International Statistical Review

International Statistical Review (2025) doi: 10.1111/insr.70012

Comparing Robust Versions of Distance Covariance: A Comment on the Biloop Approach

Dominic Edelmann[1,2]

[1]*Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany*
[2]*Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA*
**E-mail: dominic.edelmann@dkfz-heidelberg.de*

1 Introduction

I commend the authors for their significant contributions to the field of distance correlation presented in this paper. This work marks the first thorough investigation of the robustness properties of distance correlation, distance covariance, and distance standard deviation in terms of influence functions and breakdown points. These robustness aspects have previously caused considerable confusion in the literature, and this work will serve as an important and clarifying reference. Additionally, the authors introduce a novel version of distance covariance based on an innovative biloop transformation. Thanks to its redescending influence function, this version of distance covariance is robust to bivariate outliers and shows promise for real-world applications.

However, the paper does not fully capture the extensive literature on robust distance covariance measures. Criticism regarding the lack of robustness in classical distance covariance can be traced back to the 2009 discussion paper of Székely and Rizzo, in which Bruno Rémillard (2009) identifies the moment assumption as a weakness of distance covariance and suggests a rank transformation as a remedy. In the same discussion, Gretton *et al.* (2009) give a brief historic overview and present several robust distance covariance type statistics (Kankainen, 1995; Feuerverger, 1993).

In this comment, I offer a concise, albeit non-comprehensive, overview of robust versions of distance covariance. Moreover, I extend the simulation studies presented in the main paper, giving further insights into the properties of the newly proposed biloop distance covariance.

2 Alternative Measures of Robust Distance Covariance

The literature offers at least three different approaches to defining extensions of distance covariance. Each approach gives rise to dependence measures with positive breakdown points.

2.1 Alternative Distance Measures

For jointly distributed vector valued random variables X, Y with finite second moment,[1] the standard distance covariance admits the representation (Székely *et al.*, 2007; Székely & Rizzo, 2009),

$$\begin{aligned} \mathrm{dCov}(X,\ Y) = \mathbb{E}[\|X - X'\|\|Y - Y'\|] + \mathbb{E}[\|X - X'\|]\mathbb{E}[\|Y - Y'\|] \\ - 2\mathbb{E}[(\|X - X'\|)(\|Y - Y''\|)], \end{aligned} \tag{1}$$

where (X', Y') and $(X''Y'')$ are independent copies of $(X,\ Y)$. It appears natural to replace the Euclidean distance by some kind of distance measure $d(\cdot\ ,\cdot)$, yielding

$$\begin{aligned} \mathrm{dCov}_d(X,\ Y) = \mathbb{E}[d(X,\ X')d(Y,\ Y')] + \mathbb{E}[d(X,\ X')]\,\mathbb{E}[d(Y,\ Y')] \\ - 2\mathbb{E}[(d(X,\ X'))(d(Y,\ Y''))]. \end{aligned} \tag{2}$$

Extensions of this type have first been systematically studied by Russel Lyons (2013). Sejdinovic *et al.* (2013) gave a slightly more general definition and established the exact equivalence of this type of generalised distance covariance with the Hilbert–Schmidt-Independence Criterion (Gretton *et al.*, 2005). In the machine learning community, a particularly popular choice leading to a generalised distance covariance with positive breakdown point is the 'radial basis function (RBF) distance',

$$d(x_1,x_2) = 1 - \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right), \tag{3}$$

where $\sigma > 0$ is a bandwidth parameter.

Many more distance functions have been proposed in the literature; notably, this approach enables the construction of dependence measure in fairly general metric spaces and hence for various applications. The important restriction is that the used distances must be of negative type; if this condition is not satisfied, $dcov_d$ loses non-negativity and other desirable properties, and alternative measures must be sought (Móri & Székely, 2020).

