1-1-2020

Two symmetric and computationally efficient Gini correlations

Courtney Vanderford
University of Mississippi

Yongli Sang
University of Louisiana at Lafayette

Xin Dang
University of Mississippi

Recommended Citation

Vanderford, C., Sang, Y., & Dang, X. (2020). Two symmetric and computationally efficient Gini correlations. Dependence Modeling, 8(1), 373–395. https://doi.org/10.1515/demo-2020-0020

DE GRUYTER Depend. Model. 2020; 8:373–395

Research Article Open Access

Courtney Vanderford, Yongli Sang, and Xin Dang*

Two symmetric and computationally efficient Gini correlations

https://doi.org/10.1515/demo-2020-0020
Received October 7, 2020; accepted November 26, 2020

Abstract: Standard Gini correlation plays an important role in measuring the dependence between random variables with heavy-tailed distributions. It is based on the covariance between one variable and the rank of the other. Hence for each pair of random variables, there are two Gini correlations and they are not equal in general, which brings a substantial difficulty in interpretation. Recently, Sang et al (2016) proposed a symmetric Gini correlation based on the joint spatial rank function with a computation cost of $O(n^2)$ where n is the sample size. In this paper, we study two symmetric and computationally efficient Gini correlations with the computational complexity of $O(n \log n)$. The properties of the new symmetric Gini correlations are explored. The influence function approach is utilized to study the robustness and the asymptotic behavior of these correlations. The asymptotic relative efficiencies are considered to compare several popular correlations under symmetric distributions with different tail-heaviness as well as an asymmetric log-normal distribution. Simulation and real data application are conducted to demonstrate the desirable performance of the two new symmetric Gini correlations.

Keywords: Asymptotic relative efficiency, computationally efficient Gini correlation, influence function, robustness, symmetric Gini correlation.

MSC subject classification: 62G35, 62G20

1 Introduction

Measuring the strength of association and correlation between two random variables is of essential importance in many research fields. Many notions of correlations have been proposed and studied [16, 21]. Perhaps the most commonly used one is Pearson's correlation coefficient which measures the linear relationship between two random variables. Pearson's correlation is computationally efficient with a computation cost of $O(n)$ where n is the sample size. It is the most statistically efficient one for normal variables; however, it is very sensitive to outliers. Even one single outlier might have a large impact on the coefficient's value and its performance [36, 37]. An important tool to study robustness is the influence function, which measures effects due to infinitesimal perturbations of the underlying distribution [13]. It has been proven that the Pearson correlation has an unbounded influence function, indicating its lack of robustness [5].

Alternatively, rank based correlations such as Spearman and Kendall's tau are robust to outliers. Kendall's tau is a similarity measure of the ranks of two random variables [17] and Spearman's correlation is the Pearson correlation coefficient evaluated on the ranks of the two variables [39]. Both values are widely used for measuring monotonic relationships. They can be computed efficiently at a cost of $O(n \log n)$ [18], and their influence functions are bounded [3]. The tradeoff to robustness is a loss of statistical efficiency in normal settings. For the correlation parameter $\rho = 0.1, 0.5, 0.9$ in the normal distribution, the asymptotic relative

Courtney Vanderford: Department of Mathematics, University of Mississippi, E-mail: cvander2@go.olemiss.edu
Yongli Sang: Department of Mathematics, University of Louisiana at Lafayette, E-mail: yongli.sang@louisiana.edu
***Corresponding Author: Xin Dang:** Department of Mathematics, University of Mississippi, E-mail: xdang@olemiss.edu

efficiencies (ARE) of Kendall's tau to the Pearson correlation are about 91%, 89% and 84%, respectively, while the ARE of the Spearman correlation are even lower [3].

Standard Gini correlations [1] are based on the covariance between one variable and the rank of the other. More specifically, let H be the joint distribution of the random variables X and Y, and let F and G be the marginal distribution functions of X and Y, respectively. The standard Gini correlations are defined as

$$\gamma_1 = \gamma(X, Y) := \frac{\mathrm{cov}(X, G(Y))}{\mathrm{cov}(X, F(X))} \quad \text{and} \quad \gamma_2 = \gamma(Y, X) := \frac{\mathrm{cov}(Y, F(X))}{\mathrm{cov}(Y, G(Y))} \tag{1}$$

reflecting different roles of X and Y. The representation of the Gini correlations indicates that they have mixed properties of those of the Pearson and Spearman correlations [39]. As expected, the statistical efficiency and robustness of Gini correlations are between those of Pearson and Spearman correlations. In terms of balance between efficiency and robustness, Gini correlations play an important role in measuring association for variables from heavy-tailed distributions [43]. The Gini correlations are computationally efficient and can be computed at a cost of $\mathrm{O}(n \log n)$ [31]. They are not symmetric in X and Y in general [31, 32], i.e., $\gamma(X, Y) \neq \gamma(Y, X)$. In some applications, this asymmetry is natural and useful [9, 12, 33]. In other scenarios, symmetry is a desired property for dependence measures. Some researchers [21, 27] even list symmetry as one of the axioms of association measures. A symmetric Gini correlation was proposed in [4, 28], which is based on the joint rank function. It is more statistically efficient than the standard Gini correlations, but it is not computationally efficient with $O(n^2)$ complexity, which means it is prohibitive for large n. Yitzhaki and Olkin [42] proposed two symmetric Gini correlations which are the arithmetic mean and geometric mean of the standard Gini correlations, respectively.

$$r_g^{(1)} = r_g^{(1)}(X, Y) := \frac{\gamma_1 + \gamma_2}{2} \quad \text{and} \quad r_g^{(2)} = r_g^{(2)}(X, Y) := \sqrt{|\gamma_1 \gamma_2|}. \tag{2}$$

Clearly those symmetric Gini correlations inherit the computational efficiency of $O(n \log n)$. However, they have not been well studied in literature except that Xu *et al.* [41] studied $r_g^{(1)}$ under the normal settings. In this paper, we systematically study the properties of these two symmetric Gini correlations and explore their statistical efficiency. Their robustness is studied by means of their influence functions. The limiting distributions of sample symmetric Gini correlations are established. It is interesting to see that there are three kinds of asymptotical sampling distribution of the sample correlation, $\hat{r}_g^{(2)}$, depending on different cases of $r_g^{(2)}$. To our best knowledge, this is a novel result and can be applied to the geometric mean type of statistics such as the symmetrized information dependence measure defined in [26].

It is worthwhile to mention that the Gini correlations in (1) and the symmetric versions in (2) are quite different from the Gini gamma or Gini coefficient [10, 24], although the names are very similar. Gini correlation γ_1 in (1) is a natural bivariate extension of univariate Gini mean difference (GMD) from the covariance representation $GMD(F) = \mathbb{E}|X_1 - X_2| = 4Cov(X, F(X))$, where X_1, X_2 are independent copies of X from F. The Gini gamma was proposed by Gini [11]. Related to the Spearman correlation in a different way, the Gini gamma is a concordance measure which is defined based on both ranks of X and Y. It is easy to check that the Gini gamma follows all axioms of concordance stated in [30]. However, neither $r_g^{(1)}$ nor $r_g^{(2)}$ is a concordance measure, and neither hold to the coherence axiom.

The paper is organized as follows. In Section 2 we provide properties of $r_g^{(1)}$ and $r_g^{(2)}$. Their influence functions are presented in Section 3. The limiting distributions of sample correlations are established in Section 4. Statistical efficiency and computational efficiency of various correlations are compared in Subsection 4.2 and their finite sample performance comparison is conducted through a simulation study on elliptical distributions and an asymmetric bivariate log-normal distribution in Section 5. A real data application on the relationship between GDP per capita and suicide rate is presented in Section 6. Final remarks are provided in Section 7. Proofs are relegated to the Appendix.

2 Two symmetric Gini correlations

Basic properties of the two symmetric Gini correlations $r_g^{(1)}$ and $r_g^{(2)}$ in (2) are explored. Their relationships with the linear correlation parameter, ρ, in bivariate elliptical distributions and log-normal distributions are presented.

2.1 General properties

Let X and Y be two random variables from F and G, respectively, with the joint distribution H.

Proposition 2.1. *Assume that H is continuous and its first moment exists, then we have*
1. $r_g^{(1)}(X, Y) = r_g^{(1)}(Y, X)$.
2. $-1 \le r_g^{(1)}(X, Y) \le 1$.
3. *If X and Y are statistically independent, then* $r_g^{(1)}(X, Y) = 0$.
4. *If Y is a monotonic increasing (decreasing) function of X, then $r_g^{(1)}(X, Y)$ equals* $+1(-1)$.
5. $r_g^{(1)}(aX + c, bY + d) = sign(ab) r_g^{(1)}(X, Y)$ *for any constants c, d and nonzero a, b.*

Proposition 2.2. *Under the same assumptions of Proposition 2.1, we have*
1. $r_g^{(2)}(X, Y) = r_g^{(2)}(Y, X)$.
2. $0 \le r_g^{(2)}(X, Y) \le 1$.
3. *If X and Y are statistically independent, then* $r_g^{(2)}(X, Y) = 0$.
4. *If Y is a monotonic function of X, then* $r_g^{(2)}(X, Y) = 1$.
5. $r_g^{(2)}(aX + c, bY + d) = r_g^{(2)}(X, Y)$ *for any constant c, d and nonzero a, b.*

The symmetry of $r_g^{(1)}$ and $r_g^{(2)}$ is obvious noting the commutative property of addition and multiplication. Properties 2-5 in the above two propositions follow simply from the properties of the original Gini correlations γ_1 and γ_2, shown by [31]. Property 5 states that the two symmetric Gini correlations describe a linear relationship between X and Y.

Note that we assume continuous H in Propositions (2.1) and (2.2). If H is not continuous, some revisions on definitions in γ_1 and γ_2 are needed for general properties. For example, replacing $F(x)$ with $(F(x) + F(x-))/2$ and $G(x)$ with $(G(x) + G(x-))/2$ in (1) keeps γ_1 and γ_2 in the range $[-1, 1]$. For simplicity, the continuous distribution is assumed throughout the paper.

Before we study the symmetric Gini correlations in elliptical distributions and lognormal distribution, we would like to provide definitions of other measures of association that will be used and compared in the paper. For H with a finite second moment, the Pearson correlation r_p is

$$r_p := \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}.$$

The rank based Spearman and Kendall's tau correlations don't need a moment condition. The Spearman correlation is defined as the Pearson correlation on the ranks of X and Y, that is,

$$r_s := r_p(F(X), G(Y)) = 12\mathbb{E}_H(F(X)G(Y)) - 3.$$

The Kendall's tau r_τ is defined as

$$r_\tau := \mathbb{E}_H\{\text{sgn}(X_1 - X_2)(Y_1 - Y_2)\} = 2P_H(X_1 - X_2)(Y_1 - Y_2) > 0) - 1,$$

where $(X_1, Y_1)^T$ and $(X_2, Y_2)^T$ are independently distributed from H.
For $\boldsymbol{Z} = (X, Y)^T$ from H with finite first moment, the joint-rank based symmetric Gini correlation $r_g^{(s)}$ [28] is defined as

$$r_g^{(s)} := \frac{\mathbb{E}_H(XS_2(\boldsymbol{Z}))}{\sqrt{\mathbb{E}_H(XS_1(\boldsymbol{Z}))}\sqrt{\mathbb{E}_H(YS_2(\boldsymbol{Z}))}} = \frac{\mathbb{E}_H(YS_1(\boldsymbol{Z}))}{\sqrt{\mathbb{E}_H(XS_1(\boldsymbol{Z}))}\sqrt{\mathbb{E}_H(YS_2(\boldsymbol{Z}))}},$$

where $\boldsymbol{S}(\boldsymbol{z}) = (S_1(\boldsymbol{z}), S_2(\boldsymbol{z}))^T = \mathbb{E}_H \frac{\boldsymbol{z}-\boldsymbol{Z}}{\|\boldsymbol{z}-\boldsymbol{Z}\|}$ is the spatial rank of $\boldsymbol{z} = (x, y)^T$ with respect to H and the norm $\|\cdot\|$ is the Euclidean norm.

Those correlations have different properties and may have different values under the same distribution. It is preferred to consider their Fisher consistent versions so that they correspond to the same quantity or same parameter [7]. For a distribution H with a parameter ρ, ρ_r is Fisher consistent for ρ if

$$\rho_r(H) = \rho.$$

We denote the Fisher consistent versions of Pearson, Spearman and Kendall's tau correlations as ρ_p, ρ_s and ρ_τ, respectively.

Next the symmetry Gini correlations as well as each of above mentioned correlation are studied in elliptical distributions and lognormal distribution.

2.2 Gini correlations in elliptical distributions

A d-variate continuous random vector $\boldsymbol{Z}$ has an elliptical distribution H if its density function is of the form

$$f(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} g\{(\boldsymbol{z}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\}, \tag{3}$$

where $\boldsymbol{\mu}$ is the location parameter, the positive definite matrix $\boldsymbol{\Sigma}$ is the scatter parameter and the nonnegative function g is the density generating function. One important property for the elliptical distribution is that the nonnegative random variable $R = \|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{Z}-\boldsymbol{\mu})\|$ is independent of $\boldsymbol{U} = \{\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{Z}-\boldsymbol{\mu})\}/R$, where $\|\cdot\|$ is the Euclidean norm and $\boldsymbol{U}$ is uniformly distributed on the unit sphere. When $d = 1$, the class of elliptical distributions coincides with the location-scale class. For $d = 2$, let $\boldsymbol{Z} = (X, Y)^T$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$, then the corresponding *linear correlation coefficient* of X and Y is

$$\rho = \rho(X, Y) := \frac{\sigma_{12}}{\sigma_1 \sigma_2}. \tag{4}$$

Conventionally, we write the parameters of bivariate elliptical distributions as $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

If second moment of $\boldsymbol{Z}$ exists, then the covariance matrix exists and is equal to $\frac{\mathbb{E}R^2}{d}\boldsymbol{\Sigma}$. In this case, the Pearson correlationr_p is well defined and is equal to the parameter ρ. More details on the elliptical distribution family refer to [6].

