Research Article

On the Relationship between Pearson Correlation Coefficient and Kendall's Tau under Bivariate Homogeneous Shock Model

Jiantian Wang

Department of Mathematics, Kean University, Union, NJ 07083, USA

Correspondence should be addressed to Jiantian Wang, jwang@kean.edu

Received 2 March 2012; Accepted 12 April 2012

Academic Editors: M. Galea, J. Hu, and A. Pascucci

Copyright © 2012 Jiantian Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper studies the relationship between Kendall's tau and Pearson correlation coefficient under the so-called bivariate homogeneous shock (BHS) model. We find Capéraà-Genest-type inequality may not hold for general BHS model. Computational simulations suggest that the Daniels' inequality is likely to be true.

1. Introduction

Pearson's correlation coefficient r is the mostly used nonparametric measure of association for two random variables. Besides it, the Spearman's ρ and Kendall's τ are two very useful measures of association. As we all know, the Spearman's ρ is the ordinary Pearson's correlation coefficient.

The relationship between ρ and τ has received considerable attention recently. For instance, Hutchinson and Lai [1] conjecture that $-1 + \sqrt{1 + 3\tau} \leq \rho \leq \min\{3\tau/2, 2\tau - \tau^2\}$ for stochastically increasing random variables. Hürlimann [2] has shown that the entire Hutchinson and Lai conjecture holds for bivariate extreme value distributions. Munroe et al. [3] show that the Hutchinson and Lai conjecture, $1 + 3\tau \leq (1 + \rho)^2$, does not hold when X and Y are stochastically increasing. Fredricks and Nelsen [4] show that, under mild regularity conditions, the limit of the ratio ρ/τ is $3/2$ as the joint distribution of the random variables approaches to independence. Capéraà and Genest [5] have shown that $\rho \geq \tau \geq 0$ when one variable is simultaneously left-tail decreasing and the other right-tail increasing. Genest and Nešlehová [6] give a short analytical proof for classical Daniels' inequality $|3\tau - 2\rho| \leq 1$, found by Daniels [7].

However, relatively little attention has been paid on the relationships between r and τ . One example on this track is Edwardes [8], in which the author has shown that $r = \tau$ under a bivariate exponential (BVE) model introduced by Marshall and Olkin [9].

Based on a two-component series system subjected to some fatal shocks, Wang and Li [10] introduced a more general bivariate model, which is referred to as bivariate homogeneous shock (BHS) model. In this paper, we study the relationships between r and τ under BHS model. We find that the most revealed relationships between ρ and τ will no longer hold for the relationships between r and τ . However, computational simulations suggest that the Daniels-type inequality, $|3\tau - 2r| \leq 1$, is likely to be true.

2. Main Results

For any pair of random variables (X, Y) , let $F(x, y)$ be its bivariate cumulative distribution function. The classical Pearson correlation coefficient r of (X, Y) is defined as follows:

$$r = \frac{\text{Cov}(X, Y)}{\{\text{Var}(X) \text{Var}(Y)\}^{1/2}}, \quad (2.1)$$

and Kendall's τ is defined as follows:

$$\tau = 4 \iint F(x, y) dF(x, y) - 1. \quad (2.2)$$

Consider a two-component series system subjected to some fatal shocks. Assume there are three kinds of fatal shocks. Shock A governed by random variable U destroys component 1, shock B governed by random variable V destroys component 2, and shock C governed by random variable W destroys both components simultaneously. We refer to such a system as bivariate homogeneous shock (BHS) model. Clearly, under this model the life length of component 1 is $X = \min(U, W)$ and that of component 2 is $Y = \min(V, W)$. Especially, if the random variables U , V , and W are all exponential, the BHS model is just the BVE model proposed by Marshall and Olkin [9].

A prominent feature of BHS model is its singularity. More specifically, even though U , V , and W all are continuous random variables, the joint distribution of X and Y is usually discontinuous.

Denote the survival functions of U , V , and W as $u(x) = pr(U > x)$, $v(x) = pr(V > x)$, and $w(x) = pr(W > x)$, respectively. Wang and Li [10] have shown that under BHS model,

$$\tau = - \int u^2 v^2 dw^2, \quad (2.3)$$

$$r = \frac{\iint u(x)v(y)w(x \odot y) dx dy}{\{\iint (uw)(x \odot y) dx dy\}^{1/2} \{\iint (vw)(x \odot y) dx dy\}^{1/2}}, \quad (2.4)$$

where $f(x \odot y)$ is defined as $f(x \odot y) = f(x \vee y) - f(x)f(y)$ for any function $f(x)$ with $\vee$ the maximization operator.