2.2 Alternative Weight Functions in the Characteristic Function Representation

In the original distance covariance paper (Székely *et al.*, 2007), the authors start with studying dependence measures of the form

$$\mathrm{dCov}(X,\ Y,\ w) = \int_{\mathbb{R}^{p+q}} |f_{X,\,Y}(t,\ s) - f_X(t)f_Y(s)|\, w(t,\ s)\,\mathrm{d}t\,\mathrm{d}s, \tag{4}$$

where

$$f_{X,\,Y}$$

They then proceed with the choice

$$w(t,\ s) = \left(constant \times \|t\|^{1+p}\|\|s\|^{1+q}\right)^{-1},$$

which yields the usual distance covariance. Böttcher *et al.* (2018) study weight functions

$$w(t,\ s) = \mu(t)\,\nu(s),$$

where μ and μ are symmetric Lévy measures, and show that the resulting dependence measure can be written equivalently in the alternative form (2). In fact, for $(X,\ Y) \in \mathbb{R}^2$, a robust measure of the form (4) predates distance covariance.[2] Specifically, Kankainen (1995) proposed the weight function

$$\frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(t^2+s^2\right)\right),$$

which can be shown to coincide with choosing the RBF distance (Equation 3) with bandwidth $\sigma = 1$ in representation Equation (2).

2.3 Data Transformations

As discussed in the main paper, another way to derive robust versions of distance covariance is to apply some kind of data transformation. Calculating the distance covariance of the normalised ranks of each margin was (to my knowledge) first proposed by Bruno Rémillard (2009) and has since then been discussed numerous times (e.g. Chakraborty & Zhang, 2019; Székely & Rizzo, 2023). A 'normal score version' of a distance covariance type measure in the bivariate case dates back to Feuerverger's 1993 proposal (Feuerverger, 1993); the general case is treated in detail by Mai *et al.* (2023). The novel biloop distance covariance also applies a data transformation before calculating distance covariance, but differs fundamentally from the rank- and normal-score approaches; the corresponding transformation maps coordinates from $\mathbb{R}$ to $\mathbb{R}^2$. It is not hard to see that the biloop distance covariance may also be expressed using the representation in Equation (2).

Of course, these three approaches are by no means exhaustive. One could equally well compute distance covariance on truncated data or consider versions based on weighted U-statistics, similar to Edelmann *et al.* (2022).

3 Simulation Results

The simulations presented below extend those in the main paper to shed further light on the properties of biloop distance covariance. Since the results of the normal score distance covariance were very similar to the rank distance covariance, the former method has been omitted. Instead the distance covariance dCov_d in Equation (2) employing the RBF distance (Equation 2) is included; this method is denoted as *RBF distance covariance* in the following. The hyperparameter c for the biloop distance covariance was selected as in the main paper, while the bandwidth for the RBF distance covariance was set using the median heuristic (Fukumizu *et al.*, 2009). A wider range of sample sizes ($n = 50, 100, 200, 400, 800, 1600, 3200$) than in the main paper was examined in each scenario. The bivariate data $(X_1, Y_1)^t, (X_2, Y_2)^t, \ldots, (X_n, Y_n)^t$ are always independently generated. I focus on the problem of independence testing. The tests are based on $K = 500$ permutations, empirical rejection rates are determined from $N_{sim} = 1000$ simulations and a nominal level of $\alpha = 0.05$ is used. All simulations were performed using the R package dcortools available on the Comprehensive R Archive Network (CRAN).

3.1 Scenarios With Bivariate Outliers

Clean data were drawn from an independent standard bivariate normal distribution. As in section 5.3 of the main paper, $n - m$ observations followed the clean data distribution, while m observations were generated via an outlier distribution. However instead of placing all outliers at the same point $(x, x)^t$, the outlier distribution is here a four component normal mixture (equal weights, each component having unit variance) with means $(6, 6)^t$, $(6, -6)^t$, $(-6, 6,)^t$ and $(-6, -6)^t$. The first three scenarios featured $m = 1$, $m = 3$ and $m = 0.06n$ outliers generated