Note that under bivariate elliptical distributions, $\gamma_1 = \gamma_2 = \rho$ [28, 31]. Consequently, we have the relationships between $r_g^{(i)}$, $i = 1, 2$, and ρ as follows.

Proposition 2.3. *For bivariate elliptical distributions with finite first moments, we have* $r_g^{(1)} = \rho$ *and* $r_g^{(2)} = |\rho|$.

If $\sigma_1 = \sigma_2$, the joint-rank based Gini correlation $r_g^{(s)}$ proposed in [28] has the following relationship with ρ.

$$r_g^{(s)} = k(\rho) = \begin{cases} \rho, & \rho = 0, \pm 1, \\ \dfrac{1}{\rho} + \dfrac{\rho - 1}{\rho} \dfrac{EK(\frac{2\rho}{\rho+1})}{EE(\frac{2\rho}{\rho+1})}, & otherwise, \end{cases} \tag{5}$$

where $EK(x) = \int_0^{\pi/2} 1/\sqrt{1 - x^2 \sin^2\theta}\, d\theta$ and $EE(x) = \int_0^{\pi/2} \sqrt{1 - x^2 \sin^2\theta}\, d\theta$ are the complete elliptic integral of the first kind and the second kind, respectively. The Fisher consistent version of $r_g^{(s)}$ is hard to obtain an explicit form but a numerical solution is possible.

For Kendall's tau, Blomqvist [2] proved that $r_\tau = 2/\pi \arcsin(\rho)$ in the normal case. Lindskog *et al.* [20] proved that this such relationship holds under all elliptical distributions in general. Hence the Fisher consistent version of Kendall's correlation is

$$\rho_\tau = \sin\left(\frac{\pi}{2} r_\tau\right). \tag{6}$$

Under elliptical distributions, the Spearman correlation $r_s = 6/\pi \arcsin(\rho/2)$, the result obtained by [22] for the normal case. Then the Fisher consistent version of Spearman correlation is

$$\rho_s = 2\sin\left(\frac{\pi}{6} r_s\right). \tag{7}$$

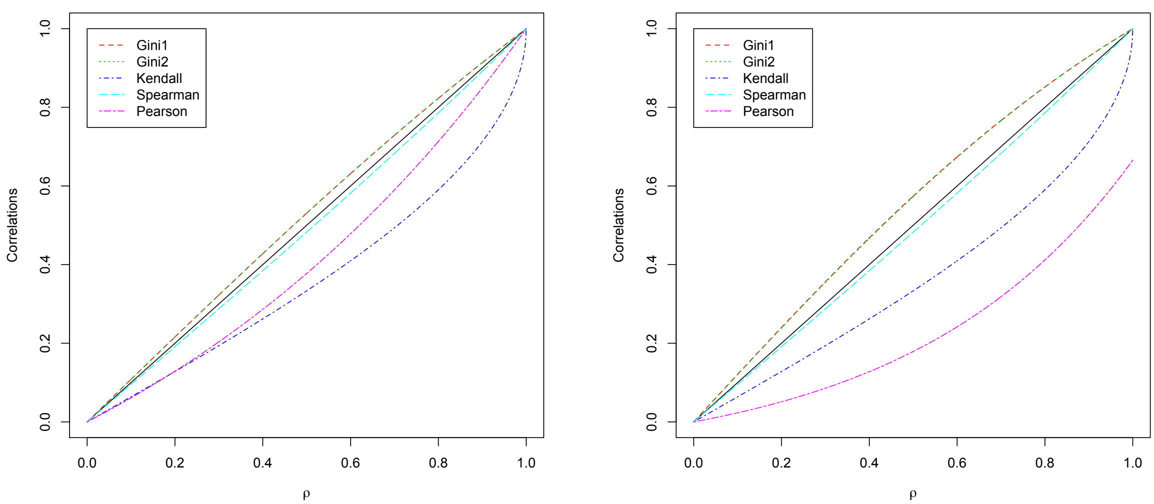


Figure 1: Various correlations in lognormal distributions with parameters $(\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho)$ in the left plot and with parameters $(\mu_1 = \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 2, \rho)$ in the right plot. The black solid line is the straight line with slope 1 passing through the origin.

2.3 Gini correlations in bivariate lognormal distribution

The random vector, $(X, Y)^T$, is said to have a bivariate lognormal distribution with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ if $(\log X, \log Y)^T$ follows a bivariate normal distribution with the same parameters.

Clearly, Kendall's tau and Spearman correlation are invariant under monotonically increasing transformations, thus equations (6) and (7) still hold. For the Pearson correlation, it is easy to have

$$r_p = \frac{\exp(\rho\sigma_1\sigma_2) - 1}{\sqrt{\{\exp(\sigma_1^2) - 1\}\{\exp(\sigma_2^2) - 1\}}}. \tag{8}$$

Then the Fisher consistent version of Pearson correlation for the parameter ρ in the lognormal distribution is

$$\rho_p = \frac{\log\left(r_p\sqrt{\exp(\sigma_1^2) - 1}\sqrt{\exp(\sigma_2^2) - 1}\right) + 1}{\sigma_1\sigma_2}. \tag{9}$$

For the two new symmetric Gini correlations, we have derived the functional relationships as below.

Proposition 2.4. *Under the bivariate lognormal distribution with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, we have*

$$r_g^{(1)} = \frac{1}{2}\left(\frac{2\Phi(\rho\sigma_1/\sqrt{2}) - 1}{2\Phi(\sigma_1/\sqrt{2}) - 1} + \frac{2\Phi(\rho\sigma_2/\sqrt{2}) - 1}{2\Phi(\sigma_2/\sqrt{2}) - 1}\right), \tag{10}$$

$$r_g^{(2)} = \sqrt{\frac{|2\Phi(\rho\sigma_1/\sqrt{2}) - 1|}{|2\Phi(\sigma_1/\sqrt{2}) - 1|}\frac{|2\Phi(\rho\sigma_2/\sqrt{2}) - 1|}{|2\Phi(\sigma_2/\sqrt{2}) - 1|}}, \tag{11}$$

where Φ is the cdf of the standard normal variable. Further, if $\sigma_1 = \sigma_2 = \sigma$, the Fisher consistent version of symmetric Gini correlations are

$$\rho_g^{(1)} = \frac{\sqrt{2}}{\sigma}\Phi^{-1}\left(\frac{r_g^{(1)}(2\Phi(\sigma/\sqrt{2}) - 1) + 1}{2}\right), \tag{12}$$

$$\rho_g^{(2)} = \frac{\sqrt{2}}{\sigma}\Phi^{-1}\left(\frac{sgn(\rho)r_g^{(2)}(2\Phi(\sigma/\sqrt{2}) - 1) + 1}{2}\right). \tag{13}$$

The proposition states that explicit forms of the Fisher consistent symmetric Gini correlations are only available for the homogeneous case. Also (13) indicates that the Fisher consistent version of $r_g^{(2)}$ requires information of the sign of ρ. If $\sigma_1 \neq \sigma_2$, we need a numerical method to approximate them.

Plots in Fig. 1 display the relationship of various correlations to the parameter ρ in the lognormal distributions. In the left plot, $\sigma_1 = \sigma_2 = 1$, we have $r_g^{(1)} = r_g^{(2)} > \rho > r_s > r_p > r_\tau$ if $0 < \rho < 1$, otherwise they are equal at 0 and 1. On the right with $\sigma_1 = 1$ and $\sigma_2 = 2$, if $0 < \rho < 1$, then $r_g^{(1)} > r_g^{(2)}$, though the differences between $r_g^{(1)}$ and $r_g^{(2)}$ are tiny and unnoticeable in the plot. Also we have $r_g^{(2)} > r_s > r_\tau > r_p$. Note that the Pearson correlation, r_p, can not reach 1 when $\sigma_1 \neq \sigma_2$. The maximum value in the plot above is 0.6642169 when $\rho = 1$. From Equation (8), it is easy to prove that $r_p < 1$ for $\rho = 1$ if $\sigma_1 \neq \sigma_2$. In other words, for a normal random variable X and a positive constant $a \neq 1$, $r_p(exp(X), exp(aX)) < 1$, meaning that the Pearson correlation is not suitable to describe nonlinear relationships.

3 Influence function

The influence function (IF) introduced by Hampel [13] is now a standard tool which serves two purposes. The first is to measure local robustness for effects on estimators due to infinitesimal perturbations of distribution functions. The second is to derive limiting distributions and asymptotic variances. See also [14]. For a cdf H on $\mathbb{R}^d$ and a functional $T : H \mapsto T(H) \in \mathbb{R}^m$ with $m \geq 1$, the IF of T at H is defined as

$$\mathrm{IF}(\boldsymbol{z}; T, H) = \lim_{\epsilon \downarrow 0} \frac{T((1-\epsilon)H + \epsilon\delta_{\boldsymbol{z}}) - T(H)}{\epsilon}, \quad \boldsymbol{z} \in \mathbb{R}^d,$$

where $\delta_{\boldsymbol{z}}$ denotes the point mass distribution at $\boldsymbol{z}$. Under regularity conditions on T (see [14, 34] for details), we have $\mathbb{E}_H\{\mathrm{IF}(\boldsymbol{Z}; T, H)\} = \boldsymbol{0}$ and the von Mises expansion

$$T(H_n) - T(H) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(\boldsymbol{z}_i; T, H) + o_p(n^{-1/2}), \tag{14}$$

where H_n denotes the empirical distribution based on a sample $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n$. This representation shows the connection between the IF and the robustness of T, observation by observation. Further, (14) yields the asymptotic m-variate normality of $T(H_n)$,

$$\sqrt{n}(T(H_n) - T(H)) \xrightarrow{d} N(\boldsymbol{0}, \mathbb{E}_H(\mathrm{IF}(\boldsymbol{Z}; T, H)\mathrm{IF}(\boldsymbol{Z}; T, H)^T)). \tag{15}$$

We first derive the influence functions for the standard Gini correlations γ_1 and γ_2, which are stated in the following proposition.

Proposition 3.1. *For any continuous bivariate distribution H with finite first moment, the influence functions of the traditional Gini correlations are given by*

$$IF((u,v)^T; \gamma_1, H) = \gamma_1\left(\frac{(u - \mathbb{E}X)[G(v) - \mathbb{E}G(Y)]}{cov(X, G(Y))} - \frac{(u - \mathbb{E}X)[F(u) - \mathbb{E}F(X)]}{cov(X, F(X))}\right),$$

$$IF((u,v)^T;\gamma_2,H) = \gamma_2\left(\frac{(v-\mathbb{E}Y)[F(u)-\mathbb{E}F(X)]}{cov(Y,F(X))} - \frac{(v-\mathbb{E}Y)[G(v)-\mathbb{E}G(Y)]}{cov(Y,G(Y))}\right).$$

The influence functions of the standard Gini correlations are approximately linear in u and v. Comparing with the quadratic effects of the Pearson correlation coefficient [5],

$$\mathrm{IF}((u,v)^T;r_p,H) = \frac{(u-\mathbb{E}X)(v-\mathbb{E}Y)}{\sigma_X\sigma_Y} - \frac{1}{2}r_p\left[\frac{(u-\mathbb{E}X)^2}{\sigma_X^2} + \frac{(v-\mathbb{E}Y)^2}{\sigma_Y^2}\right],$$

γ_1 and γ_2 are more robust than the Pearson correlation. However, they are not strictly robust since their influence functions are unbounded. Kendall's tau, r_τ, and Spearman correlation, r_s, have bounded influence functions [3], which are

$$\begin{aligned}\mathrm{IF}((u,v)^T;r_\tau,H) &= 2\{2P_H[(u-X)(v-Y)>0]-1-r_\tau\},\\ \mathrm{IF}((u,v)^T;r_s,H) &= -3r_s-9+12(F(u)G(v)+\mathbb{E}(F(X)I(Y>v))+\mathbb{E}(G(Y)I(X\geq u)).\end{aligned}$$

In this sense, the standard Gini correlations are more robust than r_p but less robust than r_τ and r_s.

