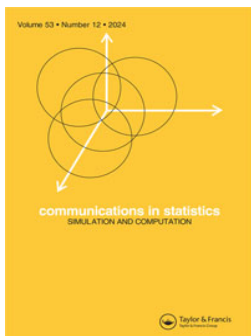


Notice: This material may be protected
by copyright law (title 17 U.S. code).

If this material has missing or unreadable pages, please contact the
Georgia State University Library Interlibrary Loan Office
at libill@langate.gsu.edu or (404) 413-2790.

Thank you for using Interlibrary Loan and ILLiad.



A revisit to Pearson correlation coefficient under multiplicative distortions

Siming Deng, Jun Zhang, Yingcong Huang, Jiongtao Zhong & Xiaozhen Yang

To cite this article: Siming Deng, Jun Zhang, Yingcong Huang, Jiongtao Zhong & Xiaozhen Yang (02 Apr 2024): A revisit to Pearson correlation coefficient under multiplicative distortions, Communications in Statistics - Simulation and Computation, DOI: [10.1080/03610918.2024.2333352](https://doi.org/10.1080/03610918.2024.2333352)

To link to this article: <https://doi.org/10.1080/03610918.2024.2333352>



Published online: 02 Apr 2024.



Submit your article to this journal [↗](#)



Article views: 126



View related articles [↗](#)




View Crossmark data [↗](#)



Citing articles: 4 View citing articles [↗](#)



A revisit to Pearson correlation coefficient under multiplicative distortions

Siming Deng, Jun Zhang , Yingcong Huang, Jiongtao Zhong, and Xiaozhen Yang

School of Mathematical Sciences, Shenzhen University, Shenzhen, China

ABSTRACT

We consider the estimation of Pearson correlation coefficient when two continuous variables can not be directly observed but measured with multiplicative distortion measurement errors. Different from the identifiability conditions for the distortion functions in literature, we propose a new estimation method of Pearson correlation coefficient by transforming the linear model between two variables into varying coefficient models, a moment-based estimator of the Pearson correlation coefficient is proposed. This method can deal with the non-independence condition between the confounding variable and the unobserved variables. We study the asymptotic results of these proposed estimators, and we make some comparisons among the proposed estimators through the simulation. These methods are applied to analyze a real dataset for an illustration.

ARTICLE HISTORY

Received 22 September 2023
Accepted 16 March 2024

KEYWORDS

Correlation coefficient;
Distance correlation coefficient;
Local linear smoothing;
Multiplicative distortion measurement errors

MATHEMATICS SUBJECT CLASSIFICATION (2000)

62G05; 62G08; 62G20

1. Introduction

Measurement error usually happens when there is a difference between a measured value of quantity and its true value. It may arise from various sources, for example, instrumental contamination or biological variation. Additional complications can arise in practical collected data analytic situations when one or more of the covariates is measured with error, and statistical properties of a specific model or method may be misleading if one ignores the presence of measurement errors such as the attenuation bias (Fuller 1987; Carroll et al. 2006) even when the sample size is large. To obtain the correct results under the measurement errors scenario, various methods have been developed in recent decades such as the regressions calibration methods (Liang, Härdle, and Carroll 1999; Carroll et al. 2006), empirical likelihood based methods (Yang, Li, and Tong 2015), simulation extrapolation (Yang, Tong, and Li 2019), dimension reduction based methods (Li and Yin 2007). In this paper, we consider the following multiplicative distortion measurement errors models

$$\tilde{Y} = \phi(U)Y, \quad \tilde{X} = \psi(U)X, \quad (1.1)$$

where (Y, X) are unobservable continuous variables, (\tilde{Y}, \tilde{X}) are the observed and distorted variables. The confounding variable $U \in \mathbb{R}^1$ is continuous and observable. The multiplicative distortion functions $\phi(\cdot), \psi(\cdot)$ are unknown smooth functions. It is noted that $\phi(\cdot), \psi(\cdot)$ distort unobserved Y, X in multiplicative fashions.

To estimate the Pearson correlation coefficient $\rho = \frac{\text{Cov}(Y, X)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$ of (Y, X) in model (1.1), Şentürk and Müller (2005b) used the binning technique by transforming it into a varying

coefficient model, then they proposed a covariate adjusted correlation (Cadcor) estimator of ρ . Another way to estimate ρ in model (1.1) is to estimate distortion functions $\phi(\cdot)$ and $\psi(\cdot)$ at first. To estimate unknown distortion functions, some identifiability conditions should be assumed and result in various calibration estimation methods, such as the conditional mean calibration methods (Zhang, Li, and Feng 2015), the conditional absolute mean calibration methods (Delaigle, Hall, and Zhou 2016; Zhang, Lin, and Feng 2020; Zhang 2021; Zhang, Chen, and Wei 2023), the conditional variance calibration methods (Zhang 2019; Zhang, Lin, and Li 2019) and the logarithmic calibration methods (Zhang, Yang, et al. 2020; Zhang, Yang, and Li 2020; Zhang and Cui 2021). Based on these estimation methods for unknown distortion functions, researchers estimated the unobserved variables as $\hat{Y}_i = \frac{\tilde{Y}_i}{\phi(U_i)}$, $\hat{X}_i = \frac{\tilde{X}_i}{\psi(U_i)}$ and then used the calibrated variables $\{\hat{Y}_i, \hat{X}_i\}_{i=1}^n$ as the “observed” variables to estimate ρ .

Except for the Cadcor estimation method (Şentürk and Müller 2005b), it is remarkable that the above calibration methods can obtain the asymptotically efficient estimation procedures to estimate unknown ρ . This is because the asymptotic variance of the Cadcor estimator of ρ has two additive terms. One term is classical asymptotic variance of the moment estimator of ρ (Ferguson, 1996 Section 8) when the i.i.d samples $\{X_i, Y_i\}_{i=1}^n$ are available without distortions. The other term in the asymptotic variance of Cadcor estimator is related with the unknown distortion functions, that is, the unknown distortion functions exist in its asymptotic variance. It is remarkable that the conditional mean calibration estimator (Zhang, Feng, and Zhou 2014), the conditional variance calibration estimator (Zhang and Lin 2023), the conditional absolute logarithmic calibration estimator (Zhang, Xu, and Wei 2023) and the parametric calibration estimator (Zhang 2022) are all asymptotic efficient since the asymptotic variances of these estimators are the same with the moment estimator of ρ (Ferguson, 1996 Section 8), and the distortion functions do not exist in the asymptotic variances. No matter which the existing estimators has been used to estimate ρ in literature, there is a fundamental assumption of that the confounding variable U is independent of unknown observed (X, Y) . As claimed in Zhang, Lin, and Feng (2020), the question to be answered is whether or not the independence condition helps to define the interpretable unobserved variables of interest from their observable counter parts. By far, the independence condition is commonly assumed in the literature because of the identifiability problems for the unknown distortion functions and unobserved variables. If the independence condition fails, other assumptions on the models should be considered instead. Zhang, Lin, and Zhou (2024) opened a new chapter for the multiplicative distortion models and attempted to solve this problem by considering the multiplicative distortions linear regression models without the independence condition. Zhang, Lin, and Zhou (2024) transformed the multiplicative distortions linear regression models into a varying coefficient model and introduced a new identifiability condition of distortion functions. This method can be applicable to the un-independence between the confounding variable and the unobserved response variable or covariates.

In this paper, we proposed a new estimation method for Pearson correlation coefficient without the independence condition between the confounding variable U and the unobserved variables. The first new idea is about new identifiability conditions for the distortion functions. These identifiability conditions are different from those in literature such as the “no distortion average” conditions by assuming known expectations of distortion functions (Şentürk and Müller 2005b; Zhang, Li, and Feng 2015; Zhang 2019; Zhang and Cui 2021). Under the new identifiability conditions, we proposed a varying coefficient moment based estimator for the estimation of ρ in model (1.1), and studied the asymptotic results of the estimator. Under the new identifiability conditions, the independence condition between the confounding variable and unobserved variables is not required for estimating the Pearson correlation coefficient ρ . Monte Carlo simulation experiments are carried out to examine these comparisons. We also analyzed a dataset created from a higher education institution as an illustration.

The paper is structured as follows. In [Sec. 2](#), we propose the new identifiability conditions, the varying coefficient moment based estimator and derive related asymptotic results. In [Sec. 3](#), we report the results of simulation studies. In [Sec. 4](#), a real data is analyzed. All technical proofs of the asymptotic results are given in Appendix.

2. Estimation method and asymptotic results

2.1. New identifiability conditions

We assume the following regression relations between the unobserved variables Y and X :

$$Y = \alpha + \beta X + \varepsilon, \quad X = \delta + \gamma Y + \epsilon, \quad (2.1)$$

where (ε, ϵ) are the model errors such that $E(\varepsilon) = E(\epsilon) = 0$, $E(\varepsilon^2) < \infty$ and $E(\epsilon^2) < \infty$. Under the conditions $E(\varepsilon|X, U) = E(\epsilon|Y, U) = 0$, the unknown coefficients (β, γ) in the models [\(2.1\)](#) are connected with the Pearson correlation coefficient $\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$:

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \rho \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}, \quad (2.2)$$

$$\gamma = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} = \rho \sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}}. \quad (2.3)$$

Based on [\(2.2\)](#) and [\(2.3\)](#), Şentürk and Müller (2005a) proposed to use the relation $\rho^2 = \beta\gamma$ to estimate ρ . Different from the estimation methods in Şentürk and Müller (2005a), we propose the moment-based varying coefficient estimator of ρ under the following [Assumption A](#):

Assumption A: $\phi(u)\psi(u) \neq 0$, for all $u \in [\mathcal{U}_L, \mathcal{U}_R]$, where $[\mathcal{U}_L, \mathcal{U}_R]$ ($\mathcal{U}_L < \mathcal{U}_R$) denotes the compact support of U . Moreover, $E\left[\frac{\phi(U)}{\psi(U)}\right] > 0$.

Together with models [\(1.1\)–\(2.1\)](#), we have the following varying coefficient models

$$\tilde{Y} = \alpha\phi(U) + \beta\frac{\phi(U)}{\psi(U)}\tilde{X} + \varepsilon\phi(U) = \zeta(U) + \eta(U)\tilde{X} + \tilde{\varepsilon}, \quad (2.4)$$

$$\tilde{X} = \delta\psi(U) + \gamma\frac{\psi(U)}{\phi(U)}\tilde{Y} + \epsilon\psi(U) = \xi(U) + \lambda(U)\tilde{Y} + \tilde{\epsilon}. \quad (2.5)$$

Motivated from these relations [\(2.2\)–\(2.5\)](#), under [Assumption A](#), we have

$$\rho^2 = E[\eta(U)\lambda(U)], \quad \text{sgn}(\rho) = \text{sgn}(\beta) = \text{sgn}(E[\eta(U)]). \quad (2.6)$$

At the population level, the sign of ρ is the same with $E[\eta(U)]$, we can estimate ρ^2 at first through the expectation $E[\eta(U)\lambda(U)]$. It is noted that the [Assumption A](#) only requires that the sign of the expectation $E\left[\frac{\phi(U)}{\psi(U)}\right]$ is positive. This is different from the identifiability conditions in the literature, such as the “no average distortions” $E[\phi(U)] = E[\psi(U)] = 1$ (Li, Zhang, and Feng 2016; Zhao and Xie 2018; Feng, Gai, and Zhang 2019; Zhang, 2019 2021, 2023; Zhang, Lin, and Li 2019; Zhang, Gai, et al. 2020; Zhang, Lin, and Feng 2020; Zhang, Zhu, et al. 2020; Zhang, Li, and Yang 2022; Zhang, Chen, and Wei 2023), the “logarithmic absolute mean conditions” $E[\log(\phi(U))] = E[\log(\psi(U))] = 0$ (Zhang, Yang, et al. 2020; Zhang, Yang, and Li 2020; Zhang and Cui 2021; Zhang, Xu, and Wei 2023), and the “quotient mean condition” $E\left[\frac{\phi(U)}{\psi(U)}\right] = 1$ (Zhang, Lin, and Zhou 2024). These conditions for unknown distortion functions are analogous

to the classical additive measurement errors: $E(e) = 0$ for $W = X + e$, where W is error-prone and X is error-free (Li, Zhang, and Feng 2016; Tomaya and de Castro 2018; de Castro and Vidal 2019; Yang, Tong, and Li 2019). The relations in (2.6) and the Assumption A are more mild since we did not require the strong condition: the confounding variable U is independent of (X, Y) . For example, the independent condition and the “no average conditions” are used to obtain $E(\tilde{Y}|U = u) = \phi(u)E(Y) = \phi(u)E(\tilde{Y})$. The unknown distortion function $\phi(u)$ is estimated by the Nadaraya-Watson estimators (Nadaraya 1964; Watson 1964) or local linear estimators (Fan and Gijbels 1996) in the distortion measurement errors literature, and then, the unknown variable is calibrated such as $\hat{Y} = \frac{\tilde{Y}}{\hat{\phi}(u)}$ for further estimation or statistical inferences. Our estimation method related in relations (2.6) does not use the independent condition, and this is different from all the methods in the distortion measurement errors models.

