Asymptotic Properties of Pearson's Rank-Variate Correlation Coefficient under Contaminated Gaussian Model

Rubao Ma[1], Weichao Xu[1]*, Yun Zhang[1], Zhongfu Ye[2]

1 Department of Automatic Control, School of Automation, Guangdong University of Technology, Guangzhou, Guangdong, China, 2 Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui, China

Abstract

This paper investigates the robustness properties of Pearson's rank-variate correlation coefficient (PRVCC) in scenarios where one channel is corrupted by impulsive noise and the other is impulsive noise-free. As shown in our previous work, these scenarios that frequently encountered in radar and/or sonar, can be well emulated by a particular bivariate contaminated Gaussian model (CGM). Under this CGM, we establish the asymptotic closed forms of the expectation and variance of PRVCC by means of the well known Delta method. To gain a deeper understanding, we also compare PRVCC with two other classical correlation coefficients, i.e., Spearman's rho (SR) and Kendall's tau (KT), in terms of the root mean squared error (RMSE). Monte Carlo simulations not only verify our theoretical findings, but also reveal the advantage of PRVCC by an example of estimating the time delay in the particular impulsive noise environment.

Introduction

Correlation coefficients are indices that depict the strength of statistical relationship between two random variables obeying a joint probability distribution [1]. In general, correlation coefficients should be large and positive if there is a high probability that large (small) values of one variable occur in conjunction with large (small) values of another; and it should be large and negative if the direction reverses [2]. Due to their theoretical and algorithmic advantages, correlation coefficients have been widely used in many sub-areas of signal processing [3–18]. Among many methods of correlation analysis in practice, Pearson's product moment correlation coefficient (PPMCC), Kendall's tau (KT) and Spearman's rho (SR) are perhaps the most prevalent ones [19].

There are many advantages and disadvantages to these three classical coefficients. PPMCC is optimal under the bivariate normal model (BNM), and is appropriate mainly for characterizing linear correlations. However, it will output misleading results if nonlinearity is involved in the data. On the other hand, the two rank-based coefficients, SR and KT, are not as powerful as PPMCC when the data follows bivariate normal distributions. Nevertheless, they are invariant under increasing monotone transformations, which makes them more suitable for many

nonlinear cases in practice. Moreover, theoretical and empirical results indicate that SR and KT surpass PPMCC when data are corrupted by impulsive noise [12,20]. Besides these three classical coefficients, some methods such as Pearson's rank-variate correlation coefficient (PRVCC) [21], Gini correlation (GC) [22], and order statistics correlation coefficient (OSCC) [23–25] among others [26–29], have also been proposed in the literature.

It was known long before that PPMCC is an optimal estimator of the population correlation coefficient, in the sense of unbiasedness and approaching the Cramer-Rao lower bound for large samples under bivariate normal models [30]. Despite its desired properties just mentioned, PPMCC might not be applicable under the circumstances where the data are corrupted by impulsive noise, that is, the distribution of data deviates from the BNM. Consider the following scenario that frequently occurs in radar, sonar or communication. We have a prescribed signal, whether deterministic or stochastic, whose statistical property is priorly known. Our purpose is to estimate the correlation between this "clean" signal and the associated distorted version from the receiver that might be corrupted by a tiny fraction of impulsive noise (outliers with very large variance [31–33]). To deal with such case, one might adopt the conventional strategy, that is, ranking

the cardinal variable(s) and resorting afterwards to SR or KT [22], which are robust against both nonlinearity and impulsive noise [20]. However, using only ranks of the two variables, we unavoidably lose useful information embedded in the variates of the "clean" variable. A better strategy would be to rely on coefficients, such as PRVCC [21], that accommodate both ordinal and cardinal information contained in the samples. Our purpose in this work is thus to investigate the properties of the historical PRVCC by both theoretical and empirical means.

The contribution in this work is twofold. Firstly, we establish the asymptotic closed forms of the expectation and variance of PRVCC under a contaminated Gaussian model that emulates a frequently encountered scenario in practice. Secondly, we demonstrate the superiority of PRVCC over PPMCC, SR and KT, in terms of the root mean squared error, by an example of estimating the time delay in the particular impulsive noise environment. These theoretical and empirical findings might be helpful to rejuvenate the historical PRVCC, which has long been forgotten in the literature due to insufficient understanding on its theoretical properties.

For convenience of later discussion, we employ symbols $\mathbb{E}(\cdot)$, $\mathbb{V}(\cdot)$, $\mathbb{C}(\cdot,\cdot)$ and $\mathrm{corr}(\cdot,\cdot)$ to denote the mean, variance, covariance and correlation of (between) random variables, respectively. Univariate and bivariate normal distributions are denoted by $\mathcal{N}(\mu,\sigma^2)$ and $\mathcal{N}(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, respectively. The sign $\simeq$ reads "is approximately equal to", whereas the sign $\triangleq$ stands for "is defined as". The notation $u(t)=\mathcal{O}(v(t))$ denotes that $u(t)/v(t)\to 0$ as $t\to L$(might be infinite) [34]. The symbol $n^{[m]}$ stands for the product $n(n-1)\cdots(n-m+1)$. Other notations will be defined where it first enters the text.

Methods

This section presents the definitions of PRVCC as well as a particular CGM model simulating the impulsive noise environment mentioned in the previous section. Moreover, some auxiliary results are also established for further theoretical analysis.

1 Definitions of PRVCC

Let X and Y be two random variables following a continuous bivariate distribution. Denote by F_X and F_Y the marginal distributions of X and Y, respectively. Then, according to a historical paper of Pearson [21], one of the population versions of PRVCC can be defined, in modern notation, by

$$r_{\mathcal{H}}^P(Y,X) \triangleq \mathrm{corr}[F_X(X),Y] = \frac{\mathbb{C}[F_X(X),Y]}{\sqrt{\mathbb{V}[F_X(X)]\mathbb{V}(Y)}}. \quad (1)$$

Exchanging the roles of X and Y in (1) yields the other version

$$r_{\mathcal{H}}^P(X,Y) \triangleq \mathrm{corr}[F_Y(Y),X] = \frac{\mathbb{C}[F_Y(Y),X]}{\sqrt{\mathbb{V}[F_Y(Y)]\mathbb{V}(X)}}.$$

Let $\{(X_i,Y_i)\}_{i=1}^n$ be n independent and identically distributed (i.i.d.) data pairs drawn from a continuous bivariate population. After rearranging $\{X_i\}_{i=1}^n$ in ascending order, we get a new sequence $X_{(1)} < \cdots < X_{(n)}$, which is termed the order statistics of

X [35–37]. Suppose that X_j is at the kth position in the sorted sequence. The integer $k\in[1,n]$ is named the rank of X_j and is denoted by $P_j\in\{P_j\}_{j=1}^n$. Let $\bar{\xi}$ represent the arithmetic mean of n data points $\{\xi_i\}_{i=1}^n$. Then, based on (1), the sample version with respect to $r_{\mathcal{H}}^P(Y,X)$ can be constructed as

$$r_{\mathcal{H}}(Y_i,X_i) \triangleq \frac{\sum_{i=1}^n \left[\hat{\boldsymbol{F}}_X(X_i)-\bar{\hat{\boldsymbol{F}}}_X(X_i)\right](Y_i-\bar{Y})}{\left\{\sum_{i=1}^n\left[\hat{\boldsymbol{F}}_X(X_i)-\bar{\hat{\boldsymbol{F}}}_X(X_i)\right]^2\sum_{i=1}^n(Y_i-\bar{Y})^2\right\}^{\frac{1}{2}}} \quad (2)$$

where $\{\hat{\boldsymbol{F}}_X(X_i)\}_{i=1}^n$ is the empirical cdf of $\{X_i\}_{i=1}^n$. Substituting the relationship $\hat{\boldsymbol{F}}_X(X_i)=P_i/n$ into (2) along with some simplifications leads to

$$r_{\mathcal{H}}(Y_i,X_i) = \left[\frac{3n}{n+1}\right]^{\frac{1}{2}}\frac{\frac{1}{n(n-1)}\sum_{i=1}^n(2P_i-1-n)Y_i}{\left[\frac{1}{n-1}\sum_{i=1}^n(Y_i-\bar{Y})^2\right]^{\frac{1}{2}}}. \quad (3)$$

Exchanging the role of X and Y in (3) gives another sample version $r_{\mathcal{H}}(X_i,Y_i)$ with respect to $r_{\mathcal{H}}^P(X,Y)$. Note that in general $r_{\mathcal{H}}(X,Y)\neq r_{\mathcal{H}}(Y,X)$. The choice between $r_{\mathcal{H}}(X_i,Y_i)$ and $r_{\mathcal{H}}(Y_i,X_i)$ depends on different roles played by X and Y in the scenario mentioned in the previous section. To avoid redundancy, we will focus on the properties of $r_{\mathcal{H}}(Y_i,X_i)$ which is abbreviated as $r_{\mathcal{H}}$ in the sequel unless ambiguity occurs.

2 Contaminated Gaussian Model

To simulate the specific circumstance remarked in Section **Introduction**, throughout we utilize the following CGM representing the joint probability density function (pdf) of two random variables X and Y [20]

$$(1-\varepsilon)\mathcal{N}(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)+\varepsilon\mathcal{N}(\mu'_x,\mu'_y,\sigma'^2_x,\sigma'^2_y,\rho') \quad (4)$$

where $0<\varepsilon\ll 1$, $\mu_x=\mu'_x=\mu_y=\mu'_y$, $\sigma'_y=\sigma_y$, $\sigma'_x\gg\sigma_x$ and $\sigma'_x\gg\sigma_y$. Under this specific CGM, it is obvious that the marginal distribution of Y is $\mathcal{N}(\mu_y,\sigma_y^2)$, whereas the marginal distribution of X is $(1-\varepsilon)\mathcal{N}(\mu_x,\sigma_x^2)+\varepsilon\mathcal{N}(\mu_x,\sigma'^2_x)$. In other words, under Model (4), Y stands for a "clean" normal variable while X stands for a "dirty" variable corrupted by a tiny fraction of Gaussian component with vary large variance (might tending to infinity). In this model, the parameter ρ, which is considered of interest, is what we aim at estimating as accurate as possible, while ρ' and ε are interferences we seek to suppress. For the reason why Model (4) can happen in practice, please see Appendix A in our previous work [20].

3 Auxiliary Results

To establish our major results of Theorem 1 in the next subsection, some auxiliary results summarized in Lemma 1 below are mandatory.

Table 1. Quantities for Evaluation of B_1 in Eq. (36).

