linear estimator $\hat{\psi}_2(u)$ is used to estimate $\psi(u)$ by

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

(2.3)

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

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

$$
\begin{align*}
\hat{f}_1(x) &= \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{\hat{X}_i^{[1]} - x}{h_2}\right), \\
\hat{F}_1(x) &= \frac{1}{n}\sum_{i=1}^{n} I \left\{\hat{X}_i^{[1]} \leq x\right\},
\end{align*}
$$

(2.4)

$$
\begin{align*}
\hat{f}_2(x) &= \frac{1}{nh_3}\sum_{i=1}^{n} K\left(\frac{\hat{X}_i^{[2]} - x}{h_3}\right), \\
\hat{F}_2(x) &= \frac{1}{n}\sum_{i=1}^{n} I \left\{\hat{X}_i^{[2]} \leq x\right\},
\end{align*}
$$

(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 $\rho_k$ are proposed as, for $s = 1, 2$,

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

(A2) The distorting function \( \psi(u) \) has three continuous derivatives on \([U_L, U_R]\).

(A3) The density functions \( f(x) \) of \( X \) and \( \tilde{f}(\tilde{x}) \) of \( \tilde{X} \) 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, \( \int_{-C}^{C} t^2 K(t)dt \neq 0 \) and \( \int_{-C}^{C} |t| K(t)dt < \infty \) for \( j = 1, 2, 3 \).

(A5) As \( n \to \infty, n h^4_s \to 0 \), \( \log n n h^4_s \to 0 \) 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 \( h_s, s = 1, 2, 3, 4 \) are used in Section 2.1 and bandwidth \( h_4 \) is used in Section 2.2.

In the following, we define \( \widehat{f}^k F_s = \frac{1}{n} \sum^n_{i=1} \hat{f}^k_s (\hat{X}^k_s) \hat{F}_s (\hat{X}^k_s) \), \( \tilde{f}^k_s = \frac{1}{n} \sum^n_{i=1} \tilde{f}^k (\tilde{X}^k) \) and \( \hat{\sigma}_{s, f^k} = \left\{ \frac{1}{n} \sum^n_{i=1} \left[ \hat{f}^k_s (\hat{X}^k_s) - \hat{f}^k \right] \right\}^{1/2} \), \( E(f^k F) = E[f^k(X)F(X)], E(f^k) = E[f^k(X)] \). Moreover, let \( \mu_{f^k} = \int f^k(x)dx \) and

\[
\delta_k(t, f, F) = \begin{pmatrix} \int f^{k+1}(x)dx + (k+1) f^k(t) \int t f^{k+1}(x)dx \int f^{2k}(t) + \frac{(k+1) \mu_{f^{k+1}}}{\sigma_{f^k}} f^{k}(t) \end{pmatrix}, \eta_k = \begin{pmatrix} E(f^k F) \\ E(f^k) \end{pmatrix}.
\]

**Theorem 2.1.** Assume that conditions (A1)-(A5) hold,
Remark 1. The term $k^2 \eta_k^T \sigma_k \text{Var}(\psi(U)) \left( \frac{I_{s=1}}{|E(X)|^2} + \frac{I_{s=2}}{|E(|X|)|^2} \right)$ 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 $\text{Var}(\psi(U)) = 0$ i.e., $\psi(u) \equiv 1$. When $\text{Var}(\psi(U)) > 0$, the increment of the asymptotic variance of $\hat{\rho}_k^{[s]}$ is caused by using the calibrated variables $\{X_i^{[s]} \mid i = 1, \ldots, n\}$ instead of unobserved $\{X_i \mid i = 1, \ldots, n\}$, i.e., the effect of distorting function $\psi(U)$ exists.

Remark 2. It is known that $|E(X)| \leq E(|X|)$, and $\frac{1}{|E(X)|^2} \geq \frac{1}{E(|X|)^2}$ 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, i.e., $P(X < 0) = 0$, 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 $\bar{X} = O_P(n^{-1/2})$ 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 $\rho_k$. 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,
\documentclass{imsart-bjps}

\begin{document}

\section*{Theoretical Results}

\begin{theorem}
\label{thm:asymptotic_variance}
Assume that conditions (A1)-(A5) hold, \( \hat{\rho}_n^{[s]}(\rho_k) \) converges to a centered chi-squared distribution with degree of freedom one.
\end{theorem}

From Theorem \ref{thm:asymptotic_variance}, an EL confidence interval for \( \rho_k \) is constructed as

\[ I_{\rho_k} = \left\{ \hat{\rho}_k^{[s]} : \hat{\ell}_n^{[s]}(\rho_k) \leq c_\kappa \right\}, \]

where \( c_\kappa \) denotes the \( \kappa \) quantile of the chi-squared distribution with degree of freedom one.

\end{document}
2.2 A special setting

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

$$\hat{F}(x) = EI\{\tilde{X} \leq x\} = E[E\{X \leq x/\psi(U)\}|U] = E[F(x/\psi(U))|U]$$

$$= E[1 - F(2\gamma - x/\psi(U))|U] = E[1 - EI\{X \leq 2\gamma - x/\psi(U)\}|U]$$

$$= E\left[1 - EI\{\tilde{X} \leq 2\gamma\psi(U) - x\}|U]\right]. \quad (2.6)$$

From (2.6), if $X$ is symmetric about zero ($\gamma = 0$), then, $\tilde{F}(x) = 1 - \hat{F}(-x)$. This implies that $\tilde{X}$ 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 $\tilde{X}$ about zero. Similar to (2.4) and (2.5), we propose the estimators of the density function of $\tilde{X}$, $\hat{f}(\tilde{x})$ and distribution function of $\tilde{X}$, $\tilde{F}(\tilde{x})$ as

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

where $h_4$ is a positive-valued bandwidth.