Proposition 3.2. *For any continuous distribution H with finite first moment, the influence functions of $r_g^{(1)}$ and $r_g^{(2)}$ are given by*

$$IF((u,v)^T;r_g^{(1)},H) = \frac{1}{2}IF((u,v)^T;\gamma_1,H) + \frac{1}{2}IF(u,v)^T;\gamma_2,H)$$

$$IF((u,v)^T;r_g^{(2)},H) = \begin{cases}\frac{sgn(\gamma_1\gamma_2)}{2r_g^{(2)}}(\gamma_2 IF((u,v)^T;\gamma_1,H)+\gamma_1 IF((u,v)^T;\gamma_2,H)) & if \quad r_g^{(2)}\neq 0;\\ does\ not\ exist, & if \quad r_g^{(2)}=0.\end{cases}$$

Since the square root function is not differentiable at zero, the influence function of $r_g^{(2)}$ does not exist when $r_g^{(2)}=0$. This brings difficulty in deriving the limiting distribution of sample $\hat{r}_g^{(2)}$ when $r_g^{(2)}=0$, as explained further in a later section. The influence function of $r_g^{(1)}$ and that of nonzero $r_g^{(2)}$ are linear combinations of the influence functions of γ_1 and γ_2, and hence are approximately linear in u and v. The symmetric Gini correlation $r_g^{(s)}$ proposed in [28] also has an approximately linear influence function. We expect that the newly studied Gini correlations and the symmetric one based on the joint rank perform similarly in terms of robustness and statistical efficiency.

In Figure 2, we demonstrate the influence functions of r_p, r_τ, $r_g^{(s)}$ and $r_g^{(1)}$ and $r_g^{(2)}$ under the bivariate normal distribution with $\mu_1=\mu_2=0$, $\sigma_1=\sigma_2=1$ and $\rho=0.5$. Since we know that $r_g^{(1)}=\rho$ and $r_g^{(2)}=|\rho|$ for bivariate normal distributions, the influence functions for the two Gini correlations are identical for $\rho=0.5$, and thus share the same plot in Figure 2. Indeed under a general elliptical distribution, $\mathrm{IF}((u,v)^T;r_g^{(2)},H)=\mathrm{IF}((u,v)^T;r_g^{(1)},H)$ for $\rho>0$ and $\mathrm{IF}((u,v)^T;r_g^{(2)},H)=-\mathrm{IF}((u,v)^T;r_g^{(1)},H)$ for $\rho<0$. Note that scales of the value of the influence functions in the four plots are quite different.

4 Estimation

Estimation of the two new symmetric Gini correlations can be done easily by plugging in estimators $\hat{\gamma}_1$ and $\hat{\gamma}_2$ of γ_1 and γ_2, respectively. Given a random sample $\mathcal{Z}=\{\boldsymbol{Z}_1,\boldsymbol{Z}_2,\ldots,\boldsymbol{Z}_n\}$ with $\boldsymbol{Z}_i=(X_i,Y_i)^T$, the traditional Gini correlations γ_1 and γ_2 can be estimated by a ratio of U-statistics. That is,

$$\hat{\gamma}_1 = \frac{U_1}{U_2} = \frac{2/[n(n-1)]\sum_{1\leq i<j\leq n}h_1((X_i,Y_i),(X_j,Y_j))}{2/[n(n-1)]\sum_{1\leq i<j\leq n}h_2((X_i,Y_i),(X_j,Y_j))}, \tag{16}$$

$$\hat{\gamma}_2 = \frac{U_3}{U_4} = \frac{2/[n(n-1)]\sum_{1\leq i<j\leq n}h_3((X_i,Y_i),(Y_j,X_j))}{2/[n(n-1)]\sum_{1\leq i<j\leq n}h_4((X_i,Y_i),(X_j,Y_j))}, \tag{17}$$

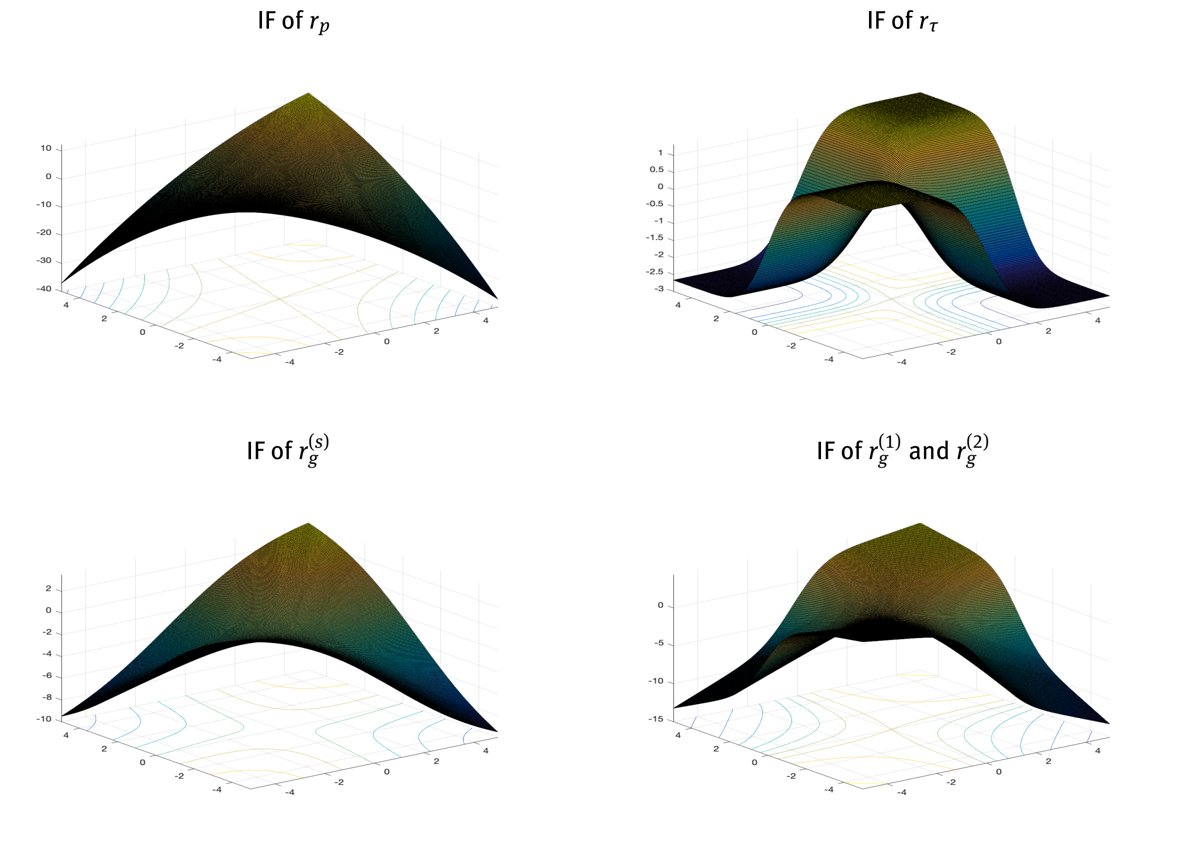


Figure 2: Influence functions of correlation coefficients r_p, r_τ, $r_g^{(s)}$, $r_g^{(1)}$ and $r_g^{(2)}$ for the bivariate normal distribution with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0.5$.

where $h_1\big((x_1, y_1), (x_2, y_2)\big) = h_3\big((y_1, x_1), (y_2, x_2)\big) = 1/4[(x_1 - x_2)I(y_1 > y_2) + (x_2 - x_1)I(y_2 > y_1)]$ and $h_2\big((x_1, y_1), (x_2, y_2)\big) = h_4\big((y_1, x_1), (y_2, x_2)\big) = 1/4|x_1 - x_2|$. Schechtman and Yitzhaki [31] applied U-statistics theorem to establish consistency and asymptotic normality of $\hat{\gamma}_1$ and $\hat{\gamma}_2$. The same result can be reached through the influence function approach which is derived in Proposition 3.1. More specifically, for H with finite second moment,

$$\sqrt{n}(\hat{\gamma}_1 - \gamma_1) \xrightarrow{d} N(0, v_{\gamma_1}) \quad \text{and} \quad \sqrt{n}(\hat{\gamma}_2 - \gamma_2) \xrightarrow{d} N(0, v_{\gamma_2}) \qquad \text{as } n \to \infty,$$

where the asymptotic variances v_{γ_1} and v_{γ_2} are $\mathbb{E}[\mathrm{IF}((X, Y)^T; \gamma_1, H)^2]$ and $\mathbb{E}[\mathrm{IF}((X, Y)^T; \gamma_2, H)^2]$, respectively. For a bivariate normal distribution, Xu *et al.* [41] provided an explicit formula $v_{\gamma_1} = v_{\gamma_2} = \pi/3 + (\pi/3 + 4\sqrt{3})\rho^2 - 4\rho \arcsin(\rho/2) - 4\rho^2\sqrt{4 - \rho^2}$.

Note that a direct computation of U-statistics in (16) and (17) is time-intensive with complexity $O(n^2)$. Rewriting U_1 and U_2 as linear combinations of order statistics reduces the computation to $O(n \log n)$ [31]. That is,

$$U_1 = \frac{1}{4\binom{n}{2}} \sum_{i=1}^{n} (2i - 1 - n) X_{(Y_{(i)})} \quad \text{and} \quad U_2 = \frac{1}{4\binom{n}{2}} \sum_{i=1}^{n} (2i - 1 - n) X_{(i)},$$

where $X_{(i)}$ is the i^{th} order statistic of $X_1, X_2, \ldots, X_n$ and $X_{(Y_{(i)})}$ is the X corresponding to the order statistic $Y_{(i)}$. Similarly, U_3 and U_4 are linear combinations of order statistics. This provides computational efficiency for $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

Thus, we have computationally efficient estimators for $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(2)}$; $\hat{r}_g^{(1)}$ is the arithmetic mean of $\hat{\gamma}_1$ and $\hat{\gamma}_2$, while $\hat{r}_g^{(2)}$ is the geometric mean of $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

$$\hat{r}_g^{(1)} = \frac{\hat{\gamma}_1 + \hat{\gamma}_2}{2}, \quad \hat{r}_g^{(2)} = \sqrt{|\hat{\gamma}_1 \hat{\gamma}_2|}, \tag{18}$$

which are continuous functions of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ and they can be efficiently calculated in $O(n \log n)$ of time. The strong consistency of $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(2)}$ follows directly from the strong consistency of $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

Proposition 4.1. *Let $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$ be a random sample from a continuous bivariate distribution H with finite first moment. Then $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(2)}$ given in (18) converge almost surely to $r_g^{(1)}$ and $r_g^{(2)}$, respectively.*

4.1 Limiting distributions

To simplify the presentation, we denote

$$\begin{aligned}
\delta_1(X, Y) &= \frac{(X - \mathbb{E}X)[G(Y) - \mathbb{E}G(Y)]}{cov(X, G(Y))} - \frac{(X - \mathbb{E}X)[F(X) - \mathbb{E}F(X)]}{cov(X, F(X))}, \\
\delta_2(X, Y) &= \frac{(Y - \mathbb{E}Y)[F(X) - \mathbb{E}F(X)]}{cov(Y, F(X))} - \frac{(Y - \mathbb{E}Y)[G(Y) - \mathbb{E}G(Y)]}{cov(Y, G(Y))}.
\end{aligned}$$

With the influence function derived in Proposition 3.2, we can easily obtain the asymptotic normality of $\hat{r}_g^{(1)}$.

Proposition 4.2. *Let $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$ be a random sample from 2-dimensional distribution H with finite second moment. As $n \to \infty$,*

$$\sqrt{n}(\hat{r}_g^{(1)} - r_g^{(1)}) \xrightarrow{d} N(0, v_{g_1}),$$

where $v_{g_1} = \mathbb{E}[IF((X, Y)^T; r_g^{(1)}, H)^2] = 1/4\mathbb{E}\left[\{\gamma_1 \delta_1(X, Y) + \gamma_2 \delta_2(X, Y)\}^2\right]$.

For a bivariate normal distribution, Xu *et al.* [41] provided an explicit formula of v_{g_1} to be $v_{g_1} = (1 - \rho^2)(\pi/6 - \rho \arcsin(\rho/2) + (1 - \rho^2))/\sqrt{4 - \rho^2}$, which is smaller than v_{γ_1}, the asymptotic variance of $\hat{\gamma}_1$. This means that the symmetric Gini correlation is more statistically efficient than the standard Gini correlation under normal distributions.

Under the lognormal distribution, asymptotic normality of the Fisher consistent estimator $\hat{\rho}_g^{(1)}$ is obtained by the Delta method. Its asymptotic variance is $k_1(\rho)^{-2} v_{g_1}$, where

$$k_1(\rho) = \frac{\partial r_g^{(1)}}{\partial \rho} = \frac{\psi(\rho\sigma_1/\sqrt{2})\sigma_1/\sqrt{2}}{2\Phi(\sigma_1/\sqrt{2}) - 1} + \frac{\psi(\rho\sigma_2/\sqrt{2})\sigma_2/\sqrt{2}}{2\Phi(\sigma_2/\sqrt{2}) - 1}$$

with ψ and Φ being the pdf and cdf of the standard normal random variable, respectively.

To study the asymptotic behavior of $\hat{r}_g^{(2)}$, we have to overcome the difficulty brought about by the nonexistence of the influence function when $r_g^{(2)} = 0$. It is interesting to see that there are three different limiting distributions of $\hat{r}_g^{(2)}$, corresponding to three cases of $r_g^{(2)}$. We present the results in the following two propositions.

For $r_g^{(2)} \neq 0$, the influence function of $r_g^{(2)}$ exists and can be used to establish the asymptotic normality of $\hat{r}_g^{(2)}$ and calculate its asymptotic variance.

Proposition 4.3. *Let $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$ be a random sample from 2-dimensional distribution H with finite second moment. When $r_g^{(2)} \neq 0$ and as $n \to \infty$,*

$$\sqrt{n}(\hat{r}_g^{(2)} - r_g^{(2)}) \xrightarrow{d} N(0, v_{g_2}),$$

where $v_{g_2} = \mathbb{E}[IF((X, Y)^T; r_g^{(2)}, H)^2] = \frac{|\gamma_1\gamma_2|}{4}\mathbb{E}[\{\delta_1(X, Y) + \delta_2(X, Y)\}^2]$.