In the same manner as Chen et al. [11], we define a proportional hazard model as a submodel of BHS. We say a BHS model is a proportional hazards model if there exist some constants λ , α , and β , such that

$$u(x) = u_0^\lambda(x), \quad v(x) = u_0^\alpha(x), \quad w(x) = u_0^\beta(x), \quad (2.5)$$

for some baseline survival function $u_0(x)$. Clearly, the BVE model is a proportional hazard model with $u_0(x) = e^{-x}, x \geq 0$.

We refer the constants λ , α , and β as to shape constants since they are determined by the structural representation of system. If an association measure depends only on the structural representation of system, we then say it is fully structure determined.

Under proportional hazard model, by (2.3), we can easily obtain

$$\tau = \frac{\beta}{(\lambda + \alpha + \beta)}. \quad (2.6)$$

Thus, the Kendall's τ is fully structure determined. The correlation coefficient r , as we can verify, is not fully structure determined.

The exact expression of r is not so easy to obtain in general, except for two special cases when $u_0(x) = e^{-x}, x \geq 0$, or $u_0(x) = 1 - x, 0 \leq x \leq 1$, that is, when the baseline distribution is exponential or uniform. For our convenience, we refer to the first model as model A and the second one as model B. Clearly, model A is just the BVE model. Now we list the main results in the following theorem.

Theorem 2.1. *Let r and τ be the Pearson's correlation coefficient and Kendall's tau. Then under models A or B, one has the following.*

- (i) $r \geq \tau$.
- (ii) *The limit of the ratio of r/τ can be any number larger than 1 when the model approaches to independence situation.*
- (iii) $|2r - 3\tau| \leq 1$.

Proof. The results of Theorem 2.1 in model A are just that of Edwardes [8]. So, we only need to focus on model B. Denote $g(\theta) = \iint u^\theta(x \odot y) dx dy$, and $f(\lambda, \alpha, \beta) = \iint u^\lambda(x) u^\alpha(y) u^\beta(x \odot y) dx dy$. When $u(x) = 1 - x, 0 \leq x \leq 1$, we have,

$$g(\theta) = \iint u^\theta(x \odot y) dx dy = 2 \int_0^1 (1-x)^\theta dx \int_0^x (1-(1-y)^\theta) dy = \frac{\theta}{(\theta+1)^2(\theta+2)}. \quad (2.7)$$

$$\begin{aligned}
f(\lambda, \alpha, \beta) &= \iint u^\lambda(x) u^\alpha(y) u^\beta(x \odot y) dx dy \\
&= \iint_{x \geq y} u^\lambda(x) u^\alpha(y) u^\beta(x) \{1 - u^\beta(y)\} dx dy \\
&\quad + \iint_{x < y} u^\lambda(x) u^\alpha(y) u^\beta(y) \{1 - u^\beta(x)\} dx dy \\
&= \frac{1}{\alpha+1} \cdot \frac{1}{\lambda+\beta+1} - \frac{1}{\alpha+1} \cdot \frac{1}{\lambda+\alpha+\beta+2} \\
&\quad - \frac{1}{\lambda+\beta+1} \cdot \frac{1}{\alpha+\beta+1} + \frac{1}{\alpha+\beta+1} \cdot \frac{1}{\lambda+\alpha+2\beta+2} \\
&\quad + \frac{1}{\lambda+1} \cdot \frac{1}{\alpha+\beta+1} - \frac{1}{\lambda+1} \cdot \frac{1}{\lambda+\alpha+\beta+2} \\
&\quad - \frac{1}{\lambda+\beta+1} \cdot \frac{1}{\alpha+\beta+1} + \frac{1}{\lambda+\beta+1} \cdot \frac{1}{\lambda+\alpha+2\beta+2} \\
&= \frac{\beta}{(\alpha+\beta+1)(\lambda+\beta+1)(\lambda+\alpha+\beta+2)}.
\end{aligned} \tag{2.8}$$

Thus, we obtain,

$$r = \frac{\beta}{(\lambda+\alpha+\beta+2)} \cdot \sqrt{\frac{(\lambda+\beta+2)(\alpha+\beta+2)}{(\lambda+\beta)(\alpha+\beta)}}. \tag{2.9}$$

□

Denote

$$\Gamma = \frac{r}{\tau} = \frac{\lambda+\alpha+\beta}{\lambda+\alpha+\beta+2} \cdot \sqrt{\frac{(\lambda+\beta+2)(\alpha+\beta+2)}{(\lambda+\beta)(\alpha+\beta)}}. \tag{2.10}$$