Subsets	# of terms	W_1	W_3	W_4	ϱ_{13}	ϱ_{14}	ϱ_{34}
$i \neq k = j$	$n_1^{[2]}$	$\mathcal{X}_i - \mathcal{X}_k$	$\mathcal{Y}_i$	$\mathcal{Y}_j$	$\rho/\sqrt{2}$	$-\rho/\sqrt{2}$	0
$i = j \neq k$	$n_1^{[2]}$	$\mathcal{X}_i - \mathcal{X}_k$	$\mathcal{Y}_i$	$\mathcal{Y}_j$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	1
$i \neq j \neq k$	$n_1^{[3]}$	$\mathcal{X}_i - \mathcal{X}_k$	$\mathcal{Y}_i$	$\mathcal{Y}_j$	$\rho/\sqrt{2}$	0	0
$i = j, k'$	$n_1 n_2$	$\mathcal{X}_i - \mathcal{X}'_{k'}$	$\mathcal{Y}_i$	$\mathcal{Y}_j$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	1
$i \neq j, k'$	$n_1^{[2]} n_2$	$\mathcal{X}_i - \mathcal{X}'_{k'}$	$\mathcal{Y}_i$	$\mathcal{Y}_j$	$\rho\sigma_x/\sqrt{\lambda}$	0	0
$i \neq k, j'$	$n_1^{[2]} n_2$	$\mathcal{X}_i - \mathcal{X}_k$	$\mathcal{Y}_i$	$\mathcal{Y}_{j'}$	$\rho\sqrt{2}$	0	0
$j' = k', i$	$n_1 n_2$	$\mathcal{X}_i - \mathcal{X}'_{k'}$	$\mathcal{Y}_i$	$\mathcal{Y}_{j'}$	$\rho\sigma_x/\sqrt{\lambda}$	$-\rho'\sigma_x/\sqrt{\lambda}$	0
$j' \neq k', i$	$n_1 n_2^{[2]}$	$\mathcal{X}_i - \mathcal{X}'_{k'}$	$\mathcal{Y}_i$	$\mathcal{Y}_{j'}$	$\rho\sigma_x/\sqrt{\lambda}$	0	0
$j = k, i'$	$n_1 n_2$	$\mathcal{X}'_{i'} - \mathcal{X}_k$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_j$	$\rho'\sigma_x/\sqrt{\lambda}$	$-\rho\sigma_x/\sqrt{\lambda}$	0
$j \neq k, i'$	$n_1^{[2]} n_2$	$\mathcal{X}'_{i'} - \mathcal{X}_k$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_j$	$\rho'\sigma_x/\sqrt{\lambda}$	0	0
$i' \neq k' j$	$n_1 n_2^{[2]}$	$\mathcal{X}'_{i'} - \mathcal{X}'_{k'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_j$	$\rho'/\sqrt{2}$	0	0
$i' = j', k$	$n_1 n_2$	$\mathcal{X}'_{i'} - \mathcal{X}_k$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{j'}$	$\rho'\sigma_x/\sqrt{\lambda}$	$\rho'\sigma_x/\sqrt{\lambda}$	1
$i' \neq j', k$	$n_1 n_2^{[2]}$	$\mathcal{X}'_{i'} - \mathcal{X}_k$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{j'}$	$\rho'\sigma_x/\sqrt{\lambda}$	0	0
$i' \neq k' = j'$	$n_2^{[2]}$	$\mathcal{X}''_{i'} - \mathcal{X}'_{k'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{j'}$	$\rho'/\sqrt{2}$	$-\rho'/\sqrt{2}$	0
$i' = j' \neq k'$	$n_2^{[2]}$	$\mathcal{X}'_{i'} - \mathcal{X}'_{k'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{j'}$	$\rho'/\sqrt{2}$	$\rho'/\sqrt{2}$	1
$i' \neq j' \neq k'$	$n_2^{[3]}$	$\mathcal{X}'_{i'} - \mathcal{X}'_{k'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{j'}$	$\rho'/\sqrt{2}$	0	0

[1] In the second column $n_p^{[m]} \triangleq n_p(n_p - 1) \ldots (n_p - m + 1)$ for $p=1,2$.
[2] In the columns for $\varrho_{(\cdot)}$, $\lambda \triangleq \sigma_x^2 + \sigma_{x'}^2$.

Table 2. Quantities for Evaluating $A=\sum_{i,j,k,l}^{n}\mathbb{E}\{H(X_i-X_k)H(X_j-X_l)Y_iY_j\}$ in Eq. (37).

ℓ	Subsets	$\alpha_\ell(n_1,n_2)$	W_1	W_2	W_3	W_4	ϱ_{12}	ϱ_{13}	ϱ_{14}	ϱ_{23}	ϱ_{24}	ϱ_{34}
1	$i=j\neq k=l$	$n_1^{[2]}$	X_i-X_k	X_j-X_l	Y_i	Y_j	1	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	1
2	$l=i\neq j=k$	$n_1^{[2]}$	X_i-X_k	X_j-X_l	Y_i	Y_j	-1	$\rho/\sqrt{2}$	$-\rho/\sqrt{2}$	$-\rho/\sqrt{2}$	$\rho/\sqrt{2}$	0
3	$i=j\neq k\neq l$	$n_1^{[3]}$	X_i-X_k	X_j-X_l	Y_i	Y_j	$1/2$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	1
4	$k=j\neq i\neq l$	$n_1^{[3]}$	X_i-X_k	X_j-X_l	Y_i	Y_j	$-1/2$	$\rho/\sqrt{2}$	$-\rho/\sqrt{2}$	0	$\rho/\sqrt{2}$	0
5	$l=i\neq j\neq k$	$n_1^{[3]}$	X_i-X_k	X_j-X_l	Y_i	Y_j	$-1/2$	$\rho/\sqrt{2}$	0	$-\rho/\sqrt{2}$	$\rho/\sqrt{2}$	0
6	$l=k\neq i\neq j$	$n_1^{[3]}$	X_i-X_k	X_j-X_l	Y_i	Y_j	$1/2$	$\rho/\sqrt{2}$	0	0	$\rho/\sqrt{2}$	0
7	$i\neq j\neq k\neq l$	$n_1^{[4]}$	X_i-X_k	X_j-X_l	Y_i	Y_j	0	$\rho/\sqrt{2}$	0	0	$\rho/\sqrt{2}$	0
8	$i=j\neq k,l'$	$n_1^{[2]}n_2$	X_i-X_k	$X_j-X'_{l'}$	Y_i	Y_j	$\sigma_x/\sqrt{2\lambda}$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	1
9	$i\neq j=k,l'$	$n_1^{[2]}n_2$	X_i-X_k	$X_j-X'_{l'}$	Y_i	Y_j	$-\sigma_x/\sqrt{2\lambda}$	$\rho/\sqrt{2}$	$-\rho/\sqrt{2}$	0	$\rho\sigma_x/\sqrt{\lambda}$	0
10	$i\neq j\neq k,l'$	$n_1^{[3]}n_2$	X_i-X_k	$X_j-X'_{l'}$	Y_i	Y_j	0	$\rho/\sqrt{2}$	0	0	$\rho\sigma_x/\sqrt{\lambda}$	0
11	$i=j\neq l,k'$	$n_1^{[2]}n_2$	$X_i-X'_{k'}$	X_j-X_l	Y_i	Y_j	$\sigma_x/\sqrt{2\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho/\sqrt{2}$	$\rho/\sqrt{2}$	1
12	$i=l\neq j,k'$	$n_1^{[2]}n_2$	$X_i-X'_{k'}$	X_j-X_l	Y_i	Y_j	$\sigma_x/\sqrt{2\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	0	$-\rho/\sqrt{2}$	$\rho/\sqrt{2}$	0
13	$i\neq j\neq l,k'$	$n_1^{[3]}n_2$	$X_i-X'_{k'}$	X_j-X_l	Y_i	Y_j	0	$\rho\sigma_x/\sqrt{\lambda}$	0	0	$\rho/\sqrt{2}$	0
14	$j=k\neq l,i'$	$n_1^{[2]}n_2$	$X'_{i'}-X_k$	X_j-X_l	$Y'_{i'}$	Y_j	$\sigma_x/\sqrt{2\lambda}$	$\rho'\sigma'_x/\sqrt{\lambda}$	$-\rho\sigma_x/\sqrt{\lambda}$	0	$\rho/\sqrt{2}$	0
15	$j\neq k=l,i'$	$n_1^{[2]}n_2$	$X'_{i'}-X_k$	X_j-X_l	$Y'_{i'}$	Y_j	$\sigma_x/\sqrt{2\lambda}$	$\rho'\sigma'_x/\sqrt{\lambda}$	0	0	$\rho/\sqrt{2}$	0
16	$j\neq k\neq l,i'$	$n_1^{[3]}n_2$	$X'_{i'}-X_k$	X_j-X_l	$Y'_{i'}$	Y_j	0	$\rho'\sigma'_x/\sqrt{\lambda}$	0	0	$\rho/\sqrt{2}$	0
17	$i\neq k=l,j'$	$n_1^{[2]}n_2$	X_i-X_k	$X'_{j'}-X_l$	Y_i	$Y'_{j'}$	$\sigma_x/\sqrt{2\lambda}$	$\rho/\sqrt{2}$	0	0	$\rho'\sigma'_x/\sqrt{\lambda}$	0
18	$i=k\neq l,j'$	$n_1^{[2]}n_2$	X_i-X_k	$X'_{j'}-X_l$	Y_i	$Y'_{j'}$	$\sigma_x/\sqrt{2\lambda}$	$\rho/\sqrt{2}$	0	$-\rho\sigma_x/\sqrt{\lambda}$	$\rho'\sigma'_x/\sqrt{\lambda}$	0
19	$i\neq k\neq l,j'$	$n_1^{[3]}n_2$	X_i-X_k	$X'_{j'}-X_l$	Y_i	$Y'_{j'}$	0	$\rho/\sqrt{2}$	0	0	$\rho'\sigma'_x/\sqrt{\lambda}$	0
20	$i=j,k'=l'$	$n_1 n_2$	$X_i-X'_{k'}$	$X_j-X'_{l'}$	Y_i	Y_j	$(\sigma_x\sigma_x+\sigma'_x\sigma'_x)/\lambda$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	1
21	$i=j,k'\neq l'$	$n_1 n_2$	$X_i-X'_{k'}$	$X_j-X'_{l'}$	Y_i	Y_j	$\sigma_x\sigma_x/\lambda$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	1
22	$i\neq j,k'=l'$	$n_1^{[2]}n_2$	$X_i-X'_{k'}$	$X_j-X'_{l'}$	Y_i	Y_j	$\sigma'_x\sigma'_x/\lambda$	$\rho\sigma_x/\sqrt{\lambda}$	0	0	$\rho\sigma_x/\sqrt{\lambda}$	0
23	$i\neq j,k'\neq l'$	$n_1^{[2]}n_2^{[2]}$	$X_i-X'_{k'}$	$X_j-X'_{l'}$	Y_i	Y_j	0	$\rho\sigma_x/\sqrt{\lambda}$	0	0	$\rho\sigma_x/\sqrt{\lambda}$	0
24	$i'\neq k',j\neq l$	$n_1^{[2]}n_2^{[2]}$	$X'_{i'}-X'_{k'}$	X_j-X_l	$Y'_{i'}$	Y_j	0	$\rho'/\sqrt{2}$	0	0	$\rho/\sqrt{2}$	0
25	$i'=l',j=k$	$n_1 n_2$	$X'_{i'}-X_k$	$X_j-X'_{l'}$	$Y'_{i'}$	Y_j	$-(\sigma_x\sigma_x+\sigma'_x\sigma'_x)/\lambda$	$\rho'\sigma'_x\sqrt{\lambda}$	$-\rho\sigma_x/\sqrt{\lambda}$	$-\rho'\sigma'_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	0
26	$i'\neq l',j=k$	$n_1 n_2^{[2]}$	$X'_{i'}-X_k$	$X_j-X'_{l'}$	$Y'_{i'}$	Y_j	$\sigma_x\sigma_x/\lambda$	$\rho'\sigma'_x/\sqrt{\lambda}$	$-\rho\sigma_x/\sqrt{\lambda}$	0	$\rho\sigma_x/\sqrt{\lambda}$	0
27	$i'=l',j\neq k$	$n_1^{[2]}n_2$	$X'_{i'}-X_k$	$X_j-X'_{l'}$	$Y'_{i'}$	Y_j	$\sigma'_x\sigma'_x/\lambda$	$\rho'\sigma'_x/\sqrt{\lambda}$	0	$-\rho'\sigma'_x/\sqrt{\lambda}$	$\rho\sigma_x/\sqrt{\lambda}$	0
28	$i'\neq l',j\neq k$	$n_1^{[2]}n_2^{[2]}$	$X'_{i'}-X_k$	$X_j-X'_{l'}$	$Y'_{i'}$	Y_j	0	$\rho'\sigma'_x/\sqrt{\lambda}$	0	0	$\rho\sigma_x/\sqrt{\lambda}$	0

[1]In the third column $n_p^{[m]}\triangleq n_p(n_p-1)\ldots(n_p-m+1)$ for $p=1,2$.
[2]In the columns for $\varrho_{(\cdot)}$ $\lambda\triangleq\sigma_x^2+\sigma_x'^2$.