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

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

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{f}^k(\tilde{X}_i) - \frac{1}{n} \sum_{i=1}^{n} \hat{f}^k(\tilde{X}_i) \right]^2 \right\}^{1/2},$$

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

$$\delta_k(t, \hat{f}, \hat{F}) = \begin{pmatrix}
(k + 1)\hat{f}^k(t)\hat{F}(t) + \int_t^\infty \hat{f}^{k+1}(\tilde{x})d\tilde{x} \\
(k + 1)\hat{f}^k(t) + \frac{2k+1}{2\sigma_{\hat{f}^k}}\hat{f}^k(t) + \frac{(k+1)\mu_{\hat{f}^k+1}}{\sigma_{\hat{f}^k}}\hat{f}^k(t)
\end{pmatrix}. $$
Theorem 2.3. Assume that conditions (A1)-(A5) hold, (a) let
\[ \tilde{\Sigma}_k = \text{Cov}(\delta_k(\tilde{X}, \tilde{f}, \tilde{F})) , \]
we have
\[ \frac{\sqrt{n} }{ L - \to N(0, \tilde{\Sigma}_k) } . \]
(b) \[ \sqrt{n} ( \hat{\rho}_k - \tilde{\rho}_k ) \overset{L}{\to} N(0, \tilde{\theta}_k^T \tilde{\Sigma}_k \tilde{\theta}_k) \], where
\[ \tilde{\theta}_k = \frac{1}{\hat{\sigma}_k} \tilde{f}_k \]
\[ \frac{1}{\sqrt{12}} \left( \hat{\sigma}_k \right) = \frac{1}{\sqrt{12}} \left( \hat{\sigma}_k \right) . \]
The correlation coefficient \( \hat{\rho}_k \) can be estimated through the estimating
equation in the population level:
\[ E \left\{ \sqrt{12} \left( \frac{\hat{f}_k(\tilde{X}) - \frac{1}{2}}{\hat{f}_k(\tilde{X})} \right) - \tilde{\rho}_k \left( \frac{\hat{f}_k(\tilde{X}) - E[\tilde{f}_k(\tilde{X})]}{\hat{f}_k(\tilde{X})} \right) \right\} = 0. \]
The empirical log-likelihood ratio function is defined as
\[ \hat{\ell}_n(\tilde{\rho}_k) = -2 \max \left\{ \sum_{i=1}^{n} \log(n\hat{\rho}_i) : \hat{\rho}_i \geq 0, \sum_{i=1}^{n} \hat{\rho}_i = 1, \sum_{i=1}^{n} \hat{\rho}_i \hat{\Theta}_{n,i}(\tilde{\rho}_k) = 0 \right\} , \]
where, for \( i = 1, \ldots, n \),
\[ \hat{\Theta}_{n,i}(\tilde{\rho}_k) = \frac{1}{\sqrt{12}} \left( \frac{\hat{f}_k(\tilde{X}_i) - \frac{1}{2}}{\hat{f}_k(\tilde{X}_i)} \right) - \tilde{\rho}_k \left( \frac{\hat{f}_k(\tilde{X}_i) - E[\tilde{f}_k(\tilde{X}_i)]}{\hat{f}_k(\tilde{X}_i)} \right) \]
\[ \times \left( \frac{\hat{f}_k(\tilde{X}_i) - \hat{f}_k}{\hat{f}_k(\tilde{X}_i)} \right) . \]
Using Lagrange multiplier, we have \( \hat{\ell}_n(\tilde{\rho}_k) = 2 \sum_{i=1}^{n} \log \left\{ 1 + \tilde{\lambda} \hat{\Theta}_{n,i}(\tilde{\rho}_k) \right\} \),
where \( \tilde{\lambda} \) is determined by the equation
\[ \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{n,i}(\tilde{\rho}_k) = 0. \]
Theorem 2.4. Assume that conditions (A1)-(A5) hold, \( \hat{\ell}_n(\tilde{\rho}_k) \) 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
\[
\tilde{I}_{\tilde{\rho}_k} = \left\{ \tilde{\rho}_k' : \tilde{L}(\tilde{\rho}_k') \leq c_k \right\},
\]
where \( c_k \) denotes the \( k \) quantile of the chi-squared distribution with degree of freedom one. From (2.6), the symmetry of \( \tilde{X} \) and \( X \) about \( \gamma \) is equivalent when \( \psi(U) \equiv 1 \). We can also conduct a hypothesis \( H_0 : \psi(u) = 1, u \in [U_L, U_R] \) and \( H_1 : \psi(u) \neq 1 \), for some \( u \in [U_L, U_R] \) at first. If the hypothesis \( H_0 \) holds, one can directly use the distorted variable \( \tilde{X} \), estimator \( \tilde{\rho}_k' \) and Theorem (2.4) to conclude whether \( \tilde{X} \) is symmetric or not. If \( \tilde{X} \) is symmetric, then \( X \) is symmetric because \( \psi(U) \equiv 1 \). Otherwise, if \( \tilde{X} \) is asymmetric, the symmetry of \( X \) can be determined through the proposed estimators \( \tilde{\rho}_k[l] \), \( l = 1, 2 \), 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 \( H_0 \). One can use one of them to test \( H_0 \). 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

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

where \( \vartheta \) stands for the median, \( \sigma^2 \) stands for the variance and \( q_\delta \) stands for the \( \delta \)-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 \( \{(\tilde{X}_i, U_i), i = 1, \ldots, n\} \) 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 \( \{X_1, X_2, \ldots, X_n\} \) are observable, Cabilio and Cabilio (1996) considered to use \( \tilde{S}_P = \tilde{X} - \text{med}(X) \) for testing
symmetry versus skewness, where $\bar{X}$ and $\text{me}(X)$ 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

$$\hat{S}_{pl} = \frac{\bar{X} - \text{me}(\hat{X}^l)}{s^l}, \quad l = 1, 2,$$

where, $\text{me}(\hat{X}^l)$ is the median of calibrated variables $\{\hat{X}_1^l, \ldots, \hat{X}_n^l\}$, $s^l$ is the estimator of $\sigma^2$, and $s^l_l$ is defined as $s^l_l = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{X}_i - E(\hat{X}_i))^2}{\psi^2(U_i)}$, $l = 1, 2$.

To estimate $S_p$, we have

$$\sqrt{n} \left( \hat{S}_{pl} - S_p \right) \xrightarrow{L} N(0, \sigma^2_{S_{p,1}}),$$

where

$$\sigma^2_{S_{p,1}} = \frac{1}{\sigma^2} \text{Var} \left( \frac{\vartheta(2\bar{X} - X)}{E(X)} \right) = \frac{1}{\sigma^2} \left( \frac{E(X) - \vartheta}{2\sigma^2} [X - E(X)]^2 \right),$$

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

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 $|\bar{X}| = |X|\psi(U) = X\psi(U) = \bar{X}$. 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 $|\bar{X}| = |X|\psi(U) = -X\psi(U) = -\bar{X}$. 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 $\vartheta$, it is easily seen that
Measuring Symmetry and Asymmetry

E(X) = \vartheta. Then, the asymptotic variances \( \sigma^2_{S_{P,1}}, \sigma^2_{S_{P,2}} \) reduce to