Remark 4.1. *If $\gamma_1 = \gamma_2 \neq 0$, we have $v_{g_1} = v_{g_2}$, meaning that two estimators $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(2)}$ have the same statistical efficiency.*

If $r_g^{(2)} = 0$, the influence function of $r_g^{(2)}$ does not exist, and hence we have to rely on U-statistic theory to derive the limiting distributions of $\hat{r}_g^{(2)}$. There are two different cases resulting from $r_g^{(2)} = 0$, depending on whether or not both γ_1 and γ_2 are zero. Without loss of generality, we assume $\gamma_1 = 0$ and the two cases correspond to $\gamma_2 = 0$ and $\gamma_2 \neq 0$, respectively.

Proposition 4.4. *Let $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$ be a random sample from 2-dimensional distribution H with finite second moment. When $r_g^{(2)} = 0$, we have*

1. *If $\gamma_2 \neq 0$, $\hat{r}_g^{(2)}$ converges to the square root of a folded normal random variable. That is,*

$$n^{1/4}\hat{r}_g^{(2)} \xrightarrow{d} \sqrt{|Z|},$$

where Z is a normal random variable with mean zero and variance given in the proof.

2. *If $\gamma_2 = 0$, we have*

$$\sqrt{n}\hat{r}_g^{(2)} \xrightarrow{d} \frac{24}{\sqrt{\Delta_1\Delta_2}}\sqrt{\left|\sum_{s=1}^{\infty}\lambda_s(\chi_{1s}^2 - 1)\right|},$$

where $\Delta_1 = 4Cov(X, F(X))$ and $\Delta_2 = 4Cov(Y, G(Y))$ are Gini's mean differences for F and G, respectively, χ_{1s}^2 $(s = 1, 2, \ldots)$ are independent χ_1^2 variables and $\{\lambda_s\}$ $(s = 1, 2, \ldots)$ are coefficients given in the proof.

4.2 Asymptotic relative efficiency

We compare the asymptotic efficiency of the symmetric Gini correlations with other correlations under elliptical distributions and lognormal distributions. We consider Fisher consistent estimators. Note that the purpose here is not to estimate parameter ρ, which is usually provided by likelihood inference. Rather, the Fisher consistent correlation coefficients estimate the same parameter and hence their asymptotic variances and statistical efficiencies are comparable. Denote $\hat{\rho}_g^{(1)}, \hat{\rho}_g^{(2)}, \hat{\rho}_g^{(s)}, \hat{\rho}_\gamma, \hat{\rho}_\tau$ and $\hat{\rho}_p$ as corresponding estimators of symmetric Gini, standard Gini γ_1, Kendall's tau and Pearson correlations. The asymptotic variances of those estimators are derived by the Delta method.

We consider three elliptical distributions with the same parameters $(\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0.1, 0.5, 0.9)$ but different fatness on the tail regions, which are

- normal distribution with $g(t) = 1/(2\pi)e^{-t/2}$,
- t-distributions with $g(t) = 1/(2\pi)(1 + t/\nu)^{-\nu/2-1}$, where $\nu = 5, 15$ is the degrees of freedom,
- Kotz type distribution with $g(t) = 1/(2\pi)e^{-\sqrt{t}}$.

Bivariate lognormal distributions with parameters $(\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0.1, 0.5, 0.9)$ and $(\mu_1 = \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 2, \rho = 0.1, 0.5, 0.9)$ are also considered.

We compute the asymptotic variances (ASV) of the Pearson estimators $\hat{\rho}_p$, and asymptotic relative efficiencies (ARE) of estimators $\hat{\rho}_g^{(1)}$, $\hat{\rho}_g^{(2)}$, $\hat{\rho}_g^{(s)}$, $\hat{\rho}_\gamma$, and $\hat{\rho}_\tau$ relative to $\hat{\rho}_p$, which are reported in the first part of Table 1. The asymptotic relative efficiency (ARE) of one estimator $\hat{\rho}_1$ with respect to another $\hat{\rho}_2$ is defined by

$$ARE(\hat{\rho}_1, \hat{\rho}_2) = ASV(\hat{\rho}_2)/ASV(\hat{\rho}_1).$$

The second part of Table 1 lists ASV of all correlations under the lognormal distribution with $\sigma_1 = 1$ and $\sigma_2 = 2$. In this case, the Pearson correlation has extremely large asymptotic variances, the result agreeing well with [19, 23]. The asymptotic variance of $\hat{r}_p$ involves the fourth moment and is given by Witting and Müller-Funk ([40]) as follows.

$$v_p = (1 + \frac{r_p^2}{2})\frac{\sigma_{22}}{\sigma_{20}\sigma_{02}} + \frac{r_p^2}{4}(\frac{\sigma_{40}}{\sigma_{20}^2} + \frac{\sigma_{04}}{\sigma_{02}^2} - \frac{4\sigma_{31}}{\sigma_{11}\sigma_{20}} - \frac{4\sigma_{13}}{\sigma_{11}\sigma_{02}}),$$

Dist	ρ	$\hat\rho_g^{(1)}$	$\hat\rho_g^{(2)}$	$\hat\rho_g^{(s)}$	$\hat\rho_{\gamma_1}$	$\hat\rho_\tau$	ASV($\hat\rho_p$)
	0.1	0.9776	0.9776	0.9321	0.9558	0.9111	0.9816
Normal	0.5	0.9570	0.9570	0.9769	0.9398	0.8915	0.5631
	0.9	0.9053	0.9053	0.9601	0.9004	0.8439	0.0361
	0.1	1.0505	1.0505	1.0182	1.0304	1.0146	1.1558
$t(15)$	0.5	1.0230	1.0230	1.0560	0.9852	0.9896	0.6643
	0.9	0.9564	0.9564	1.0289	0.9468	0.8804	0.0427
	0.1	2.0233	2.0233	2.0095	1.9502	2.2586	2.8800
$t(5)$	0.5	1.8646	1.8646	1.9795	1.7666	2.1060	1.5961
	0.9	1.5665	1.5665	1.8629	1.5346	1.7940	0.1019
	0.1	1.3539	1.3539	1.2081	1.1385	1.2171	1.6382
Kotz	0.5	1.0732	1.0732	1.1850	1.0854	1.1510	0.9378
	0.9	0.9882	0.9882	1.1599	0.9789	1.0256	0.0602
	0.1	4.1136	4.1136	N/A	2.1825	3.4713	3.7341
Lognormal	0.5	9.8414	9.8414	N/A	6.0519	9.7859	6.1741
($\sigma_1 = \sigma_2 = 1$)	0.9	14.630	14.630	N/A	12.491	21.052	0.9006
		ASV($\hat\rho_g^{(1)}$)	ASV($\hat\rho_g^{(2)}$)	ASV($\hat\rho_g^{(s)}$)	ASV($\hat\rho_{\gamma_1}$)	ASV($\hat\rho_\tau$)	ASV($\hat\rho_p$)
	0.1	4.9484	4.2810	N/A	1.7109	1.0774	18862
Lognormal	0.5	2.7026	2.4453	N/A	1.0202	0.6316	227142
($\sigma_1 = 1; \sigma_2 = 2$)	0.9	0.1858	0.1820	N/A	0.0721	0.0428	379866

Table 1: Asymptotic relative efficiencies (ARE) of estimators $\hat\rho_g^{(1)}, \hat\rho_g^{(2)}, \hat\rho_g^{(s)}, \hat\rho_{\gamma_1}$ and $\hat\rho_\tau$ relative to $\hat\rho_p$ for different distributions, with asymptotic variance (ASV($\hat\rho_p$)) of Pearson estimator $\hat\rho_p$. The second part is ASV of all correlations under the lognormal distribution with $\sigma_1 = 1$ and $\sigma_2 = 2$.

where $\sigma_{kl} = \mathbb{E}[(X-\mathbb{E}X)^k(Y-\mathbb{E}Y)^l]$. For the lognormal case of $\rho = 0.5$, $\sigma_1 = 1$ and $\sigma_2 = 2$, we have $v_p = 72895.7$ and using the Delta method, the ASV of the Fisher consistent Pearson correlation $\hat\rho_p$ is v_p multiplied by 3.12.

Since we have yet to determine the relationship between ρ and $r_g^{(s)}$ for the lognormal distribution, the asymptotic relative efficiencies of $\hat\rho_g^{(s)}$ under the lognormal distribution are not presented in this paper. Note that by Remark 4.1, we have $\gamma_1 = \gamma_2$ and hence the ASV's of $\hat\rho_g^{(1)}$ and $\hat\rho_g^{(2)}$ are same for all cases except for the second setup of the lognormal distribution. In that case, $\hat\rho_g^{(2)}$ is 15%, 10% and 2% more efficient than $\hat\rho_g^{(1)}$ for $\rho = 0.1, 0.5$ and 0.9, respectively.

Table 1 shows that the asymptotic variances of $\hat\rho_p, \hat\rho_g^{(1)}, \hat\rho_g^{(2)}, \hat\rho_g, \hat\rho_\gamma$, and $\hat\rho_\tau$ all decrease as ρ increases in elliptical distributions. Asymptotic variances increase for t distributions as the degrees of freedom ν decrease. Under normal distributions, the Pearson correlation estimator is the maximum likelihood estimator of ρ, thus is the most efficient asymptotically. The two proposed symmetric Gini estimators $\hat\rho_g^{(1)}, \hat\rho_g^{(2)}$ are both high in efficiency with ARE's greater than 90 percent; thus, more efficient than Kendall's estimator $\hat\rho_\tau$ and the traditional Gini correlation estimator $\hat\rho_\gamma$. For heavy-tailed elliptical distributions, symmetric Gini estimators $\hat\rho_g^{(1)}$ and $\hat\rho_g^{(2)}$ are more efficient than Pearson's estimator $\hat\rho_p$. They are also more efficient than the traditional Gini correlation in all elliptical distributions. The rank based symmetric Gini correlation $\hat\rho_g^{(s)}$ has a similar efficiency as $\hat\rho_g^{(1)}$ and $\hat\rho_g^{(2)}$, but it has a slight advantage when $\rho = 0.5$ and 0.9. Under the lognormal distribution with $\sigma_1 = \sigma_2 = 1$, $\hat\rho_g^{(1)}$ and $\hat\rho_g^{(2)}$ are competitive with Kendall's tau. Under the case of $\sigma_1 = 1, \sigma_2 = 2$ however, a large variation in Y will degrade the performance of $\hat\gamma_2$ and consequently $\hat\rho_g^{(1)}$ and $\hat\rho_g^{(2)}$. ASV of Kendall's tau is the most efficient in this case.

5 Empirical Results

We first conduct a small simulation to compare computational efficiency of each correlation. Then we compare finite sample statistical efficiency of these methods.

5.1 Computational efficiency

To study the computational efficiency of these methods among finite samples, we perform a small simulation to compare the calculation times of the two symmetric Gini correlation estimators $\hat{r}_g^{(1)}$, $\hat{r}_g^{(2)}$ with Kendall's tau $\hat{r}_\tau$, Spearman $\hat{r}_s$, and Pearson $\hat{r}_p$ correlation estimators, as well as the symmetric Gini correlation estimator $\hat{r}_g^{(s)}$. Samples of sizes $n = 10000$, 50000 and 100000 were drawn from a bivariate Normal distribution with parameters ($\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0$). For each sample, the computation times of each correlation measure were recorded. The procedure is then repeated 30 times to procure the mean and standard deviation of computation times for each measure. In Table 2, we display the mean and standard deviation (in parenthesis) of calculation times for $\hat{r}_g^{(1)}$, $\hat{r}_g^{(2)}$, $\hat{r}_g^{(s)}$, $\hat{\tau}$, $\hat{r}_s$, and $\hat{r}_p$. The values in Table 2 were achieved on a Windows PC with an Intel ® Core™ i7-9700K CPU @ 3.60GHz, 8 cores. The R package "pcaPP" is used for fast computation of Kendall's tau correlation.

n	$r_g^{(1)}$	$r_g^{(2)}$	$r_g^{(s)}$	r_τ	r_s	r_p
10, 000	.004 (.0072)	.001 (.0040)	.390 (.0061)	.000 (.0000)	.002 (.0063)	.000 (.0000)
50, 000	.007 (.0079)	.008 (.0086)	9.75 (.0286)	.005 (.0078)	.011 (.0084)	.000 (.0000)
100, 000	.016 (.0061)	.013 (.0076)	39.6 (.4872)	.008 (.0083)	.024 (.0093)	.002 (.0053)

Table 2: The mean and standard deviation (in parenthesis) of calculation times for $\hat{r}_g^{(1)}$, $\hat{r}_g^{(2)}$, $\hat{r}_g^{(s)}$, $\hat{\tau}$, $\hat{r}_s$, and $\hat{r}_p$ under a bivariate Normal distribution.