The varying coefficient models (2.4)–(2.5) have been well studied *via* various estimation methods, see for example, Fan and Huang (2005), Fan and Zhang (1999), Feng, Hu, and Xue (2016), Feng, Li, et al. (2021), Feng, Tian et al. (2021), Feng and Xue (2016), Hastie and Tibshirani (1993), Lian (2015), Lian, Lai, and Liang (2013), Yang, Li, and Peng (2014), Zhang et al. (2018), Zhang and Peng (2010), and Zhao and Lian (2016). In this paper, we adopt the local linear regression technique (Fan and Huang 2005) to estimate $(\zeta(u), \eta(u))$ in model (2.4) and $(\xi(u), \lambda(u))$ in model (2.5). Suppose that we have an ‘i.i.d’ sample $\{\tilde{Y}_i, \tilde{X}_i, U_i\}_{i=1}^n$ from model (1.1). For each u in a neighborhood of u_0 , we approximate $\zeta(u_0)$ and $\eta(u_0)$ by $\zeta(u) + \zeta'(u)(u_0 - u)$, $\eta(u) + \eta'(u)(u_0 - u)$, respectively. Then, the local estimators of $\{\zeta(u), \eta(u)\}$ and derivatives $\{\zeta'(u), \eta'(u)\}$ are obtained by minimizing the local least squares criterion (2.7) with respect to $\mathbf{a} = (a_0, a_1)^\top$ and $\mathbf{b} = (b_0, b_1)^\top$,

$$\sum_{i=1}^n \left\{ \tilde{Y}_i - \tilde{\mathbf{X}}_i^\top \mathbf{a} - \tilde{\mathbf{X}}_i^\top \mathbf{b}(U_i - u) \right\}^2 K_h(U_i - u), \quad \tilde{\mathbf{X}}_i = (1, \tilde{X}_i)^\top, \quad (2.7)$$

where, $K_h(U_i - u) = \frac{1}{h} K\left(\frac{U_i - u}{h}\right)$ with $K(\cdot)$ being a kernel density function and h being a bandwidth. A simple calculation gives that

$$\begin{aligned} & \left((\hat{\zeta}(u), \hat{\eta}(u))^\top, (h\hat{\zeta}'(u), h\hat{\eta}'(u))^\top \right)^\top \\ & \stackrel{\text{def}}{=} (\hat{\mathbf{a}}^\top, h\hat{\mathbf{b}}^\top)^\top = (\tilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,u} \tilde{\mathbf{D}}_{n,u})^{-1} \tilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,u} \tilde{\mathbf{V}}_n, \end{aligned} \quad (2.8)$$

where $\tilde{\mathbf{V}}_n = (\tilde{Y}_1, \dots, \tilde{Y}_n)^\top$, $\mathbf{W}_{n,u} = \text{diag}(K_h(U_1 - u), \dots, K_h(U_n - u))$, and

$$\tilde{\mathbf{D}}_{n,u} = \begin{pmatrix} \tilde{\mathbf{X}}_1^\top & \frac{U_1 - u}{h} \tilde{\mathbf{X}}_1^\top \\ \vdots & \vdots \\ \tilde{\mathbf{X}}_n^\top & \frac{U_n - u}{h} \tilde{\mathbf{X}}_n^\top \end{pmatrix},$$

For each u in a neighborhood of u_0 , we approximate $\xi(u_0)$ and $\lambda(u_0)$ by $\xi(u) + \xi'(u)(u_0 - u)$, $\lambda(u) + \lambda'(u)(u_0 - u)$, respectively. Then, the local estimators of $\{\xi(u), \lambda(u)\}$ and derivatives $\{\xi'(u), \lambda'(u)\}$ are obtained by minimizing the local least squares criterion (2.9) with respect to $\mathbf{c} = (c_0, c_1)^\top$ and $\mathbf{d} = (d_0, d_1)^\top$,

$$\sum_{i=1}^n \left\{ \tilde{X}_i - \tilde{\mathbf{Y}}_i^\top \mathbf{c} - \tilde{\mathbf{Y}}_i^\top \mathbf{d}(U_i - u) \right\}^2 K_h(U_i - u), \quad \tilde{\mathbf{Y}}_i = (1, \tilde{Y}_i)^\top. \quad (2.9)$$

A simple calculation gives that

$$\begin{aligned} & \left((\hat{\xi}(u), \hat{\lambda}(u))^{\top}, (h\hat{\xi}'(u), h\hat{\lambda}'(u))^{\top} \right)^{\top} \\ & \stackrel{\text{def}}{=} (\hat{\mathbf{c}}^{\top}, h\hat{\mathbf{d}}^{\top})^{\top} = (\tilde{\mathbf{G}}_{n,u}^{\top} \mathbf{W}_{n,u} \tilde{\mathbf{G}}_{n,u})^{-1} \tilde{\mathbf{G}}_{n,u}^{\top} \mathbf{W}_{n,u} \tilde{\mathbf{Q}}_n, \end{aligned} \quad (2.10)$$

where $\tilde{\mathbf{Q}}_n = (\tilde{X}_1, \dots, \tilde{X}_n)^{\top}$, and

$$\tilde{\mathbf{G}}_{n,u} = \begin{pmatrix} \tilde{\mathbf{Y}}_1^{\top} & \frac{U_1 - u}{h} \tilde{\mathbf{Y}}_1^{\top} \\ \vdots & \vdots \\ \tilde{\mathbf{Y}}_n^{\top} & \frac{U_n - u}{h} \tilde{\mathbf{Y}}_n^{\top} \end{pmatrix},$$

Using (2.6) and estimators (2.8) and (2.10), the estimator of ρ is obtained as

$$\hat{\rho} = \text{sgn} \left(\frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i) \right) \left[\left(\frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i) \hat{\lambda}(U_i) \right)^2 \right]^{1/4}. \quad (2.11)$$

It is noted that the estimator $\hat{\rho}$ in (2.11) is different from the binning estimator proposed in Şentürk and Müller (2005a). Şentürk and Müller (2005a) assumed the identifiability conditions $E[\phi(U)] = 1, E[\psi(U)] = 1$, the non-zero mean conditions $E(X)E(Y) \neq 0$ (or $E(\tilde{X})E(\tilde{Y}) \neq 0$) and obtained the relations $\beta = \frac{E[\eta(U)\tilde{X}]}{E[\tilde{X}]}, \gamma = \frac{E[\lambda(U)\tilde{Y}]}{E[\tilde{Y}]}, \rho = \text{sgn}(\beta)\sqrt{\beta\gamma}$ at the population level. In detail, the estimators of (β, γ) are obtained as $\hat{\beta}_S = \frac{\sum_{i=1}^m L_i \hat{\eta}_{i,b} \bar{X}_{i,b}}{\sum_{i=1}^m \bar{X}_i}$ and $\hat{\gamma}_S = \frac{\sum_{i=1}^m L_i \hat{\lambda}_{i,b} \bar{Y}_{i,b}}{\sum_{i=1}^m \bar{Y}_i}$, where $\hat{\eta}_{i,b}, \hat{\lambda}_{i,b}$ are binning least squares estimators in each bin B_i , L_i is the number of $\{U_k\}_{k=1}^n$ in B_i (i.e. the number of $\{k, U_k \in B_i\}$), and $\bar{X}_{i,b} = \frac{1}{L_i} \sum_{k \in B_i} \tilde{X}_k, \bar{Y}_{i,b} = \frac{1}{L_i} \sum_{k \in B_i} \tilde{Y}_k$. The binning estimator $\hat{\rho}_b$ of ρ is defined as $\hat{\rho}_b = \text{sgn}(\hat{\beta}_S) \sqrt{\hat{\beta}_S \hat{\gamma}_S I\{\text{sgn}(\hat{\beta}_S) = \text{sgn}(\hat{\gamma}_S)\}}, I\{\cdot\}$ is the indicator function. Both estimators $\hat{\rho}$ and $\hat{\rho}_b$ need the sign function but the newly proposed estimator $\hat{\rho}$ requires less restriction conditions used in $\hat{\rho}_b$. Neither identifiability conditions $E[\phi(U)] = 1, E[\psi(U)] = 1$ nor the non-zero mean conditions $E(X)E(Y) \neq 0$ are imposed to obtain estimator $\hat{\rho}$. Moreover, the binning estimator $\hat{\rho}_b$ also need the independence condition between U and (Y, X) , while the estimator $\hat{\rho}$ does not need this independence condition either. In conclusions, the estimator $\hat{\rho}$ can be used in a general setting when one of the identifiability conditions, the non-zero mean conditions and the independence conditions do not hold in practice.

We define $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^{\top}$ for any matrix or vector \mathbf{A} . We now list the conditions needed in the following theorems.

- (C1) The distortion functions $\phi(u) \neq 0$ and $\psi(u) \neq 0$ for all $u \in [\mathcal{U}_L, \mathcal{U}_R]$. Moreover, the distortion functions $\phi(u), \psi(u)$ have three continuous derivatives. 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 $[\mathcal{U}_L, \mathcal{U}_R]$.
- (C2) For some $s \geq 4, E(|Y|^s) < \infty, E(|X|^s) < \infty$. The matrices $\mathbf{\Sigma}(u)$ and $\mathbf{\Omega}(u)$ defined in Theorem 1 are positive-definite for all $u \in [\mathcal{U}_L, \mathcal{U}_R]$.
- (C3) The kernel function $K(\cdot)$ is a symmetric bounded density function supported on $[-A, A]$ satisfying a Lipschitz condition. $K(\cdot)$ also has second-order continuous bounded derivatives, satisfying $\mu_2 = \int_{-A}^A s^2 K(s) ds \neq 0$ and $\mu_{K,2} = \int_{-A}^A K^2(s) ds \neq 0$.
- (C4) As $n \rightarrow \infty$, the bandwidth h satisfies $\frac{\log^2 n}{nh^2} \rightarrow 0$ and $nh^4 \rightarrow 0$.

These are mild conditions that are satisfied in most practical situations. Condition (C1) is recently used in the study of distortion measurement error models, see Delaigle, Hall, and Zhou (2016), Gai, Zhang, and Zhou (2024), Zhang (2023); Zhang et al. (2023), Zhao and Xie (2018), and Zhong et al. (2023). The bounded assumption for the support of U entails no loss of generality for practical usage. The bounded assumption of U is also used in the Cadcor estimator of the correlation coefficient (Şentürk and Müller 2005a). Condition (C2) is used in the proof of theorems. Condition (C3) is the usual condition for the kernel function $K(\cdot)$. The Epanechnikov kernel satisfies this condition. Condition (C4) is the condition for the bandwidth h in the nonparametric kernel smoothing.

Let $\hat{\boldsymbol{\kappa}}(u) = (\hat{\zeta}(u), \hat{\eta}(u))^\top$, $\boldsymbol{\kappa}(u) = (\zeta(u), \eta(u))^\top$, $\boldsymbol{\kappa}''(u) = (\zeta''(u), \eta''(u))^\top$, $\boldsymbol{\psi}(u) = \text{diag}(1, \psi(u))$, $\hat{\boldsymbol{\omega}}(u) = (\hat{\xi}(u), \hat{\lambda}(u))^\top$, $\boldsymbol{\omega}(u) = (\xi(u), \lambda(u))^\top$, $\boldsymbol{\omega}''(u) = (\xi''(u), \lambda''(u))^\top$, $\boldsymbol{\phi}(u) = \text{diag}(1, \phi(u))$, and

$$\begin{aligned}\Lambda(u) &= E(\varepsilon^2 \mathbf{X}^{\otimes 2} | U = u), \quad \Sigma(u) = E(\mathbf{X}^{\otimes 2} | U = u), \quad \mathbf{X} = (1, X)^\top, \\ \Delta(u) &= E(\varepsilon^2 \mathbf{Y}^{\otimes 2} | U = u), \quad \Omega(u) = E(\mathbf{Y}^{\otimes 2} | U = u), \quad \mathbf{Y} = (1, Y)^\top, \\ F(u) &= [\boldsymbol{\psi}(u)]^{-1} [\Sigma(u)]^{-1} \Lambda(u) [\Sigma(u)]^{-1} [\boldsymbol{\psi}(u)]^{-1}, \\ \boldsymbol{\Upsilon}(u) &= [\boldsymbol{\phi}(u)]^{-1} [\Omega(u)]^{-1} \Delta(u) [\Omega(u)]^{-1} [\boldsymbol{\phi}(u)]^{-1}.\end{aligned}$$

Theorem 1. Under conditions (C1)–(C3), as $n \rightarrow \infty$, $h \rightarrow 0$, $nh \rightarrow \infty$, we have

$$\begin{aligned}\sqrt{nh} \left(\hat{\boldsymbol{\kappa}}(u) - \boldsymbol{\kappa}(u) - \frac{\mu_2 h^2}{2} \boldsymbol{\kappa}''(u) \right) &\xrightarrow{L} N \left(0, \frac{\mu_{K,2} \phi^2(u)}{f_U(u)} F(u) \right), \\ \sqrt{nh} \left(\hat{\boldsymbol{\omega}}(u) - \boldsymbol{\omega}(u) - \frac{\mu_2 h^2}{2} \boldsymbol{\omega}''(u) \right) &\xrightarrow{L} N \left(0, \frac{\mu_{K,2} \psi^2(u)}{f_U(u)} \boldsymbol{\Upsilon}(u) \right).\end{aligned}$$

Let $\mathbf{e}_2 = (0, 1)^\top$, $\sigma^2 = E[\mathbf{m}^2(X, Y, U, \varepsilon, \epsilon)]$, where

$$\mathbf{m}(X, Y, U, \varepsilon, \epsilon) = \frac{\mathbf{e}_2^\top}{2} \left[\sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}} [\Sigma(U)]^{-1} \mathbf{X} \varepsilon + \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}} [\Omega(U)]^{-1} \mathbf{Y} \epsilon \right].$$

Theorem 2. Under conditions (C1)–(C4) and *Assumption A*, as $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{L} N(0, \sigma^2), \quad \rho \neq 0;$$

The asymptotic results of *Theorem 2* indicate that the proposed estimator $\hat{\rho}$ has a faster convergent rate than the root- n convergence rate when $\rho = 0$ holds. In this case, under the model assumptions $E(\varepsilon|X, U) = 0$ and $E(\epsilon|Y, U) = 0$, $\rho = 0$ entails that $\beta = \gamma = 0$ and also $\text{Cov}(X, Y) = \text{Cov}(X, \varepsilon) = \text{Cov}(Y, \epsilon) = 0$. When U is independent of (Y, X) , and $\varepsilon = Y - \alpha - \beta X$, $\epsilon = X - \delta - \gamma Y$, we have that U is independent of (ε, ϵ) and $\Sigma(U) = E(\mathbf{X}^{\otimes 2})$, $\Omega(U) = E(\mathbf{Y}^{\otimes 2})$. Define $X_{\mathcal{O}} = X - E(X)$, $Y_{\mathcal{O}} = Y - E(Y)$, we have $\varepsilon = Y_{\mathcal{O}} - \beta X_{\mathcal{O}}$, $\epsilon = X_{\mathcal{O}} - \gamma Y_{\mathcal{O}}$, and the expression $\mathbf{m}(X, Y, U, \varepsilon, \epsilon)$ becomes to

$$\begin{aligned}\mathbf{m}(X, Y, U, \varepsilon, \epsilon) &= \frac{[X - E(X)]\varepsilon + [Y - E(Y)]\epsilon}{2\sqrt{\text{Var}(X)\text{Var}(Y)}} \\ &= \frac{2X_{\mathcal{O}}Y_{\mathcal{O}} - \rho\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}X_{\mathcal{O}}^2 - \rho\sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}}Y_{\mathcal{O}}^2}{2\sqrt{\text{Var}(X_{\mathcal{O}})\text{Var}(Y_{\mathcal{O}})}},\end{aligned}$$