Lemma 1. *Assume that the random vector $[W_1\ W_2\ W_3\ W_4]^\top$ follows a quadrivariate normal distribution with $\mathbb{E}(W_r)=0$, $\mathbb{V}(W_r)=\sigma_r^2$, and $\mathrm{corr}(W_r,W_s)=\varrho_{rs}$ for $r,s=1,\ldots,4$. Write $H(t)=1$ for $t>0$ and $H(t)=0$ for $t\leq 0$. Then*

$$\mathcal{I}\triangleq\mathbb{E}\{H(W_1)H(W_2)W_3W_4\}$$
$$=\frac{1}{2\pi}\frac{\sigma_3\sigma_4}{\sqrt{1-\varrho_{12}^2}}[\varrho_{13}\varrho_{24}+\varrho_{14}\varrho_{23}-\varrho_{12}(\varrho_{13}\varrho_{14}+\varrho_{23}\varrho_{24})]$$
$$+\sigma_3\sigma_4\varrho_{34}\left(\frac{1}{4}+\frac{\sin^{-1}\varrho_{12}}{2\pi}\right),\qquad(5)$$

$$\mathcal{J}\triangleq\mathbb{E}\{H(W_1)W_2W_3W_4\}$$
$$=\frac{\sigma_2\sigma_3\sigma_4}{\sqrt{2\pi}}(\varrho_{12}\varrho_{34}+\varrho_{13}\varrho_{24}+\varrho_{14}\varrho_{23}-\varrho_{12}\varrho_{13}\varrho_{14}),\qquad(6)$$

$$\mathcal{K}\triangleq\mathbb{E}\{H(W_1)W_2W_3\}=\frac{\sigma_2\sigma_3\varrho_{23}}{2},\qquad(7)$$

and

Table 3. Quantities for Evaluation of K in Eq. (45).

Subsets	# of terms	W_1	W_2	W_3	ρ_{12}	ϱ_{13}	ϱ_{23}
$i \neq j = l$	$n_1^{[2]}$	$\mathcal{X}_i - \mathcal{X}_j$	$\mathcal{Y}_i$	$\mathcal{Y}_l - \bar{Y}$	$\rho/\sqrt{2}$	$-\rho/\sqrt{2\vartheta}$	$-1/n\sqrt{\vartheta}$
$i = k \neq j$	$n_1^{[2]}$	$\mathcal{X}_i - \mathcal{X}_j$	$\mathcal{Y}_i$	$\mathcal{Y}_l - \bar{Y}$	$\rho/\sqrt{2}$	$\rho/\sqrt{2\vartheta}$	$\sqrt{\vartheta}$
$i \neq j \neq l$	$n_1^{[3]}$	$\mathcal{X}_i - \mathcal{X}_j$	$\mathcal{Y}_i$	$\mathcal{Y}_l - \bar{Y}$	$\rho\sqrt{2}$	0	$-1/n\sqrt{\vartheta}$
$j = l, i'$	$n_1 n_2$	$\mathcal{X}_{i'} - \mathcal{X}_j$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_l - \bar{Y}$	$\rho\sigma_{x'}/\sqrt{\lambda}$	$-\dfrac{n\rho\sigma\vartheta + \rho'\sigma'_x}{n\sqrt{\lambda\vartheta}}$	$-1/n\sqrt{\vartheta}$
$j \neq l, i'$	$n_1^{[2]} n_2$	$\mathcal{X}_{i'} - \mathcal{X}_j$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_l - \bar{Y}$	$\rho\sigma_{x'}/\sqrt{\lambda}$	$\dfrac{\rho\sigma - \rho'\sigma'_x}{n\sqrt{\lambda\vartheta}}$	$-1/n\sqrt{\vartheta}$
$i = l, j'$	$n_1 n_2$	$\mathcal{X}_i - \mathcal{X}_{j'}$	$\mathcal{Y}_i$	$\mathcal{Y}_l - \bar{Y}$	$\rho\sigma/\sqrt{\lambda}$	$\dfrac{n\rho\sigma\vartheta + \rho'\sigma'_x}{n\sqrt{\lambda\vartheta}}$	$\sqrt{\vartheta}$
$i \neq l, j'$	$n_1^{[2]} n_2$	$\mathcal{X}_i - \mathcal{X}_{j'}$	$\mathcal{Y}_i$	$\mathcal{Y}_l - \bar{Y}$	$\rho\sigma_{x'}/\sqrt{\lambda}$	$\dfrac{\rho'\sigma'_x - \rho\sigma}{n\sqrt{\lambda\vartheta}}$	$-1/n\sqrt{\vartheta}$
$i \neq j, l'$	$n_1^{[2]} n_2$	$\mathcal{X}_i - \mathcal{X}_j$	$\mathcal{Y}_i$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho/\sqrt{2}$	0	$-1/n\sqrt{\vartheta}$
$l, l' \neq j'$	$n_1 n_2^{[2]}$	$\mathcal{X}_{i'} - \mathcal{X}_{j'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho'/\sqrt{2}$	0	$-1/n\sqrt{\vartheta}$
$j, i' = l'$	$n_1 n_2$	$\mathcal{X}_{i'} - \mathcal{X}_j$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho\sigma_{x'}/\sqrt{\lambda}$	$\dfrac{n\rho'\sigma'_x\vartheta + \rho\sigma}{n\sqrt{\lambda\vartheta}}$	$\sqrt{\vartheta}$
$j, i' \neq l'$	$n_1^{[2]} n_2$	$\mathcal{X}_{i'} - \mathcal{X}_j$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho\sigma_{x'}/\sqrt{\lambda}$	$\dfrac{\rho\sigma - \rho'\sigma'_x}{n\sqrt{\lambda\vartheta}}$	$-1/n\sqrt{\vartheta}$
$i, j' = l'$	$n_1 n_2$	$\mathcal{X}_i - \mathcal{X}_{j'}$	$\mathcal{Y}_i$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho\sigma/\sqrt{\lambda}$	$-\dfrac{n\rho'\sigma'_x\vartheta + \rho\sigma}{n\sqrt{\lambda\vartheta}}$	$-1/n\sqrt{\vartheta}$
$i, j' \neq l'$	$n_1 n_2^{[2]}$	$\mathcal{X}_i - \mathcal{X}_{j'}$	$\mathcal{Y}_i$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho\sigma/\sqrt{\lambda}$	$\dfrac{\rho'\sigma'_x - \rho\sigma}{n\sqrt{\lambda\vartheta}}$	$-1/n\sqrt{\vartheta}$
$i' \neq j' = l'$	$n_2^{[2]}$	$\mathcal{X}_{i'} - \mathcal{X}_{j'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho'/\sqrt{2}$	$-\rho'/\sqrt{2\vartheta}$	$-1/n\sqrt{\vartheta}$
$i' = l' \neq j'$	$n_2^{[2]}$	$\mathcal{X}_{i'} - \mathcal{X}_{j'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho'/\sqrt{2}$	$\rho'/\sqrt{2\vartheta}$	$\sqrt{\vartheta}$
$i' \neq j' \neq l'$	$n_2^{[3]}$	$\mathcal{X}_{i'} - \mathcal{X}_{j'}$	$\mathcal{Y}_{i'}$	$\mathcal{Y}_{l'} - \bar{Y}$	$\rho'/\sqrt{2}$	0	$-1/n\sqrt{\vartheta}$

[1] In the second column $n_p^{[m]} \triangleq n_p(n_p - 1)\cdots(n_p - m + 1)$ for $p = 1,2$.
[2] In the columns for $\varrho_{(\cdot)}$ $\lambda \triangleq \sigma_x^2 + \sigma_{x'}^2$ and $\vartheta \triangleq 1 - \frac{1}{n}$.
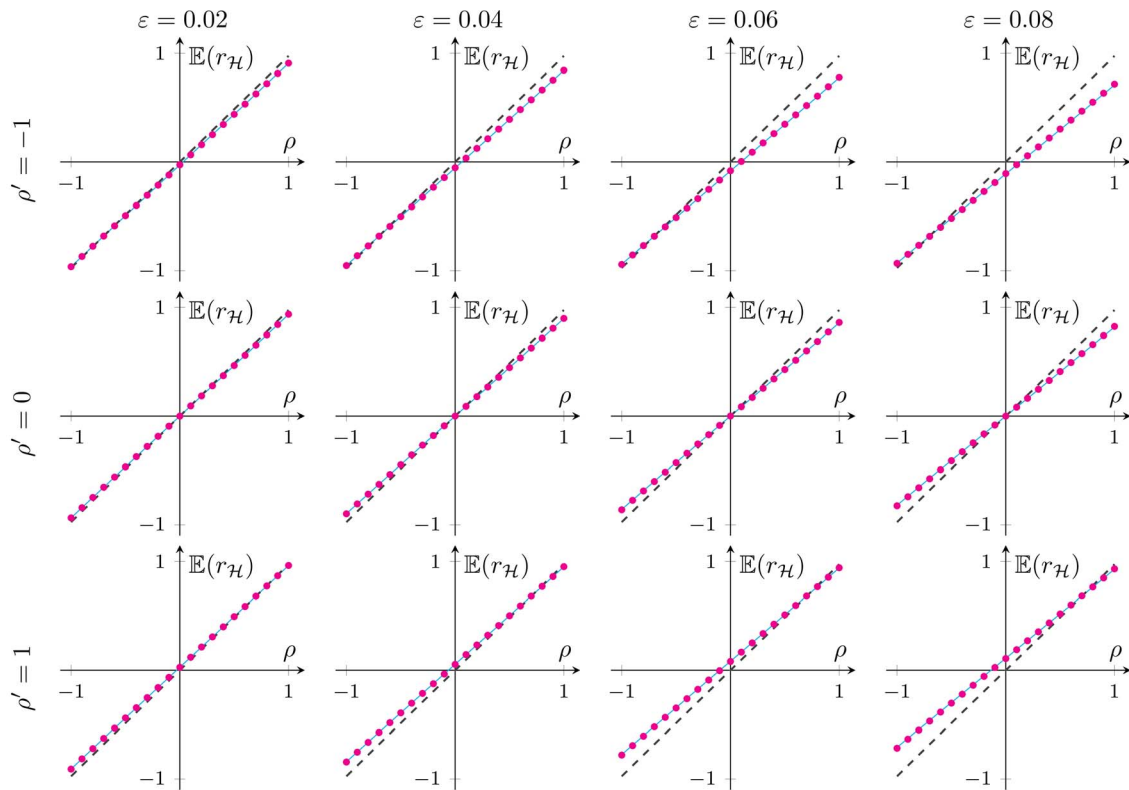
Figure 1. The numerical verification of (21), the expectation of $r_{\mathcal{H}}$ in Theorem 1. The number of samples is chose as $n=100$. In the vertically up direction, ρ' is decreasing following $\{1,0,-1\}$ respectively; whereas ε corresponds to a increasing trend in the horizontally right direction, following $\{0.02,0.04,0.06,0.08\}$ respectively. It shows a good agreement between the simulation result (circles) and the theoretical computation (solid lines) in each subplot. As a reference, the contamination-free version (49) is also posted together (see dashed curves).
doi:10.1371/journal.pone.0112215.g001

$$\mathcal{L} \stackrel{\Delta}{=} \mathbb{E}\{H(W_1)W_2\} = \frac{\sigma_2 \varrho_{12}}{\sqrt{2\pi}}. \tag{8}$$