We want to show $\Gamma > 1$. Denote $\alpha + \beta = u$, $\lambda + \beta = v$, and $\lambda + \alpha + \beta = w$. Then,

$$\begin{aligned}
\Gamma &= \frac{r}{\tau} = \frac{w}{w+2} \sqrt{\frac{u+2}{u} \cdot \frac{v+2}{v}} \\
&= \frac{1}{1+(2/w)} \sqrt{\left(1 + \frac{2}{u}\right) \left(1 + \frac{2}{v}\right)}.
\end{aligned} \tag{2.11}$$

Clearly, we have, $\max\{u, v\} \leq w \leq u + v$. With a little bit notational confusion, we relabel $2/u, 2/v$ and $2/w$ as u, v , and w , respectively. Then, we have, $w \leq \min\{u, v\}$. Without loss of generality, we assume $u \geq v$, and then,

$$\begin{aligned}\Gamma &= \frac{1}{1+w} \sqrt{(1+u)(1+v)} \\ &\geq \frac{1}{1+v} \sqrt{(1+v)(1+v)} = 1.\end{aligned}\tag{2.12}$$

As we can see, the equality holds only when $u = v = w$, that is, $\alpha = \lambda = \beta = 0$. Hence, when the three parameters are not all zero, $\Gamma > 1$, and thus $r > \tau$.

Consider

$$\Phi = \lim_{\beta \rightarrow 0^+} \frac{r}{\tau} = \frac{\lambda + \alpha}{\lambda + \alpha + 2} \sqrt{\left(1 + \frac{2}{\lambda}\right) \left(1 + \frac{2}{\alpha}\right)}.\tag{2.13}$$

In a similar way, we can show that $\Phi > 1$. We can show that Φ can be any number that is larger than 1. Let $\alpha = k\lambda$, then,

$$\Phi = \frac{\sqrt{(1+\lambda)(1+k\lambda)}}{1 + (k/(1+k))\lambda}.\tag{2.14}$$

As $k \rightarrow 0$, $\Phi \rightarrow \sqrt{1+\lambda}$, which can be any number that is larger than 1.

Denote $\Psi = 2r - 3\tau$, then,

$$\Psi = \frac{2\beta}{(\lambda + \alpha + \beta + 2)} \sqrt{\frac{(\lambda + \beta + 2)(\alpha + \beta + 2)}{(\lambda + \beta)(\alpha + \beta)}} - \frac{3\beta}{\lambda + \alpha + \beta}.\tag{2.15}$$

Since Ψ is symmetric about α and λ , the minimum or maximum of Ψ will be attained on $\alpha = \lambda$. So we just need to show that the minimum or maximum of Ψ will be between -1 and 1 .

When $\alpha = \lambda$, Ψ becomes

$$\begin{aligned}\Psi &= \frac{2\beta}{(2\alpha + \beta + 2)} \frac{\alpha + \beta + 2}{\alpha + \beta} - \frac{3\beta}{2\alpha + \beta} \\ &= \frac{2\beta(\alpha + \beta + 2)(2\alpha + \beta) - 3\beta(2\alpha + \beta + 2)(\alpha + \beta)}{(\alpha + \beta)(2\alpha + \beta)(2\alpha + \beta + 2)} \\ &= \frac{N}{D},\end{aligned}\tag{2.16}$$

Table 1

	r	τ	r/τ	$2r - 3\tau$
Case 1	0.3984	0.3312	1.2032	-0.1966
Case 2	0.1774	0.1788	0.9919	-0.1817
Case 3	0.5801	0.4271	1.3581	-0.1212
Case 4	0.6208	0.3924	1.5821	0.0644
Case 5	0.1100	0.1788	0.6151	-0.3165

where

$$N = -\left[\beta^3 + \beta^2(3\alpha + 2) + \beta(2\alpha^2 - 2\alpha)\right], \quad (2.17)$$

$$D = \beta^3 + \beta^2(5\alpha + 2) + \beta(8\alpha^2 + 6\alpha) + 4\alpha^3 + 4\alpha.$$

We have,

$$D + N = 2\alpha\beta^2 + (6\alpha^2 + 8\alpha)\beta + 4\alpha^2(\alpha + 1) > 0, \quad (2.18)$$

$$D - N = 2\beta^3 + (8\alpha + 4)\beta^2 + (10\alpha^2 + 4\alpha)\beta + 4\alpha^2(\alpha + 1) > 0.$$

Hence, $\Psi = N/D$ will be between -1 and 1 .