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

\[
\sigma^2_{S_{P,2}} = \frac{\text{Var}(\bar{X})}{\sigma^2} + \frac{\vartheta^2E(\bar{X}^2)}{E|X|^2\sigma^2} \frac{\text{Var}(\psi(U))}{\text{Var}(\psi(U))} + 1
\]

\[
+ \frac{2\vartheta E(\bar{X}^2)}{E|X|^2\sigma^2} \frac{\text{Var}(\psi(U))}{\text{Var}(\psi(U))} + \frac{E|X - \vartheta|}{\sigma^2f(\vartheta)}.
\]

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

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

\[
\hat{\sigma}^2_{S_{P,1}} = \frac{4\bar{X}^2}{s_1^2} \frac{\text{Var}_1(\psi(U))}{\text{Var}_1(\psi(U))} + \frac{1}{4s_1^2\hat{f}_2(\text{me}(\bar{X}^2))} + \frac{\hat{E}_1(|X - \vartheta|)}{s_1^2\hat{f}_1(\text{me}(\bar{X}^2))},
\]

where \( \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n \hat{X}_i^2, \hat{f}_1(\text{me}(\bar{X}^2)) \) is defined in (2.4) by substituting \( x \)
with \( \text{me}(\hat{X}^2) \), \( \text{Var}_1(\psi(U)) \) is defined as \( \text{Var}_1(\psi(U)) = \frac{1}{n} \sum_{i=1}^n \hat{\psi}_1(U_i) - \left\{ \hat{\psi}_1(U) \right\}^2, \hat{\psi}_1(U) = \frac{1}{n} \sum_{i=1}^n \hat{\psi}_1(U_i) \), and \( \hat{E}_1(|X - \vartheta|) = \frac{1}{n} \sum_{i=1}^n |\hat{X}_i| - \text{me}(\hat{X}^2) \).

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

\[
\hat{\sigma}^2_{S_{P,2}} = \frac{\text{Var}(\bar{X})}{s_2^2} + \frac{\text{me}(\bar{X}^2)\bar{X}^2}{\text{Var}_2(\psi(U))} \frac{\text{Var}_2(\psi(U))}{\text{Var}_2(\psi(U))} + 1
\]

\[
+ \frac{2\text{me}(\bar{X}^2)\bar{X}|X|}{\text{Var}_2(\psi(U))} \frac{\text{Var}_2(\psi(U))}{\text{Var}_2(\psi(U))} + 1
\]

\[
+ \frac{1}{4s_2^2\hat{f}_2(\text{me}(\bar{X}^2))} + \frac{\hat{E}_2|X - \vartheta|}{s_2^2\hat{f}_2(\text{me}(\bar{X}^2))},
\]

where, \( \text{Var}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \hat{X}_i - \bar{X})^2, \bar{X}|X| = \frac{1}{n} \sum_{i=1}^n \hat{X}_i|\hat{X}_i|, \text{Var}_2(\psi(U)) = \frac{1}{n} \sum_{i=1}^n \hat{\psi}_2(U_i) - \left\{ \hat{\psi}_2(U) \right\}^2, \hat{\psi}_2(U) = \frac{1}{n} \sum_{i=1}^n \hat{\psi}_2(U_i) \), and also \( \hat{E}_2(|X - \vartheta|) = \)
distribution, i.e., \( \Phi(\cdot | z) \), is defined in (2.5) by substituting \( x \) with \( \hat{X}^{[2]} \). In particular, if \( \vartheta = 0 \), a consistent estimator of the asymptotic variance \( \sigma_{S_{P;2}}^2 \) is defined as 
\[
\hat{\sigma}_{S_{0;P;2}}^2 = \frac{\text{Var}(\hat{X})}{s_2^2} + \frac{1}{4s_2^2} \frac{1}{f_2(0)} + \frac{|\bar{X}|}{s_2^2 f_2(0)}.
\]

**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 \( \alpha \)-level asymptotic free test of \( H_0^\vartheta : f(x + \vartheta) = f(\vartheta - x) \) when \( \vartheta \neq 0 \) is 
\[
\sqrt{n} \hat{S}_{P;1}/\hat{\sigma}_{S_{P;1}} \geq z_{1-\alpha/2}, \quad l = 1, 2,
\]
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 
\[
\sqrt{n} \hat{S}_{P;2}/\hat{\sigma}_{S_{P;2}} \geq z_{1-\alpha/2} \text{ or } \sqrt{n} \hat{S}_{0;P;2}/\hat{\sigma}_{0;S_{P;2}} \geq z_{1-\alpha/2}
\]
to test the hypothesis \( H_0^\vartheta \).

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

\[
\hat{S}_{G;l} = \frac{\bar{X} - \text{me}(\hat{X}^{[l]})}{E(|X - \vartheta|)}, \quad l = 1, 2.
\]

We have the following asymptotic result.

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

\[
\sqrt{n} (\hat{S}_{G;l} - S_G) \xrightarrow{L} N(0, \sigma_{S_{G;l}}^2),
\]

where

\[
\sigma_{S_{G;l}}^2 = \frac{1}{[E|X - \vartheta|]^2} \text{Var} \left( \frac{2\vartheta}{E(X)} \bar{X} - \frac{2\vartheta - E(X)}{E(X)} X - \frac{I\{X \leq \vartheta\}}{f(\vartheta)} \right) - \frac{E(X) - \vartheta}{E|X - \vartheta|} |X - \vartheta|,
\]

\[
\sigma_{S_{G;2}}^2 = \frac{1}{[E|X - \vartheta|]^2} \text{Var} \left( \bar{X} + \frac{2\vartheta - E(X)}{E(|X|)} (|\bar{X}| - |X|) - \frac{I\{X \leq \vartheta\}}{f(\vartheta)} \right) - \frac{E(X) - \vartheta}{E|X - \vartheta|} |X - \vartheta|.
\]
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 $|\tilde{X}| = |X|\psi(U) = X\psi(U) = \tilde{X}$. 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 $|\tilde{X}| = |X|\psi(U) = -X\psi(U) = -\tilde{X}$.

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 $\vartheta$ and $E(X) = \vartheta$, if $\vartheta \neq 0$, the asymptotic variance $\sigma^2_{S_{G,1}}$ reduces to $\sigma^2_{S_{G,1}} = \frac{\sigma^2}{E\left(\frac{X^2 - \vartheta^2}{2}\right)^2} \sigma^2_{S_{P,1}}$, then a consistent estimator of $\sigma^2_{S_{G,1}}$ is obtained as $\hat{\sigma}^2_{S_{G,1}} = \frac{\hat{\sigma}^2}{E\left(\frac{X^2 - \vartheta^2}{2}\right)^2} \hat{\sigma}^2_{S_{P,1}}$. For the estimator $\hat{S}_{G,2}$, the asymptotic variance $\sigma^2_{S_{G,2}}$ also reduces to $\sigma^2_{S_{G,2}} = \frac{\sigma^2}{E\left(\frac{X^2 - \vartheta^2}{2}\right)^2} \sigma^2_{S_{P,2}}$, and then a consistent estimator of $\sigma^2_{S_{G,2}}$ is obtained as $\hat{\sigma}^2_{S_{G,2}} = \frac{\hat{\sigma}^2}{E\left(\frac{X^2 - \vartheta^2}{2}\right)^2} \hat{\sigma}^2_{S_{P,2}}$.