Proof. Write $\mathcal{W}_r = W_r/\sigma_r$ for $r=1,2,3,4$. Then

$$\mathcal{I} = \sigma_3 \sigma_4 \mathbb{E}\{H(\mathcal{W}_1)H(\mathcal{W}_2)\mathcal{W}_3\mathcal{W}_4\}, \tag{9}$$

$$\mathcal{J} = \sigma_2 \sigma_3 \sigma_4 \mathbb{E}\{H(\mathcal{W}_1)\mathcal{W}_2\mathcal{W}_3\mathcal{W}_4\}, \tag{10}$$

$$\mathcal{K} = \sigma_2 \sigma_3 \mathbb{E}\{H(\mathcal{W}_1)\mathcal{W}_2\mathcal{W}_3\}, \tag{11}$$

$$\mathcal{L} = \sigma_2 \mathbb{E}\{H(\mathcal{W}_1)\mathcal{W}_2\}. \tag{12}$$

The results of (5) and (7) follow readily by substituting the results in [39] into the right sides of (9) and (11), respectively. Next we show that (6) and (8) also hold true. Let $i \stackrel{\Delta}{=} \sqrt{-1}$. Then, according to

[39], it follows that

$$\mathbb{E}\{H(\mathcal{W}_1)\mathcal{W}_2\mathcal{W}_3\mathcal{W}_4\} =$$

$$\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\sum_{r=0}^{\infty} \mathcal{B}_{lmnpqr} \varrho_{12}^l \varrho_{13}^m \varrho_{14}^n \varrho_{23}^p \varrho_{24}^q \varrho_{34}^r \tag{13}$$
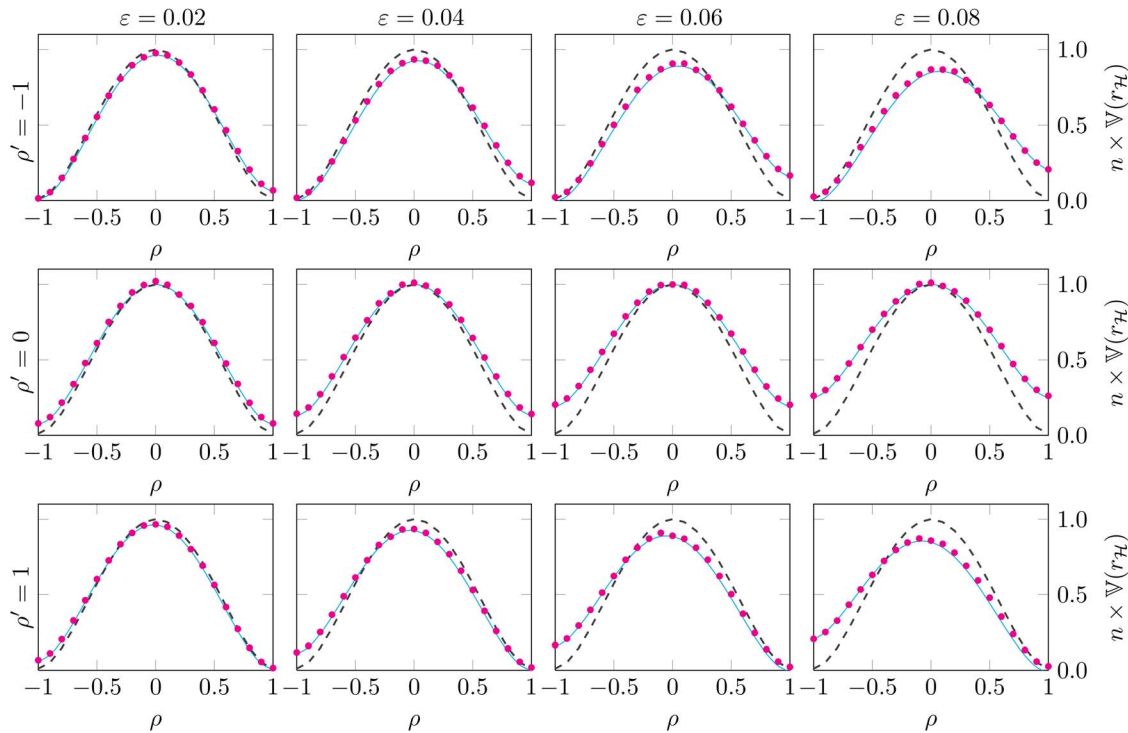
where

$$\mathcal{B}_{lmnpqr} =$$

$$(-1)^{l+m+n+p+q+r} \frac{\mathcal{G}_{l+m+n}\mathcal{H}_{l+p+q}\mathcal{H}_{m+p+r}\mathcal{H}_{n+q+r}}{l!m!n!p!q!r!}, \tag{14}$$

$$\mathcal{G}_s = \begin{cases} \frac{1}{2} & \text{for } s=0 \\ 0 & \text{for } s=2,4,6,\ldots \\ \frac{(2k)!}{i\sqrt{2\pi}2^k k!} & \text{for } s=2k+1, k=0,1,2,\ldots \end{cases}. \tag{15}$$

and

$$\mathcal{H}_s = \begin{cases} 0 & \text{for } s=0; \\ -i & \text{for } s=1; \\ 0 & \text{for } s\geq 2. \end{cases} \tag{16}$$

Figure 2. The numerical verification of (22), the variance of $r_{\mathcal{H}}$ in Theorem 1. The number of samples is chose as $n = 100$. In the vertically up direction, ρ' is decreasing following $\{1, 0, -1\}$ respectively; whereas ε corresponds to a increasing trend in the horizontally right direction, following $\{0.02, 0.04, 0.06, 0.08\}$ respectively. It shows a good agreement between the simulation result (circles) and the theoretical computation (solid lines) in each subplot. As a reference, the contamination-free version (50) is also posted together (see dashed curves).
doi:10.1371/journal.pone.0112215.g002

It is seen that $l+p+q$, $m+p+r$ and $n+q+r$ are all subscripts of the $\mathcal{H}$-terms in (14). Since, by (16), only $\mathcal{H}_1$ is non-null, then (14) and hence (13) are non-null only when the following conditions are satisfied, i.e.,

$$\begin{cases} l+p+q=1 \\ m+p+r=1 \\ n+q+r=1 \\ l,m,n,p,q,r\geq 0. \end{cases} \tag{17}$$

It is easy to verify that there are only four solutions to (17), as

$$\begin{cases} l=r=1, p=q=m=m=0 \\ l=m=n=1, p=q=r=0 \\ m=q=1, l=p=r=n=0 \\ n=p=1, l=m=r=q=0. \end{cases} \tag{18}$$

Substituting (18) into (13) and using (15) and (16) thereafter produce

$$\mathbb{E}\{H(\mathcal{W}_1)\mathcal{W}_2\mathcal{W}_3\mathcal{W}_4\} = (-1)^2 \mathcal{G}_1(-i)^3 (\varrho_{12}\varrho_{34}+\varrho_{13}\varrho_{24}+\varrho_{14}\varrho_{23})$$
$$+ (-1)^3 \mathcal{G}_3(-i)^3 \varrho_{12}\varrho_{13}\varrho_{14} \tag{19}$$
$$= \frac{1}{\sqrt{2\pi}}(\varrho_{12}\varrho_{34}+\varrho_{13}\varrho_{24}+\varrho_{14}\varrho_{23}-\varrho_{12}\varrho_{13}\varrho_{14})$$

which along with (10) leads to (6). By a similar argument, we have

$$\mathbb{E}\{H(\mathcal{W}_1)\mathcal{W}_2\} = (-1)^1 \frac{\mathcal{G}_1\mathcal{H}_1\varrho_{12}^1}{1!} = \frac{\varrho_{12}}{\sqrt{2\pi}} \tag{20}$$

which together with (12) yields (8). This completes the proof of the lemma. $\qquad\square$

4 Asymptotic Mean and Variance of PRVCC Under CGM (4)

By applying the *delta method* [38] with Lemma 1, we are ready to establish the closed forms of the mean and variance of PRVCC for samples generated by CGM (4).

Theorem 1. *Let $\{(\mathcal{X}_j, \mathcal{Y}_j)\}_{j=1}^{n_1}$ be n_1 i.i.d. data pairs drawn from a bivariate normal population $\mathcal{N}(0, 0, \sigma_x^2, \sigma_y^2, \rho)$ and $\{(\mathcal{X}_{j'}', \mathcal{Y}_{j'}')\}_{j'=1}^{n_2}$ be n_2 i.i.d. data pairs drawn from another bivariate normal population $\mathcal{N}(0, 0, \sigma_x'^2, \sigma_y'^2, \rho')$. Assume that $(\mathcal{X}, \mathcal{Y})$ and $(\mathcal{X}', \mathcal{Y}')$ are mutually independent. Write $n = n_1 + n_2$ and $\varepsilon = n_2/n$. Denote by $\{(X_i, Y_i)\}_{i=1}^{n}$ the union of $\{(\mathcal{X}_j, \mathcal{Y}_j)\}_{j=1}^{n_1}$ and $\{(\mathcal{X}_{j'}', \mathcal{Y}_{j'}')\}_{j'=1}^{n_2}$. Then, as n large, ε small, $\sigma_y' = \sigma_y$, $\sigma_x' \gg \sigma_x$ and $\sigma_x' \gg \sigma_y$, the expectation and variance of PRVCC defined in (3) are*

$$\mathbb{E}(r_{\mathcal{H}}) \simeq \sqrt{\frac{3}{\pi}}\rho\left[(1-2\varepsilon)\left(1-\frac{3}{4n}+\frac{\rho^2}{2n}\right)\right]$$
$$+ \sqrt{\frac{6}{\pi}}\varepsilon\rho'\left(1+\frac{1}{4n}-\frac{1}{\sqrt{2}n}+\frac{\rho'^2}{2n}\right) \tag{21}$$

Table 4. RMSE of Three Estimators for $n=100$, $\varepsilon=\{0.02, 0.04, 0.06, 0.08\}$ and $\rho'=\{-1,0\}$.