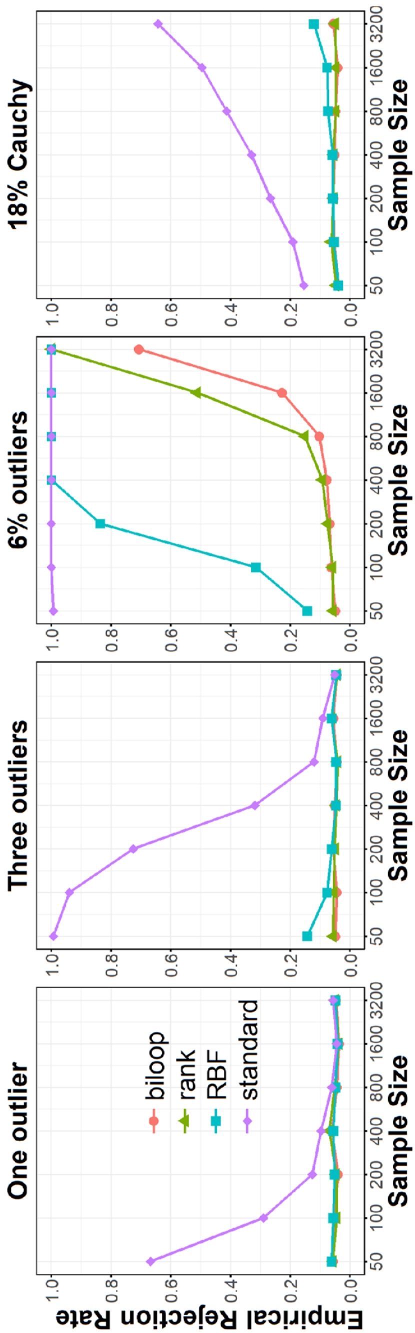

Figure 1. *Empirical rejection rates at level* $\alpha = 0.05$ *for scenarios with bivariate outliers. The results are based on* $N_{sim} = 1\ 000$ *simulations.*

by this normal mixture distribution. For the last scenario, both coordinates of $m = 0.18n$ observations are drawn independently from standard Cauchy distributions.

Across all scenarios (see Figure 1), biloop distance covariance yielded the lowest rejection rates, followed by rank and RBF distance covariance. Notably, in the 6% outliers scenarios, the rejection rates of the biloop distance covariance are substantially larger than the nominal level for higher sample sizes. This shows that the robustness property of biloop distance covariance demonstrated in figure 16 of the main paper only holds for small sample sizes; this is in fact not surprising since biloop distance covariance is consistent against all alternatives and bivariate outliers induce a dependence between X and Y. The biloop transformation maps both the X and Y coordinates of the outliers to values very close to the respective median and this nonlinear dependence is detected by distance covariance. Furthermore, differences in rejection rates between the biloop and the rank distance covariance are substantially smaller than in the main paper; likely because the point-mass outliers of the original setup induced a strictly monotone dependence, whereas the mixture here does not.

3.2 Scenarios With Univariate Outliers

For creating univariate outliers, two sets of indices $I^X = \left\{i_1^X, \ldots i_m^X\right\}$ and $I^Y = \left\{i_1^Y, \ldots i_m^Y\right\}$ were sampled without replacement from $\{1, \ldots, n\}$. The clean data was generated by a standard bivariate normal with correlation 0.2. The X coordinates with indices in I^X and the Y coordinates i-th indices in I^Y were replaced by outliers. The first plot in Figure 2 demonstrates the power of each method for clean data. In the second and third scenario, the outliers ($m = 3$ and $m = 0.06n$) were drawn from a two-component normal mixture distribution with standard deviation 1 and means 6 and -6. In the last scenario, the outliers ($m = 0.18n$) were drawn from a standard Cauchy distribution.

Overall, the biloop and the rank distance covariance show the best performance with only slight disadvantages compared to standard distance covariance in the uncontaminated setting and clear power advantages under heavy contamination. The rank distance covariance performed marginally better when few outliers were present, while the biloop distance covariance is better under heavy contamination. The performance of the RBF distance covariance remains largely unaffected by outliers. However, as previously observed and demonstrated in the clean data setting, this method has generally lower power in linear-dependence settings. The standard distance covariance suffered most from the contamination by outliers; however, the differences in power are rather moderate. This may come as a surprise considering the stark differences observed in Figure 1.