The asymptotic variance σ^2 in [Theorem 2](#) becomes to

$$\begin{aligned}\sigma^2 &= \frac{1}{4\text{Var}(X_O)\text{Var}(Y_O)} \text{Var} \left(2X_O Y_O - \rho \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}} X_O^2 - \rho \sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}} Y_O^2 \right) \\ &= \frac{\rho^2}{4} \left\{ \frac{\text{Var}(Y_O^2)}{\{\text{Var}(Y_O)\}^2} + 2 \frac{\text{Cov}(Y_O^2, X_O^2)}{\text{Var}(Y_O)\text{Var}(X_O)} + \frac{\text{Var}(X_O^2)}{\{\text{Var}(X_O)\}^2} \right\} + \frac{\text{Var}(Y_O X_O)}{\text{Var}(Y_O)\text{Var}(X_O)} \\ &\quad - \rho \frac{\text{Cov}(Y_O^2, Y_O X_O)}{\{\text{Var}(Y_O)\}^{3/2} \{\text{Var}(X_O)\}^{1/2}} - \rho \frac{\text{Cov}(X_O^2, X_O Y_O)}{\{\text{Var}(X_O)\}^{3/2} \{\text{Var}(Y_O)\}^{1/2}}.\end{aligned}\quad (2.12)$$

It is worth noting that the estimator $\hat{\rho}$ is asymptotical efficient when U is independent of (Y, X) . In other words, the asymptotic variance σ^2 in (2.12) is the same as the classical asymptotic variance of the sample correlation coefficient $\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ when the variables $\{X_i, Y_i\}_{i=1}^n$ are observed without multiplicative distortions (see, for example [Ferguson \(1996\)](#) [Section 8, Theorem 8]). The estimator $\hat{\rho}$ automatically eliminates the effect caused by the multiplicative distortions $\phi(U), \psi(U)$, i.e. the effect of distortion measurement errors vanishes. Similar phenomenon has been found in [Zhang \(2022\)](#), [Zhang, Feng, and Zhou \(2014\)](#), [Zhang and Lin \(2023\)](#), [Zhang, Xu, and Wei \(2023\)](#), and [Zhang, Yang, and Feng \(2024\)](#).

When (X, Y) follows a bivariate normal distribution, [Theorem 7](#) of [Székely, Rizzo, and Bakirov \(2007\)](#) showed that the relation between the distance correlation coefficient and ρ is presented as

$$\text{dCov}^2(X, Y) = \frac{\rho \arcsin(\rho) + \sqrt{1 - \rho^2} - \rho \arcsin(\rho/2) - \sqrt{4 - \rho^2} + 1}{1 + \pi/3 - \sqrt{3}}.$$

Under the bivariate normal setting, the estimator of the distance correlation coefficient is defined as

$$\widehat{\text{dCov}}^2(X, Y) = \frac{\hat{\rho} \arcsin(\hat{\rho}) + \sqrt{1 - \hat{\rho}^2} - \hat{\rho} \arcsin(\hat{\rho}/2) - \sqrt{4 - \hat{\rho}^2} + 1}{1 + \pi/3 - \sqrt{3}}.$$

Directly using the Delta theorem, we can have that $\sqrt{n}(\widehat{\text{dCov}}^2(X, Y) - \text{dCov}^2(X, Y)) \xrightarrow{L} N(0, \sigma_d^2)$, $\sigma_d^2 = \frac{[\arcsin(\rho) - \arcsin(\rho/2)]^2}{(1 + \pi/3 - \sqrt{3})^2} \sigma^2$. This asymptotic result shows that $\sigma_d^2 = 0$ when $\rho = 0$. In other words, both estimators $\widehat{\text{dCov}}^2(X, Y), \hat{\rho}$ are all asymptotically superior with a fast convergent rate $o(n^{-1/2})$ when (X, Y) are jointly distributed as bivariate normal with $\rho = 0$ ((X, Y) are independent of each other).

To construct confidence intervals of ρ , we estimate asymptotic covariance σ^2 in [Theorem 2](#) as follows. Define that

$$\begin{aligned}\hat{\epsilon}_i &= \tilde{Y}_i - \tilde{\mathbf{X}}_i^\top \hat{\boldsymbol{\kappa}}(U_i), \quad \hat{\epsilon}_i = \tilde{X}_i - \tilde{\mathbf{Y}}_i^\top \hat{\boldsymbol{\omega}}(U_i), \quad i = 1, \dots, n, \\ \hat{\mathbf{S}}(u) &= \frac{\sum_{i=1}^n K_h(U_i - u) \tilde{\mathbf{X}}_i^{\otimes 2}}{\sum_{i=1}^n K_h(U_i - u)}, \quad \hat{\mathbf{Q}}(u) = \frac{\sum_{i=1}^n K_h(U_i - u) \tilde{\mathbf{Y}}_i^{\otimes 2}}{\sum_{i=1}^n K_h(U_i - u)}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_2^\top \left\{ \hat{\lambda}(U_i) [\hat{\mathbf{S}}(U_i)]^{-1} \tilde{\mathbf{X}}_i \hat{\epsilon}_i + \hat{\eta}(U_i) [\hat{\mathbf{Q}}(U_i)]^{-1} \tilde{\mathbf{Y}}_i \hat{\epsilon}_i \right\}^{\otimes 2} \mathbf{e}_2.\end{aligned}$$

The $(1 - \alpha) \times 100\%$ ($0 < \alpha < 1$) confidence interval for $\rho \neq 0$ is

$$\left(\hat{\rho} - \sqrt{\frac{\hat{\sigma}_d^2}{n}} z_{\alpha/2}, \quad \hat{\rho} + \sqrt{\frac{\hat{\sigma}_d^2}{n}} z_{\alpha/2} \right),$$

where, $z_{\alpha/2}$ is the quantile satisfying $P(N(0, 1) \geq z_{\alpha/2}) = \alpha/2$. When (X, Y) follows a bivariate normal distribution, the $(1 - \alpha) \times 100\%$ confidence interval for $\text{dCov}^2(X, Y)$ ($\rho \neq 0$) is

$$\left(\widehat{\text{dCov}^2}(X, Y) - \sqrt{\frac{\hat{\sigma}_d^2}{n}} z_{\alpha/2}, \quad \widehat{\text{dCov}^2}(X, Y) + \sqrt{\frac{\hat{\sigma}_d^2}{n}} z_{\alpha/2} \right),$$

where, $\hat{\sigma}_d^2 = \frac{[\arcsin(\hat{\rho}) - \arcsin(\hat{\rho}/2)]^2}{(1 + \pi/3 - \sqrt{3})^2} \hat{\sigma}^2$.

3. Implementation

Simulation studies are made in this section to show the performance of our proposed methods. The bandwidth h and the kernel function $K(t)$ do not have impact on the asymptotic variance of the proposed estimator of ρ . Then, we use the Epanechnikov kernel $K(t) = 0.75(1 - t^2)I\{|t| < 1\}$ here and $h = \hat{\sigma}_U n^{-1/3}$, and $\hat{\sigma}_U = \sqrt{\frac{1}{n} \sum_{i=1}^n (U_i - \bar{U})^2}$, $\bar{U} = \frac{1}{n} \sum_{i=1}^n U_i$. The rule of thumb bandwidth selection h is fairly effective and easy to apply in practice. Since the optimal bandwidth for h cannot be obtained because under-smoothing ($nh^4 \rightarrow 0$ in Condition (C4)) for the non-parametric estimates is necessary. Our numerical experience suggests that the numerical results in this section were stable when we shifted several values around this data-driven bandwidth. In the following, the first three examples are conducted under the independence condition between U and unobserved (X, Y) . We conduct a simulation in Example 4 when the independence condition does not hold.

Example 1. In this simulation, the true unobserved variables (X, Y) were generated from a bivariate normal distribution with mean vector $(-1, 1)$, $\text{Var}(X) = \text{Var}(Y) = 1$, and the Pearson correlation coefficient ρ is set to be $-0.9, -0.5, 0, 0.75$. The confounding variable U is generated from Uniform $(0, 1)$, independent with (X, Y) . The distortion functions are chosen as $\phi(u) = 1.25 - 3(u - 0.5)^2$, $\psi(u) = 1 + 0.5 \cos(2\pi u)$. 1000 realizations are generated and sample sizes are $n = 300, n = 500, n = 1000$ and $n = 2000$, respectively. In this example, the expectations of (X, Y) are nonzero and U is independent of (X, Y) , so the conditional mean calibration (Cui et al. 2009; Zhang, Feng, and Zhou 2014; Zhang 2022) based estimator $\hat{\rho}_C$, the conditional absolute mean calibration (Delaigle, Hall, and Zhou 2016; Feng, Gai, and Zhang 2019; Zhang, Gai, et al. 2020; Zhang, Lin, and Feng 2020; Zhang, Zhu, et al. 2020; Zhang 2021) based estimator $\hat{\rho}_A$, the conditional variance calibration (Zhang 2019; Zhang, Li, and Yang, 2022; Zhang, Lin, and Li 2019) based estimator $\hat{\rho}_V$ can be used here to estimate the Pearson correlation coefficient ρ . The true

estimator $\hat{\rho}_T = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ (using simulated dataset

$\{X_i, Y_i\}_{i=1}^n$), and the naive estimator $\hat{\rho}_N = \frac{\sum_{i=1}^n (\tilde{X}_i - \bar{\tilde{X}})(\tilde{Y}_i - \bar{\tilde{Y}})}{\sqrt{\sum_{i=1}^n (\tilde{X}_i - \bar{\tilde{X}})^2 \sum_{i=1}^n (\tilde{Y}_i - \bar{\tilde{Y}})^2}}$, $\bar{\tilde{X}} = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i$, $\bar{\tilde{Y}} = \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i$,

are conducted in this example to make comparisons among these estimators.

We report the mean, standard errors and mean squared errors (MSE) for the estimators $\hat{\rho}_T, \hat{\rho}, \hat{\rho}_C, \hat{\rho}_A, \hat{\rho}_V$ and $\hat{\rho}_N$. In Table 1, it is seen that all the mean values for the proposed estimator $\hat{\rho}$ are close to the true value ρ , and the values of MSE decrease to zero as the sample size n increases. When $\rho = -0.9$, the $\hat{\rho}_C, \hat{\rho}_A$ perform better than $\hat{\rho}$, while $\hat{\rho}_V$ is the worst. When $\rho = -0.5$, $\hat{\rho}$ is the best and its MSE values are slightly smaller than those of $(\hat{\rho}_C, \hat{\rho}_A)$, $\hat{\rho}_V$ is still

Table 1. Simulation results of mean (M), Standard Error (SD) and mean Squared Error (MSE) for the true estimator $\hat{\rho}_T$, the proposed estimator $\hat{\rho}$, the conditional mean calibrated estimator $\hat{\rho}_C$, the conditional absolute mean calibrated estimator $\hat{\rho}_A$, the conditional variance calibrated estimator $\hat{\rho}_V$, and the naive estimator $\hat{\rho}_N$.

	$n = 300$			$n = 500$			$n = 1000$			$n = 2000$		
	M	SD	MSE	M	SD	MSE	M	SD	MSE	M	SD	MSE
$\rho = -0.9$												
$\hat{\rho}_T$	-0.8993	0.0113	0.1293	-0.8996	0.0083	0.0697	-0.8999	0.0058	0.0339	-0.8999	0.0042	0.0183
$\hat{\rho}$	-0.8901	0.0122	0.2473	-0.8923	0.0089	0.1387	-0.8947	0.0060	0.0639	-0.8962	0.0043	0.0329
$\hat{\rho}_C$	-0.8940	0.0144	0.2440	-0.8963	0.0091	0.0959	-0.8977	0.0061	0.0429	-0.8981	0.0043	0.0221
$\hat{\rho}_A$	-0.8962	0.0118	0.1537	-0.8972	0.0087	0.0831	-0.8981	0.0059	0.0393	-0.8983	0.0043	0.0212
$\hat{\rho}_V$	-0.8553	0.0868	9.5379	-0.8693	0.0671	5.4508	-0.8861	0.0439	2.1284	-0.8940	0.0190	0.3979
$\hat{\rho}_N$	-0.6454	0.0297	65.6662	-0.6439	0.0229	66.0873	-0.6448	0.0162	65.3820	-0.6445	0.0116	65.4091
$\rho = -0.5$												
$\hat{\rho}_T$	-0.4992	0.0430	1.8510	-0.5006	0.0329	1.0818	-0.5006	0.0250	0.6259	-0.4997	0.0160	0.2580
$\hat{\rho}$	-0.5138	0.0388	1.6975	-0.5090	0.0312	1.0534	-0.5047	0.0241	0.6054	-0.5014	0.0157	0.2502
$\hat{\rho}_C$	-0.4935	0.0440	1.9790	-0.4968	0.0334	1.1257	-0.4986	0.0251	0.6325	-0.4984	0.0160	0.2587
$\hat{\rho}_A$	-0.4958	0.0435	1.9093	-0.4981	0.0330	1.0942	-0.4991	0.0250	0.6284	-0.4986	0.0160	0.2585
$\hat{\rho}_V$	-0.4710	0.0641	4.9448	-0.4782	0.0582	3.8636	-0.4895	0.0398	1.6997	-0.4956	0.0189	0.3769
$\hat{\rho}_N$	-0.3287	0.0458	31.4491	-0.3290	0.0341	30.4196	-0.3298	0.0253	29.6091	-0.3293	0.0175	29.4434
$\rho = 0$												
$\hat{\rho}_T$	-0.0000	0.0582	3.3895	0.0014	0.0457	2.0916	-0.0002	0.0303	0.9236	-0.0004	0.0222	0.4969
$\hat{\rho}$	0.0606	0.0532	6.5112	0.0645	0.0405	5.8188	0.0606	0.0278	4.4547	0.0599	0.0201	4.0033
$\hat{\rho}_C$	-0.0001	0.0586	3.4328	0.0014	0.0458	2.0978	-0.0002	0.0303	0.9185	-0.0003	0.0223	0.5003
$\hat{\rho}_A$	0.0002	0.0586	3.4179	0.0013	0.0456	2.0790	-0.0003	0.0303	0.9208	-0.0004	0.0223	0.4977
$\hat{\rho}_V$	0.0011	0.0575	3.3088	0.0021	0.0452	2.0450	0.0004	0.0305	0.9345	0.0002	0.0221	0.4925
$\hat{\rho}_N$	0.0659	0.0511	6.9594	0.0652	0.0398	5.8301	0.0642	0.0262	4.8128	0.0640	0.0197	4.4979
$\rho = 0.75$												
$\hat{\rho}_T$	0.7495	0.0247	0.6107	0.7499	0.0197	0.3895	0.7492	0.0138	0.1923	0.7498	0.0095	0.0903
$\hat{\rho}$	0.7454	0.0248	0.6370	0.7467	0.0199	0.4067	0.7468	0.0139	0.2037	0.7480	0.0095	0.0944
$\hat{\rho}_C$	0.7289	0.0433	2.3179	0.7383	0.0218	0.6125	0.7431	0.0145	0.2590	0.7464	0.0097	0.1070
$\hat{\rho}_A$	0.7406	0.0258	0.7553	0.7439	0.0202	0.4466	0.7457	0.0140	0.2162	0.7475	0.0095	0.0979
$\hat{\rho}_V$	0.7049	0.0746	7.6030	0.7156	0.0745	6.7351	0.7350	0.0387	1.7278	0.7447	0.0136	0.2138
$\hat{\rho}_N$	0.6568	0.0261	9.3720	0.6567	0.0212	9.1529	0.6552	0.0145	9.1976	0.6559	0.0103	8.9604