ρ	$\varepsilon=0.02$			$\varepsilon=0.04$			$\varepsilon=0.06$			$\varepsilon=0.08$		
	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$
−1.0	0.0111	**0.0007**	0.0187	0.0200	**0.0023**	0.0287	0.0287	**0.0047**	0.0389	0.0369	**0.0078**	0.0488
−0.9	0.0254	**0.0233**	0.0262	0.0278	**0.0257**	0.0301	0.0306	**0.0287**	0.0346	0.0336	**0.0320**	0.0396
−0.8	0.0416	0.0402	**0.0400**	0.0411	0.0402	**0.0402**	0.0404	0.0401	0.0404	**0.0394**	0.0396	0.0405
−0.7	0.0563	0.0551	**0.0537**	0.0548	0.0538	**0.0525**	0.0532	0.0523	**0.0510**	0.0519	0.0508	**0.0498**
−0.6	0.0687	0.0678	**0.0659**	0.0670	0.0659	**0.0641**	0.0666	0.0652	**0.0635**	0.0665	0.0647	**0.0627**
−0.5	0.0793	0.0788	**0.0765**	0.0787	0.0779	**0.0757**	0.0797	0.0784	**0.0760**	0.0820	0.0803	**0.0772**
−0.4	0.0882	0.0881	**0.0856**	0.0889	0.0884	**0.0858**	0.0918	0.0910	**0.0876**	0.0972	0.0962	**0.0916**
−0.3	0.0950	0.0952	**0.0924**	0.0977	0.0978	**0.0943**	0.1037	0.1037	**0.0991**	0.1123	0.1124	**0.1058**
−0.2	0.1013	0.1018	**0.0989**	0.1057	0.1063	**0.1022**	0.1146	0.1154	**0.1094**	0.1269	0.1283	**0.1197**
−0.1	0.1051	0.1059	**0.1026**	0.1125	0.1134	**0.1087**	0.1251	0.1266	**0.1194**	0.1412	0.1436	**0.1331**
0.0	0.1074	0.1083	**0.1050**	0.1173	0.1185	**0.1133**	0.1345	0.1364	**0.1282**	0.1543	0.1573	**0.1456**
0.1	0.1085	0.1093	**0.1057**	0.1226	0.1236	**0.1181**	0.1431	0.1449	**0.1362**	0.1672	0.1702	**0.1579**
0.2	0.1082	0.1086	**0.1050**	0.1263	0.1267	**0.1213**	0.1509	0.1521	**0.1436**	0.1809	0.1832	**0.1712**
0.3	0.1058	0.1056	**0.1025**	0.1287	0.1281	**0.1236**	0.1583	0.1579	**0.1510**	0.1925	0.1930	**0.1828**
0.4	0.1021	0.1008	**0.0984**	0.1303	0.1276	**0.1251**	0.1658	0.1625	**0.1585**	0.2044	0.2015	**0.1951**
0.5	0.0979	**0.0949**	0.0942	0.1320	**0.1260**	0.1270	0.1723	**0.1645**	0.1660	0.2152	**0.2070**	0.2069
0.6	0.0924	**0.0869**	0.0890	0.1321	**0.1212**	0.1282	0.1783	**0.1639**	0.1731	0.2259	**0.2095**	0.2193
0.7	0.0860	**0.0764**	0.0837	0.1323	**0.1138**	0.1300	0.1831	**0.1584**	0.1801	0.2355	**0.2068**	0.2315
0.8	0.0789	**0.0628**	0.0784	0.1322	**0.1020**	0.1326	0.1879	**0.1475**	0.1882	0.2443	**0.1967**	0.2440
0.9	0.0726	**0.0450**	0.0750	0.1320	**0.0825**	0.1361	0.1925	**0.1259**	0.1970	0.2528	**0.1733**	0.2575
1.0	0.0660	**0.0047**	0.0752	0.1296	**0.0179**	0.1414	0.1933	**0.0391**	0.2068	0.2567	**0.0679**	0.2708
−1.0	0.0432	**0.0028**	0.0511	0.0795	**0.0095**	0.0898	0.1153	**0.0195**	0.1279	0.1508	**0.0328**	0.1652
−0.9	0.0509	**0.0342**	0.0529	0.0819	**0.0531**	0.0858	0.1133	**0.0748**	0.1189	0.1454	**0.0985**	0.1524
−0.8	0.0594	**0.0504**	0.0587	0.0840	**0.0678**	0.0849	0.1106	**0.0883**	0.1130	0.1376	**0.1106**	0.1411
−0.7	0.0696	**0.0644**	0.0673	0.0876	**0.0783**	0.0864	0.1086	**0.0956**	0.1085	0.1309	**0.1150**	0.1318
−0.6	0.0788	0.0760	**0.0759**	0.0915	0.0863	**0.0892**	0.1072	**0.1000**	0.1055	0.1243	**0.1153**	0.1234
−0.5	0.0867	0.0852	**0.0837**	0.0954	0.0927	**0.0927**	0.1063	0.1025	**0.1040**	0.1187	**0.1139**	0.1169
−0.4	0.0932	0.0927	**0.0902**	0.0985	0.0974	**0.0959**	0.1060	0.1044	**0.1035**	0.1139	0.1117	**0.1116**
−0.3	0.0985	0.0986	**0.0958**	0.1014	0.1013	**0.0988**	0.1060	0.1056	**0.1033**	0.1102	0.1096	**0.1078**
−0.2	0.1022	0.1027	**0.0996**	0.1038	0.1043	**0.1013**	0.1054	0.1058	**0.1029**	0.1072	0.1075	**0.1048**
−0.1	0.1043	0.1052	**0.1021**	0.1048	0.1055	**0.1024**	0.1049	0.1056	**0.1027**	0.1056	0.1063	**0.1033**
0.0	0.1053	0.1061	**0.1030**	0.1052	0.1061	**0.1029**	0.1050	0.1059	**0.1027**	0.1050	0.1058	**0.1028**
0.1	0.1044	0.1052	**0.1020**	0.1051	0.1059	**0.1028**	0.1051	0.1058	**0.1028**	0.1055	0.1062	**0.1032**
0.2	0.1023	0.1029	**0.0999**	0.1035	0.1039	**0.1010**	0.1050	0.1053	**0.1026**	0.1072	0.1075	**0.1048**

Table 4. Cont.

ρ	$\varepsilon=0.02$			$\varepsilon=0.04$			$\varepsilon=0.06$			$\varepsilon=0.08$		
	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$	$\hat{\rho}_S$	$\hat{\rho}_K$	$\hat{\rho}_H$
0.3	0.0984	0.0986	**0.0957**	0.1013	0.1012	**0.0987**	0.1055	0.1052	**0.1029**	0.1106	0.1100	**0.1081**
0.4	0.0930	0.0924	**0.0900**	0.0989	0.0979	**0.0962**	0.1054	0.1038	**0.1029**	0.1140	**0.1118**	0.1119
0.5	0.0865	0.0850	**0.0835**	0.0956	0.0930	**0.0929**	0.1059	**0.1022**	0.1036	0.1187	**0.1139**	0.1167
0.6	0.0788	**0.0759**	**0.0759**	0.0916	**0.0864**	0.0894	0.1069	**0.0997**	0.1055	0.1244	**0.1154**	0.1237
0.7	0.0693	**0.0642**	0.0671	0.0876	**0.0784**	0.0865	0.1083	**0.0953**	0.1083	0.1305	**0.1147**	0.1314
0.8	0.0597	**0.0506**	0.0589	0.0843	**0.0680**	0.0853	0.1109	**0.0888**	0.1131	0.1376	**0.1105**	0.1409
0.9	0.0506	**0.0341**	0.0526	0.0817	**0.0531**	0.0856	0.1136	**0.0749**	0.1192	0.1453	**0.0983**	0.1524
1.0	0.0430	**0.0028**	0.0508	0.0792	**0.0094**	0.0896	0.1155	**0.0196**	0.1281	0.1506	**0.0328**	0.1652

In the upper panel are RMSEs for $\rho'=-1$; whereas in the lower panel are RMSEs for $\rho'=0$. In each of the eight blocks, the minima of RMSE with respect to $\hat{\rho}_S$, $\hat{\rho}_K$ and $\hat{\rho}_H$ are highlighted in bold font in a rowwise manner.

doi:10.1371/journal.pone.0112215.t004

and

$$\mathbb{V}(r_{\mathcal{H}}) \simeq \frac{1}{n}\left[1+\frac{3\rho^2}{\pi}\left(2\sqrt{3}-\frac{11}{2}\right)+\frac{3\rho^4}{\pi}\right]$$
$$-\frac{6\varepsilon}{n\pi}\left[\left(2\sqrt{3}-7\right)\rho^2+\rho'^2+2\rho^4\right]$$
$$-\frac{3\sqrt{2}\varepsilon\rho\rho'}{n\pi}\left(3-\rho^2-\rho'^2\right)$$

(22)

Proof. From (3), it is easy to verify that $r_{\mathcal{H}}$ is shift invariant. Therefore, we lose no generality by assuming that $\mu_x=\mu_y=0$ hereafter. For convenience, write

$$U = \frac{1}{n(n-1)}\sum_{i=1}^{n}(2P_i-1-n)Y_i$$

(23)

$$V = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2.$$

(24)

Then, by the well known *delta method* [38], it follows that

$$\mathbb{E}(r_{\mathcal{H}}) \simeq \sqrt{\frac{3n}{n+1}}\frac{\mathbb{E}(U)}{\sqrt{\mathbb{E}(V)}}\left[1-\frac{1}{2}\frac{\mathbb{C}(U,V)}{\mathbb{E}(U)\mathbb{E}(V)}+\frac{3}{8}\frac{\mathbb{V}(V)}{\mathbb{E}^2(V)}\right]$$

(25)

and

$$\mathbb{V}(r_{\mathcal{H}}) \simeq \frac{3n}{n+1}\frac{\mathbb{E}^2(U)}{\mathbb{E}(V)}\left[\frac{\mathbb{V}(U)}{\mathbb{E}^2(U)}+\frac{1}{4}\frac{\mathbb{V}(V)}{\mathbb{E}^2(V)}-\frac{\mathbb{C}(U,V)}{\mathbb{E}(U)\mathbb{E}(V)}\right].$$