From the complexity study, we know that $\hat{r}_g^{(1)}$, $\hat{r}_g^{(2)}$, $\hat{\tau}$, and $\hat{r}_s$ all have calculation times of $O(n \log n)$, $\hat{r}_g^{(s)}$ has a calculation time of $O(n^2)$, and $\hat{r}_p$ has a calculation time of $O(n)$. In Table 2, we can see that $\hat{r}_p$ is the most computationally efficient, with $\hat{r}_g^{(1)}$, $\hat{r}_g^{(2)}$, $\hat{\tau}$, and $\hat{r}_s$ being only slightly less efficient. It is clear from Table 2 that all of $\hat{r}_g^{(1)}$, $\hat{r}_g^{(2)}$, $\hat{r}_p$, $\hat{\tau}$, and $\hat{r}_s$ would perform well with most all sample sizes, however, $\hat{r}_g^{(s)}$ would not perform well with large samples.

5.2 Finite sample efficiency

In order to study the efficiency of these methods among finite samples, we conduct a small simulation comparing the two symmetric Gini correlations with Kendall's τ, Spearman, and Pearson correlation estimators. Samples of sizes $n = 30$ and $n = 300$ were drawn from 4 t-distributions with degrees of freedom 1, 5, 15, and ∞, and from the Kotz and Lognormal distributions. Let $\boldsymbol{\mu} = (0, 0)^T$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ be the parameters. The R Package "mnormt" was used to generate data from the multivariate t distributions, bivariate normal distribution and the lognormal distribution by taking the exponential transformation of a bivariate normal random sample. We generate data from the Kotz distribution by first obtaining uniformly distributed random vectors on the unit circle by $\mathbf{u} = (\cos\theta, \sin\theta)^T$ with θ in $[0, 2\pi]$, then generate r from a Gamma distribution with shape parameter α=2, and scale parameter β=1. Thus, we obtain $\Sigma^{1/2} r\mathbf{u} + \mu$, a sample from a bivariate Kotz(μ, Σ) distribution.

An estimator $\hat{\rho}^{(m)}$ is computed for the m^{th} sample and the root mean squared error (RMSE) is used for a criterion for assessing estimators, which is defined as

$$RMSE(\hat{\rho}) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\rho}^{(m)} - \rho)^2}.$$

In our experiment, M is set to be 3000. The procedure is then repeated 30 times to procure the mean and standard deviation of $\sqrt{n}$RMSE. In Table 3, we display the mean and standard deviation (in parenthesis) of $\sqrt{n}$RMSE of $\hat{\rho}_g^{(1)}$, $\hat{\rho}_g^{(2)}$, $\hat{\rho}_\tau$, $\hat{\rho}_s$, and $\hat{\rho}_p$.

We notice a decreasing trend in $\sqrt{n}$RMSEs as ρ increases for each sample size and an increasing trend as degrees of freedom, ν, decrease for t distributions. Under the normal distribution, $\sqrt{n}$RMSEs of both proposed symmetric Gini estimators, $\hat{\rho}_g^{(1)}$ and $\hat{\rho}_g^{(2)}$, are highly competitive with $\sqrt{n}$RMSE of $\hat{\rho}_p$. For $\rho = 0.1$, $\hat{\rho}_g^{(2)}$ outperforms $\hat{\rho}_p$ in all distributions. We include the heavy-tailed distribution, $t(1)$, to demonstrate the behavior of Pearson and Gini estimators when their asymptotic variances may not exist. We observe that for large sample size, $\hat{\rho}_p$ is around twice as large as both $\hat{\rho}_g^{(1)}$ and $\hat{\rho}_g^{(2)}$. When the sample size is small ($n = 30$), and degree of freedom ν is large (15, ∞) $\hat{\rho}_g^{(2)}$ performs the best. For the lognormal distribution, when ρ is small, we see $\hat{\rho}_g^{(2)}$ outperforms $\hat{\rho}_\tau$ and $\hat{\rho}_s$. For the remaining cases in the lognormal distribution both proposed symmetric Gini estimators have a smaller $\sqrt{n}$RMSE than the Pearson correlation estimator. As expected, Kendall's tau and Spearman's correlation estimator produced similar $\sqrt{n}$RMSE's under Normal and log-normal distributions.

5.3 Robustness

We also conduct a simulation with contaminated data to demonstrate robustness and show how contamination affects the performance of each correlations. We generate contaminated data of sizes ($n = 300, 1000$) from the following mixture normal model with contamination rates ($\varepsilon = 1\%, 5\%$).

$$(1-\varepsilon)N(\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0.9) + \varepsilon N(\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = \sigma, \rho = -0.9),$$

where $\sigma = 2, 4$. The majority of the data is highly positively correlated with a contamination by a small portion of negatively correlated outliers. The same criterion $\sqrt{n}RMSE$ is used to evaluate the difference between each correlation estimator and the true parameter value 0.9. M and the number of repetitions are the same as the previous subsection: 3000 and 30, respectively. The result is listed in Table 4.

In each case above, the Pearson correlation has the highest RMSE. This indicates the Pearson correlation's sensitivity to contamination and the high level of degradation those outliers have on its performance. The most robust correlation is the Kendall's tau. The performance of the Gini correlations are between those of the Pearson and Kendall's correlations. This result supports our findings from the derived influence functions in Section 3. The two symmetric Gini correlations $\hat{\rho}_g^{(1)}$ and $\hat{\rho}_g^{(2)}$ perform very similarly, but they are less robust than the joint rank based Gini correlation $\hat{\rho}_g^{(s)}$.

6 Real data analysis

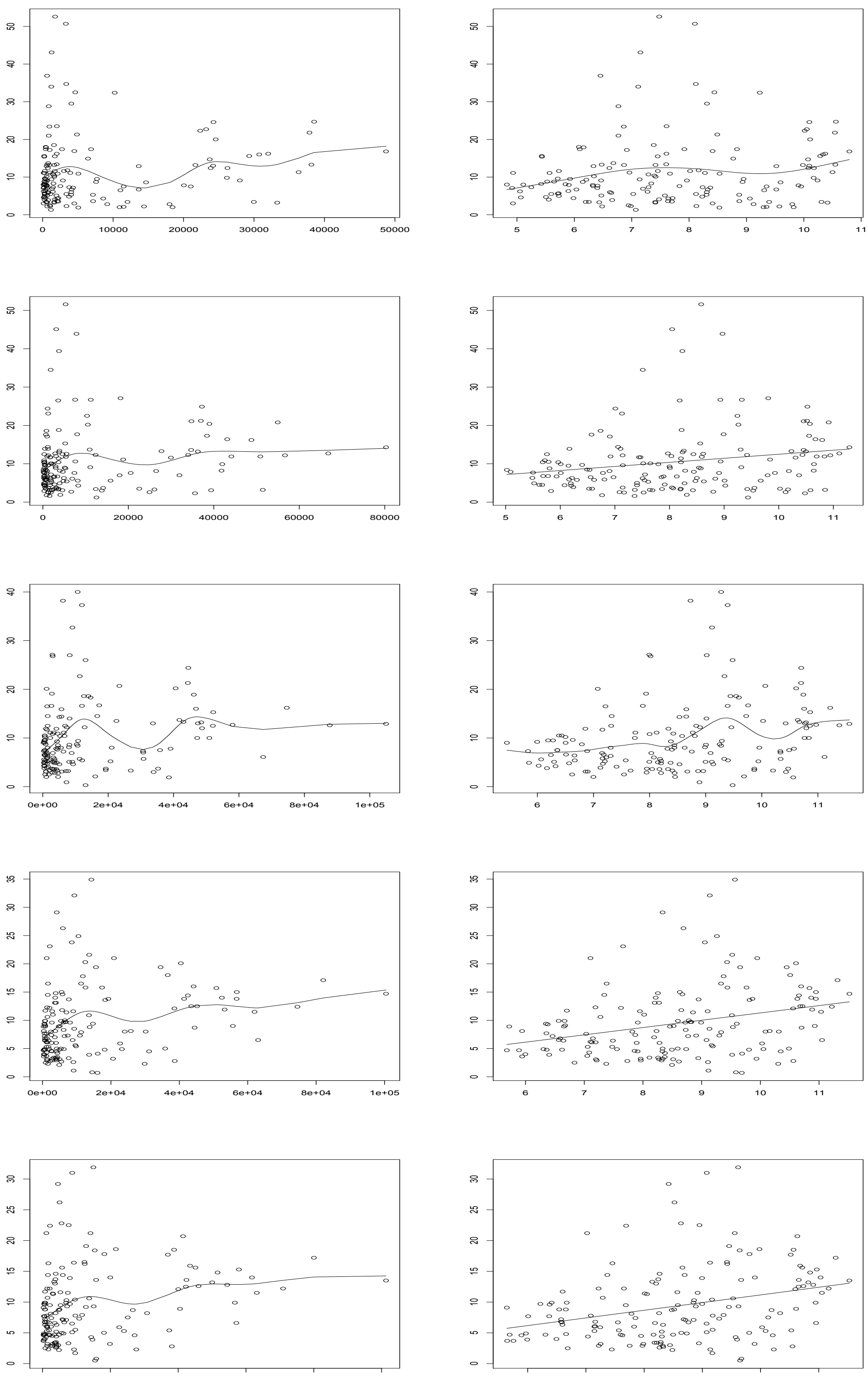
For the purpose of illustration, we apply the developed Gini correlations to the "GDP per captia and Suicide rates" data which is available on Kaggle. Many factors (mental health issues, weather, culture, etc.) affect suicide. We would like to explore whether or not an economic factor, such as GDP, relates to suicide rate by measuring the correlation using several correlation coefficients.

The data contains information from 160 countries around the world from the years 2000, 2005, 2010, 2015 and 2016. There are 2 missing values in 2000 data and 5 missing values in other years. We drop those countries with missing values and consider only the complete data for each year. We analyze how GDP and crude suicide rates are related and how the relationship changes through years. The crude suicide rate is