Thus, the proposed $\alpha$-level asymptotic free test of $H_0^g: f(x + \vartheta) = f(\vartheta - x)$ is

$$\sqrt{n}\hat{S}_{G,1}/\hat{\sigma}_{S_{G,1}} \geq z_{1 - \alpha/2}, \quad l = 1, 2.$$ In particular, if $\vartheta = 0$, a consistent estimator of the asymptotic variance $\sigma^2_{S_{G,2}}$ is defined as $\sigma^2_{S_{G,2}} = \frac{\hat{\sigma}^2}{E\left(\frac{X^2 - \vartheta^2}{2}\right)^2} \hat{\sigma}^2_{S_{P,2}}$.

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 $\hat{S}_M = \frac{\sum^n_{i=1} (X_i - \tilde{X})^3}{\left|\sum^n_{i=1} (X_i - \tilde{X})^2\right|^{3/2}}$, 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

$$\hat{S}_{M,l} = \frac{n^{-1} \sum^n_{i=1} \left(\tilde{X}_i - \hat{E}(\tilde{X}_i | U_i)\right)^3}{\left[n^{-1} \sum^n_{i=1} \hat{\psi}_i^3(U_i)\right]^{3/l}}.$$

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

\[ \sqrt{n} \left( \frac{\hat{S}_{M,l} - S_M}{\tilde{S}_{M,l} - S_M} \right) \xrightarrow{L} N(0, \sigma_{S_M}^2), \quad l = 1, 2, \]

where,

\[ \sigma_{S_M}^2 = \text{Var} \left( \frac{\psi^3(U)(X - E(X))^3}{E[\psi^3(U)]\sigma^3} - \frac{S_M}{E[\psi^3(U)]}\psi^3(U) - \frac{3S_M}{2\sigma^2}(X - E(X))^2 \right), \]

**Remark 5.** The asymptotic variance \( \sigma_{S_M}^2 \) is expressed as \( \sigma_{S_M}^2 = \frac{\delta_\psi(6)\mu_X(6)}{\delta_\psi(3)^2\sigma^6} \), \( S_M^2\frac{\delta_\psi(6)}{\delta_\psi(3)^2} + S_M^2\frac{\mu_X(4)}{\delta_\psi} - 3S_M\frac{\mu_X(5)}{\delta_\psi} + \frac{3S_M^2}{4}\), 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 \( \sigma_{S_M}^2 \) reduces to \( \sigma_{S_M}^2 = \frac{\delta_\psi(6)\mu_X(6)}{\delta_\psi(3)^2\sigma^6} \). It is known that when \( X \) is symmetric and observed without measurement errors, the asymptotic variance of \( \hat{S}_M \) is shown to be \( \frac{\mu_X(6)}{\sigma^6} \) (Gupta; 1967). The Cauchy-inequality entails that \( \delta_\psi(6) \geq \delta_\psi(3) \), and then \( \sigma_{S_M}^2 \geq \frac{\mu_X(6)}{\sigma^6} \) when \( X \) is symmetric. This implies the multiplicative distortion function \( \psi(U) \) increases the classical asymptotic variance of sample skewness.

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

### 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

\[ \hat{S}_{Q,\delta} = \frac{\hat{S}_{Q,\delta}}{q_{1-\delta} - q_{\delta}}, \quad \delta \in (0,0.5), \quad l = 1, 2, \]

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

\[
\sqrt{n} \left( \frac{\tilde{S}_{Q,\delta}^{[l]} - S_{Q,\delta}^{[l]}}{\hat{S}_{Q,\delta}^{[l]}} \right) \xrightarrow{L} N(0, \sigma_{S_{Q,\delta}}^2), \quad l = 1, 2,
\]

where, let \( \omega_\delta = \frac{q_{1-\delta} - \vartheta}{q_{1-\delta} - q_\delta} \), and

\[
\sigma_{S_{Q,\delta}}^2 = \frac{4\delta(1 - \delta)}{(q_{1-\delta} - q_\delta)^2} \left[ \frac{\omega_\delta^2}{f^2(q_\delta)} + \frac{(1 - \omega_\delta)^2}{f^2(q_{1-\delta})} \right] + \frac{1}{(q_{1-\delta} - q_\delta)^2 f^2(\vartheta)} \left[ \frac{8\delta^2 \omega_\delta (1 - \omega_\delta)}{(q_{1-\delta} - q_\delta)^2 f(q_\delta) f(q_{1-\delta})} - \frac{4\delta \omega_\delta}{(q_{1-\delta} - q_\delta)^2 f(q_\delta) f(\vartheta)} \right]
\]

Remark 6. Note that estimators \( \tilde{S}_{Q,\delta}^{[l]} \) of skewness are efficient. In other words, the proposed estimator \( S_{Q,\delta} \) eliminates the effect caused by the additive adjusted covariate \( \psi(U) \), i.e., 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 \( \sigma_{S_{Q,\delta}}^2 \) reduces to \( \sigma_{S_{Q,\delta}}^2 = \frac{1}{(q_{1-\delta} - q_\delta)^2} \left[ \frac{2\delta}{f^2(q_\delta)} + \frac{1}{f^2(\vartheta)} \right] - \frac{4\delta}{f^2(q_\delta) f^2(\vartheta)} \), and a consistent estimator of \( \sigma_{S_{Q,\delta}}^2 \) is proposed as

\[
\left( \hat{\sigma}_{S_{Q,\delta}}^{[l]} \right)^2 = \frac{1}{(q_{1-\delta} - q_\delta)^2} \left[ \frac{2\delta}{f^2(q_\delta)} + \frac{1}{f^2(\text{me}(X^{[l]}))} \right] - \frac{4\delta}{f^2(q_\delta) f^2(\text{me}(X^{[l]}))} \]

The \( \alpha \)-level asymptotic free test of \( H_0^* : f(x + \vartheta) = f(\vartheta - x) \), can be constructed as \( \sqrt{n} \tilde{S}_{Q,\delta}^{[l]} / \hat{\sigma}_{S_{Q,\delta}}^{[l]} \geq z_{1-\alpha/2} \), where \( \hat{\sigma}_{S_{Q,\delta}}^{[l]} = \sqrt{\left( \hat{\sigma}_{S_{Q,\delta}}^{[l]} \right)^2} \), \( l = 1, 2 \). If \( E(X) = 0 \), the \( \alpha \)-level asymptotic free test is \( \sqrt{n} \tilde{S}_{Q,\delta}^{[2]} / \hat{\sigma}_{S_{Q,\delta}}^{[2]} \geq z_{1-\alpha/2} \).