(26)
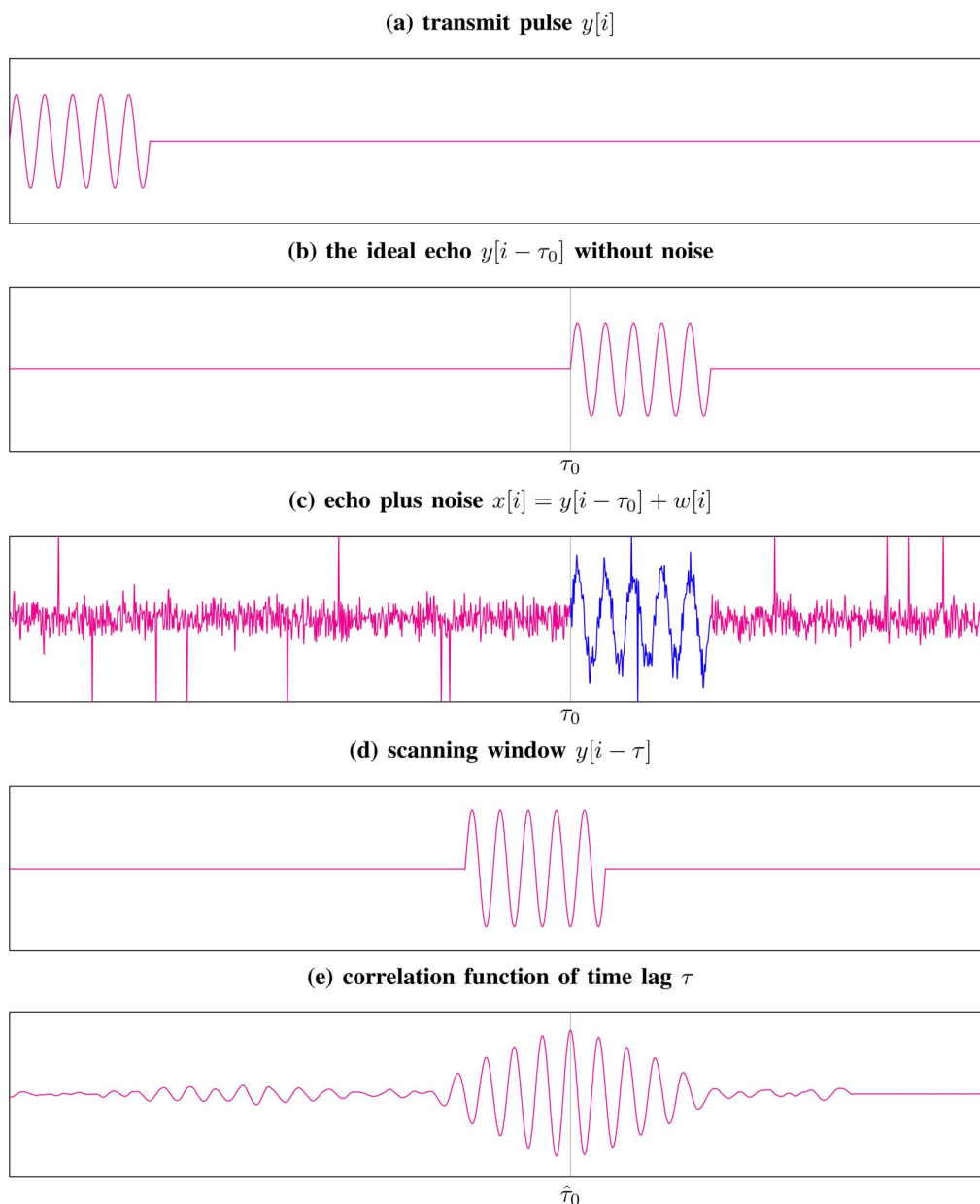
Obviously, we only need to evaluate $\mathbb{E}(U)$, $\mathbb{E}(V)$, $\mathbb{V}(U)$, $\mathbb{V}(V)$ and $\mathbb{C}(U,V)$ in order to work out (25) and (26). From the theorem assumption, both $\{\mathcal{Y}_j\}_{j=1}^{n_1}$ and $\{\mathcal{Y}_j\}_{j'=1}^{n_2}$ follow the same normal distribution $\mathcal{N}(0,\sigma_y^2)$, then $(n-1)V/\sigma_y^2$ obeys a χ_{n-1}^2 distribution, which means that

$$\mathbb{E}(V) = \sigma_y^2$$

(27)

$$\mathbb{V}(V) = \frac{2\sigma_y^4}{n-1}.$$

(28)

Using Lemma 1 with some tedious algebra, we can obtain $\mathbb{E}(U)$, $\mathbb{V}(U)$ and $\mathbb{C}(U,V)$.

We first derive $\mathbb{E}(U)$. From the definition (23) and the relationships $P_i=\sum_{j=1}^{n}H(X_i-X_j)+1$ [40],

(a) transmit pulse $y[i]$

(b) the ideal echo $y[i - \tau_0]$ **without noise**

τ_0

(c) echo plus noise $x[i] = y[i - \tau_0] + w[i]$

τ_0

(d) scanning window $y[i - \tau]$

(e) correlation function of time lag τ

$\hat{\tau}_0$

Figure 3. Schematic illustration of estimating the time-delay τ_0 in Model (4). The time-shift $\hat{\tau}_0$ with respect to the maximum of the correlation function in the bottom panel is considered as an estimate of the true time-delay τ_0.
doi:10.1371/journal.pone.0112215.g003

$$\mathbb{E}(U) = \frac{1}{n^{[2]}} \mathbb{E}\left\{ 2 \sum_{i \neq j=1}^{n} \sum H(X_i - X_j)Y_i - (n-1)\sum_{i=1}^{n} Y_i \right\}. \quad (29)$$

Expanding and recalling that $\mathbb{E}(Y) = 0$, it follows that

$$\mathbb{E}(U) = \frac{2}{n^{[2]}} \sum_{i \neq j=1}^{n} \sum \mathbb{E}\{H(X_i - X_j)Y_i\}. \quad (30)$$

Since, by definition, $\{(X_i, Y_i)\}_{i=1}^{n}$ is a mixture of $\{(\mathcal{X}_j, \mathcal{Y}_j)\}_{j=1}^{n_1}$ and $\{(\mathcal{X}'_{j'}, \mathcal{Y}'_{j'})\}_{j'=1}^{n_2}$, (30) can be expanded as

$$
\begin{aligned}
\mathbb{E}(U) \quad &= \frac{2}{n^{[2]}} \sum_{i=1}^{n_1} \sum_{k=1}^{n_1} \mathbb{E}\{H(\mathcal{X}_i - \mathcal{X}_k)\mathcal{Y}_i\} \\
&+ \frac{2}{n^{[2]}} \sum_{i=1}^{n_1} \sum_{k'=1}^{n_2} \mathbb{E}\{H(\mathcal{X}_i - \mathcal{X}'_{k'})\mathcal{Y}_i\} \\
&+ \frac{2}{n^{[2]}} \sum_{i'=1}^{n_2} \sum_{k=1}^{n_1} \mathbb{E}\{H(\mathcal{X}'_{i'} - \mathcal{X}_k)\mathcal{Y}'_{i'}\} \\
&+ \frac{2}{n^{[2]}} \sum_{i'=1}^{n_2} \sum_{k'=1}^{n_2} \mathbb{E}\{H(\mathcal{X}'_{i'} - \mathcal{X}'_{k'})\mathcal{Y}'_{i'}\}
\end{aligned}
\quad (31)
$$

Table 5. Performance comparison of $r_{\mathcal{H}}$, r_K and $r_{\mathcal{S}}$ for $y[i]$ being a segment of sin wave.

SNR	$r_{\mathcal{H}}$	r_K	$r_{\mathcal{S}}$
0	1000.0 ± 0.96	1000.0 ± 1.01	1000.0 ± 1.03
−1	1000.0 ± 1.05	1000.0 ± 1.27	1000.0 ± 1.12
−2	1000.0 ± 1.13	1000.0 ± 1.23	1000.0 ± 1.23
−3	1000.0 ± 1.26	1000.0 ± 1.71	1000.0 ± 1.70
−4	1000.0 ± 2.22	1000.0 ± 2.08	1000.0 ± 2.31
−5	1000.0 ± 3.37	1000.0 ± 3.57	1000.0 ± 3.41
−6	1000.0 ± 7.34	1000.1 ± 7.21	1000.0 ± 7.45
−7	1000.1 ± 9.22	1000.1 ± 9.29	1000.1 ± 9.23
−8	999.9 ± 14.44	1000.2 ± 14.99	1000.1 ± 15.30
−9	1000.2 ± 19.75	1000.2 ± 20.82	1000.1 ± 21.05
−10	999.7 ± 29.72	999.7 ± 30.69	999.8 ± 30.66

which becomes (32) after some straightforward algebra along with the assistance of (8) in Lemma 1.

$$\mathbb{E}(U) = \frac{\sigma_y}{\sqrt{\pi}} \frac{2}{n^{[2]}} \Big[\frac{n_1^{[2]} \rho}{2} + \frac{n_1 n_2 \sigma_y \rho}{\sqrt{2(\sigma_x^2 + \sigma_x'^2)}} + \frac{n_1 n_2 \sigma_x' \rho'}{\sqrt{2(\sigma_x^2 + \sigma_x'^2)}} + \frac{n_2^{[2]} \rho'}{2} \Big]. \tag{32}$$

Next we evaluate $\mathbb{V}(U)$, which can be written as

$$\mathbb{V}(U) = \mathbb{C}(U,U) = \frac{4A - 2(n-1)B + (n-1)^2 C - D}{n^2(n-1)^2} \tag{33}$$

where

$$D = n^2(n-1)^2 \mathbb{E}(U)\mathbb{E}(U) \tag{34}$$

$$C = \sum_{i,j}^{n} \mathbb{E}(Y_i Y_j) = n\sigma_y^2 \tag{35}$$

$$B = \sum_{i,j,k}^{n} \mathbb{E}\{H(X_i - X_k)Y_i Y_j\} + \sum_{i,j,l}^{n} \mathbb{E}\{H(X_j - X_l)Y_i Y_j\} \tag{36}$$

and

$$A = \sum_{i,j,k,l}^{n} \mathbb{E}\{H(X_i - X_k)H(X_j - X_l)Y_i Y_j\}. \tag{37}$$

The expression of D is easily obtained by substituting (31) into (34).

For convenience, denote by B_1 and B_2 the two triple summations of B in (36). Then it follows that B_1 is decomposable into eight sub-triple summations which can be further partitioned into 16 disjoint and exhaustive subsets that listed in Table 1. An application of (7) to Table 1 leads directly to

$$B_1 = n_1^{[2]} \frac{\sigma_y^2}{2} + n_1 n_2 \frac{\sigma_y^2}{2} + n_1 n_2 \frac{\sigma_y^2}{2} + n_2^{[2]} \frac{\sigma_y^2}{2}.$$

Similarly we also have

$$B_2 = n_1^{[2]} \frac{\sigma_y^2}{2} + n_1 n_2 \frac{\sigma_y^2}{2} + n_1 n_2 \frac{\sigma_y^2}{2} + n_2^{[2]} \frac{\sigma_y^2}{2}.$$

Thus

$$B = B_1 + B_2 = n(n-1)\sigma_y^2.$$

The quadruple summation in (37) can be decomposed as

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} = \Big(\sum_{i=1}^{n_1} + \sum_{i'=1}^{n_2} \Big) \Big(\sum_{j=1}^{n_1} + \sum_{j'=1}^{n_2} \Big) \times \Big(\sum_{k=1}^{n_1} + \sum_{k'=1}^{n_2} \Big) \Big(\sum_{l=1}^{n_1} + \sum_{l'=1}^{n_2} \Big) \tag{38}$$

Expanding (38) according to different suffixes of $(i\,j\,k\,l)$ and $(i'\,j'\,k'\,l')$, we obtain 16 sub-quadruple summations which can be further partitioned into 56 disjoint and exhaustive subsets. In other words, A is a summation of 56 integrals of the form

$\mathbb{E}\{H(W_1)H(W_2)W_3W_4\}$, i.e., the $\mathcal{I}_\ell$-terms, weighted by corresponding subset cardinality, i.e., the α_ℓ-terms. By substituting into (5) the corresponding parameters tabulated in Table 2 as well as exploiting the symmetry of (38), we obtain the expression of A. Substituting the expressions of A, B, C and D into (33) and tidying up lead to the expression of $\mathbb{V}(U)$ in (39).