MSE is in the scale of 10^{-3} .

the worst. When $\rho = 0$, $\hat{\rho}_C$, $\hat{\rho}_A$ and $\hat{\rho}_V$ perform well but $\hat{\rho}$ is the worst. This is not surprised since the estimator $\hat{\rho}$ is obtained through the estimators of varying coefficient based estimators $\hat{\eta}(U_i)$, $\hat{\lambda}(U_i)$. In this case, (X, Y) are independent of each other when $\rho = 0$, and the estimators $\hat{\eta}(U_i)$, $\hat{\lambda}(U_i)$ may perform not well. When $\rho = 0.75$, $\hat{\rho}$ is the best and its MSE values are slightly smaller than those of $(\hat{\rho}_C, \hat{\rho}_A)$, $\hat{\rho}_V$ is still the worst. In general, the values of MSE for $\hat{\rho}_C$, $\hat{\rho}_A$ are more stable than the proposed estimator $\hat{\rho}$, but the MSE values of $\hat{\rho}$ can be almost the same with those of $\hat{\rho}_T$, $\hat{\rho}_C$, $\hat{\rho}_A$. As we indicated in Theorem 2, the proposed estimator $\hat{\rho}$ can be asymptotically efficient when U is independent of (X, Y) , and $(\hat{\rho}_C, \hat{\rho}_A)$ are also asymptotically efficient (Zhang, Feng, and Zhou 2014; Zhang 2022). As indicated in Zhang and Lin (2023), although the conditional variance calibrated estimator $\hat{\rho}_V$ are shown asymptotically efficient, but it needs much more large sample to achieve this especially when $\rho = -0.9$. Similar phenomenon also exists for $\hat{\rho}$ when $\rho = 0$, and the conditional variance calibrated estimator $\hat{\rho}_V$ and $\hat{\rho}$ (for $\rho = 0$) may perform not stable sometimes for the finite sample size. For the naive estimator $\hat{\rho}_N$, its absolute mean value of biases are (0.25, 0.17, 0.06, 0.09) when $\rho = -0.9$, $\rho = -0.5$, $\rho = 0$ and $\rho = 0.75$, respectively. These biases do not decrease to zero even when the sample size 2000, and the values of MSE do not decrease to zero either. It implies that the consequence of ignoring multiplicative distortion functions definitely result in wrong estimators with large biases.

Example 2. In this simulation, the confounding variable U , the distortion functions $\phi(u)$, $\psi(u)$ and the Pearson correlation coefficient ρ are the same with Example 1. The true unobserved variables (X, Y) were generated from a bivariate normal distribution with mean vector $(0, 0)$, $\text{Var}(X) = \text{Var}(Y) = 1$, independent of U . 1000 realizations are generated and sample sizes are

$n = 300$, $n = 500$, $n = 1000$ and $n = 2000$, respectively. It is noted that $E(Y) = E(X) = 0$, so the conditional mean calibrated estimator $\hat{\rho}_C$ is not workable in this example.

In Table 2, we report the mean, standard errors and mean squared errors (MSE) for the estimators $\hat{\rho}_T, \hat{\rho}, \hat{\rho}_A, \hat{\rho}_V$ and $\hat{\rho}_N$. In this table, all the mean values for the proposed estimator $\hat{\rho}$ are close to the true value ρ , and the values of MSE decrease to zero as the sample size n increases. In general, $\hat{\rho}_A$ performs the best, $\hat{\rho}$ is the second best, and $\hat{\rho}_V$ is not better than $\hat{\rho}_A$ and $\hat{\rho}$. When the sample size n is 2000, the MSE values of $\hat{\rho}_A, \hat{\rho}$ are close to those of $\hat{\rho}_T$, and the conditional variance calibrated estimator $\hat{\rho}_V$ also has a better performance when $\rho \neq -0.9$. When $\rho = 0$, the estimator $\hat{\rho}$ performs well and has much smaller MSE values than those in Table 1. For the naive estimator $\hat{\rho}_N$, it works well when $\rho = 0$. The bivariate normal distribution of (X, Y) with $\rho = 0$ entails that (X, Y) is independent of each other. In this case, $\frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{Y}_i \xrightarrow{P} E(\tilde{X}\tilde{Y}) = E[\phi(U)\psi(U)]E(X)E(Y) = 0, \bar{X} \xrightarrow{P} E[\psi(U)]E(X) = 0, \bar{Y} \xrightarrow{P} E[\phi(U)]E(Y) = 0, \bar{X} \bar{Y} \xrightarrow{P} 0$, and $\frac{1}{n} \sum_{i=1}^n (\tilde{X}_i - \bar{X})(\tilde{Y}_i - \bar{Y}) \xrightarrow{P} 0$, and the estimator $\hat{\rho}_N$ can be used to estimate $\rho = 0$ in this example.

Example 3. In this simulation, the confounding variable U and the Pearson correlation coefficient ρ are the same with Example 1. The distortion functions are considered as $\phi(u) = \exp(3u - 1.5), \psi(u) = \exp(0.5 \sin(2\pi u))$. The true unobserved variables (X, Y) were generated from a bivariate normal distribution with mean vector $(0, 0)$, $\text{Var}(X) = \text{Var}(Y) = 1$, independent of U . 1000 realizations are generated and sample size are $n = 300, n = 500, n = 1000$ and $n = 2000$, respectively. In this example, $E[\log(\phi(U))] = 0, E[\log(\psi(U))] = 0$, the conditional absolute logarithmic calibration estimation procedure (Zhang, Yang, et al. 2020; Zhang, Yang, and Li 2020; Zhang and Cui 2021; Zhang, Xu, and Wei 2023) can be used here to estimate ρ , but the conditional absolute mean calibrated estimator $\hat{\rho}_A$ and conditional variance calibrated estimator $\hat{\rho}_V$ do not work due to $E[\phi(U)] = 1.4198 \neq 1, E[\psi(U)] = 1.0635 \neq 1$. We denote the conditional absolute logarithmic calibrated estimator as $\hat{\rho}_L$ in this example.

In Table 3, we report the mean, standard errors and mean squared errors (MSE) for the estimators $\hat{\rho}_T, \hat{\rho}, \hat{\rho}_L$ and $\hat{\rho}_N$. All the mean values for the proposed estimator $\hat{\rho}$ are close to the true value ρ , and the values of MSE decrease to zero as the sample size n increases. In general, $\hat{\rho}_L$ performs slightly better than $\hat{\rho}$ when $\rho \neq 0$. As the sample size n increases to 2000, the MSE values of $\hat{\rho}, \hat{\rho}_L$ are close to those of $\hat{\rho}_T$. When $\rho = 0$, the estimator $\hat{\rho}$ performs the best and has much smaller MSE values than those in Table 1. For the naive estimator $\hat{\rho}_N$, it still works well when $\rho = 0$ in this example. As (X, Y) is independent of each other with $E(X) = E(Y) = 0$, we have $\frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{Y}_i \xrightarrow{P} 0, \bar{X} \xrightarrow{P} 0, \bar{Y} \xrightarrow{P} 0, \bar{X} \bar{Y} \xrightarrow{P} 0$, and $\frac{1}{n} \sum_{i=1}^n (\tilde{X}_i - \bar{X})(\tilde{Y}_i - \bar{Y}) \xrightarrow{P} 0$, and then the estimator $\hat{\rho}_N$ can be used to estimate $\rho = 0$ in this example.

Example 4. In this example, 1000 realizations are generated and sample size are $n = 300, n = 500, n = 1000$ and $n = 2000$, respectively. The confounding variable $U \sim \text{Unif}(0, 1)$, and the variables (X, Y) are correlated with U in the following three cases:

- Case 1** $X|U \sim N(-U - 1, \sqrt{1 + U^3}), Y|U \sim N(U^2, 1 + 0.5U)$, the distortion functions $\phi(u), \psi(u)$ are $\phi(u) = \frac{\exp(u)}{\exp(1)-1}$ and $\psi(u) = 1 + 0.5 \cos(2\pi u)$ and the true value ρ is calculated as $\rho = -0.0557$;
- Case 2** $X|U \sim N(2U - 1, 1 - 0.5U^2), Y|U \sim N(3U^2 - 1, \sqrt{1 - u})$, the distortion functions $\phi(u), \psi(u)$ are the same with case 1, and the true value ρ is calculated as $\rho = 0.4287$;
- Case 3** $X|U \sim N(U^2 + 1, \sqrt{1 + U^3}), Y|U \sim N(2U - 1, \sqrt{1 - 0.5U^3})$, the distortion functions $\phi(u), \psi(u)$ are $\phi(u) = \exp(0.5 - u)$ and $\psi(u) = \exp(0.5 \cos(2\pi u))$ and the true value ρ is calculated as $\rho = 0.1310$.

Table 2. Simulation results of mean (M), Standard Error (SD) and mean Squared Error (MSE) for the true estimator $\hat{\rho}_T$, the proposed estimator $\hat{\rho}$, the conditional absolute mean calibrated estimator $\hat{\rho}_A$, the conditional variance calibrated estimator $\hat{\rho}_V$, and the naive estimator $\hat{\rho}_N$.

	n = 300			n = 500			n = 1000			n = 2000		
	M	SD	MSE	M	SD	MSE	M	SD	MSE	M	SD	MSE
$\rho = -0.9$												
$\hat{\rho}_T$	-0.9001	0.0112	0.1252	-0.9004	0.0086	0.0734	-0.8998	0.0058	0.0335	-0.8999	0.0043	0.0184
$\hat{\rho}$	-0.8940	0.0122	0.1848	-0.8958	0.0089	0.0961	-0.8968	0.0060	0.0457	-0.8977	0.0044	0.0243
$\hat{\rho}_A$	-0.8975	0.0116	0.1416	-0.8986	0.0087	0.0778	-0.8986	0.0059	0.0362	-0.8989	0.0043	0.0199
$\hat{\rho}_V$	-0.8768	0.0570	3.7813	-0.8844	0.0452	2.2823	-0.8917	0.0273	0.8154	-0.8957	0.0210	0.4610
$\hat{\rho}_N$	-0.7666	0.0165	18.0556	-0.7666	0.0135	17.9800	-0.7652	0.0095	18.2524	-0.7651	0.0065	18.2465
$\rho = -0.5$												
$\hat{\rho}_T$	-0.4980	0.0444	1.9749	-0.4984	0.0330	1.0948	-0.5000	0.0236	0.5549	-0.5001	0.0167	0.2794
$\hat{\rho}$	-0.5151	0.0401	1.8356	-0.5085	0.0311	1.0395	-0.5052	0.0230	0.5567	-0.5031	0.0163	0.2746
$\hat{\rho}_A$	-0.4951	0.0445	2.0083	-0.4963	0.0332	1.1182	-0.4989	0.0236	0.5584	-0.4994	0.0167	0.2797
$\hat{\rho}_V$	-0.4830	0.0565	3.4853	-0.4865	0.0447	2.1836	-0.4951	0.0293	0.8821	-0.4984	0.0168	0.2854
$\hat{\rho}_N$	-0.4237	0.0398	7.4074	-0.4242	0.0302	6.6494	-0.4256	0.0216	6.0023	-0.4253	0.0153	5.8138
$\rho = 0$												
$\hat{\rho}_T$	0.0009	0.0572	3.2758	-0.0017	0.0440	1.9409	-0.0008	0.0316	0.9989	-0.0001	0.0221	0.4906
$\hat{\rho}$	-0.0009	0.0537	2.8940	-0.0004	0.0441	1.9529	0.0003	0.0328	1.0782	-0.0001	0.0236	0.5573
$\hat{\rho}_A$	0.0006	0.0573	3.2885	-0.0018	0.0439	1.9341	-0.0009	0.0317	1.0049	-0.0001	0.0222	0.4913
$\hat{\rho}_V$	0.0014	0.0570	3.2548	-0.0015	0.0439	1.9250	-0.0001	0.0316	0.9962	-0.0001	0.0221	0.4902
$\hat{\rho}_N$	0.0006	0.0496	2.4632	-0.0013	0.0380	1.4486	-0.0002	0.0275	0.7545	-0.0001	0.0192	0.3675
$\rho = 0.75$												
$\hat{\rho}_T$	0.7491	0.0251	0.6313	0.7497	0.0203	0.4111	0.7502	0.0137	0.1876	0.7497	0.0100	0.0996
$\hat{\rho}$	0.7455	0.0252	0.6562	0.7468	0.0204	0.4250	0.7480	0.0138	0.1953	0.7483	0.0100	0.1022
$\hat{\rho}_A$	0.7457	0.0258	0.6861	0.7473	0.0205	0.4277	0.7488	0.0137	0.1897	0.7488	0.0100	0.1014
$\hat{\rho}_V$	0.7320	0.0428	2.1533	0.7361	0.0390	1.7124	0.7441	0.0231	0.5690	0.7470	0.0118	0.1470
$\hat{\rho}_N$	0.6387	0.0270	13.1159	0.6379	0.0216	13.0215	0.6382	0.0146	12.7189	0.6375	0.0108	12.7827

MSE is in the scale of 10^{-3} .