3. Computational Simulations

In order to investigate the relationships between r and τ under general BHS model, we conduct some computational simulations. For the sample data $\{(x_i, y_i), i = 1, 2, \dots, n\}$, the sample's r is computed as follows:

$$\hat{r} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}. \quad (3.1)$$

While the sample's Kendall's τ is computed as follows:

$$\hat{\tau} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} a_{ij} b_{ij}, \quad (3.2)$$

where $a_{ij} = 1$ if $x_i < x_j$, and -1 if $x_i > x_j$, and $b_{ij} = 1$ if $y_i < y_j$, and -1 if $y_i > y_j$. We compute the sample values of r , τ , r/τ , and $2r - 3\tau$ under several cases. The sample size is set as $n = 100$, and for each computation, the iteration is 100. Table 1 gives the results.

In case 1, we set U , V , and W as uniform variables on $[0, 1]$. By (2.6) and (2.9), we obtain, $r = 0.4$ and $\tau = 1/3$, and the ratio $r/\tau = 1.2$. In case 2, the variables U , V , and W all follow exponential distributions with means 1, 2, and 3, respectively. In this case,

$r = \tau = (1/3)/(1 + 1/2 + 1/3) = 2/11 = 0.1818$, so the ratio is exactly 1. Based on the computation results for these two cases, we can see that the numerical computations are quite precise. In case 3, we set all variables to follow standard normal distribution. In case 4, we set $U = Z^2$, $V = (Z + 0.2)^2$, and $W = (Z + 1.0)^2$, where Z is the variable of standard normal. In case 5, we set $U = E_1^2$, $V = E_2^2$, and $W = E_3^2$, where $E_1 \sim \exp(1)$, $E_2 \sim \exp(2)$, and $E_3 \sim \exp(3)$. Surprisingly, in this case, the ratio r/τ is less than 1. In all the cases, we find the Daniels' inequality $|2r - 3\tau| \leq 1$ holds.

4. Concluding Remarks

The relationship between Spearman's ρ , which is the ordinary Pearson correlation coefficient, and Kendall's τ has been of interest for a long time. However, little attention has been paid on the relationship of Pearson correlation coefficient and Kendall's τ . In this paper, we investigate their relationship under the so-called BHS model. We find that even though for some typical BHS models, the Capéraà-Genest type inequality, $r \geq \tau \geq 0$, holds, but for general BHS model, the inequality may not hold. Our simulation studies suggest that the Daniels-type inequality, $-1 \leq 2r - 3\tau \leq 1$, will hold under BHS model. We thus conjecture that the Daniels-type inequality will be valid in general. However, theoretical confirmation for such conjecture merits further study.

References

- [1] T. P. Hutchinson and C. D. Lai, *Continuous Multivariate Distributions, Emphasising Applications*, Rumsby Scientific Publishing, Adelaide, Australia, 1990.
- [2] W. Hürlimann, "Hutchinson-Lai's conjecture for bivariate extreme value copulas," *Statistics & Probability Letters*, vol. 61, no. 2, pp. 191–198, 2003.
- [3] P. Munroe, T. Ransford, and C. Genest, "Un contre-exemple à une conjecture de Hutchinson et Lai," *Les Comptes Rendus de l'Académie des Sciences de Paris, Series I*, vol. 348, no. 5–6, pp. 305–310, 2010.
- [4] G. A. Fredricks and R. B. Nelsen, "On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables," *Journal of Statistical Planning and Inference*, vol. 137, no. 7, pp. 2143–2150, 2007.
- [5] P. Capéraà and C. Genest, "Spearman's ρ is larger than Kendall's τ for positively dependent random variables," *Journal of Nonparametric Statistics*, vol. 2, no. 2, pp. 183–194, 1993.
- [6] C. Genest and J. Nešlehová, "Analytical proofs of classical inequalities between Spearman's ρ and Kendall's τ ," *Journal of Statistical Planning and Inference*, vol. 139, no. 11, pp. 3795–3798, 2009.
- [7] H. E. Daniels, "Rank correlation and population models," *Journal of the Royal Statistical Society Series B*, vol. 12, pp. 171–181, 1950.
- [8] M. D. Edwards, "Kendall's τ is equal to the correlation coefficient for the BVE distribution," *Statistics & Probability Letters*, vol. 17, no. 5, pp. 415–419, 1993.
- [9] A. W. Marshall and I. Olkin, "A multivariate exponential distribution," *Journal of the American Statistical Association*, vol. 62, pp. 30–44, 1967.
- [10] J. Wang and Y. Li, "Dependency measures under bivariate homogeneous shock models," *Statistics*, vol. 39, no. 1, pp. 73–80, 2005.
- [11] Y. Y. Chen, M. Hollander, and N. A. Langberg, "Small-sample results for the Kaplan-Meier estimator," *Journal of the American Statistical Association*, vol. 77, no. 377, pp. 141–144, 1982.