3.3 Scenarios in Which Biloop Distance Covariance Has Low Rejection Rate

In the first two sets of simulations, the rank distance covariance and the biloop distance covariance showed a similar performance over all settings. For the last set of simulations, I have purposely selected scenarios to highlight cases where biloop distance covariance has markedly lower rejection rates than competing methods (Figure 3). In the first two settings, a fraction of the sample (6% and 18%, respectively) was drawn from a bivariate normal distribution with mean $(6, 6)^T$ and unit variances, while the remainder followed a standard bivariate normal. In the third scenario, 82% of the observations were generated via an independent bivariate standard normal distribution and 18% were drawn from a bivariate standard normal distribution with standard deviations $\sqrt{20}$, mean 0 and correlation 0.5. In the last scenario, observations were drawn from a two-component normal mixture distribution with weights 0.82 and 0.18. The first

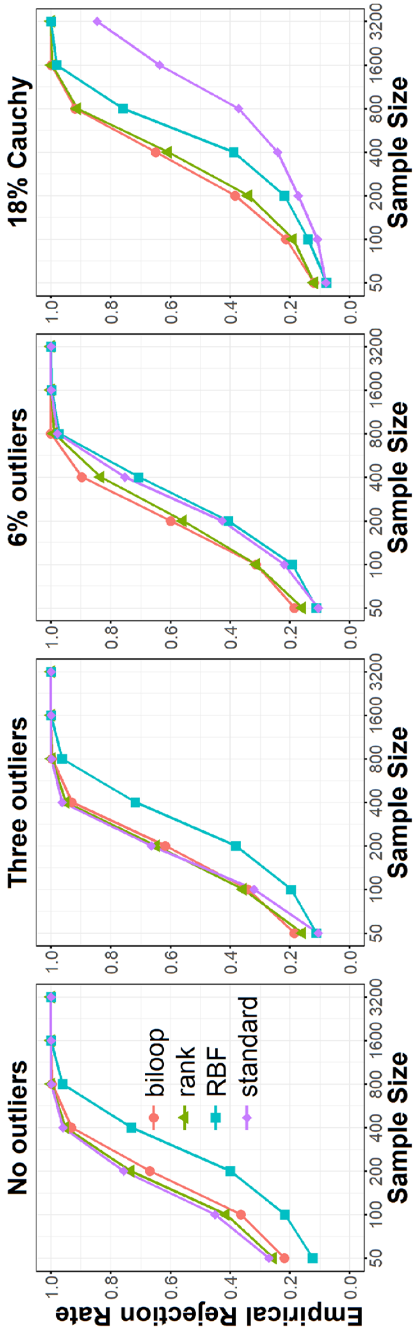

Figure 2. *Empirical rejection rates at level $\alpha = 0.05$ for scenarios with univariate outliers. The results are based on $N_{sim} = 1000$ simulations.*