Table 3. Simulation results of mean (M), Standard Error (SD) and mean Squared Error (MSE) for the true estimator $\hat{\rho}_T$, the proposed estimator $\hat{\rho}$, the conditional absolute logarithmic calibrated estimator $\hat{\rho}_L$, and the naive estimator $\hat{\rho}_N$.

	n = 300			n = 500			n = 1000			n = 2000		
	M	SD	MSE	M	SD	MSE	M	SD	MSE	M	SD	MSE
$\rho = -0.9$												
$\hat{\rho}_T$	-0.8999	0.0111	0.1229	-0.9001	0.0088	0.0773	-0.8997	0.0062	0.0384	-0.8999	0.0041	0.0167
$\hat{\rho}$	-0.8944	0.0119	0.1742	-0.8960	0.0091	0.0991	-0.8970	0.0064	0.0498	-0.8980	0.0041	0.0211
$\hat{\rho}_L$	-0.8935	0.0125	0.1990	-0.8961	0.0093	0.1004	-0.8976	0.0063	0.0459	-0.8987	0.0042	0.0188
$\hat{\rho}_N$	-0.5364	0.0232	132.7398	-0.5360	0.0177	132.7979	-0.5347	0.0123	133.5967	-0.5350	0.0087	133.2493
$\rho = -0.5$												
$\hat{\rho}_T$	-0.5010	0.0443	1.9700	-0.4996	0.0340	1.1541	-0.4989	0.0237	0.5639	-0.4995	0.0172	0.2962
$\hat{\rho}$	-0.5175	0.0405	1.9498	-0.5095	0.0320	1.1146	-0.5045	0.0228	0.5411	-0.5025	0.0169	0.2936
$\hat{\rho}_L$	-0.4946	0.0449	2.0450	-0.4958	0.0344	1.2022	-0.4968	0.0238	0.5753	-0.4984	0.0172	0.3011
$\hat{\rho}_N$	-0.2987	0.0363	41.8417	-0.2980	0.0275	41.5431	-0.2968	0.0195	41.6836	-0.2971	0.0139	41.3341
$\rho = 0$												
$\hat{\rho}_T$	-0.0028	0.0587	3.4550	0.0030	0.0439	1.9324	0.0007	0.0329	1.0838	-0.0005	0.0230	0.5309
$\hat{\rho}$	0.0018	0.0440	1.9458	-0.0001	0.0375	1.4081	0.0002	0.0270	0.7301	0.0005	0.0216	0.4685
$\hat{\rho}_L$	-0.0022	0.0591	3.4951	0.0030	0.0441	1.9503	0.0006	0.0330	1.0872	-0.0004	0.0231	0.5343
$\hat{\rho}_N$	-0.0017	0.04228	1.7903	0.0018	0.0311	0.9724	0.0004	0.0241	0.5829	-0.0003	0.0166	0.2758
$\rho = 0.75$												
$\hat{\rho}_T$	0.7486	0.0259	0.6762	0.7505	0.0193	0.3713	0.7497	0.0138	0.1917	0.7496	0.0097	0.0943
$\hat{\rho}$	0.7459	0.0262	0.7061	0.7481	0.0195	0.3839	0.7478	0.0138	0.1964	0.7483	0.0098	0.0988
$\hat{\rho}_L$	0.7410	0.0269	0.8054	0.7459	0.0199	0.4133	0.7472	0.0139	0.2030	0.7484	0.0097	0.0977
$\hat{\rho}_N$	0.4472	0.0278	92.4546	0.4469	0.0222	92.3714	0.4460	0.0152	92.6462	0.4454	0.0106	92.8866

MSE is in the scale of 10^{-3} .

In case 1, the $E(X) = -\frac{3}{2}, E(Y) = \frac{1}{3}, E[\phi(U)] = E[\psi(U)] = 1$ but (X, Y) is not independent of U . The simulation results in Table 4 show that the proposed estimator $\hat{\rho}$ works well in case 1, and the mean values of $\hat{\rho}$ are close to the true estimator $\hat{\rho}_T$. However, the estimators $\hat{\rho}_C, \hat{\rho}_A$ and $\hat{\rho}_V$ fail to estimate ρ as the their mean values of absolute biases are around 0.0560, 0.0509, 0.0346.

Table 4. Simulation results of mean (M), Standard Error (SD) and mean Squared Error (MSE) for the true estimator $\hat{\rho}_T$, the proposed estimator $\hat{\rho}$, the conditional mean calibrated estimator $\hat{\rho}_C$, the conditional absolute mean calibrated estimator $\hat{\rho}_A$, the conditional variance calibrated estimator $\hat{\rho}_V$, and the naive estimator $\hat{\rho}_N$.

	$n = 300$			$n = 500$			$n = 1000$			$n = 2000$		
	M	SD	MSE	M	SD	MSE	M	SD	MSE	M	SD	MSE
Case 1: $\rho = -0.0557$												
$\hat{\rho}_T$	-0.0548	0.0616	3.7995	-0.0552	0.0464	2.1590	-0.0573	0.0329	1.0901	-0.0559	0.0242	0.5857
$\hat{\rho}$	-0.0561	0.0834	6.9651	-0.0557	0.0645	4.1649	-0.0554	0.0332	1.1029	-0.0554	0.0206	0.4245
$\hat{\rho}_C$	0.0016	0.0660	7.2775	0.0001	0.0522	5.8339	0.0001	0.0367	4.4563	0.0003	0.0264	3.8397
$\hat{\rho}_A$	-0.0038	0.0605	6.3570	-0.0036	0.0447	4.7165	-0.0065	0.0324	3.4685	-0.0048	0.0228	3.1100
$\hat{\rho}_V$	-0.0202	0.0619	5.0942	-0.0199	0.0464	3.4340	-0.0229	0.0330	2.1689	-0.0211	0.0236	1.7537
$\hat{\rho}_N$	-0.1185	0.0683	8.6117	-0.1188	0.0541	6.9185	-0.1197	0.0375	5.5056	-0.1178	0.0281	4.6531
Case 2: $\rho = 0.4287$												
$\hat{\rho}_T$	0.4278	0.0448	2.0079	0.4272	0.0369	1.3644	0.4288	0.0244	0.5964	0.4274	0.0173	0.3019
$\hat{\rho}$	0.4272	0.0381	1.4532	0.4258	0.0312	0.9796	0.4279	0.0209	0.4376	0.4257	0.0149	0.2282
$\hat{\rho}_C$	-0.0039	0.0899	194.6028	-0.0007	0.0448	185.7130	0.0004	0.0149	182.9243	-0.0001	0.0054	183.2398
$\hat{\rho}_A$	0.3400	0.0493	10.1566	0.3404	0.0403	9.2791	0.3418	0.0266	8.1164	0.3405	0.0190	7.9876
$\hat{\rho}_V$	0.4603	0.0911	9.3672	0.4623	0.0816	7.8457	0.4570	0.0800	7.2525	0.4626	0.0701	6.1244
$\hat{\rho}_N$	0.5107	0.0387	8.3721	0.5101	0.0315	7.7670	0.5111	0.0211	7.3761	0.5104	0.0150	7.0353
Case 3: $\rho = 0.1310$												
$\hat{\rho}_T$	0.1303	0.0552	3.0561	0.1310	0.0421	1.7740	0.1314	0.0310	0.9666	0.1312	0.0221	0.4927
$\hat{\rho}$	0.1313	0.0339	1.1494	0.1371	0.0263	0.7312	0.1360	0.0179	0.3465	0.1330	0.0127	0.1678
$\hat{\rho}_L$	0.0069	0.0549	18.4060	0.0088	0.0413	16.6504	0.0075	0.0297	16.1381	0.0080	0.0214	15.5869
$\hat{\rho}_N$	0.0867	0.0541	4.8952	0.0891	0.0421	3.5314	0.0891	0.0306	2.6910	0.0896	0.0219	2.1900

MSE is in the scale of 10^{-3} .

It indicates that the conditional mean (or absolute mean) calibration method and conditional variance method can not be used to estimate the target parameter ρ when then independence condition between U and (X, Y) does not hold. In case 2, $E(X) = E(Y) = 0, E[\phi(U)] = E[\psi(U)] = 1$ but (X, Y) is not independent of U . The proposed estimator $\hat{\rho}$ still works well in case 2 as its mean values of $\hat{\rho}$ are close to the true estimator $\hat{\rho}_T$. It is not surprised that $\hat{\rho}_C$ does not work in case 2 since $E(X) = E(Y) = 0$. But the estimators $\hat{\rho}_A$ and $\hat{\rho}_V$ still fail to estimate ρ as the their mean values of absolute biases are around 0.0882, 0.0339. It indicates that the conditional absolute mean calibration method and conditional variance method can not be used to estimate ρ when the independence condition between U and (X, Y) does not hold. In case 3, $E[\log(\phi(U))] = E[\log(\psi(U))] = 0$ but (X, Y) is not independent of U . The proposed estimator $\hat{\rho}$ works well in general, but the conditional absolute logarithmic calibrated estimator as $\hat{\rho}_L$ does not provide the correct estimate since its mean values of absolute biases are around 0.1230. These three cases show that the existing estimators $\hat{\rho}_C, \hat{\rho}_A$ and $\hat{\rho}_V$ relied on the independence condition between the U and unobserved (X, Y) , while the newly proposed estimator $\hat{\rho}$ does not need it. For the naive estimator $\hat{\rho}_N$, it still produces large biases and does not work in three cases. In Table 5, we report the simulation results of 95% confidence intervals. When the sample size n is greater than or equal to 500, we see that the asymptotic confidence intervals show satisfactory performances both in terms of average length of the confidence intervals and of the coverage probabilities. Generally, the asymptotic intervals are recommended to construct asymptotic confidence intervals when the sample size is large for the practical usage.

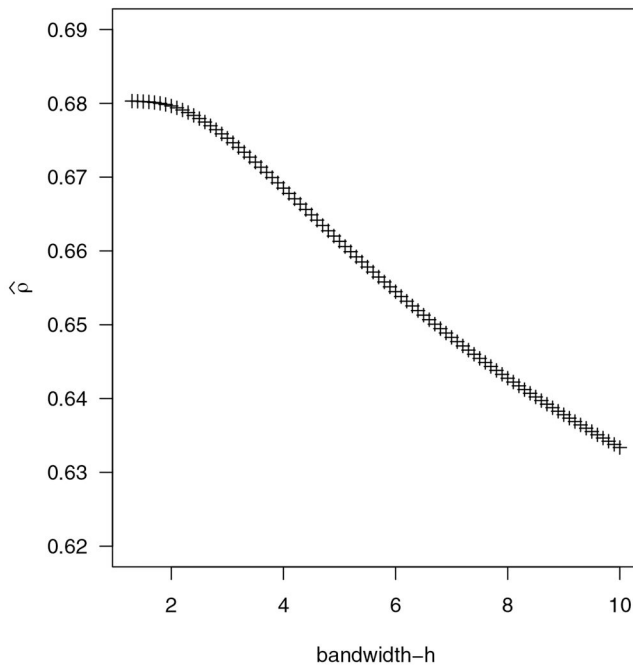
4. Real data analysis

In this section, we apply our methods to analyze a dataset created from a higher education institution, which is related to 4424 students enrolled in different undergraduate degrees, such as agronomy, design, education, nursing, journalism, management, social service, and technologies. The dataset is available online: <https://archive-beta.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success>. Here, we are interested in analyzing the relationship between the underlying previous qualification grade ($Y = \text{PQG}$) and the underlying admission grade

Table 5. Simulation results of confidence intervals of $\hat{\rho}$.

		$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
Case 1: $\rho = -0.0557$					
$\hat{\rho}$	Lower	-0.0815	-0.0716	-0.0585	-0.0590
	Upper	-0.0363	-0.042	-0.0409	-0.0488
	AL	0.0452	0.0296	0.0175	0.0102
	CP	93.9%	94.2%	94.6%	94.8%
Case 2: $\rho = 0.4287$					
$\hat{\rho}$	Lower	0.3592	0.3744	0.3905	0.3999
	Upper	0.4953	0.4805	0.4656	0.4531
	AL	0.1360	0.1061	0.0751	0.0532
	CP	93.9%	94.1%	94.5%	94.8%
Case 3: $\rho = 0.1310$					
$\hat{\rho}$	Lower	0.0357	0.0526	0.0751	0.0897
	Upper	0.2271	0.2216	0.2002	0.1761
	AL	0.1913	0.1690	0.1251	0.0864
	CP	94.3%	94.8%	95.1%	95.2%

"Lower" stands for the lower bound, "upper" stands for upper bound, "AL" stands for average length, "CP" stands for the coverage probabilities.

**Figure 1.** The pattern between the bandwidth h and the proposed estimator $\hat{\rho}$.

($X = AG$), and we suggest the variable (age) as the confounding variable U in the multiplicative distortion functions.