Dist	ρ	n	$\hat{\rho}_g^{(1)}$	$\hat{\rho}_g^{(2)}$	$\hat{\rho}_g^{(s)}$	$\hat{\rho}_\tau$	$\hat{\rho}_s$	$\hat{\rho}_p$
Norm	0	30	1.0342 (.0122)	1.0232 (.0124)	1.0071 (.0116)	1.0874 (.0129)	1.0604 (.0131)	1.0188 (.0117)
		300	1.0182 (.0130)	1.0091 (.0130)	1.0117 (.0142)	1.0553 (.0136)	1.0526 (.0136)	1.0058 (.0130)
	0.1	30	1.0240 (.0114)	0.7766 (.0112)	0.9999 (.0149)	1.0764 (.0110)	1.0505 (.0112)	1.0086 (.0121)
		300	1.0039 (.0143)	0.9430 (.0126)	1.0041 (.0126)	1.0413 (.0150)	1.0387 (.0151)	0.9913 (.0134)
	0.5	30	0.8022 (.0120)	0.8007 (.0117)	0.7828 (.0119)	0.8505 (.0128)	0.8428 (.0135)	0.7821 (.0110)
		300	0.7708 (.0097)	0.7710 (.0097)	0.7590 (.0105)	0.8004 (.0096)	0.8095 (.0099)	0.7536 (.0094)
	0.9	30	0.2249 (.0050)	0.2251 (.0050)	0.2188 (.0041)	0.2529 (.0057)	0.2787 (.0066)	0.2094 (.0047)
		300	0.2013 (.0021)	0.2013 (.0021)	0.1943 (.0023)	0.2104 (.0025)	0.2252 (.0031)	0.1911 (.0017)
$t(15)$	0.1	30	1.0654 (.0133)	0.8088 (.0107)	1.0379 (.0123)	1.1050 (.0142)	1.0713 (.0138)	1.0739 (.0122)
		300	1.0520 (.0154)	0.9776 (.0135)	1.0431 (.0126)	1.0745 (.0162)	1.0623 (.0160)	1.0758 (.0155)
	0.5	30	0.8349 (.0101)	0.8324 (.0082)	0.8179 (.0100)	0.8735 (.0110)	0.8642 (.0113)	0.8335 (.0099)
		300	0.8053 (.0099)	0.8056 (.0099)	0.7965 (.0094)	0.8238 (.0099)	0.8344 (.0097)	0.8132 (.0101)
	0.9	30	0.2387 (.0047)	0.2391 (.0047)	0.2284 (.0047)	0.2636 (.0049)	0.2965 (.0058)	0.2266 (.0042)
		300	0.2148 (.0026)	0.2148 (.0026)	0.2036 (.0029)	0.2208 (.0026)	0.2478 (.0032)	0.2078 (.0030)
$t(5)$	0.1	30	1.1753 (.0166)	0.8914 (.0126)	1.1533 (.0122)	1.1534 (.0162)	1.1039 (.0163)	1.2961 (.0169)
		300	1.1919 (.0169)	1.0703 (.0117)	1.1772 (.0139)	1.1312 (.0144)	1.1009 (.0145)	1.5167 (.0244)
	0.5	30	0.9374 (.0108)	0.9277 (.0097)	0.9107 (.0079)	0.9228 (.0096)	0.9144 (.0107)	1.0264 (.0166)
		300	0.9286 (.0118)	0.9297 (.0118)	0.9039 (.0125)	0.8801 (.0130)	0.9080 (.0133)	1.1631 (.0212)
	0.9	30	0.2807 (.0077)	0.2814 (.0077)	0.2598 (.0056)	0.2857 (.0070)	0.3395 (.0082)	0.2904 (.0098)
		300	0.2589 (.0042)	0.2589 (.0042)	0.2314 (.0039)	0.2414 (.0039)	0.3011 (.0049)	0.3011 (.0082)
$t(1)$	0.1	30	2.6127 (.0208)	2.1841 (.0211)	2.4072 (.0303)	1.3646 (.0181)	1.2460 (.0177)	3.3060 (.0228)
		300	6.4427 (.1052)	5.2414 (.0986)	6.0693 (.0827)	1.3753 (.0134)	1.2719 (.0145)	10.2563 (.1223)
	0.5	30	2.2951 (.0456)	1.5163 (.0150)	2.1902 (.0405)	1.1217 (.0142)	1.1390 (.0158)	2.9658 (.0441)
		300	5.5947 (.1303)	4.0462 (.0418)	5.4037 (.1292)	1.1027 (.0107)	1.4174 (.0123)	9.1351 (.1800)
	0.9	30	1.1631 (.0556)	0.9047 (.0282)	1.1915 (.0494)	0.4023 (.0117)	0.6295 (.0155)	1.5994 (.0695)
		300	2.8601 (.1712)	2.2857 (.0754)	2.7902 (.1719)	0.3471 (.0042)	0.9411 (.0089)	4.9999 (.2216)
Kotz	0.1	30	1.1821 (.0114)	0.8978 (.0115)	1.1503 (.0151)	1.1884 (.0135)	1.1291 (.0125)	1.2439 (.0113)
		300	1.1753 (.0186)	1.0615 (.0148)	1.1644 (.0165)	1.1660 (.0161)	1.1248 (.0157)	1.2714 (.0201)
	0.5	30	0.9324 (.0130)	0.9251 (.0107)	0.9089 (.0135)	0.9451 (.0138)	0.9355 (.0142)	0.9692 (.0139)
		300	0.9159 (.0104)	0.9167 (.0104)	0.8881 (.0133)	0.9103 (.0104)	0.9525 (.0109)	0.9705 (.0109)
	0.9	30	0.2765 (.0041)	0.2769 (.0041)	0.2601 (.0075)	0.2939 (.0055)	0.3538 (.0057)	0.2697 (.0044)
		300	0.2511 (.0038)	0.2511 (.0038)	0.2282 (.0030)	0.2515 (.0039)	0.3453 (.0045)	0.2484 (.0032)
Log-norm	0.1	30	1.0905 (.0117)	0.8027 (.0114)	N/A	1.0783 (.0126)	1.0522 (.0121)	1.6570 (.0241)
		300	1.1377 (.0136)	1.0186 (.0123)	N/A	1.0401 (.0120)	1.0375 (.0121)	1.7446 (.0298)
	0.5	30	0.9048 (.0134)	0.9160 (.0128)	N/A	0.8522 (.0133)	0.8434 (.0132)	1.2171 (.0159)
		300	0.9107 (.0152)	0.9134 (.0152)	N/A	0.7977 (.0102)	0.8065 (.0107)	1.6509 (.0272)
	0.9	30	0.2855 (.0041)	0.2860 (.0041)	N/A	0.2552 (.0035)	0.2815 (.0038)	0.3287 (.0055)
		300	0.2584 (.0032)	0.2584 (.0032)	N/A	0.2115 (.0031)	0.2261 (.0035)	0.5059 (.0079)
Log-norm $\sigma_1 = 1$ $\sigma_2 = 2$	0.1	30	0.8479 (.0139)	0.8203 (.0134)	N/A	1.0777 (.0121)	1.0524 (.0115)	1.6443 (.0146)
		300	1.3134 (.0155)	1.1014 (.0161)	N/A	1.0393 (.0108)	1.0368 (.0107)	1.9501 (.0313)
	0.5	30	0.9925 (.0124)	1.0078 (.0114)	N/A	0.8476 (.0098)	0.8397 (.0108)	1.4877 (.0142)
		300	1.2022 (.0173)	1.1943 (.0188)	N/A	0.8021 (.0092)	0.8114 (.0094)	3.0405 (.0353)
	0.9	30	0.3722 (.0059)	0.3721 (.0059)	N/A	0.2531 (.0044)	0.2795 (.0052)	0.5643 (.0101)
		300	0.3556 (.0054)	0.3534 (.0054)	N/A	0.2117 (.0025)	0.2265 (.0030)	1.9782 (.0258)

Table 3: The mean and standard deviation (in parenthesis) of $\sqrt{n}$RMSE of $\hat{\rho}_g^{(1)}$, $\hat{\rho}_g^{(2)}$, $\hat{\rho}_\tau$, $\hat{\rho}_s$, and $\hat{\rho}_p$ under different distributions.

Year GDP vs Suicide Rate Log(GDP) vs Suicide Rate



2000

2005

2010

2015

2016

Figure 3: Scatter plots between GDP and Suicide Rate and log(GDP) and Suicide Rate in different years. A cubic smoothing spline fitting curve is added in each plot.

σ	ε	n	$\hat{\rho}_g^{(1)}$	$\hat{\rho}_g^{(2)}$	$\hat{\rho}_g^{(s)}$	$\hat{\rho}_\tau$	$\hat{\rho}_s$	$\hat{\rho}_p$
2	1%	300	0.9111 (.0102)	0.9114 (.0102)	0.4538 (.0057)	0.3632 (.0040)	0.6120 (.0051)	1.5256 (.0270)
		1000	1.4962 (.0095)	1.4963 (.0095)	0.6917 (.0055)	0.5393 (.0044)	0.9796 (.0056)	2.4065 (.0223)
	5%	300	3.7800 (.0128)	3.7811 (.0128)	2.1315 (.0094)	1.5171 (.0072)	2.5729 (.0088)	5.6020 (.0241)
		1000	6.7915 (.0147)	6.7920 (.0147)	3.7512 (.0106)	2.7032 (.0066)	4.6069 (.0082)	10.0111 (.0289)
4	1%	300	1.8163 (.0128)	1.8178 (.0128)	0.8741 (.0080)	0.4059 (.0048)	0.7394 (.0058)	4.9148 (.0491)
		1000	3.1001 (.0121)	3.1009 (.0121)	1.4171 (.0070)	0.6319 (.0041)	1.2340 (.0050)	8.3176 (.0501)
	5%	300	7.1781 (.0239)	7.1841 (.0239)	4.8545 (.0242)	1.8511 (.0057)	3.2721 (.0081)	14.1536 (.0539)
		1000	12.100 (.0270)	13.003 (.0270)	8.5919 (.0261)	3.3165 (.0070)	5.8925 (.0088)	25.960 (.0633)

Table 4: The mean and standard deviation (in parenthesis) of $\sqrt{n}$RMSE of each correlation estimator in the contaminated Normal data.

the number of suicide deaths in a year, divided by the population and multiplied by 100,000. The countries with the highest suicide rates are Russia and Lithuania. Their suicide rates range from 32 to 52 per 100000 people. Luxembourg is the country with the highest GDP per captia of \$48736 in 2000 and \$101305 in 2016. Ethiopia, Burundi, and Somalia are countries with the lowest GDP of \$124 in 2000 and \$282 in 2016. There is a high degree of positive skewness in the distribution of GDP, hence we also consider the log transformation of GDP data to handle the asymmetry. We draw the scatterplot between GDP per capita and SR as well as the scatterplot between log(GPD) and SR per year in Figure 3. We also add a cubic smoothing spline fitting curve in each plot. We used default values of parameters of smooth.spline in R to fit the curves. We can see that the fitted curves demonstrate non-linear relationship between GDP per capita and SR, but almost linear relationships between log(GDP) and suicide rate except for the year 2010.

		Year				
Variables	Method	2000	2005	2010	2015	2016
(GDP, SR)	$\hat{r}_g^{(1)}$	.1970(.0784)	.2767(.0738)	.3374(.0636)	.3596(.0640)	.3642(.0631)
	$\hat{r}_g^{(2)}$	.1930(.0772)	.2757(.0727)	.3368(.0632)	.3580(.0638)	.3616(.0632)
	$\hat{\gamma}_1$	.2360(.0984)	.2998(.0948)	.3583(.0753)	.3933(.0733)	.4070(.0696)
	$\hat{\gamma}_2$	.1579(.0700)	.2535(.0641)	.3165(.0610)	.3259(.0635)	.3213(.0646)
	$\hat{r}_g^{(s)}$	.1005(.0676)	.1541(.0627)	.2041(.0578)	.2386(.0569)	.2511(.0587)
	$\hat{r}_\tau$	.0874(.0509)	.1371(.0487)	.1857(.0448)	.2016(.0447)	.2051(.0453)
	$\hat{r}_p$	.1074(.0700)	.1404(.0645)	.1936(.0583)	.2441(.0580)	.2584(.0593)
(log(GDP), SR)	$\hat{r}_g^{(1)}$	.1486(.0714)	.2371(.0666)	.3028(.0618)	.3213(.0630)	.3225(.0634)
	$\hat{r}_g^{(2)}$	.1483(.0717)	.2366(.0671)	.3025(.0620)	.3213(.0630)	.3225(.0634)
	$\hat{\gamma}_1$	.1393(.0775)	.2207(.0742)	.2891(.0672)	.3168(.0659)	.3236(.0653)
	$\hat{\gamma}_2$	.1579(.0700)	.2535(.0641)	.3165(.0610)	.3259(.0635)	.3213(.0646)
	$\hat{r}_g^{(s)}$	.1376(.0643)	.2110(.0600)	.2709(.0565)	.3022(.0594)	.3074(.0604)
	$\hat{r}_\tau$	.0874(.0509)	.1371(.0487)	.1857(.0448)	.2016(.0447)	.2051(.0453)
	$\hat{r}_p$	.1299(.0606)	.2037(.0543)	.2674(.0518)	.2950(.0557)	.2985(.0571)

Table 5: All types of correlations for (GDP, SR) and (log(GDP), SR), respectively. The standard deviations are in parenthesis.

We have calculated the symmetric Gini correlations for (GDP, SR) and (log(GDP), SR), as well as other correlations presented for comparison in Table 5. We utilize the jackknife method to provide an estimation of the variation of the sample correlations. Let $\hat{r}_{(-i)}$ be the jackknife pseudo value of a correlation estimator $\hat{r}$

based on the sample with the i^{th} observation deleted. Then the jackknife variance is

$$\hat{v}_r = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{r}_{(-i)} - \bar{\hat{r}}_{(\cdot)})^2 \tag{19}$$

where $\bar{\hat{r}}_{(\cdot)} = 1/n\sum_{i=1}^{n}\hat{r}_{(-i)}$. See [35] for more details. Table 5 lists the jackknife standard deviations in parentheses.

From Table 5, we observe that all the listed correlations between GDP per capita and SR are less than .5000, which indicates a weak or moderate association between GDP per capita and SR and is consistent with Figure 3. However, with each year, we notice an increasing trend in the correlations between GDP and SR. The data suggest that the correlations between the two become more significant as time passes. Values of $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(2)}$ are close to each other, but there is a visible difference between the regular Gini correlations, $\hat{\gamma}_1$ and $\hat{\gamma}_2$. After the log transformation on GDP, the difference becomes less significant. The monotonic transformation does not change the rank of the GDP. Kendall's τ and $\hat{\gamma}_2$ should maintain the same values before and after the transformation, which agrees with the values we have shown in Table 5.

		$\hat{r}_g^{(1)}$	$\hat{r}_g^{(2)}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{r}_g^{(s)}$	$\hat{r}_\tau$	$\hat{r}_p$
2015	complete	.3213(.0630)	.3213(.0630)	.3168(.0659)	.3259(.0635)	.3022(.0594)	.2016(.0447)	.2950(.0557)
	deleted	.3347(.0718)	.3344(.0719)	.3148(.0734)	.3511(.0718)	.3446(.0727)	.2015(.0508)	.3531(.0685)
2016	complete	.3225(.0634)	.3225(.0634)	3236(.0653)	.3213(.0646)	.3074(.0604)	.2051(.0453)	.2985(.0571)
	deleted	.3578(.0689)	.3574(.0690)	.3416(.0705)	.3740(.0691)	.3656(.0685)	.2186(.0499)	.3736(.0646)

Table 6: Correlations between log(GDP) and SA for the complete data and the deleted data in 2015 and 2016. The standard deviations are in parenthesis.

To demonstrate robustness, we delete some outliers and compare the differences of each correlation estimator in the complete data and in the edited data. We expect the Pearson correlation to show the largest difference, the Kendall's τ correlation to demonstrate the smallest, and the Gini correlations to be somewhere in-between. We consider log(GDP) and SA data from 2015 and 2016. We delete all countries with $SR > 20$. The results listed in Table 6 confirm what we expect. In 2015, the Pearson correlation estimator changes from 0.295 to 0.353, while symmetric Gini correlations only have a slight change from 0.321 to 0.335. The Kendall's tau correlation is the most stable. A similar conclusion can be drawn for the 2016 data. This experiment illustrates that Pearson correlation is not robust and may not be a good measure of association even though the cubic smoothing spline fitting lines in the scatter plots in Fig 3 are almost linear in 2015 and 2016, suggesting the usage of the Pearson correlation. Other correlations are more preferred in this example.