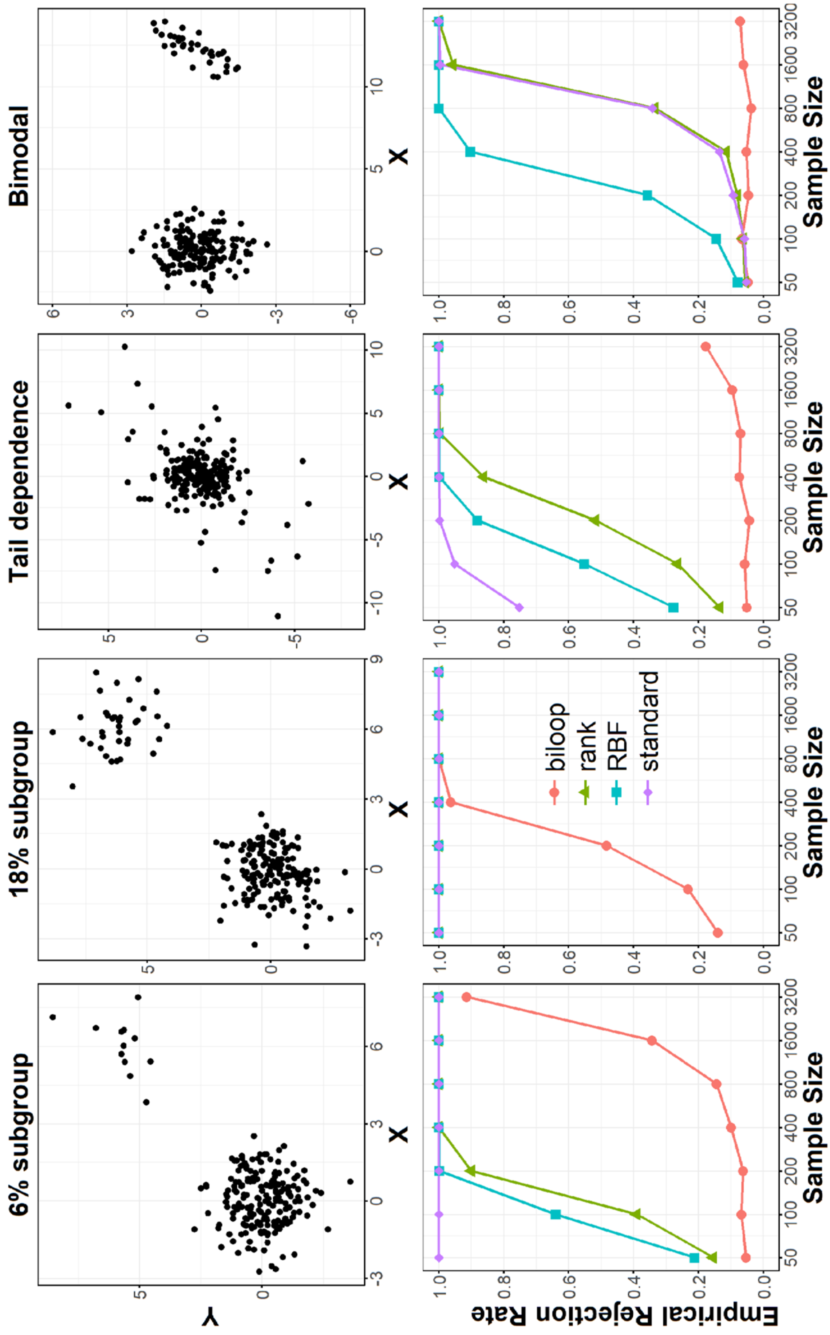

Figure 3. *Scenarios in which biloop distance covariance has low rejection rate. The upper plots illustrates the scenarios for sample size $n = 200$. The lower plot provide the empirical rejection rates at level $\alpha = 0.05$ based on $N_{sim} = 1\ 000$ simulations.*

component is an independent bivariate standard normal distribution, whereas the second component has mean $(12, 0)^t$, standard deviations 1 and correlation 0.8.

The first scenario is very similar to the one displayed in the first plot of figure 17 in the main paper. The induced monotone dependence leads to substantially higher rejection rates for the competing methods than in comparable scenarios in Figure 1, whereas the biloop distance covariance is less impacted. The second plot shows that the biloop distance covariance also has lower power in settings where the second subset contains a substantial fraction of the observations, however the rejection rate is not near the nominal level of $\alpha = 0.05$ for any of the considered sample sizes. The last two scenarios demonstrate that the biloop distance covariance has low rejection rate when dependence is confined to coordinates that are far from the median. In the last scenario, the power of the biloop distance covariance is particularly low—the X coordinates of all observations in the smaller subgroup are transformed to values very close to the median, which makes it very hard detect to the dependence in this subgroup. The RBF distance covariance has substantially larger rejection rate than all other methods in the last scenario demonstrating that this method is more sensitive to local dependence patterns.

4 Concluding Remarks

From my perspective, distance covariance methods hold great promise for the analysis of high-dimensional data. When testing a large number of bivariate dependencies, distance covariance provides a computationally efficient omnibus independence test—crucial in settings where visual inspection is impossible and where classical correlation measures may miss nonlinear relationships. Likewise, because outliers can lurk undetected in high dimensions, it is often advisable to employ a robust variant that automatically down-weights or excludes extreme observations.