We conducted the varying coefficient based estimation method to estimate ρ by using the Gaussian kernel function $K(t)$. In Figure 1, the pattern between the bandwidth h and $\hat{\rho}$ shows a decreasing trend, and $\hat{\rho} \in (0.6334, 0.6803)$ when $h \in (1.2, 10)$. Using the rule of thumb suggested by Silverman (1986), the bandwidth h was chosen as $h = 3\hat{\sigma}_U n^{-1/3} = 1.3866$, and the estimator of ρ is obtained as $\hat{\rho} = 0.6803$. While the naive estimator $\hat{\rho}_N$ is obtained as $\hat{\rho}_N = 0.5804$, and the difference between $\hat{\rho}$ and $\hat{\rho}_N$ is large. The 95% confidence interval of ρ is $(0.6549, 0.7056)$, which also excludes the naive estimator $\hat{\rho}_N$. Together with the nonlinear

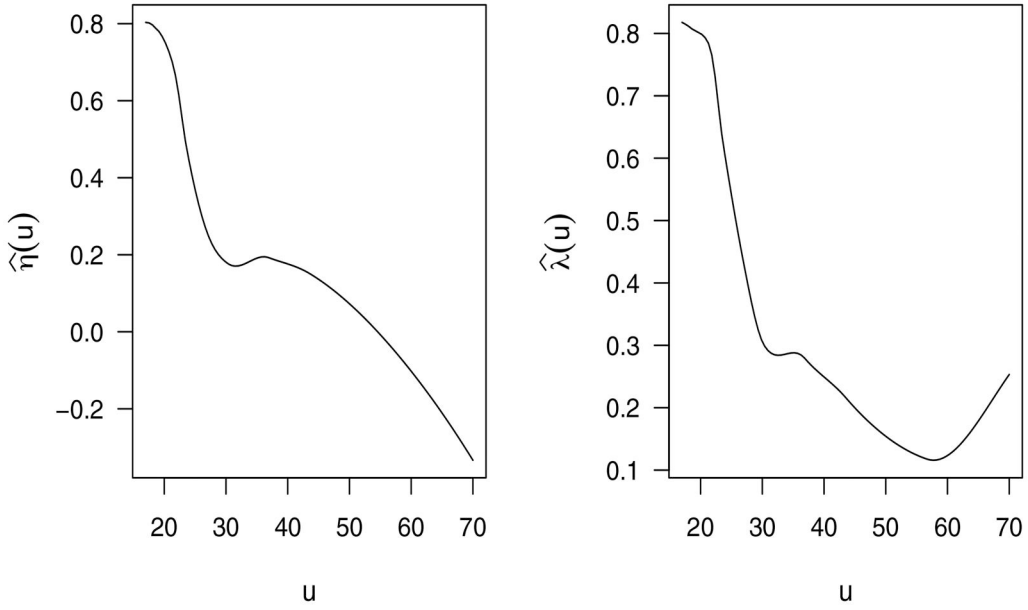


Figure 2. The patterns between the $\hat{\eta}(u), \hat{\lambda}(u)$ against the confounding variable.

patterns of $\hat{\eta}(u)$ and $\hat{\lambda}(u)$ in Figure 2, the confounding variable U has impact on the observed previous qualification grade (\tilde{Y}) and the observed admission grade (\tilde{X}).

If we used the conditional mean calibration method to analyze this dataset, the estimator $\hat{\rho}_C$ is obtained as $\hat{\rho}_C = 0.5715$, which is also not included in the 95% confidence interval of ρ above. And the value of the estimator $\hat{\rho}_C$ is close to the naive estimator $\hat{\rho}_N$. However, the conditional mean calibrated estimator $\hat{\rho}_C$ can not be used here. We explain it as follows. From the models (2.4) and (2.5), if $\alpha\delta \neq 0$, we have $\frac{\tilde{Y}}{\xi(U)} = 1 + \frac{\beta}{\alpha}X + \frac{\epsilon}{\alpha}$ and $\frac{\tilde{X}}{\xi(U)} = 1 + \frac{\gamma}{\delta}Y + \frac{\epsilon}{\delta}$. Let $g_X(u) = E\left[\frac{\tilde{Y}}{\xi(U)}|U=u\right] = 1 + \frac{\beta}{\alpha}E(X|U=u)$ and $w_Y(u) = E\left[\frac{\tilde{X}}{\xi(U)}|U=u\right] = 1 + \frac{\gamma}{\delta}E(Y|U=u)$. Appealing to the local linear estimators (Fan and Gijbels 1996), we used the calibrated variables $\left\{\frac{\tilde{Y}_i}{\xi(U_i)}, U_i\right\}_{i=1}^n$ to estimate $g_X(u)$, and we used $\left\{\frac{\tilde{X}_i}{\xi(U_i)}, U_i\right\}_{i=1}^n$ to estimate $w_Y(u)$. The patterns of $\hat{g}_X(u)$ and $\hat{w}_Y(u)$ are presented in Figure 3. The plots in Figure 3 indicate $E(X|U=u)$ and $E(Y|U=u)$ are both nonlinear functions of u , and it implies the independence condition between the unobserved variables (Y, X) and U is not valid. The conditional mean calibration estimation method can not be used here, but the proposed estimator $\hat{\rho}$ is workable. As a consequence, this dataset should be analyzed under the non-independence condition between (Y, X) and U .

5. Discussions and further research

We consider the estimation and confidence intervals of the Pearson correlation coefficient under the multiplicative distortions for the unobserved continuous random variables. A varying coefficient moment based estimator is proposed. The newly proposed estimator improves three conditions in literature: the identifiability conditions of unknown distortion functions, the non-zero mean conditions of unobserved variables and the independence condition between the confounding variable and the unobserved variables. The newly proposed estimation procedure can deal with the independence cases in literature and achieve the optimal property: the effects of

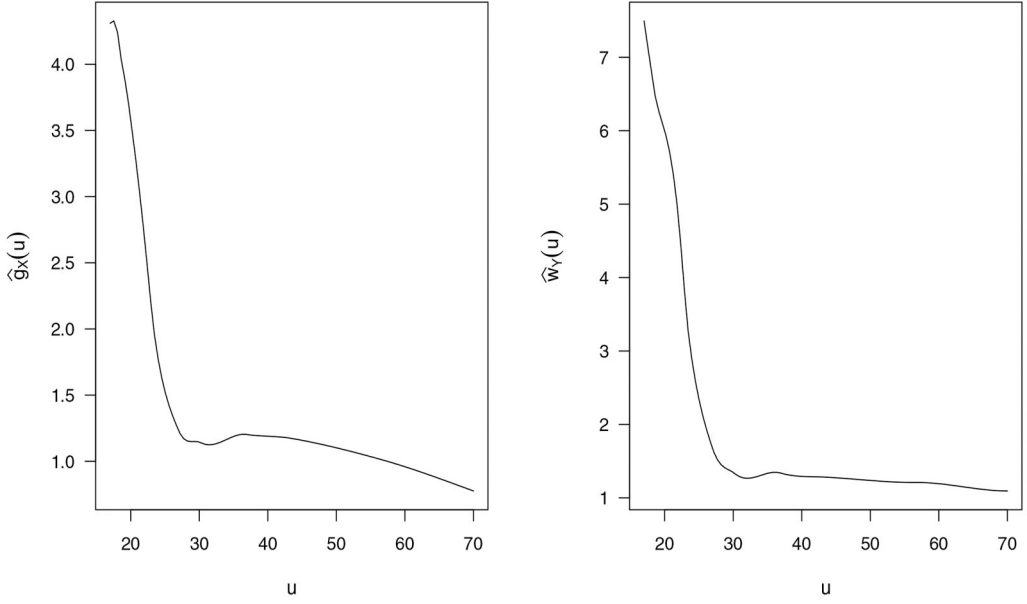


Figure 3. The patterns between the $\hat{g}_X(u), \hat{w}_Y(u)$ against the confounding variable.

distortion functions vanish in an asymptotic way as if the distortion functions are nonexistent. Some comparison between different estimation procedures are discussed through simulation. In future work, the proposed method can be used for the semi-parametric models with multiplicative distortions, such as partial linear models, partial linear single-index models. One can also take into account the case of single-index distortion functions under multivariate confounding variables and un-independence conditions. The Pearson correlation coefficient under additive distortion without the independence condition can also be incorporated into a plan. The research of this topic is in progress.

6. Appendix

6.1. A technical lemma

Lemma 1. Suppose $E|V|^3 < \infty, E(V|W = w) = m(w)$ and its derivatives up to second order are bounded for all $w \in [\mathcal{W}_L, \mathcal{W}_R]$, $\sup_w \int |v|^s f(w, v) dv < \infty$ for an $s > 2$, where $f(w, v)$ is the joint density of W and V , and $[\mathcal{W}_L, \mathcal{W}_R]$ ($\mathcal{W}_L < \mathcal{W}_R$) denotes the compact support of W . Let $(W_i, V_i), i = 1, 2, \dots, n$ be independent and identically distributed (i.i.d.) samples from (W, V) . If $n^{2\epsilon-1}h \rightarrow \infty$ for $\epsilon < 1 - s^{-1}$, then

$$\sup_{w \in [\mathcal{W}_L, \mathcal{W}_R]} \left| \frac{1}{n} \sum_{i=1}^n K_h(W_i - w) V_i - f_W(w) m(w) - \frac{1}{2} [f_W(w) m(w)]'' \mu_{K,2} h^2 \right| = O(\tau_{n,h}), a.s.$$

where, $f_W(w)$ is the density function of W , $\tau_{n,h} = h^3 + \sqrt{\frac{\log n}{nh}}$ and $\mu_{K,2} = \int K(w) w^2 dw$.

Proof. Lemma 1 can be immediately proved from the result obtained by Mack and Silverman (1982).

6.2. Proof of Theorem 1

Proof. Step 1 We define $\widehat{\boldsymbol{\kappa}}(u) = (\widehat{\zeta}(u), \widehat{\eta}(u))^\top$, $\widehat{\boldsymbol{\kappa}}'(u) = (\widehat{\zeta}'(u), \widehat{\eta}'(u))^\top$, and $\boldsymbol{\kappa}(u) = (\zeta(u), \eta(u))^\top$, $\boldsymbol{\kappa}'(u) = (\zeta'(u), \eta'(u))^\top$, moreover, $\boldsymbol{\kappa}''(u) = (\zeta''(u), \eta''(u))^\top$. Define that $\kappa_{n,h} = h^2 + \sqrt{\frac{\log n}{nh}}$. Using Lemma 1, uniformly in $u \in [\mathcal{U}_L, \mathcal{U}_R]$, the Lemma A.2 in Fan and Huang (2005) entails that

$$\begin{aligned} & \frac{1}{n} \widetilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,z} \widetilde{\mathbf{D}}_{n,u} \\ &= \begin{pmatrix} f_U(u) \boldsymbol{\psi}(u) \boldsymbol{\Sigma}(u) \boldsymbol{\psi}(u), & \mathbf{0} \\ \mathbf{0}, & \mu_2 f_U(u) \boldsymbol{\psi}(u) \boldsymbol{\Sigma}(u) \boldsymbol{\psi}(u) \end{pmatrix} \{1 + O_P(\kappa_{n,h})\}, \end{aligned} \quad (\text{A.1})$$

where $\boldsymbol{\psi}(u)$ and $\boldsymbol{\Sigma}(u)$ are defined in Theorem 1. Moreover, we have

$$\begin{aligned} & \begin{pmatrix} \widehat{\boldsymbol{\kappa}}(u) - \boldsymbol{\kappa}(u) \\ h \widehat{\boldsymbol{\kappa}}'(u) - h \boldsymbol{\kappa}'(u) \end{pmatrix} \\ &= \left(\widetilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,u} \widetilde{\mathbf{D}}_{n,u} \right)^{-1} \widetilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,u} \left(\widetilde{\mathbf{V}}_n - \widetilde{\mathbf{D}}_{n,u} \begin{pmatrix} \boldsymbol{\kappa}(u) \\ h \boldsymbol{\kappa}'(u) \end{pmatrix} \right). \end{aligned} \quad (\text{A.2})$$

Similar to (A.1), using Lemma 1, we have

$$\begin{aligned} & \frac{1}{n} \widetilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,u} \left(\widetilde{\mathbf{V}}_n - \widetilde{\mathbf{D}}_{n,u} \begin{pmatrix} \boldsymbol{\kappa}(u) \\ h \boldsymbol{\kappa}'(u) \end{pmatrix} \right) \\ &= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \widetilde{\mathbf{X}}_i \phi(U_i) \varepsilon_i \\ \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \left(\frac{U_i - u}{h} \right) \widetilde{\mathbf{X}}_i \phi(U_i) \varepsilon_i \end{pmatrix} \\ &+ \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \widetilde{\mathbf{X}}_i^{\otimes 2} (\boldsymbol{\kappa}(U_i) - \boldsymbol{\kappa}(u) - \boldsymbol{\kappa}'(u)(U_i - u)) \\ \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \left(\frac{U_i - u}{h} \right) \widetilde{\mathbf{X}}_i^{\otimes 2} (\boldsymbol{\kappa}(U_i) - \boldsymbol{\kappa}(u) - \boldsymbol{\kappa}'(u)(U_i - u)) \end{pmatrix}. \end{aligned} \quad (\text{A.3})$$

Directly using Lemma 1, we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \widetilde{\mathbf{X}}_i^{\otimes 2} (\boldsymbol{\kappa}(U_i) - \boldsymbol{\kappa}(u) - \boldsymbol{\kappa}'(u)(U_i - u)) \\ &= \frac{\mu_2 h^2}{2} f_U(u) \boldsymbol{\psi}(u) \boldsymbol{\Sigma}(u) \boldsymbol{\psi}(u) \boldsymbol{\kappa}''(u) + O_P(h^2 \kappa_{n,h}). \end{aligned} \quad (\text{A.4})$$

Let \mathbf{I}_2 be an identity matrix of size 2 and $\mathbf{0}_{2 \times 2}$ be an symmetric zero matrix of size 2. From (A.1)–(A.4), we have

$$\begin{aligned}
& \widehat{\boldsymbol{\kappa}}(u) - \boldsymbol{\kappa}(u) \\
&= (\mathbf{I}_2, \mathbf{0}_{2 \times 2}) \begin{pmatrix} \widehat{\boldsymbol{\kappa}}(u) - \boldsymbol{\kappa}(u) \\ h\widehat{\boldsymbol{\kappa}}'(u) - h\boldsymbol{\kappa}'(u) \end{pmatrix} \\
&= (\mathbf{I}_2, \mathbf{0}_{2 \times 2}) \left(\widetilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,u} \widetilde{\mathbf{D}}_{n,u} \right)^{-1} \widetilde{\mathbf{D}}_{n,u}^\top \mathbf{W}_{n,u} \left(\widetilde{\mathbf{V}}_n - \widetilde{\mathbf{D}}_{n,u} \begin{pmatrix} \boldsymbol{\kappa}(u) \\ h\boldsymbol{\kappa}'(u) \end{pmatrix} \right) \\
&= [\boldsymbol{\psi}(u)]^{-1} [\boldsymbol{\Sigma}(u)]^{-1} [\boldsymbol{\psi}(u)]^{-1} \frac{1}{nf_U(u)} \sum_{i=1}^n K_h(U_i - u) \widetilde{\mathbf{X}}_i \phi(U_i) \varepsilon_i \\
&\quad + \frac{\mu_2 h^2}{2} \boldsymbol{\kappa}''(u) + O_P(\kappa_{n,h}^2).
\end{aligned} \tag{A.5}$$