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
generally, when the sample size $n$ coincided with the Remark 2 in Theorem 2.1, the estimator $\hat{\rho}$ better than $\hat{\rho}_{s}$, but has much smaller values of MSE. When the sample size $n$ increases, the performances of $\hat{\rho}$ are better both in terms of average length of the confidence intervals and the coverage probabilities. In Table 1, we see that when the sample size $n$ is 300 or 500, the performances of $\hat{\rho}_{B,5}$, $\hat{\rho}_{s,5}$, $s = 1, 2$ are better than $\hat{\rho}_{k,5}$ and $\hat{\rho}_{k,5}$, $k = 1, 2, 3$ because the values of mean of $\hat{\rho}_{s,5}$ and $\hat{\rho}_{B,5}$ are close to zero and the values of MSE are smaller. We find that the performances of $\hat{\rho}_{s,5}$, $s = 1, 2$ are almost the same as $\hat{\rho}_{k,5}$ when the sample size $n$ gets larger. We see that the empirical likelihood confidence intervals of $\hat{\rho}_{s,5}$, $s = 1, 2$ and $\hat{\rho}_{B,5}$ show satisfactory performances, and the coverage probabilities of $\hat{\rho}_{s,5}$, $\hat{\rho}_{B,5}$ are much better than $\hat{\rho}_{k,5}$ and $\hat{\rho}_{k,5}$, $k = 1, 2, 3$, which indicate that a larger value of $k$ may cause lower coverage probabilities with shorter average lengths but has much smaller values of MSE. When the sample size $n$ increases, the performances of $\hat{\rho}_{s,5}$, $s = 1, 2$ and $\hat{\rho}_{B,5}$ become better both in terms of average lengths 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 $\hat{\rho}_{0,5}$ are all smaller than those of $\hat{\rho}_{0,5}$. This is not surprised and coincided with the Remark 2 in Theorem 2.1, the estimator $\hat{\rho}_{0,5}^{[2]}$ performs better than $\hat{\rho}_{0,5}^{[1]}$ in estimation because $|E(X)| = 1$, $E(|X|) = 1.166$ and $\frac{1}{E(X)^{\alpha}} > \frac{1}{E(X)^{\alpha}}$ holds true. Besides, the naive estimator $\hat{\rho}_{k}$ fails to recover the symmetry of the underlying unobserved variable $X$. The values of mean of $\hat{\rho}_{k}$ all depart from zero, and the confidence 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 $\hat{\rho}^{[1]}_{k}$ fails but the estimator $\hat{\rho}^{[2]}_{k}$ works. We investigate the performance of $\hat{\rho}^{B}_{k}, \hat{\rho}^{[2]}_{k}$ and $\hat{\rho}_{k}$ for $k = 0.5, 1, 2, 3$. Note that $X$ is symmetric about zero, then the naive estimator $\hat{\rho}_{k}$ works 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 $\hat{\rho}^{B}_{0.5}, \hat{\rho}^{[2]}_{0.5}$ and $\hat{\rho}_{0.5}$ are better than $\hat{\rho}^{B}_{k}, \hat{\rho}^{[2]}_{k}$ and $\hat{\rho}_{k}, k = 1, 2, 3$, because the values of mean of $\hat{\rho}^{[2]}_{0.5}$ and $\hat{\rho}^{B}_{0.5}$ are close to zero and the values of MSE are smaller, and coverage probabilities of empirical likelihood confidence intervals of $\hat{\rho}^{B}_{0.5}, \hat{\rho}^{[2]}_{0.5}$ and $\hat{\rho}_{0.5}$ 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 lengths even it results in smaller values of MSE. When the sample size $n$ increases, the performances of $\hat{\rho}^{[2]}_{0.5}$ are close to the benchmark estimator $\hat{\rho}^{B}_{0.5}$ 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 $\hat{\rho}^{[1]}_{k}$ also fails and the estimator $\hat{\rho}^{[2]}_{k}$ works. We investigate the performances of $\hat{\rho}^{B}_{k}, \hat{\rho}^{[2]}_{k}$ and $\hat{\rho}_{k}$ 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 $\hat{\rho}^{B}_{k}$ and $\hat{\rho}_{k}^{[2]}$ are better than $\hat{\rho}^{B}_{k}$ 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, $\hat{\rho}^{B}_{0.5}$ and $\hat{\rho}_{0.5}$ 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 $\hat{\rho}_{k}$ fails and has larger bias compared with $\hat{\rho}^{B}_{k}$ and $\hat{\rho}_{k}^{[2]}$. It is seen that the right confidence intervals of $\hat{\rho}^{B}_{k}$ and $\hat{\rho}_{k}^{[2]}$ exclude the left confidence intervals of $\hat{\rho}_{k}$ when the sample size $n$ is 300 and 500. This implies that the bias of $\hat{\rho}_{k}$ 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 $X \sim N(-2, 1), X \sim \chi^{2}_{5} - 5$ (a 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.
The means (MEAN), standard errors (SE), mean squared errors (MSE) and the 95% confidence intervals for $\hat{\rho}_b$, $\hat{\rho}_s^1$, $\hat{\rho}_s^2$ and $\hat{\rho}_b$. “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 $10^{-3}$.