Several robust variants of distance covariance have been proposed; among them, the biloop distance correlation stands out because its redescending influence function substantially diminishes the effect of any observation whose marginal distance is far from the median. As demonstrated in this comment and the main paper, this property makes the biloop method robust to outliers, even when these outliers induce a monotone dependence. In scenarios with only univariate outliers or bivariate outliers that do not induce monotone dependence, its performance closely matches that of rank-based distance covariance. A possible caveat is that, by construction, the biloop distance covariance is virtually blind to certain very strong dependence patterns (see Figure 3). Thus, while the method is promising, we must carefully weigh its robustness benefits against potential losses in power for specific dependence structures. Targeted simulations and theoretical analyses are needed to precisely quantify this trade-off.

ACKNOWLEDGEMENTS

Open Access funding enabled and organized by Projekt DEAL.

Endnotes

[1]The definition extends readily to distributions with only finite first moments; see Edelmann and Goeman (2022).

[2]However, it should be noted in this context that the principle of energy distance underlying distance covariance has been introduced by Székely between 1984 and 1985 (Rizzo & Székely, 2016).

References

Böttcher, B., Keller-Ressel, M., & Schilling, R. L. (2018). Detecting independence of random vectors: Generalized distance covariance and gaussian covariance. *Modern Stochastics: Theory and Applications*, **5**(3), 353–383

Chakraborty, S. & Zhang, X. (2019). Distance metrics for measuring joint dependence with application to causal inference. *Journal of the American Statistical Association*.

Edelmann, D. & Goeman, J. (2022). A regression perspective on generalized distance covariance and the Hilbert–Schmidt independence criterion. *Statistical Science*, **37**(4), 562–579.

Edelmann, D., Welchowski, T., & Benner, A. (2022). A consistent version of distance covariance for right-censored survival data and its application in hypothesis testing. *Biometrics*, **78**(3), 867–879.

Feuerverger, A. (1993). A consistent test for bivariate dependence. *International Statistical Review/Revue Internationale De Statistique*, 419–433.

Fukumizu, K., Gretton, A., Lanckriet, G., Schölkopf, B., & Sriperumbudur, B. K. (2009). Kernel choice and classifiability for RKHS embeddings of probability distributions. *Advances in Neural Information Processing Systems*, **22**.

Gretton, A., Bousquet, O., Smola, A., & Schölkopf, B. (2005). Measuring statistical dependence with hilbert-schmidt norms. In: Jain, S., Simon, H.U., & Tomita, E. (Eds.) *Algorithmic Learning Theory. ALT 2005. Lecture Notes in Computer Science*, vol. **3734**. Springer, pp. 63–77.

Gretton, A., Fukumizu, K., & Sriperumbudur, B. K. (2009). Discussion of: Brownian distance covariance. *The Annals of Applied Statistics*, **3**(4), 1285–1294.

Kankainen, A. (1995). Consistent testing of total independence based on the empirical characteristic function. PhD thesis, University of Jyväskylä.

Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab.*, **41**(5), 3284–3305.

Mai, Q., He, D., & Zou, H. (2023). Coordinatewise gaussianization: Theories and applications. *Journal of the American Statistical Association*, **118**(544), 2329–2343.

Móri, T. F. & Székely, G. J. (2020). The earth mover's correlation. *Annales Universitatis Scientiarum Budapestinensi*, **50**, 249–269.

R'emillard, B. (2009). Discussion of: Brownian distance covariance. *Annals of Applied Statistics*, **3**(4), 1295–1298.

Rizzo, M. L. & Székely, G. J. (2016). Energy distance. *Wiley Interdisciplinary Reviews: Computational Statistics*, **8**(1), 27–38.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *Annals of Statistics*, **41**(5), 2263–2291.

Székely, G. J. & Rizzo, M. L. (2009). Brownian distance covariance. *Annals of Applied Statistics*, **3**(4), 1236–1265.

Székely, G. J. & Rizzo, M. L. (2023). *The Energy of Data and Distance Correlation*. Chapman and Hall/CRC.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, **35**(6), 2769–2794.