From (A.5), we have

$$\begin{aligned}
& \sqrt{nh} \left(\widehat{\boldsymbol{\kappa}}(u) - \boldsymbol{\kappa}(u) - \frac{\mu_2 h^2}{2} \boldsymbol{\kappa}''(u) \right) \\
& \xrightarrow{L} N \left(0, \frac{\mu_{K,2} \phi^2(u) [\boldsymbol{\psi}(u)]^{-1} [\boldsymbol{\Sigma}(u)]^{-1} \boldsymbol{\Lambda}(u) [\boldsymbol{\Sigma}(u)]^{-1} [\boldsymbol{\psi}(u)]^{-1}}{f_U(u)} \right).
\end{aligned} \tag{A.6}$$

Step 2 We define $\widehat{\boldsymbol{\omega}}(u) = (\widehat{\xi}(u), \widehat{\lambda}(u))^\top$, $\widehat{\boldsymbol{\omega}}'(u) = (\widehat{\xi}'(u), \widehat{\lambda}'(u))^\top$, $\boldsymbol{\omega}(u) = (\xi(u), \lambda(u))^\top$, $\boldsymbol{\omega}'(u) = (\xi'(u), \lambda'(u))^\top$, $\boldsymbol{\omega}''(u) = (\xi''(u), \lambda''(u))^\top$. Similar to (A.1), we have

$$\begin{aligned}
& \frac{1}{n} \widetilde{\mathbf{G}}_{n,u}^\top \mathbf{W}_{n,u} \widetilde{\mathbf{G}}_{n,u} \\
&= \begin{pmatrix} f_U(u) \phi(u) \boldsymbol{\Omega}(u) \phi(u), & \mathbf{0} \\ \mathbf{0}, & \mu_2 f_U(u) \phi(u) \boldsymbol{\Omega}(u) \phi(u) \end{pmatrix} \{1 + O_P(\kappa_{n,h})\},
\end{aligned} \tag{A.7}$$

Moreover, we have

$$\begin{aligned}
& \begin{pmatrix} \widehat{\boldsymbol{\omega}}(u) - \boldsymbol{\omega}(u) \\ h\widehat{\boldsymbol{\omega}}'(u) - h\boldsymbol{\omega}'(u) \end{pmatrix} \\
&= \left(\widetilde{\mathbf{G}}_{n,u}^\top \mathbf{W}_{n,u} \widetilde{\mathbf{G}}_{n,u} \right)^{-1} \widetilde{\mathbf{G}}_{n,u}^\top \mathbf{W}_{n,u} \left(\widetilde{\mathbf{Q}}_n - \widetilde{\mathbf{G}}_{n,u} \begin{pmatrix} \boldsymbol{\omega}(u) \\ h\boldsymbol{\omega}'(u) \end{pmatrix} \right).
\end{aligned} \tag{A.8}$$

Similar to (A.1) and (A.7), using Lemma 1, we have

$$\begin{aligned}
& \frac{1}{n} \widetilde{\mathbf{G}}_{n,u}^\top \mathbf{W}_{n,u} \left(\widetilde{\mathbf{Q}}_n - \widetilde{\mathbf{G}}_{n,u} \begin{pmatrix} \boldsymbol{\omega}(u) \\ h\boldsymbol{\omega}'(u) \end{pmatrix} \right) \\
&= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \widetilde{\mathbf{Y}}_i \boldsymbol{\psi}(U_i) \varepsilon_i \\ \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \left(\frac{U_i - u}{h} \right) \widetilde{\mathbf{Y}}_i \boldsymbol{\psi}(U_i) \varepsilon_i \end{pmatrix} \\
&\quad + \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \widetilde{\mathbf{Y}}_i^{\otimes 2} (\boldsymbol{\omega}(U_i) - \boldsymbol{\omega}(u) - \boldsymbol{\omega}'(u)(U_i - u)) \\ \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \left(\frac{U_i - u}{h} \right) \widetilde{\mathbf{Y}}_i^{\otimes 2} (\boldsymbol{\omega}(U_i) - \boldsymbol{\omega}(u) - \boldsymbol{\omega}'(u)(U_i - u)) \end{pmatrix}.
\end{aligned} \tag{A.9}$$

Directly using [Lemma 1](#), we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n K_h(U_i - u) \tilde{Y}_i^{\otimes 2} (\varpi(U_i) - \varpi(u) - \varpi'(u)(U_i - u)) \\ &= \frac{\mu_2 h^2}{2} f_U(u) \phi(u) \Omega(u) \phi(u) \varpi''(u) + O_P(h^2 \kappa_{n,h}). \end{aligned} \quad (\text{A.10})$$

From [\(A.7\)–\(A.10\)](#), we have

$$\begin{aligned} & \widehat{\varpi}(u) - \varpi(u) \\ &= (\mathbf{I}_2, \mathbf{0}_{2 \times 2}) \begin{pmatrix} \widehat{\varpi}(u) - \varpi(u) \\ h \widehat{\varpi}'(u) - h \varpi'(u) \end{pmatrix} \\ &= (\mathbf{I}_2, \mathbf{0}_{2 \times 2}) \left(\tilde{\mathbf{G}}_{n,u}^\top \mathbf{W}_{n,u} \tilde{\mathbf{G}}_{n,u} \right)^{-1} \tilde{\mathbf{G}}_{n,u}^\top \mathbf{W}_{n,u} \begin{pmatrix} \tilde{\mathbf{Q}}_n - \tilde{\mathbf{G}}_{n,u} \begin{pmatrix} \varpi(u) \\ h \varpi'(u) \end{pmatrix} \end{pmatrix} \\ &= [\phi(u)]^{-1} [\Omega(u)]^{-1} [\phi(u)]^{-1} \frac{1}{n f_U(u)} \sum_{i=1}^n K_h(U_i - u) \tilde{Y}_i \psi(U_i) \epsilon_i \\ & \quad + \frac{\mu_2 h^2}{2} \varpi''(u) + O_P(\kappa_{n,h}^2). \end{aligned} \quad (\text{A.11})$$

From [\(A.11\)](#), we have

$$\begin{aligned} & \sqrt{nh} \left(\widehat{\varpi}(u) - \varpi(u) - \frac{\mu_2 h^2}{2} \varpi''(u) \right) \\ & \xrightarrow{L} N \left(0, \frac{\mu_{K,2} \psi^2(u) [\phi(u)]^{-1} [\Omega(u)]^{-1} \Delta(u) [\Omega(u)]^{-1} [\phi(u)]^{-1}}{f_U(u)} \right). \end{aligned} \quad (\text{A.12})$$

Based on [\(A.6\)](#) and [\(A.12\)](#), we have completed the proof of [Theorem 1](#).

6.3. Proof of Theorem 2

Step 1 Recalling that $\mathbf{e}_2 = (0, 1)^\top$, from [\(A.5\)](#), we have

$$\begin{aligned} \hat{\eta}(u) - \eta(u) &= \mathbf{e}_2^\top (\hat{\kappa}(u) - \kappa(u)) \\ &= \mathbf{e}_2^\top [\psi(u)]^{-1} [\Sigma(u)]^{-1} [\psi(u)]^{-1} \frac{1}{n f_U(u)} \sum_{i=1}^n K_h(U_i - u) \tilde{\mathbf{X}}_i \phi(U_i) \epsilon_i \\ & \quad + \frac{\mu_2 h^2}{2} \mathbf{e}_2^\top \kappa''(u) + O_P(\kappa_{n,h}^2) \\ &= \mathbf{e}_2^\top [\Sigma(u)]^{-1} [\psi(u)]^{-1} \frac{1}{n f_U(u) \psi(u)} \sum_{i=1}^n K_h(U_i - u) \tilde{\mathbf{X}}_i \phi(U_i) \epsilon_i \\ & \quad + \frac{\mu_2 h^2}{2} \eta''(u) + O_P(\kappa_{n,h}^2). \end{aligned} \quad (\text{B.1})$$

Similar to [\(B.1\)](#), we have

$$\begin{aligned} \hat{\lambda}(u) - \lambda(u) &= \mathbf{e}_2^\top (\widehat{\varpi}(u) - \varpi(u)) \\ &= \mathbf{e}_2^\top [\Omega(u)]^{-1} [\phi(u)]^{-1} \frac{1}{n f_U(u) \phi(u)} \sum_{i=1}^n K_h(U_i - u) \tilde{Y}_i \psi(U_i) \epsilon_i \\ & \quad + \frac{\mu_2 h^2}{2} \lambda''(u) + O_P(\kappa_{n,h}^2). \end{aligned} \quad (\text{B.2})$$

As $nh^4 \rightarrow 0$ and $\frac{(\log n)^2}{nh^2} \rightarrow 0$, the projection of U -statistic (Serfling 1980) entails that

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n [\hat{\eta}(U_i) - \eta(U_i)] \lambda(U_i) \\
&= \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \mathbf{e}_2^\top [\boldsymbol{\Sigma}(U_i)]^{-1} [\boldsymbol{\psi}(U_i)]^{-1} \frac{\lambda(U_i)}{f_U(U_i) \psi(U_i)} K\left(\frac{U_i - U_j}{h}\right) \tilde{\mathbf{X}}_j \phi(U_j) \epsilon_j \\
&\quad + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Sigma}(U_i)]^{-1} [\boldsymbol{\psi}(U_i)]^{-1} \frac{\lambda(U_i)}{\psi(U_i)} \tilde{\mathbf{X}}_i \phi(U_i) \epsilon_i + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Sigma}(U_i)]^{-1} [\boldsymbol{\psi}(U_i)]^{-1} \frac{\lambda(U_i)}{\psi(U_i)} \boldsymbol{\psi}(U_i) \mathbf{X}_i \phi(U_i) \epsilon_i + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Sigma}(U_i)]^{-1} \frac{\lambda(U_i) \phi(U_i)}{\psi(U_i)} \mathbf{X}_i \epsilon_i + o_P(n^{-1/2}).
\end{aligned} \tag{B.3}$$

Similar to (B.3), we have

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n [\hat{\lambda}(U_i) - \lambda(U_i)] \eta(U_i) \\
&= \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \mathbf{e}_2^\top [\boldsymbol{\Omega}(U_i)]^{-1} [\boldsymbol{\phi}(U_i)]^{-1} \frac{\eta(U_i)}{f_U(U_i) \phi(U_i)} K\left(\frac{U_i - U_j}{h}\right) \tilde{\mathbf{Y}}_j \psi(U_j) \epsilon_j \\
&\quad + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Omega}(U_i)]^{-1} [\boldsymbol{\phi}(U_i)]^{-1} \frac{\eta(U_i)}{\phi(U_i)} \tilde{\mathbf{Y}}_i \psi(U_i) \epsilon_i + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Omega}(U_i)]^{-1} [\boldsymbol{\phi}(U_i)]^{-1} \frac{\eta(U_i)}{\phi(U_i)} \boldsymbol{\phi}(U_i) \mathbf{Y}_i \psi(U_i) \epsilon_i + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Omega}(U_i)]^{-1} \frac{\eta(U_i) \psi(U_i)}{\phi(U_i)} \mathbf{Y}_i \epsilon_i + o_P(n^{-1/2}).
\end{aligned} \tag{B.4}$$

Using $\eta(U) = \beta \frac{\phi(U)}{\psi(U)}$ and $\lambda(U) = \gamma \frac{\psi(U)}{\phi(U)}$, the asymptotic expressions of (B.3) and (B.4) can be re-written as

$$\frac{1}{n} \sum_{i=1}^n [\hat{\eta}(U_i) - \eta(U_i)] \lambda(U_i) = \frac{\gamma}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Sigma}(U_i)]^{-1} \mathbf{X}_i \epsilon_i + o_P(n^{-1/2}), \tag{B.5}$$

$$\frac{1}{n} \sum_{i=1}^n [\hat{\lambda}(U_i) - \lambda(U_i)] \eta(U_i) = \frac{\beta}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Omega}(U_i)]^{-1} \mathbf{Y}_i \epsilon_i + o_P(n^{-1/2}). \tag{B.6}$$

Using (B.5) and (B.6), and $\beta\gamma = \rho^2$, we have

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i) \hat{\lambda}(U_i) - \rho^2 \\
&= \frac{1}{n} \sum_{i=1}^n [\hat{\eta}(U_i) - \eta(U_i)] \lambda(U_i) + \frac{1}{n} \sum_{i=1}^n [\hat{\lambda}(U_i) - \lambda(U_i)] \eta(U_i) \\
&\quad + \frac{1}{n} \sum_{i=1}^n [\eta(U_i) \lambda(U_i) - \rho^2] + O_P(\kappa_{n,h}^2) \\
&= \frac{\gamma}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Sigma}(U_i)]^{-1} \mathbf{X}_i \varepsilon_i + \frac{\beta}{n} \sum_{i=1}^n \mathbf{e}_2^\top [\boldsymbol{\Omega}(U_i)]^{-1} \mathbf{Y}_i \epsilon_i + o_P(n^{-1/2}) \\
&= \frac{\rho}{n} \sum_{i=1}^n \mathbf{e}_2^\top \left[\sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}} [\boldsymbol{\Sigma}(U_i)]^{-1} \mathbf{X}_i \varepsilon_i + \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}} [\boldsymbol{\Omega}(U_i)]^{-1} \mathbf{Y}_i \epsilon_i \right] + o_P(n^{-1/2}).
\end{aligned} \tag{B.7}$$

From (B.7), we have

$$\begin{aligned}
& \left[\frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i) \hat{\lambda}(U_i) \right]^2 - \rho^4 \\
&= \frac{2\rho^3}{n} \sum_{i=1}^n \mathbf{e}_2^\top \left[\sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}} [\boldsymbol{\Sigma}(U_i)]^{-1} \mathbf{X}_i \varepsilon_i + \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}} [\boldsymbol{\Omega}(U_i)]^{-1} \mathbf{Y}_i \epsilon_i \right] + o_P(n^{-1/2}).
\end{aligned} \tag{B.8}$$

When $\rho \neq 0$, we have

$$\begin{aligned}
& \left(\left[\frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i) \hat{\lambda}(U_i) \right]^2 \right)^{1/4} - \rho \\
&= \frac{1}{2n} \sum_{i=1}^n \mathbf{e}_2^\top \left[\sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}} [\boldsymbol{\Sigma}(U_i)]^{-1} \mathbf{X}_i \varepsilon_i + \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}} [\boldsymbol{\Omega}(U_i)]^{-1} \mathbf{Y}_i \epsilon_i \right] + o_P(n^{-1/2}).
\end{aligned} \tag{B.9}$$

From (B.1), we have $\frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i) \xrightarrow{P} E[\eta(U)]$ and $\text{sgn}(\frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i)) \xrightarrow{P} \text{sgn}(\rho)$ under the **Assumption A**. The asymptotic expression (B.8) entails that $\left[\frac{1}{n} \sum_{i=1}^n \hat{\eta}(U_i) \hat{\lambda}(U_i) \right]^2 = o_P(n^{-1/2})$ when $\rho = 0$. We have complete the proof of **Theorem 2**.