7 Conclusion

We have systematically studied two symmetric Gini correlations $r_g^{(1)}$ and $r_g^{(2)}$, which are the arithmetic and geometric means of the traditional Gini correlations γ_1 and γ_2. We studied basic properties of $r_g^{(1)}$ and $r_g^{(2)}$, as well as their relationships to the correlation parameter in the elliptical distributions and log-normal distribution. Such relationships enable us to obtain Fisher consistent versions of each correlation. We derived their influence functions in order to gauge robustness. They are more robust than the Pearson correlation but less robust than Kendall's tau and Spearman correlations. We established asymptotic distributions of the sample correlations. Usual asymptotic normality holds for $\hat{r}_g^{(1)}$ as well as for $\hat{r}_g^{(2)}$ as long as $r_g^{(2)} \neq 0$. Their asymptotic variances are obtained through the influence function approach. For $r_g^{(2)} = 0$, $\hat{r}_g^{(2)}$ has two different limiting distributions, depending on whether or not both γ_1 and γ_2 equal 0. We compared their computational efficiency and statistical efficiency with the rank-based symmetric Gini, Kendall's tau and the Pearson correlation. $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(2)}$ can be efficiently calculated with a computational complexity of $O(nlogn)$. Asymptotic efficiency and finite sample efficiency of each correlation are obtained under various elliptical distributions and asymmetric lognormal distributions. In summary, the two symmetric Gini correlations balance well among statistical efficiency, robustness, and computational efficiency.

Continuations of this work could advance in several directions. The jackknife empirical likelihood (JEL) method proposed by Jing *et al.* [15] has been proven to be effective and reliable in dealing with U-statistics. Sang *et al.* [29] have applied JEL to the classical Gini correlations. It could be beneficial to develop JEL for the two symmetric Gini correlations. In the current work, comparisons among correlations are made in elliptical distributions and lognormal distributions. It would be worthwhile to explore the comparisons in wide families of bivariate distributions such as copula family and Farlie-Gumbel-Morgenstern models. Fontanari *et al.* [8] proposed a new Archimedean copulas based on the Lorenz curve that is highly related to Gini index and Gini correlations. It is interesting to study correlations in this family. Dang *et al.* [4] extended the Gini mean difference in one dimension to the Gini covariance matrix (GCM) in high dimensions. However, its computation cost is $O(n^2)$. It would be worthwhile to study the GCM based on $r_g^{(1)}$ or $r_g^{(2)}$ which should be more computationally efficient.

Appendix

Proof of Proposition 2.4. The results of (10) and (11) directly follow from

$$\gamma_1 = \frac{2\Phi(\rho\sigma_1/\sqrt{2})-1}{2\Phi(\sigma_1/\sqrt{2})-1} \quad \text{and} \quad \gamma_2 = \frac{2\Phi(\rho\sigma_2/\sqrt{2})-1}{2\Phi(\sigma_2/\sqrt{2})-1},$$

which are given by [31], although their (2.3) has a typo. The authors didn't provide a proof in [31], thus for the purpose of completeness, we provide a detailed proof here.

Related to $\Phi(x)$, the cdf of a standard normal variable, the error function is defined as

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt.$$

From the tables of integrals of the error functions [25], two identities we will use in the proof are listed below.

$$erf(x) = 2\Phi(x\sqrt{2})-1 \implies \Phi(x) = \frac{1}{2}[erf(\frac{x}{\sqrt{2}})+1] \qquad (p3, \text{Eq}(6) \text{ in } [25]) \tag{20}$$

$$\int_{-\infty}^{\infty} erf(x)e^{-(ax+b)^2}\,dx = -\frac{\sqrt{\pi}}{a}erf(\frac{b}{\sqrt{a^2+1}}) \qquad (p8, \text{Eq}(13) \text{ in } [25]) \tag{21}$$

We will use the following equation throughout the remainder of the proof:

$$\int_{-\infty}^{\infty}\Phi(z)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-\mu)^2}\,dz = \Phi(\frac{\mu}{\sqrt{2}}). \tag{22}$$

This is because

$$\begin{aligned}
\int_{-\infty}^{\infty}\Phi(z)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-\mu)^2}\,dz &= \int_{-\infty}^{\infty}\frac{1}{2}[erf(\frac{z}{\sqrt{2}})+1]\frac{1}{\sqrt{2\pi}}e^{-(\frac{z}{\sqrt{2}}-\frac{\mu}{\sqrt{2}})^2}\,dz \qquad \text{by (20)}\\
&= \frac{1}{2}+\frac{1}{2}\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}erf(x)e^{-(x-\frac{\mu}{\sqrt{2}})^2}\,dx\\
&= \frac{1}{2}+\frac{1}{2}erf(\frac{\mu}{2}) \qquad \text{by (21)}\\
&= \Phi(\frac{\mu}{\sqrt{2}}).
\end{aligned}$$

Let $(X, Y)^T$ follow a lognormal distribution with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Then the marginal distributions are $F(x) = \Phi((\log x - \mu_1)/\sigma_1)$ and $G(y) = \Phi((\log y - \mu_2)/\sigma_2)$, respectively. $\mathbb{E}X = \exp(\mu_1 + \sigma_1^2/2)$ and $\mathbb{E}(X|Y) = \exp(\mu_1 + \rho\sigma_1(\log Y - \mu_2)/\sigma_2 + \sigma_1^2(1-\rho^2)/2)$ by [23]. We have

$$\begin{aligned}\mathbb{E}XF(X) &= \int_0^\infty x\Phi\Big(\frac{\ln x - \mu_1}{\sigma_1}\Big)\frac{1}{\sqrt{2\pi}\sigma_1 x}e^{-\frac{(\ln x-\mu_1)^2}{2\sigma_1^2}}dx = \int_{-\infty}^\infty \Phi(z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}e^{\sigma_1 z+\mu_1}dz\\ &= e^{\mu_1+\sigma_1^2/2}\int_{-\infty}^\infty \Phi(z)\frac{1}{\sqrt{2\pi}}e^{-(z-\sigma_1)^2/2}dz = e^{\mu_1+\sigma_1^2/2}\Phi(\sigma_1/\sqrt{2}).\end{aligned}$$

The last equation is due to (22). Also

$$\begin{aligned}\mathbb{E}XG(Y) &= \mathbb{E}[G(Y)\mathbb{E}(X|Y)] = \mathbb{E}[G(Y)\exp(\mu_1 + \rho\sigma_1(\log Y - \mu_2)/\sigma_2 + \sigma_1^2(1-\rho^2)/2)]\\ &= e^{\mu_1+\sigma_1^2(1-\rho^2)/2}\int_0^\infty \Phi\Big(\frac{\ln y - \mu_2}{\sigma_2}\Big)e^{\rho\sigma_1(\log y-\mu_2)/\sigma_2}\frac{1}{\sqrt{2\pi}\sigma_2 y}e^{-(\log y-\mu_2)^2/(2\sigma_2^2)}dy\\ &= e^{\mu_1+\sigma_1^2(1-\rho^2)/2}\int_{-\infty}^\infty \Phi(z)e^{\rho\sigma_1 z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz\\ &= e^{\mu_1+\sigma_1^2(1-\rho^2)/2}e^{\rho^2\sigma_1^2/2}\int_{-\infty}^\infty \Phi(z)e^{-(z-\rho\sigma_1)^2/2}dz\\ &= e^{\mu_1+\sigma_1^2/2}\Phi(\rho\sigma_1/\sqrt{2}).\end{aligned}$$

Thus,

$$\begin{aligned}\gamma_1 &= \frac{cov(X, G(Y))}{cov(X, F(X))} = \frac{\mathbb{E}[XG(Y)] - \mathbb{E}[X]\mathbb{E}[G(Y)]}{\mathbb{E}[XF(X)] - \mathbb{E}[X]\mathbb{E}[F(X)]} = \frac{e^{\mu_1+\sigma_1^2/2}\Phi(\rho\sigma_1/\sqrt{2}) - e^{\mu_1+\sigma_1^2/2}/2}{e^{\mu_1+\sigma_1^2/2}\Phi(\sigma_1/\sqrt{2}) - e^{\mu_1+\sigma_1^2/2}/2}\\ &= \frac{2\Phi(\rho\sigma_2/\sqrt{2}) - 1}{2\Phi(\sigma_2/\sqrt{2}) - 1}.\end{aligned}$$

Similar arguments for γ_2 complete the proof. □

Proof of Proposition 3.1. To find influence function of Gini correlation, let $T_1(H) = cov(X, G(Y))$, $T_2(H) = cov(X, F(X))$, $T_3(H) = cov(Y, F(X))$, $T_4(H) = cov(Y, G(Y))$ and $h(t_1, t_2) = t1/t2$. Then $\gamma_1 = h(T_1, T_2)$ and $\gamma_2 = h(T_3, T_4)$. Denote the influence function of T_i as $L_i(u, v) = IF((u, v)^T; T_i, H)$, for $i = 1, 2, 3, 4$. Let $\tilde{H} = (1-\varepsilon)H + \varepsilon\delta_{(u,v)^T}$, then

$$\begin{aligned}T_1(\tilde{H}) &= \mathbb{E}_{\tilde{H}}XG(Y) - \mathbb{E}_{\tilde{H}}X\mathbb{E}_{\tilde{H}}G(Y)\\ &= (1-\varepsilon)\mathbb{E}XG(Y) + \varepsilon uG(v) - [(1-\varepsilon)\mathbb{E}X + \varepsilon u][(1-\varepsilon)\mathbb{E}G(Y) + \varepsilon G(v)]\\ &= (1-\varepsilon)\mathbb{E}XG(Y) - (1-\varepsilon)^2\mathbb{E}X\mathbb{E}G(Y) + \varepsilon uG(v) - \varepsilon(1-\varepsilon)(u\mathbb{E}G(Y) + G(v)\mathbb{E}X) - \varepsilon^2 uG(v).\end{aligned}$$

Hence,

$$\begin{aligned}L_1(u, v) &= \lim_{\varepsilon\to 0}\frac{T_1(\tilde{H}) - T_1(H)}{\varepsilon}\\ &= -\mathbb{E}XG(Y) + 2\mathbb{E}X\mathbb{E}G(Y) + uG(v) - u\mathbb{E}G(Y) - G(v)\mathbb{E}X\\ &= (u - \mathbb{E}X)(G(v) - \mathbb{E}G(Y)) - \mathrm{cov}(X, G(Y)).\end{aligned}$$

Similarly, we have

$$\begin{aligned}L_2(u, v) &= (u - \mathbb{E}X)(F(u) - \mathbb{E}F(X)) - \mathrm{cov}(X, F(X)),\\ L_3(u, v) &= (v - \mathbb{E}Y)(F(u) - \mathbb{E}F(X)) - \mathrm{cov}(Y, F(X)),\end{aligned}$$

$$L_4(u,v) = (v - \mathbb{E}Y)(G(v) - \mathbb{E}G(Y)) - \text{cov}(Y, G(Y)).$$

Hence,

$$\begin{aligned}\text{IF}((u,v);\gamma_1,H) &= \sum_{i=1}^{2} \frac{\partial h}{\partial t_i}\Big|_{h(T_1,T_2)} L_i(u,v) = \frac{1}{T_2}L_1(u,v) - \frac{T_1}{T_2^2}L_2(u,v) \\ &= \gamma_1\left(\frac{(u-\mathbb{E}X)[G(v)-\mathbb{E}G(Y)]}{cov(X,G(Y))} - \frac{(u-\mathbb{E}X)[F(u)-\mathbb{E}F(X)]}{cov(X,F(X))}\right)\end{aligned}$$

Similar arguments on $IF((u,v);\gamma_2,H)$ complete the proof. □

Proof of Proposition 3.2. Define $g_1(t_1,t_2) = \dfrac{t_1+t_2}{2}$ and $g_2 = \sqrt{|t_1t_2|}$. Then $r_g^{(1)} = g_1(\gamma_1,\gamma_2)$ and $r_g^{(2)} = g_2(\gamma_1,\gamma_2)$. Let $\text{IF}_1(u,v)$ and $\text{IF}_2(u,v)$ denote the influence functions for γ_1 and γ_2, respectively. For any $r_g^{(1)}$, we have

$$\text{IF}((u,v)^T; r_g^{(1)}, H) = \frac{\partial g_1(\gamma_1,\gamma_2)}{\partial\gamma_1}\text{IF}_1(u,v) + \frac{\partial g_1(\gamma_1,\gamma_2)}{\partial\gamma_2}\text{IF}_2(u,v) = \frac{1}{2}\text{IF}_1(x,y) + \frac{1}{2}\text{IF}_2(x,y).$$

Since g_2 is not differentiable at $\gamma_1 = 0$ and/or $\gamma_2 = 0$, the influence function of $r_g^{(2)}$ does not exist for $r_g^{(2)} = 0$. For nonzero $r_g^{(2)}$, we have

$$\begin{aligned}\text{IF}((u,v)^T; r_g^{(2)}, H) &= \frac{\partial g_2(\gamma_1,\gamma_2)}{\partial\gamma_1}\text{IF}_1(u,v) + \frac{\partial g_2(\gamma_1,\gamma_2)}{\partial\gamma_2}\text{IF}_2(u,v) \\ &= \frac{\text{sgn}(\gamma_1\gamma_2)\gamma_2}{2\sqrt{|\gamma_1\gamma_2|}}\text{IF}_1(u,v) + \frac{\text{sgn}(\gamma_1\gamma_2)\gamma_1}{2\sqrt{|\gamma_1\gamma_2|}}\text{IF}_2(u,v)\end{aligned}$$

□

Proof of Proposition 4.1. A proof of the proposition follows directly from the fact of strong consistency of U-statistics U_1, U_2, U_3, U_4 by the U-statistics theorem [34] and the fact that $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(1)}$ are continuous functions of U_1, U_2, U_3, U_4. By the continuous mapping theorem [34], the strong consistency of $\hat{r}_g^{(1)}$ and $\hat{r}_g^{(2)}$ holds. □