$$
\begin{aligned}
\mathbb{V}(U) \quad &= \frac{\sigma_y^2}{n^2(n-1)^2}\{\frac{n(n^2-1)}{3} \\
&+ \frac{2\rho^2}{\pi}\left[\sqrt{3}n_1^{[3]} - n_1^{[2]}(2n_1-3)\right] \\
&+ \frac{2\rho'^2}{\pi}[\sqrt{3}n_2^{[3]} - 2\sqrt{2}n_1 n_2^{[2]} \\
&- n_2(n_1^2 - 2n_1 n_2 + 2n_2^2 + 2n_1 - 5n_2 + 3)]\}.
\end{aligned}
\tag{39}
$$

Finally we deal with $\mathbb{C}(U,V)$. Write

$$
U_1 \triangleq \frac{2}{n(n-1)}\sum_{i\neq j=1}^{n}\sum H(X_i - X_j)Y_i
\tag{40}
$$

and

$$
U_2 \triangleq \frac{-1}{n}\sum_{i=1}^{n} Y_i.
\tag{41}
$$

Then $U = U_1 + U_2$. Now

$$
\begin{aligned}
\mathbb{C}(U,V) &= \mathbb{C}(U_1,V) - \underbrace{\mathbb{C}(U_2,V)}_{=0} \\
&= \mathbb{E}(U_1 V) - \mathbb{E}(U_1)\mathbb{E}(V) \\
&= \mathbb{E}(U_1 V) - \mathbb{E}(U)\mathbb{E}(V).
\end{aligned}
\tag{42}
$$

Since we have $\mathbb{E}(U)$ in (32) and $\mathbb{E}(V)$ in (27), the second term in (42) can be easily obtained, as

$$
\begin{aligned}
\mathbb{E}(U)\mathbb{E}(V) &= \frac{\sigma_y^3}{\sqrt{\pi}}\frac{2}{n^{[2]}}[\frac{n_1^{[2]}\rho}{2} + \frac{n_1 n_2 \sigma_y \rho}{\sqrt{2(\sigma_x^2 + \sigma_x'^2)}} \\
&+ \frac{n_1 n_2 \sigma'_x \rho'}{\sqrt{2(\sigma_x^2 + \sigma_x'^2)}} + \frac{n_2^{[2]}\rho'}{2}],
\end{aligned}
\tag{43}
$$

whereas the first term, $\mathbb{E}(U_1 V)$, can be written as

$$
\mathbb{E}(U_1 V) = \frac{2K}{n(n-1)^2}
\tag{44}
$$

where

$$
K = \mathbb{E}\left\{\sum_{i\neq j=1}^{n}\sum_{l=1}^{n} H(X_i - X_j)Y_i(Y_l - \bar{Y})(Y_l - \bar{Y})\right\}.
\tag{45}
$$

It is evident that $(X_i - X_j)$, Y_i, $(Y_l - \bar{Y})$ and $(Y_l - \bar{Y})$ can be regarded as W-terms in Lemma 1, i.e.,

$$
[(X_i - X_j) \ \ Y_i \ \ (Y_l - \bar{Y}) \ \ (Y_l - \bar{Y})]^\top \quad = [W_1 \ \ W_2 \ \ W_3 \ \ W_4]^\top.
$$

As mentioned above, X is a union of $\mathcal{X}$ and $\mathcal{X}'$, and Y is a union of $\mathcal{Y}$ and $\mathcal{Y}'$, the triple summation in (45) can be split into eight terms as

$$
\begin{aligned}
\sum_{i\neq j=1}^{n}\sum\sum_{l=1}^{n} &= \left(\sum_{i\neq j=1}^{n_1}\sum\sum_{l=1}^{n_1}\right) + \left(\sum_{i'=1}^{n_2}\sum_{j=1}^{n_1}\sum_{l=1}^{n_1}\right) + \\
&\left(\sum_{i=1}^{n_1}\sum_{j'=1}^{n_2}\sum_{l=1}^{n_1}\right) + \left(\sum_{i\neq j=1}^{n_1}\sum\sum_{l'=1}^{n_2}\right) + \\
&\left(\sum_{i'\neq j'=1}^{n_2}\sum\sum_{l=1}^{n_1}\right) + \left(\sum_{i'=1}^{n_2}\sum_{j=1}^{n_1}\sum_{l'=1}^{n_2}\right) + \\
&\left(\sum_{i=1}^{n_1}\sum_{j'=1}^{n_2}\sum_{l'=1}^{n_2}\right) + \left(\sum_{i'\neq j'=1}^{n_2}\sum\sum_{l'=1}^{n_2}\right).
\end{aligned}
\tag{46}
$$

Using (6) in Lemma 1 along with the corresponding parameters tabulated in Table 3, we can work out each sub-triple summation in (46). A series of straightforward algebra leads readily to (47).

$$
\begin{aligned}
\mathbb{C}(U,V) &= \frac{2}{n(n-1)^2}\{\frac{n_1^{[2]}(n-1)\sigma_y^3\rho}{n\sqrt{\pi}}\left[\frac{n+2}{2} + \frac{1}{n-1} - \frac{\rho^2 n}{2(n-1)}\right] \\
&+ \frac{n_2^{[2]}(n-1)\sigma_y^3\rho'}{n\sqrt{\pi}}\left[\frac{n+2}{2} + \frac{1}{n-1} - \frac{\rho'^2 n}{2(n-1)}\right] \\
&+ \frac{n_1 n_2(n+1)\sigma_y^3}{\sqrt{2\pi}}\frac{\rho'\sigma'_x + \sigma_x\rho}{\sqrt{\sigma_x'^2 + \sigma_x^2}} \\
&+ \frac{n_1 n_2\sigma_y^3}{\sqrt{2\pi}}\frac{\rho'\sigma'_x + \sigma_x\rho}{\sqrt{\sigma_x'^2 + \sigma_x^2}(\sigma_x'^2 + \sigma_x^2)} \\
&\times \left[\frac{1}{n}(\rho'\sigma'_x - \rho\sigma_x)^2 - (\rho'^2\sigma_x'^2 + \rho^2\sigma_x^2)\right]\} \\
&- \frac{\sigma_y^3}{\sqrt{\pi}}\frac{2}{n^{[2]}}[\frac{n_1^{[2]}\rho}{2} + \frac{n_1 n_2\sigma_y\rho}{\sqrt{2(\sigma_x^2 + \sigma_x'^2)}} + \frac{n_1 n_2\sigma'_x\rho'}{\sqrt{2(\sigma_x^2 + \sigma_x'^2)}} + \frac{n_2^{[2]}\rho'}{2}].
\end{aligned}
\tag{47}
$$

Substituting (27),(28),(32),(39) and (47) into (25) and (26), respectively, letting $\varepsilon = n_2/n$ and $\sigma'_x \to \infty$, and omitting $\mathcal{O}(n^{-2})$ terms thereafter, we finally arrive at (21) and (22), respectively. The theorem thus follows. $\square$

Remark 1. *Letting $n \to \infty$ in (21), $\mathbb{E}(r_{\mathcal{H}})$ simplifies to a neater form of*

$$
\mathbb{E}(r_{\mathcal{H}}) = \sqrt{\frac{3}{\pi}}\left[(1-2\varepsilon)\rho + \sqrt{2}\varepsilon\rho'\right]
\tag{48}
$$

which manifests how ε and ρ' affect the accuracy of the estimate of ρ by PRVCC. Moreover, as $\varepsilon \to 0$, (21) and (22) reduce respectively to

$$
\mathbb{E}(r_{\mathcal{H}}) \simeq \sqrt{\frac{3}{\pi}}\rho\left(1 - \frac{3}{4n} + \frac{\rho^2}{2n}\right)
\tag{49}
$$

and

$$\mathbb{V}(r_{\mathcal{H}}) \simeq \frac{1}{n}\left[1 + \frac{3\rho^2}{\pi}\left(2\sqrt{3} - \frac{11}{2}\right) + \frac{3\rho^4}{\pi}\right] \qquad (50)$$

which are the contamination-free versions of PRVCC.

Results and Discussion

In this section we verify the correctness of Theorem 1 by Monte Carlo simulations. To gain a further insight about PRVCC, we also compare it with the two classical correlation coefficients, i.e., SR and KT, in terms of RMSE. At last, we will provide an examples of time-delay estimation under CGM (4). Since the theoretical results in Theorem 1 only hold true for large sample size n and small ε, in this section we set the sample size $n=100$ and $0.02 \leq \varepsilon \leq 0.1$. All samples are generated by suitable functions in the Matlab environment. For the sake of accuracy, the number of Monte Carlo trials is set to be 10^5 unless otherwise stated.

1 Verification of Theorem 1

Figure 1. verifies the correctness of the mean of PRVCC under CGM (4) for large samples and small ε. Specifically, in Figure 1. we plot the simulation results (circles) and the theoretical results of (21) (solid lines), and the contamination-free version (49) (dashed lines) under different combinations of $\rho' \in \{-1,0,1\}$ and $\varepsilon \in \{0.02,0.04,0.06,0.08\}$. Good agreements are observed between the simulation results and the theoretical counterparts. It can also be observed that the larger the contamination fraction ε and difference between ρ and ρ', the bigger the bias between $\mathbb{E}(r_{\mathcal{H}})$ and the ideal dashed curve corresponding to $\varepsilon=0$.

Figure 2. verifies the correctness of the variance of PRVCC, by plotting the simulation results (circles) and the theoretical results of (22) (solid lines) concerning $\mathbb{V}(r_{\mathcal{H}})$ in the same scenarios as in Figure 1. For the purpose of comparison, the contamination-free version (50) (dashed lines) is also included in each subplot to highlight the effects of ρ' and ε. Note that we have multiplied $\mathbb{V}(r_{\mathcal{H}})$ by n for a better visual effect. This figure shows good agreements between the simulation results and the corresponding theoretical ones. Moreover, it is seen that when $\rho'=0$, the curves are symmetric and the magnitude of $\mathbb{V}(r_{\mathcal{H}})$ increase with ε, especially for $|\rho|$ large. On the other hand, when $\rho' \neq 0$, the curves are no longer asymmetric. Specifically, for $|\rho|$ large, $\mathbb{V}(r_{\mathcal{H}})$ increases if ρ and ρ' have opposite signs; and it decreases if ρ and ρ' have the same signs. When ε is fixed, $\mathbb{V}(r_{\mathcal{H}})|_{\rho'<0}$ is the reversal of $\mathbb{V}(r_{\mathcal{H}})|_{\rho'>0}$.

From these two figures, it follows that, although derived based on the assumptions $n \to \infty$ and $\varepsilon \to 0$, our theoretical results established in Theorem 1 are sufficiently accurate for n as small as 100 and ε as large as 0.08. This means that, PRVCC is applicable to many situations in practice, such as radar and biomedical engineering, where the sample size n is much larger than 100 and the fraction of impulsive interference ε is much lower than 0.08.

2 RMSE Comparison of PRVCC with SR amd KT

To deepen the understanding of PRVCC, in this subsection we compare in terms of RMSE the performance of PRVCC with SR and KT, which are also robust under the CGM (4) as shown in our previous work [20].