<table>
<thead>
<tr>
<th>n = 100</th>
<th>k = 0.5</th>
<th>$\bar{\rho}$</th>
<th>SE</th>
<th>MSE</th>
<th>Lower</th>
<th>Upper</th>
<th>AL</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_b$</td>
<td>0.0795</td>
<td>0.0777</td>
<td>0.1050</td>
<td>0.1050</td>
<td>94.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^1$</td>
<td>0.0850</td>
<td>0.0867</td>
<td>0.1050</td>
<td>0.1050</td>
<td>95.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^2$</td>
<td>0.3165</td>
<td>0.2815</td>
<td>0.1857</td>
<td>0.1857</td>
<td>95.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_b$</td>
<td>0.0473</td>
<td>0.0545</td>
<td>0.0739</td>
<td>0.0739</td>
<td>91.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^1$</td>
<td>0.0324</td>
<td>0.0297</td>
<td>0.0352</td>
<td>0.0352</td>
<td>92.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^2$</td>
<td>0.2948</td>
<td>0.2972</td>
<td>0.1538</td>
<td>0.1538</td>
<td>92.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 300</th>
<th>k = 0.5</th>
<th>$\bar{\rho}$</th>
<th>SE</th>
<th>MSE</th>
<th>Lower</th>
<th>Upper</th>
<th>AL</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_b$</td>
<td>0.0252</td>
<td>0.0243</td>
<td>0.0161</td>
<td>0.0161</td>
<td>94.4%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^1$</td>
<td>0.0414</td>
<td>0.0323</td>
<td>0.0787</td>
<td>0.0787</td>
<td>95.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^2$</td>
<td>0.2690</td>
<td>0.1510</td>
<td>0.1571</td>
<td>0.1571</td>
<td>90.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_b$</td>
<td>0.0151</td>
<td>0.0299</td>
<td>0.0150</td>
<td>0.0150</td>
<td>90.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^1$</td>
<td>0.0220</td>
<td>0.0200</td>
<td>0.0196</td>
<td>0.0196</td>
<td>90.4%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^2$</td>
<td>0.2696</td>
<td>0.1511</td>
<td>0.0779</td>
<td>0.0779</td>
<td>91.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 500</th>
<th>k = 0.5</th>
<th>$\bar{\rho}$</th>
<th>SE</th>
<th>MSE</th>
<th>Lower</th>
<th>Upper</th>
<th>AL</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_b$</td>
<td>0.0212</td>
<td>0.0199</td>
<td>0.0150</td>
<td>0.0150</td>
<td>95.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^1$</td>
<td>0.0264</td>
<td>0.0252</td>
<td>0.0150</td>
<td>0.0150</td>
<td>95.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^2$</td>
<td>0.2618</td>
<td>0.2572</td>
<td>0.1261</td>
<td>0.1261</td>
<td>95.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_b$</td>
<td>0.0073</td>
<td>0.0227</td>
<td>0.0096</td>
<td>0.0096</td>
<td>91.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^1$</td>
<td>0.0168</td>
<td>0.0168</td>
<td>0.0150</td>
<td>0.0150</td>
<td>93.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}_s^2$</td>
<td>0.2674</td>
<td>0.2686</td>
<td>0.1261</td>
<td>0.1261</td>
<td>93.8%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 (using the true covariate $X$ in the simulation) $\hat{S}_P = \frac{\bar{X} - me(X)}{n^{-1} \sum_{i=1}^{n} |X_i - me(X)|^2}^{1/2}$,
2 $\hat{S}_G = \frac{\bar{X} - me(X)}{n^{-1} \sum_{i=1}^{n} |X_i - me(X)|}$, $\hat{S}_M = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^3}{\left[\sum_{i=1}^{n} |X_i - \bar{X}|^2\right]^{3/2}}$, $\hat{S}_{Q, 0.25} = \frac{\bar{X} - me(X)}{q_{0.75} + \bar{X} - me(X) - 2q_{0.25} - 2me(X)}$.
Table 2. The means (MEAN), standard errors (SE), mean squared errors (MSE) and the 95% confidence intervals for $\hat{p}_k^{(1)}$, $\hat{p}_k^{(2)}$ and $\hat{p}_k$.” 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 $10^{-3}$.

<table>
<thead>
<tr>
<th>n = 100</th>
<th>k = 0.5</th>
<th>$\hat{p}_k^{(1)}$</th>
<th>$\hat{p}_k^{(2)}$</th>
<th>$\hat{p}_k$</th>
<th>MEAN</th>
<th>SE</th>
<th>MSE</th>
<th>Lower</th>
<th>Upper</th>
<th>AL</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0807</td>
<td>0.1419</td>
<td>26.6534</td>
<td>-0.2577</td>
<td>0.2594</td>
<td>0.5173</td>
<td>95.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0797</td>
<td>0.1407</td>
<td>27.3862</td>
<td>-0.2578</td>
<td>0.2604</td>
<td>0.5183</td>
<td>95.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0704</td>
<td>0.1241</td>
<td>19.7466</td>
<td>-0.2524</td>
<td>0.2576</td>
<td>0.5084</td>
<td>96.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0491</td>
<td>0.1513</td>
<td>25.3132</td>
<td>-0.2505</td>
<td>0.2473</td>
<td>0.4979</td>
<td>93.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0500</td>
<td>0.1543</td>
<td>26.3073</td>
<td>-0.2508</td>
<td>0.2489</td>
<td>0.4998</td>
<td>92.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0417</td>
<td>0.1296</td>
<td>18.5500</td>
<td>-0.2439</td>
<td>0.2416</td>
<td>0.4855</td>
<td>96.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0497</td>
<td>0.1417</td>
<td>21.7812</td>
<td>-0.2449</td>
<td>0.2467</td>
<td>0.4958</td>
<td>94.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0338</td>
<td>0.1653</td>
<td>28.4590</td>
<td>-0.2325</td>
<td>0.2306</td>
<td>0.4632</td>
<td>86.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The distorted estimators $\tilde{S}_P$ = $\frac{X - \text{me}(\bar{X})}{\sqrt{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}^{1/2}$, $\tilde{S}_G = \frac{X - \text{me}(\bar{X})}{\sqrt{n-1} \sum_{i=1}^{n} |X_i - \text{me}(X)|}$, $\tilde{S}_M = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^3}{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{3/2}}$, $\tilde{S}_{Q,0.25} = \frac{\tilde{q}_{0.75} + 0.25 - 2 \text{me}(\bar{X})}{\tilde{q}_{0.75} - 0.25}$. Here, me$(X)$ and $\tilde{q}_{0.75}$ are the median and δ-quantile of $\{X_1, \ldots, X_n\}$, and me$(\bar{X})$ and $\tilde{q}_{0.75}$ are the median and δ-quantile of $\{\bar{X}_1, \ldots, \bar{X}_n\}$. Note that the true value of $(\tilde{S}_P, \tilde{S}_G, \tilde{S}_M, \tilde{S}_{Q,0.25})$ is (0, 0, 0, 0) for normal distribution $N(-2, 1)$ and (0.2051, 0.2718, 1.2649, 0.1512) for chi-squared distribution $\chi^2_5-1$. In Table 4, it is seen that the distorted estimators $\tilde{S}_P, \tilde{S}_G, \tilde{S}_M$ and $\tilde{S}_{Q,0.25}$ have non-ignorable bias, which results in large values of MSE compared with benchmark estimators and proposed estimators.