Proof of Proposition 4.2 and 4.3. The asymptotical normality of $\hat{r}_g^{(1)}$ and the asymptotical normality $\hat{r}_g^{(2)}$ when $r_g \neq 0$ are an immediate result from the application of the influence function approach [14]. □

Proof of Proposition 4.4. We have

$$\hat{r}_g^{(2)} = \sqrt{\left|\frac{U_1}{U_2}\frac{U_3}{U_4}\right|} = \frac{\sqrt{U_1U_3}}{\sqrt{U_2U_4}},$$

since U_2 and U_4 are always positive. The denominator $\sqrt{U_2U_4}$ converges to $\dfrac{\sqrt{\Delta_1\Delta_2}}{4}$ almost surely by the U statistics theorem and the continuous mapping theorem [34]. We need to explore the limiting distribution of $\sqrt{|U_1U_3|}$. Then by Slutsky's theorem [38], the limiting distribution of $\hat{r}_g^{(2)}$ follows. Now consider U_1U_3, the product of two U statistics. We have

$$\begin{aligned}U_1U_3 &= \binom{n}{2}^{-2} \sum_{1\le i<j\le n} h_1(\boldsymbol{Z}_i,\boldsymbol{Z}_j) \sum_{1\le k<l\le n} h_3(\boldsymbol{Z}_k,\boldsymbol{Z}_l) \\ &= 6\binom{n}{2}^{-2} \sum_{1\le i<j<k<l\le n} g(\boldsymbol{Z}_i,\boldsymbol{Z}_j,\boldsymbol{Z}_k,\boldsymbol{Z}_l) + R_n,\end{aligned}$$

where $R_n = o_p(n^{-1})$ and the symmetric kernel $g(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3,\boldsymbol{z}_4) = 1/4!\sum_p h_1(\boldsymbol{z}_{i1},\boldsymbol{z}_{i2})h_3(\boldsymbol{z}_{i3},\boldsymbol{z}_{i4})$ with $\sum_p$ denoting summation over the 4! permutations (i_1,i_2,i_4,i_4) of $(1,2,3,4)$.

Define the new U statistic $U_n = \binom{n}{4}^{-1}\sum_{1\le i<j<k<l\le n} g(\boldsymbol{Z}_i,\boldsymbol{Z}_j,\boldsymbol{Z}_k,\boldsymbol{Z}_l)$. It is easy to check that U_1U_3 is asymptotically equivalent to U_n. Now consider the first order and second order projections of the kernel g. We define

$$g_1(\boldsymbol{z}) = \mathbb{E}g(\boldsymbol{z},\boldsymbol{Z}_2,\boldsymbol{Z}_3,\boldsymbol{Z}_4) = \frac{1}{2}\mathbb{E}h_1(\boldsymbol{z},\boldsymbol{Z}_2)\mathbb{E}h_3(\boldsymbol{Z}_3,\boldsymbol{Z}_4),$$

$$g_2(\boldsymbol{z}_1, \boldsymbol{z}_2) = \mathbb{E}g(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{Z}_3, \boldsymbol{Z}_4)$$
$$= \frac{1}{6}h_1(\boldsymbol{z}_1, \boldsymbol{z}_2)\mathbb{E}h_3(\boldsymbol{Z}_3, \boldsymbol{Z}_4) + \frac{1}{3}\mathbb{E}h_1(\boldsymbol{z}_1, \boldsymbol{Z}_3)\mathbb{E}h_3(\boldsymbol{z}_2, \boldsymbol{Z}_4) + \frac{1}{3}\mathbb{E}h_1(z_2, \boldsymbol{Z}_3)\mathbb{E}h_3(\boldsymbol{z}_1, \boldsymbol{Z}_4).$$

The second equations in g_1 and g_2 are due to $\gamma_1 = 0$, implying $\mathbb{E}h_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2) = 0$.

Case 1: $\gamma_2 \neq 0$.

Let $\sigma_g^2 = var[g_1(\boldsymbol{Z})] = (\gamma_2\Delta_2/8)^2 var[\mathbb{E}h_1(\boldsymbol{Z}, \boldsymbol{Z}_2)|\boldsymbol{Z}]$. Then by Hoeffding decomposition, we have $U_n = \frac{4}{n}\sum_{i=1}^n g_1(\boldsymbol{Z}_i) + o_p(n^{-1/2})$. By [34],

$$\sqrt{n}U_n \xrightarrow{d} N(0, 16\sigma_g^2).$$

Therefore,

$$\sqrt{n}\frac{U_1U_3}{U_2U_4} \xrightarrow{d} N(0, \frac{16^2\sigma_g^2}{\Delta_1\Delta_2}).$$

Hence $\sqrt{n}|\frac{U_1U_3}{U_2U_4}|$ converges to a folded normal. Finally, $n^{1/4}\hat{r}_g^{(2)}$ converges to the square root of the folded normal random variable.

Case 2: $\gamma_2 = 0$

In this case, $\mathbb{E}h_3(\boldsymbol{Z}_3, \boldsymbol{Z}_4) = 0$ and hence $g_1(\boldsymbol{z}) = 0$, meaning that U_n is a degenerate U-statistic. In the mean time, $g_2(\boldsymbol{z}_1, \boldsymbol{z}_2)$ is simplified to be $1/3\mathbb{E}h_1(\boldsymbol{z}_1, \boldsymbol{Z}_3)\mathbb{E}h_3(\boldsymbol{z}_2, \boldsymbol{Z}_4) + 1/3\mathbb{E}h_1(z_2, \boldsymbol{Z}_3)\mathbb{E}h_3(\boldsymbol{z}_1, \boldsymbol{Z}_4)$. Therefore,

$$nU_n = \frac{12}{n-1}\sum_{1 \le i < j \le n} g_2(\boldsymbol{Z}_1, \boldsymbol{Z}_2) + o_p(1).$$

Define $g_2(\boldsymbol{z}_1, \boldsymbol{z}_2) = \sum_{s=1}^{\infty}\lambda_s\phi_s(\boldsymbol{z}_1)\phi_s(\boldsymbol{z}_2)$, where

$$\int_{\mathbb{R}^2} g_2(\boldsymbol{z}_1, \boldsymbol{z}_2)\phi_s(\boldsymbol{z}_2)dH(\boldsymbol{z}_2) = \lambda_s\phi_s(\boldsymbol{z}_1).$$

By Theorem of Section 5.5.2 of Serfling (1980) [34],

$$nU_n \xrightarrow{d} 6\sum_{s=1}^{\infty}\lambda_s(\chi_{1s}^2 - 1),$$

where χ_{1s}^2 (s=1,2,...) are independent χ_1^2 variables. Therefore,

$$\sqrt{n}\sqrt{|U_1U_3|} \xrightarrow{d} 6\sqrt{|\sum_{k=1}^{\infty}\lambda_k(\chi_{1k}^2 - 1)|},$$

and hence

$$\sqrt{n}\hat{r}_g^{(2)} \xrightarrow{d} \frac{24\sqrt{\sum_{k=1}^{\infty}\lambda_k(\chi_{1k}^2 - 1)}}{\sqrt{\Delta_1\Delta_2}}.$$

This completes the proof. □

References

[1] Blitz, R.C. and J.A. Brittain (1964). An extension of the Lorenz diagram to the correlation of two variables. *Metron 23*(1-4), 137–143.

[2] Blomqvist, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist. 21*(4), 593–600.

[3] Croux, C. and C. Dehon (2010). Influence function of the Spearman and Kendall correlation measures. *Stat. Methods Appl. 19*(4), 497–515.

[4] Dang, X., H. Sang, and L. Weatherall (2019). Gini covariance matrix and its affine equivariant version. *Statist. Papers 60*, 641–666.

[5] Devlin, S.J., R. Gnanadesikan, and J.R. Kettenring (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika 62*(3), 531–545.

[6] Fang, K.T. and T. W. Anderson (1990). *Statistical Inference in Elliptically Contoured and Related Distributions*. Allerton Press, New York.

[7] Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. A 222*, 309–368.

[8] Fontanari, A., P. Cirillo, and C.W. Oosterlee (2020). Lorenz-generated bivariate Archimedean copulas. *Depend. Model. 8*, 186–209.

[9] Furman, E. and R. Zitikis (2017). Beyond the Pearson correlation: heavy-tailed risks, weighted Gini correlations, and a Gini-type weighted insurance pricing model. *Astin Bull. 47*(3), 919–942.

[10] Genest, C., J. Nešlehová, and N. Ghorbal (2010). Spearman's footrule and Gini's gamma: a review with complements. *J. Nonparametr. Stat. 22*(8), 937–954.

[11] Gini, C. (1914). *L'Ammontare e la Composizione della Ricchezza delle Nazioni*. UTET, Torino.

[12] Gribkova, N. and R. Zitikis (2019). Weighted allocations, their concomitant-based estimators, and asymptotics. *Ann. Inst. Statist. Math. 71*(4), 811–835.

[13] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc. 69*(346), 383–393.

[14] Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel (1986). *Robust Statistics*. Wiley, New York.

[15] Jing, B.-Y., J. Yuan, and W. Zhou (2009). Jackknife empirical likelihood. *J. Amer. Statist. Assoc. 104*(487), 1224–1232.

[16] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

[17] Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika 30*(1-2), 81–93.

[18] Knight, W.R. (1966). A computer method for calculating Kendall's tau with ungrouped data. *J. Amer. Statist. Assoc. 61*(314), 436–439.

[19] Lai, C.D., J.C.W. Rayner, and T.P. Hutchinson (1999). Robustness of the sample correlation - the bivariate lognormal case. *J. Appl. Math. Decis. Sci. 3*(1), 7–19.

[20] Lindskog, F., A. McNeil, and U. Schmock (2003). Kendall's tau for elliptical distributions. In G. Bol, G. Nakhaeizadeh, S.T. Rachev, T. Ridder, and K.-H. Vollmer (Eds.), *Credit Risk*, pp. 149–156. Springer, Heidelberg.

[21] Mari, D.D. and S. Kotz (2001). *Correlation and Dependence*. Imperial College Press, London.

[22] Moran, P.A.P. (1948). Rank correlation and permutation distributions. *Proc. Cambridge Philos. Soc. 44*(1), 142–144.

[23] Mostafa, M.D. and M.W. Mahmoud (1964). On the problem of estimation for the bivariate lognormal distribution. *Biometrika 51*(3-4), 522–527.

[24] Nelsen, R.B. (1998). Concordance and Gini's measure of association. *J. Nonparametr. Statist. 9*(3), 227–238.

[25] Ng, E.W. and M. Geller (1969). A table of integrals of the error functions. *J. Res. Nat. Bur. Standards Sect. B 73B*(1), 1–20.

[26] Reimherr, M. and D.L. Nicolae (2013). On quantifying dependence: A framework for developing interpretable measures. *Statist. Sci. 28*(1), 116–130.

[27] Rényi, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar. 10*, 441–451.

[28] Sang, Y., X. Dang, and H. Sang (2016). Symmetric Gini covariance and correlation. *Canad. J. Statist. 44*(3), 323–342.

[29] Sang, Y., X. Dang, and Y. Zhao (2019). Jackknife empirical likelihood methods for Gini correlations and their equality testing. *J. Statist. Plann. Inference 199*, 45–59.

[30] Scarsini, M. (1984). On measures of concordance. *Stochastica 8*(3), 201–218.

[31] Schechtman, E. and S. Yitzhaki (1987). A measure of association based on Gini's mean difference. *Comm. Statist. Theory Methods 16*(1), 207–231.

[32] Schechtman, E. and S. Yitzhaki (2003). A family of correlation coefficients based on the extended Gini index. *J. Econ. Inequal. 1*, 129–146.

[33] Schechtman, E., S. Yizhaki, and Y. Artsev (2008). The similarity between mean-variance and mean-Gini: Testing for equality of Gini correlations. *Adv. Invest. Anal. Portfolio Manag. 3*, 97–122.

[34] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

[35] Shao, J. and D. Tu (1995). *The Jackknife and Bootstrap*. Springer, New York.

[36] Shevlyakov G.L. and H. Oja (2016). *Robust Correlation*. Wiley, Chichester.

[37] Shevlyakov G.L. and P.O. Smirnov (2011). Robust estimation of the correlation coefficient: An attempt of survey. *Austrian J. Stat. 40*(1-2), 147–156.

[38] Slutsky, E. (1925). Über stochastische Asymptoten und Grenzwerte. *Metron 5*(3), 3–89.

[39] Spearman, C. (1904). "General intelligence", objectively determined and measured. *Amer. J. Psych. 15*(2), 201–292.

[40] Witting, H. and U. Müller-Funk (1995). *Mathematische Statistik II*. Teubner, Stuttgart.

[41] Xu, W., Y.S. Huang, M. Niranjan, and M. Shen (2010). Asymptotic mean and variance of Gini correlation for bivariate normal samples. *IEEE Trans. Signal Process. 58*(2), 522–534.

[42] Yitzhaki, S. and I. Olkin (1991). Concentration indices and concentration curves. In K. Mosler and M. Scarsini (Eds.), *Stochastic Orders and Decision Under Risk*, pp. 380–392. Institute of Mathematical Statistics, Hayward CA.

[43] Yitzhaki, S. and E. Schechtman (2013). *The Gini Methodology*. Springer, New York.