For fairness of comparison, some calibrations are necessary. From (21), it follows

$$\lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \mathbb{E}(r_{\mathcal{H}}) = \sqrt{\frac{3}{\pi}}\rho \qquad (51)$$

by which we can define an asymptotic unbiased estimator of the population correlation ρ, as

$$\hat{\rho}_H \overset{\Delta}{=} \sqrt{\frac{\pi}{3}} r_{\mathcal{H}}.$$

The other two asymptotic unbiased estimators based on KT and SR are defined as [20]

$$\hat{\rho}_K \overset{\Delta}{=} \sin\left(\frac{\pi}{2} r_K\right)$$

$$\hat{\rho}_S \overset{\Delta}{=} 2\sin\left(\frac{\pi}{6} r_S\right).$$

Given definitions of the $\hat{\rho}$ above, we can then compare their performance in terms of the popular RMSE defined by

$$\text{RMSE} \overset{\Delta}{=} \sqrt{\mathbb{E}(\hat{\rho}-\rho)^2}.$$

The CGM based on (4) is set to be

$$(1-\varepsilon)\mathcal{N}(0,0,3^2,3^2,\rho) + \varepsilon\mathcal{N}(0,0,10^8,3^2,\rho'),$$

where $\varepsilon \in \{0.02,0.04,0.06,0.08\}$, $\rho' \in \{-1,0\}$ and ρ increases from -1 to 1 by a step of 0.1. The RMSEs are listed in Table 4, where the minima with respect to $\hat{\rho}_H$, $\hat{\rho}_K$ and $\hat{\rho}_S$ are highlighted in bold font in a rowwise manner within each of the eight blocks. It appears that 1) all RMSEs are quite small, meaning that these three estimators perform similarly well under CGM (4); 2) $\hat{\rho}_H$ outperforms $\hat{\rho}_K$ and $\hat{\rho}_S$ in the cases that $|\rho|$ takes values from 0 to medium magnitudes; 3) $\hat{\rho}_K$ outperforms $\hat{\rho}_H$ for ρ around ± 1; 4) $\hat{\rho}_S$ plays an intermediate role between $\hat{\rho}_H$ and $\hat{\rho}_K$. Note that the RMSE values for $\rho'=1$ are not shown in Table 4 due to symmetry.

3 Example of Time-Delay Estimation

As remarked in Section **Introduction**, it is often encountered in radar, sonar or communication that we need to estimate the correlation between a prescribed "clean" signal with a distorted version corrupted by impulsive noise. Now we provide an example of time-delay estimation which is similar to this situation. In this example, the prescribed clean signal is a segment of sinusoidal wave

$$y[i] = \begin{cases} \sqrt{2}\sin\left(\dfrac{10\pi \times i}{500}\right) & 0 \leq i < 500 \\ 0 & \text{otherwise} \end{cases},$$

whereas the corrupted signal is $x[i]=y[i-\tau_0]+w[i]$ with $w[i]$ being a white contaminated Gaussian noise following the distribution of

$$(1-\varepsilon)\mathcal{N}(0,\sigma^2)+\varepsilon\mathcal{N}(0,\sigma'^2) \qquad (52)$$

where $\varepsilon = 0.05$ and $\sigma' = 10^4 \gg \sigma$. The time-delay τ_0 is set to be 1000 ms. Our purpose is to estimate τ_0 as accurate as possible under various signal to noise ratio $\text{SNR} \triangleq 20\log_{10}(1/\sigma) = -20\log_{10}\sigma$. As illustrated in Figure 3, the procedure of estimating τ_0 includes two steps. The first one is to construct a correlation function that corresponds to τ by each of $r_{\mathcal{H}}$, r_K and $r_{\mathcal{S}}$ with respect to $x[i]$ and $y[i-\tau]$. The second one is to locate time-shift $\hat{\tau}_0$ corresponding the maximum of the correlation function. The value of $\hat{\tau}_0$ is considered to be an estimate of τ_0 and restored for further analysis. Note that the number of Monte Carlo trials in this study is set to be 10^4.

Table 5 summarizes the estimates of τ_0 from $r_{\mathcal{H}}$, r_K and $r_{\mathcal{S}}$. It is observed that all three methods produce acceptable estimates of τ_0, for an SNR even as low as -10 dB. However, PRVCC is slightly better than the other two, in the sense of giving smaller biases and standard deviations in most cases.

Conclusions

This paper systematically investigates the statistical properties of the historical PRVCC under a particular contaminated Gaussian model. As shown in our previous work [20], this model simulates

reasonably some frequently encountered scenarios where one variable is clean and the other corrupted by a tiny fraction of impulsive noise with very large variance. Under this model, we establish the asymptotic closed forms of the expectation and variance of PRVCC by means of the well known Delta method. To gain a further insight on PRVCC, we also compare it with two other classical correlation coefficients, i.e., SR and KT, in terms of the popular RMSE. Monte Carlo simulations not only verify our theoretical findings, but also reveal the strength and weakness of PRVCC in various occasions. The theoretical and empirical findings in this work are believed to add new knowledge to the area of correlation analysis that prevails in many branches of science and engineering.

Author Contributions

Conceived and designed the experiments: RM WX. Performed the experiments: RM WX. Analyzed the data: RM WX. Contributed reagents/materials/analysis tools: RM WX YZ ZY. Wrote the paper: RM WX.

References

1. Kendall M, Gibbons JD (1990) Rank Correlation Methods. New York: Oxford University Press, 5th edition.
2. Gibbons JD, Chakraborti S (1992) Nonparametric Statistical Inference. New York: M. Dekker, 3rd edition.
3. Jacovitti G, Cusani R (1992) Performance of normalized correlation estimators for complex processes. IEEE Transactions on Signal Processing 40: 114–128.
4. Delmas J, Abeida H (2009) Asymptotic distribution of circularity coefficients estimate of complex random variables. Signal Processing 89: 2670–2675.
5. Chorti A, Hristopulos DT (2008) Nonparametric identification of anisotropic (elliptic) correlations in spatially distributed data sets. IEEE Transactions on Signal Processing 56: 4738–4751.
6. Chaux C, Duval L, Benazza-Benyahia A, Pesquet JC (2008) A nonlinear stein-based estimator for multichannel image denoising. IEEE Transactions on Signal Processing 56: 3855–3870.
7. Tao R, Zhang F, Wang Y (2008) Fractional power spectrum. IEEE Transactions on Signal Processing 56: 4199–4206.
8. Xu W, Zhao C, Ding Z (2009) Limited feedback multiuser scheduling of spatially correlated broadcast channels. Vehicular Technology, IEEE Transactions on 58: 4406–4418.
9. Girault J, Kouamé D, Ouahabi A (2010) Analytical formulation of the fractal dimension of filtered stochastic signals. Signal Processing 90: 2690–2697.
10. Lian J, Garner G, Muessig D, Lang V (2010) A simple method to quantify the morphological similarity between signals. Signal Processing 90: 684–688.
11. Li J, Chen X, He Z (2013) Adaptive stochastic resonance method for impact signal detection based on sliding window. Mechanical Systems and Signal Processing 36: 240–255.
12. Xu W, Hou Y, Hung Y, Zou Y (2013) A comparative analysis of spearman's rho and kendall's tau in normal and contaminated normal models. Signal Processing 93: 261–276.
13. Axehill D, Gunnarsson F, Hansson A (2008) A low-complexity high-performance preprocessing algorithm for multiuser detection using gold sequences. IEEE Transactions on Signal Processing 56: 4377–4385.
14. Beko M, Xavier J, Barroso V (2008) Further results on the capacity and error probability analysis of noncoherent mimo systems in the low snr regime. IEEE Transactions on Signal Processing 56: 2915–2930.
15. Greco M, Gini F, Farina A (2008) Radar detection and classification of jamming signals belonging to a cone class. IEEE Transactions on Signal Processing 56: 1984–1993.
16. Liu L, Amin M (2008) Performance analysis of gps receivers in non-gaussian noise incorporating precorrelation filter and sampling rate. IEEE Transactions on Signal Processing 56: 990–1004.
17. Gini F (1998) A radar application of a modified Cramer-Rao bound: parameter estimation in non-Gaussian clutter. IEEE Transactions on Signal Processing 46: 1945–1953.
18. Gini F, Michels J (1999) Performance analysis of two covariance matrix estimators in compound-Gaussian clutter. IEE Proceedings Part-F 146: 133–140.
19. Mari DD, Kotz S (2001) Correlation and Dependence. London: Imperial College Press.
20. Ma R, Xu W, Wang Q, Chen W (2014) Robustness analysis of three classical correlation coefficients under contaminated gaussian model. Signal Processing 104: 51–58.
21. Pearson K (1914) On an extension of the method of correlation by grades or ranks. Biometrika 10: 416–418.
22. Schechtman E, Yitzhaki S (1987) A measure of association base on gini's mean difference. Commun Statist-Theor Meth 16: 207–231.
23. Xu W, Chang C, Hung YS, Kwan SK, Fung PCW (2006) Order statistic correlation coefficient and its application to association measurement of biosignals. ICASSP 2006 Proceedings 2: II-1068-1071.
24. Xu W, Chang C, Hung YS, Kwan SK, Fung PCW (2007) Order statistics correlation coefficient as a novel association measurement with applications to biosignal analysis. IEEE Transactions on Signal Processing 55: 5552–5563.
25. Xu W, Chang C, Hung Y, Fung P (2008) Asymptotic properties of order statistics correlation coefficient in the normal cases. IEEE Transactions on Signal Processing 56: 2239–2248.
26. Gao ZK, Zhang XW, Jin ND, Donner R, Marwan N, et al. (Sept. 2013) Recurrence networks from multivariate signals for uncovering dynamic transitions of horizontal oil-water stratified flows. Europhysics Letters 103: 50004 (6 pp.).
27. Gao ZK, Zhang XW, Jin ND, Marwan N, Kurths J (2013) Multivariate recurrence network analysis for characterizing horizontal oil-water two-phase flow. Phys Rev E 88: 032910.
28. Su Z, Li L, Peng H, Kurths J, Xiao J, et al. (2014) Robustness of interrelated traffic networks to cascading failures. Scientific Reports 4: 5413.
29. Li L, Peng H, Kurths J, Yang Y, Schellnhuber HJ (2014) Chaos–order transition in foraging behavior of ants. Proceedings of the National Academy of Sciences 111: 8392–8397.
30. Fisher RA (1921) On the 'probable error' of a coefficient of correlation deduced from a small sample. Metron 1: 3–32.
31. Stein D (1995) Detection of random signals in gaussian mixture noise. IEEE Trans Inf Theory 41: 1788–1801.
32. Reznic Z, Zamir R, Feder M (2002) Joint source-channel coding of a gaussian mixture source over the gaussian broadcast channel. IEEE Trans Inf Theory 48: 776–781.
33. Chen R, Wang X, Liu J (2000) Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering. IEEE Trans Inf Theory 46: 2079–2094.
34. Serfling RJ (2002) Approximation Theorems of Mathematical Statistics. Wiley series in probability and mathematical statistics. New York: Wiley.
35. David H, Nagaraja H (2003) Order Statistics. Hoboken: Wiley-Interscience, 3rd edition.
36. Balakrishnan N, Rao CR (1998) Order Statistics: Applications. Handbook of statistics; v. 17. New York: Elsevier.

37. Balakrishnan N, Rao CR (1998) Order Statistics: Theory & Methods. Handbook of statistics; v. 16. New York: Elsevier.

38. Stuart A, Ord JK (1994) Kendall's Advanced Theory of Statistics: Volume 1 Distribution Theory. London: Edward Arnold, 6th edition.

39. Xu W, Hung YS, Niranjan M, Shen M (2010) Asymptotic mean and variance of Gini correlation for bivariate normal samples. IEEE Trans Signal Process 58: 522–534.

40. Moran PAP (1948) Rank correlation and product-moment correlation. Biometrika 35: 203–206.