imsart-bjps ver. 2014/10/16 file: Masure-Mul-2019.tex date: January 15, 2019
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 $n$ increases. It is also seen that
the estimator $\hat{S}_{M,1}$ performs not well compared with the other three esti-
mators, 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 $\tilde{S}_{P,l}$, $\tilde{S}_{G,l}$, $\tilde{S}_{M,l}$
and $\tilde{S}_{Q,0.25}$, $l = 1, 2$ based on 2000 realization. Here, we consider the data
generating process (DGP) given as $X \sim (1 - b) \ast N(-0.5, 1) + b \ast (-\chi^2_l)$,
b = 0.2, . . . , 1. It is easily seen that the density of X is asymmetric when
b ̸= 0. The sample size n is chosen as n = 100, n = 300 and n = 500. We
find that the power functions of λ1, ̂SP,1 and ̂SG,1 are generally better than
λ0,5, ̂SP,1 and ̂SG,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 λ0,5 works the best for detecting the asymmetry of the underlying
density function and is the most powerful than other four test statistics, and
̂SG,0.25 is the worst when the sample size n is 100 and gets better when the
sample size n is 300 and 500.

Table 4  The means (M), standard errors (Se), mean squared errors (MSE) of proposed
estimators (P) ̂SP, ̂SG, ̂SM, ̂SQ,0.25, l = 1, 2, and benchmark estimators (B) and
distorted estimators (D). MSE is in the scale of 10^{-3}.

<table>
<thead>
<tr>
<th>n</th>
<th>M</th>
<th>Se</th>
<th>Mse</th>
<th>M</th>
<th>Se</th>
<th>Mse</th>
<th>M</th>
<th>Se</th>
<th>Mse</th>
<th>M</th>
<th>Se</th>
<th>Mse</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=2.11</td>
<td>100</td>
<td>P1</td>
<td>-0.0145</td>
<td>0.0073</td>
<td>5.5792</td>
<td>-0.0237</td>
<td>0.0920</td>
<td>0.1234</td>
<td>-0.0107</td>
<td>0.0219</td>
<td>0.0431</td>
<td>-0.0025</td>
</tr>
<tr>
<td>B</td>
<td>0.0027</td>
<td>0.0072</td>
<td>5.3009</td>
<td>-0.0035</td>
<td>0.0003</td>
<td>0.0202</td>
<td>0.0010</td>
<td>0.0134</td>
<td>0.1478</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>-0.0171</td>
<td>0.0097</td>
<td>3.41925</td>
<td>-0.0220</td>
<td>0.0916</td>
<td>0.1264</td>
<td>-0.0078</td>
<td>0.0183</td>
<td>0.1678</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>-0.0081</td>
<td>0.0128</td>
<td>5.6391</td>
<td>-0.0123</td>
<td>0.0094</td>
<td>0.1418</td>
<td>-0.0061</td>
<td>0.0203</td>
<td>0.1411</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.0029</td>
<td>0.0048</td>
<td>2.2772</td>
<td>-0.0065</td>
<td>0.0541</td>
<td>3.6309</td>
<td>-0.0001</td>
<td>0.1779</td>
<td>32.3255</td>
<td>-0.0033</td>
<td>0.0770</td>
<td>5.9403</td>
</tr>
<tr>
<td>D</td>
<td>-0.0160</td>
<td>0.0037</td>
<td>0.3562</td>
<td>-0.0088</td>
<td>0.0019</td>
<td>0.3777</td>
<td>32.3991</td>
<td>0.0064</td>
<td>0.0878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>-0.0127</td>
<td>0.0032</td>
<td>1.5456</td>
<td>-0.0076</td>
<td>0.0421</td>
<td>2.7474</td>
<td>-0.0023</td>
<td>0.1363</td>
<td>18.5970</td>
<td>-0.0043</td>
<td>0.0584</td>
<td>4.1432</td>
</tr>
<tr>
<td>B</td>
<td>0.0021</td>
<td>0.0032</td>
<td>1.6524</td>
<td>-0.0075</td>
<td>0.0417</td>
<td>2.7824</td>
<td>-0.0027</td>
<td>0.1365</td>
<td>18.6381</td>
<td>-0.0042</td>
<td>0.0584</td>
<td>4.1523</td>
</tr>
<tr>
<td>D</td>
<td>0.0011</td>
<td>0.0033</td>
<td>1.1144</td>
<td>-0.0014</td>
<td>0.0424</td>
<td>1.0609</td>
<td>-0.0019</td>
<td>0.1056</td>
<td>11.1632</td>
<td>-0.0023</td>
<td>0.0597</td>
<td>3.5762</td>
</tr>
</tbody>
</table>

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
unobserved serum creatinine (Scr) data as an illustration of our method.
Assume that the distorted GFR is ̂X1 and the distorted Scr is ̂X2. 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 $\text{BSA}(m^2) = 0.007184 \times \text{Kg}^{0.425} \times \text{Cm}^{0.725}$.

We first present the patterns of $\hat{\psi}_{1,GFR}(u)$ and $\hat{\psi}_{1,SCr}(u)$ by using (2.2) in Figure 1 under the confounding variable-BSA. Figure 1 indicates that underlying distorting functions $\psi_{GFR}(u)$ and $\psi_{SCr}(u)$ are not a constant function, suggesting that the confounding variable-BSA definitely makes effect of GFR and SCr in this data. From the original data and the estimated curves of $\hat{\psi}_{1,GFR}(u)$ and $\hat{\psi}_{1,SCr}(u)$ in Figure 1, there is no negative values of the distorted GFR (i.e., $|\tilde{X}_1| = \tilde{X}_1$) and the distorted SCr ($|\tilde{X}_2| = \tilde{X}_2$), so the estimators $(\hat{\psi}_{1,GFR}(\tilde{X}_1), \hat{\psi}_{1,SCr}(\tilde{X}_1))$ and $(\hat{\psi}_{2,GFR}(u), \hat{\psi}_{2,SCr}(u))$(by using (2.3) with a common bandwidth) satisfy $\psi_{1,GFR}(u) = \hat{\psi}_{2,GFR}(u)$ and $\psi_{1,SCr}(u) = \hat{\psi}_{2,SCr}(u)$ for this dataset. Figure 1 implies the values of $\hat{\psi}_{1,GFR}(u)$ and $\hat{\psi}_{1,SCr}(u)$ should be positive. Because all the values of $(\tilde{X}_1, \tilde{X}_2)$, $X_1^{[1]}$, $X_2^{[1]}$