Acknowledgements

The authors thank the editor, the associate editor, and three referees for their constructive suggestions that helped us to improve the early manuscript. Yingcong Huang (ID: 2020193015) is a senior student majoring in Statistics at Shenzhen University. Siming Deng (ID: 2021193033), JiongTao Zhong (ID: 2021193047) and Xiaozhen Yang (ID: 2021193036) are all junior students majoring in Statistics at Shenzhen University. This work was done when these undergraduate students were supervised by the corresponding author.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Jun Zhang's research was supported by the National Natural Science Foundation of China (Grant No.12371448).

ORCID

Jun Zhang  <http://orcid.org/0000-0003-4332-5182>

References

- Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. 2006. *Nonlinear measurement error models, a modern perspective*. 2nd ed. New York: Chapman and Hall.
- Cui, X., W. Guo, L. Lin, and L. Zhu. 2009. Covariate-adjusted nonlinear regression. *The Annals of Statistics* 37 (4): 1839–70. doi:10.1214/08-AOS627.
- de Castro, M., and I. Vidal. 2019. Bayesian inference in measurement error models from objective priors for the bivariate normal distribution. *Statistical Papers* 60 (4):1059–78. doi:10.1007/s00362-016-0863-7.
- Delaigle, A., P. Hall, and W.-X. Zhou. 2016. Nonparametric covariate-adjusted regression. *The Annals of Statistics* 44 (5):2190–220. doi:10.1214/16-AOS1442.
- Fan, J., and I. Gijbels. 1996. *Local polynomial modelling and its applications*. London: Chapman & Hall.
- Fan, J., and T. Huang. 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* 11 (6):1031–57. doi:10.3150/bj/1137421639.
- Fan, J., and W. Zhang. 1999. Statistical estimation in varying coefficient models. *The Annals of Statistics* 27 (5): 1491–518. doi:10.1214/aos/1017939139.
- Feng, S., Y. Hu, and L. Xue. 2016. Model detection and variable selection for varying coefficient models with longitudinal data. *Acta Mathematica Sinica, English Series* 32 (3):331–50. doi:10.1007/s10114-016-4639-8.
- Feng, S., G. Li, H. Peng, and T. Tong. 2021. Varying coefficient panel data model with interactive fixed effects. *Statistica Sinica* 31:935–57. doi:10.5705/ss.202018.0248.
- Feng, S., P. Tian, Y. Hu, and G. Li. 2021. Estimation in functional single-index varying coefficient model. *Journal of Statistical Planning and Inference* 214:62–75. doi:10.1016/j.jspi.2021.01.003.
- Feng, S., and L. Xue. 2016. Partially functional linear varying coefficient model. *Statistics* 50 (4):717–32. doi:10.1080/02331888.2016.1138954.
- Feng, Z., Y. Gai, and J. Zhang. 2019. Correlation curve estimation for multiplicative distortion measurement errors data. *Journal of Nonparametric Statistics* 31 (2):435–50. doi:10.1080/10485252.2019.1580708.
- Ferguson, T. S. 1996. A course in large sample theory. *Texts in statistical science series*. London: Chapman & Hall.
- Fuller, W. A. 1987. Measurement error models. *Wiley series in probability and mathematical statistics: Probability and mathematical statistics*. New York: John Wiley & Sons Inc.
- Gai, Y., J. Zhang, and Y. Zhou. 2024. Parametric hypothesis tests for exponentiality under multiplicative distortion measurement errors data. *Communications in Statistics - Simulation and Computation* 53 (3):1594–617. doi:10.1080/03610918.2023.2238361.
- Hastie, T., and R. Tibshirani. 1993. Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)* 55 (4):757–79. With discussion and a reply by the authors. doi:10.1111/j.2517-6161.1993.tb01939.x.
- Li, B., and X. Yin. 2007. On surrogate dimension reduction for measurement error regression: An invariance law. *The Annals of Statistics* 35 (5):2143–72. doi:10.1214/009053607000000172.
- Li, G., J. Zhang, and S. Feng. 2016. *Modern measurement error models*. Beijing: Science Press.
- Lian, H. 2015. Quantile regression for dynamic partially linear varying coefficient time series models. *Journal of Multivariate Analysis* 141:49–66. doi:10.1016/j.jmva.2015.06.013.
- Lian, H., P. Lai, and H. Liang. 2013. Partially linear structure selection in cox models with varying coefficients. *Biometrics* 69 (2):348–57. doi:10.1111/biom.12024.
- Liang, H., W. Härdle, and R. J. Carroll. 1999. Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* 27 (5):1519–35. doi:10.1214/aos/1017939140.
- Mack, Y. P., and B. W. Silverman. 1982. Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 61 (3):405–15. doi:10.1007/BF00539840.
- Nadaraya, E. A. 1964. On estimating regression. *Theory of Probability & Its Applications* 9 (1):141–2. doi:10.1137/1109020.
- Şentürk, D., and H.-G. Müller. 2005a. Covariate adjusted correlation analysis via varying coefficient models. *Scandinavian Journal of Statistics. Theory and Applications* 32 (3):365–83.
- Şentürk, D., and H.-G. Müller. 2005b. Covariate-adjusted regression. *Biometrika* 92 (1):75–89. doi:10.1093/biomet/92.1.75.
- Serfling, R. J. 1980. *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons Inc.

- Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. Monographs on statistics and applied probability. London: Chapman & Hall.
- Székely, G. J., M. L. Rizzo, and N. K. Bakirov. 2007. Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35 (6):2769–94. doi:[10.1214/009053607000000505](https://doi.org/10.1214/009053607000000505).
- Tomaya, L. C., and M. de Castro. 2018. A heteroscedastic measurement error model based on skew and heavy-tailed distributions with known error variances. *Journal of Statistical Computation and Simulation* 88 (11):2185–200. doi:[10.1080/00949655.2018.1452925](https://doi.org/10.1080/00949655.2018.1452925).
- Watson, G. S. 1964. Smooth regression analysis, *Sankhyā. The Indian Journal of Statistics, Series A* 26:359–72.
- Yang, Y., G. Li, and H. Peng. 2014. Empirical likelihood of varying coefficient errors-in-variables models with longitudinal data. *Journal of Multivariate Analysis* 127:1–18. doi:[10.1016/j.jmva.2014.02.004](https://doi.org/10.1016/j.jmva.2014.02.004).
- Yang, Y., G. Li, and T. Tong. 2015. Corrected empirical likelihood for a class of generalized linear measurement error models. *Science China Mathematics* 58 (7):1523–36. doi:[10.1007/s11425-015-4976-6](https://doi.org/10.1007/s11425-015-4976-6).
- Yang, Y., T. Tong, and G. Li. 2019. Simex estimation for single-index model with covariate measurement error. *AStA Advances in Statistical Analysis* 103 (1):137–61. doi:[10.1007/s10182-018-0327-6](https://doi.org/10.1007/s10182-018-0327-6).
- Zhang, J. 2019. Partial linear models with general distortion measurement errors. *Electronic Journal of Statistics* 13 (2):5360–414. doi:[10.1214/19-EJS1654](https://doi.org/10.1214/19-EJS1654).
- Zhang, J. 2021. Estimation and variable selection for partial linear single-index distortion measurement errors models. *Statistical Papers* 62 (2):887–913. doi:[10.1007/s00362-019-01119-6](https://doi.org/10.1007/s00362-019-01119-6).
- Zhang, J. 2022. Measurement errors models with exponential parametric multiplicative distortions. *Journal of Statistical Computation and Simulation* 92 (12):2467–500. doi:[10.1080/00949655.2022.2037594](https://doi.org/10.1080/00949655.2022.2037594).
- Zhang, J. 2023. Nonlinear multiplicative distortion regression models with second-order estimation. *Communications in Statistics - Simulation and Computation* 52 (12):5894–924. doi:[10.1080/03610918.2021.2001656](https://doi.org/10.1080/03610918.2021.2001656).
- Zhang, J., A. Chen, and Z. Wei. 2023. Kernel density estimation for multiplicative distortion measurement regression models. *Communications in Statistics - Simulation and Computation* 52 (5):1733–52. doi:[10.1080/03610918.2021.1890122](https://doi.org/10.1080/03610918.2021.1890122).
- Zhang, J., and X. Cui. 2021. Logarithmic calibration for nonparametric multiplicative distortion measurement errors models. *Journal of Statistical Computation and Simulation* 91 (13):2623–44. doi:[10.1080/00949655.2021.1904240](https://doi.org/10.1080/00949655.2021.1904240).
- Zhang, J., Z. Feng, and B. Zhou. 2014. A revisit to correlation analysis for distortion measurement error data. *Journal of Multivariate Analysis*. 124:116–29. doi:[10.1016/j.jmva.2013.10.004](https://doi.org/10.1016/j.jmva.2013.10.004).
- Zhang, J., Y. Gai, X. Cui, and G. Li. 2020. Measuring symmetry and asymmetry of multiplicative distortion measurement errors data. *Brazilian Journal of Probability and Statistics* 34 (2):370–93. doi:[10.1214/19-BJPS432](https://doi.org/10.1214/19-BJPS432).
- Zhang, J., G. Li, and Z. Feng. 2015. Checking the adequacy for a distortion errors-in-variables parametric regression model. *Computational Statistics & Data Analysis* 83:52–64. doi:[10.1016/j.csda.2014.09.018](https://doi.org/10.1016/j.csda.2014.09.018).
- Zhang, J., G. Li, and Y. Yang. 2022. Modal linear regression models with multiplicative distortion measurement errors. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 15 (1):15–42. doi:[10.1002/sam.11541](https://doi.org/10.1002/sam.11541).
- Zhang, J., and B. Lin. 2023. Estimation of correlation coefficient with general distortion measurement errors. *Communications in Statistics - Simulation and Computation* 52 (9):4491–521. doi:[10.1080/03610918.2021.1963453](https://doi.org/10.1080/03610918.2021.1963453).
- Zhang, J., B. Lin, and Z. Feng. 2020. Conditional absolute mean calibration for partial linear multiplicative distortion measurement errors models. *Computational Statistics & Data Analysis* 141:77–93. doi:[10.1016/j.csda.2019.06.009](https://doi.org/10.1016/j.csda.2019.06.009).
- Zhang, J., B. Lin, and G. Li. 2019. Nonlinear regression models with general distortion measurement errors. *Journal of Statistical Computation and Simulation* 89 (8):1482–504. doi:[10.1080/00949655.2019.1586904](https://doi.org/10.1080/00949655.2019.1586904).
- Zhang, J., B. Lin, and Y. Zhou. 2024. Linear regression models with multiplicative distortions under new identifiability conditions. *Statistica Neerlandica* 78 (1):25–67. doi:[10.1111/stan.12304](https://doi.org/10.1111/stan.12304).
- Zhang, J., Z. Xu, and Z. Wei. 2023. Absolute logarithmic calibration for correlation coefficient with multiplicative distortion. *Communications in Statistics - Simulation and Computation* 52 (2):482–505. doi:[10.1080/03610918.2020.1859541](https://doi.org/10.1080/03610918.2020.1859541).
- Zhang, J., B. Yang, and Z. Feng. 2024. Estimation of correlation coefficient under a linear multiplicative distortion measurement errors model. *Communications in Statistics - Simulation and Computation* 53 (1):62–93. doi:[10.1080/03610918.2021.2004421](https://doi.org/10.1080/03610918.2021.2004421).
- Zhang, J., Y. Yang, S. Feng, and Z. Wei. 2020. Logarithmic calibration for partial linear models with multiplicative distortion measurement errors. *Journal of Statistical Computation and Simulation* 90 (10):1875–96. doi:[10.1080/00949655.2020.1750614](https://doi.org/10.1080/00949655.2020.1750614).
- Zhang, J., Y. Yang, and G. Li. 2020. Logarithmic calibration for multiplicative distortion measurement errors regression models. *Statistica Neerlandica* 74 (4):462–88. doi:[10.1111/stan.12204](https://doi.org/10.1111/stan.12204).
- Zhang, J., X. Yu, S. Deng, J. Zhong, Y. Zhou, and B. Lin. 2023. Estimation of correlation coefficient with monotone transformation and multiplicative distortions. *Communications in Statistics - Theory and Methods* :1–33. doi:[10.1080/03610926.2023.2288794](https://doi.org/10.1080/03610926.2023.2288794).
- Zhang, J., Y. Zhou, W. Xu, and G. Li. 2018. Semiparametric quantile estimation for varying coefficient partially linear measurement errors models. *Brazilian Journal of Probability and Statistics* 32 (3):616–56. doi:[10.1214/17-BJPS357](https://doi.org/10.1214/17-BJPS357).

- Zhang, J., J. Zhu, Y. Zhou, X. Cui, and T. Lu. 2020. Multiplicative regression models with distortion measurement errors. *Statistical Papers* 61 (5):2031–57. doi:[10.1007/s00362-018-1020-2](https://doi.org/10.1007/s00362-018-1020-2).
- Zhang, W., and H. Peng. 2010. Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *Journal of Multivariate Analysis* 101 (7):1656–80. doi:[10.1016/j.jmva.2010.03.003](https://doi.org/10.1016/j.jmva.2010.03.003).
- Zhao, J., and C. Xie. 2018. A nonparametric test for covariate-adjusted models. *Statistics & Probability Letters* 133: 65–70. doi:[10.1016/j.spl.2017.10.004](https://doi.org/10.1016/j.spl.2017.10.004).
- Zhao, K., and H. Lian. 2016. Sparsistent and constansistent estimation of the varying-coefficient model with a diverging number of predictors. *Communications in Statistics - Theory and Methods* 45 (21):6385–99. doi:[10.1080/03610926.2014.890224](https://doi.org/10.1080/03610926.2014.890224).
- Zhong, J., S. Deng, J. Zhang, and Z. Feng. 2023. Covariance ratio under multiplicative distortion measurement errors. *Communications in Statistics - Theory and Methods* :1–33. doi:[10.1080/03610926.2023.2295240](https://doi.org/10.1080/03610926.2023.2295240).