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\hat{\psi}_{0.5}$</th>
<th>$\hat{\psi}_{1,GFR}$</th>
<th>$\hat{\psi}_{1,SCr}$</th>
<th>$\hat{\psi}_{2,GFR}$</th>
<th>$\hat{\psi}_{2,SCr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 0.2</td>
<td>0.0585</td>
<td>0.0900</td>
<td>0.0590</td>
<td>0.0345</td>
<td>0.0570</td>
</tr>
<tr>
<td>b = 0.4</td>
<td>0.3015</td>
<td>0.5640</td>
<td>0.1565</td>
<td>0.1505</td>
<td>0.5905</td>
</tr>
<tr>
<td>b = 0.6</td>
<td>0.9430</td>
<td>0.5625</td>
<td>0.6855</td>
<td>0.5905</td>
<td>0.1845</td>
</tr>
<tr>
<td>b = 0.8</td>
<td>1.0000</td>
<td>0.9355</td>
<td>0.8075</td>
<td>0.9725</td>
<td>0.4700</td>
</tr>
<tr>
<td>b = 1.0</td>
<td>1.0000</td>
<td>0.9540</td>
<td>0.8230</td>
<td>0.5510</td>
<td>0.8390</td>
</tr>
<tr>
<td>n = 300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 0.2</td>
<td>0.1170</td>
<td>0.0100</td>
<td>0.0365</td>
<td>0.0665</td>
<td></td>
</tr>
<tr>
<td>b = 0.4</td>
<td>0.7575</td>
<td>0.3050</td>
<td>0.7130</td>
<td>0.0980</td>
<td></td>
</tr>
<tr>
<td>b = 0.6</td>
<td>0.9990</td>
<td>0.9875</td>
<td>0.9580</td>
<td>0.3730</td>
<td></td>
</tr>
<tr>
<td>b = 0.8</td>
<td>1.0000</td>
<td>0.9935</td>
<td>0.9725</td>
<td>0.8390</td>
<td></td>
</tr>
<tr>
<td>b = 1.0</td>
<td>1.0000</td>
<td>0.9940</td>
<td>0.9740</td>
<td>0.8815</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\hat{\psi}_{2,GFR}$</th>
<th>$\hat{\psi}_{2,SCr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 0.2</td>
<td>0.1315</td>
<td>0.0140</td>
</tr>
<tr>
<td>b = 0.4</td>
<td>0.9390</td>
<td>0.5690</td>
</tr>
<tr>
<td>b = 0.6</td>
<td>1.0000</td>
<td>0.9935</td>
</tr>
<tr>
<td>b = 0.8</td>
<td>1.0000</td>
<td>0.9910</td>
</tr>
<tr>
<td>b = 1.0</td>
<td>1.0000</td>
<td>0.9920</td>
</tr>
<tr>
<td>n = 300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 0.2</td>
<td>0.1105</td>
<td>0.0665</td>
</tr>
<tr>
<td>b = 0.4</td>
<td>0.7705</td>
<td>0.5165</td>
</tr>
<tr>
<td>b = 0.6</td>
<td>0.9995</td>
<td>0.9985</td>
</tr>
<tr>
<td>b = 0.8</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>b = 1.0</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

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

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

imsart-bjps ver. 2014/10/16 file: Masure-Mul-2019.tex date: January 15, 2019
and \( \hat{\psi}_{1, \text{GFR}}(u) \), \( \hat{\psi}_{1, \text{SCR}}(u) \) 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 \( \hat{S}_{P,1}^{[1]}, \hat{S}_{G,1}^{[1]}, \hat{S}_{M,1}^{[1]} \) and \( \hat{S}_{Q,0.25}^{[1]} \) to investigate the symmetry of the unobserved GFR and Scr. The 95% confidence intervals of \( \hat{\rho}_{0.5}^{[1]} \) 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 \( \hat{S}_{P,1}^{[1]}, \hat{S}_{G,1}^{[1]}, \hat{S}_{M,1}^{[1]} \) and \( \hat{S}_{Q,0.25}^{[1]} \) and associated \( p \)-values in Table 7. The plots of the histogram and density function estimate of original variables \( X_s \) 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 \( \hat{X}_2 \) in Figure 2 implies asymmetry, but the estimated \( \hat{X}^{[1]}_2 \) shows a slightly symmetric and also investigated by statistics \( \hat{S}_{P,1}^{[1]}, \hat{S}_{G,1}^{[1]} \) and \( \hat{S}_{Q,0.25}^{[1]} \). The statistic \( \hat{S}_{M,1} \) and \( \hat{\rho}_{0.5}^{[1]} \) show that Scr should be asymmetric. Together with Figure 1 and the performances of statistic \( \hat{S}_{M,1} \) and \( \hat{\rho}_{0.5}^{[1]} \) 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 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 (\( \tilde{X}_1 \)) and Scr (\( \tilde{X}_2 \)) 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 (\( \tilde{X}_1 \)) and Scr (\( \tilde{X}_2 \)) 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 (\( \hat{X}^{[2]}_2 \)) 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.

**Table 7** The estimation and \( p \)-values of estimators \( \hat{S}_{P,1}^{[1]}, \hat{S}_{G,1}^{[1]}, \hat{S}_{M,1}^{[1]} \) and \( \hat{S}_{Q,0.25}^{[1]} \).

<table>
<thead>
<tr>
<th></th>
<th>( \hat{S}_{P,1}^{[1]} )</th>
<th>( \hat{S}_{G,1}^{[1]} )</th>
<th>( \hat{S}_{M,1}^{[1]} )</th>
<th>( \hat{S}_{Q,0.25}^{[1]} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scr</strong></td>
<td>0.0719</td>
<td>0.4319</td>
<td>0.0856</td>
<td>0.4319</td>
</tr>
<tr>
<td></td>
<td>0.1619</td>
<td>0.0434</td>
<td>0.0827</td>
<td>0.0860</td>
</tr>
<tr>
<td><strong>GFR</strong></td>
<td>0.2733</td>
<td>0.0002</td>
<td>0.3618</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>1.1476</td>
<td>2.0812×10^{-8}</td>
<td>0.2733</td>
<td>4.0810×10^{-7}</td>
</tr>
</tbody>
</table>
Figure 1  The estimated curve of distorting functions $\psi_{SCr}(u)$ and $\psi_{GFR}(u)$, associated 95% pointwise confidence intervals (dotted lines).

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

Acknowledgements. 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 No. 11871001 and 11471029) and Beijing Natural Science Foundation (Grant No. 1182003).

References


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
Beijing Institute for Scientific and Engineering Computing
Beijing University of Technology
100124, Beijing
China

E-mail: yujie_gai@126.com