Journal of Multivariate Analysis 198 (2023) 105229

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A nonparametric test for paired data

Grzegorz Wyłupek

Institute of Mathematics, University of Wrocław, pl. Grunwaldzki 2, 50-384 Wrocław, Poland

A R T I C L E I N F O

Article history:
Received 19 November 2021
Received in revised form 27 July 2023
Accepted 28 July 2023
Available online 2 August 2023

AMS 2020 subject classifications:
primary 62H15
secondary 62G10

Keywords:
Bootstrap
Copula
Dependence
Empirical process
Matched pairs
Two-sample problem

A B S T R A C T

The paper proposes a weighted Kolmogorov–Smirnov type test for the two-sample problem for paired data. The asymptotic distribution of the test statistic under the null model is derived. The dependence of both the finite sample and the asymptotic distribution of the test statistic on the dependence structure of the data requires the use of the wild bootstrap technique for inference. The related wild bootstrap test turns out to be a consistent asymptotically level α test. With the finite sample correction applied, the test keeps the level well. An extensive simulation study demonstrates good finite sample behaviour of the test in comparison to the existing procedures. The main role in the proofs is played by tools from empirical processes. Additional simulation results, as well as an R code, which allow one to easily repeat the conducted numerical experiments, are attached as Supplementary Material.

1. Introduction

We consider a classical and particularly important problem in a wide variety of scientific inquires, i.e., two-sample paired data comparisons. Specifically, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be the independent identically distributed random vectors coming from the population with the cumulative distribution function $H(x, y)$ with the continuous margins $F(x)$ and $G(y)$. The random variables X_i and Y_i, $i \in \{1, \ldots, n\}$ are frequently interpreted as measurements of patients' responses, before and after treatment. Since the data is paired, X_1 and Y_1 are dependent, while the dependence structure can be expressed by the uniquely determined copula C, i.e., the two-dimensional cumulative distribution function with the uniform margins such that $H(x, y) = C(F(x), G(y))$.

The fully nonparametric testing problem is formulated as follows:

$$\mathcal{H} : F(x) = G(x), \quad \text{for all } x \in \mathbb{R},$$

$$\mathcal{A} : F(x) \neq G(x), \quad \text{for some } x \in \mathbb{R}, \tag{1}$$

in the presence of infinite-dimensional nuisance parameter C. The null hypothesis asserts that treatment has no effect on patients, while the alternative claims it has an effect on them.

Even though with a strong parametric flavour, the one-sample t-test statistic [18] based on the differences $Z_i = Y_i - X_i$, $i \in \{1, \ldots, n\}$ seems to be the most popular solution to the problem (1). However, it works well under two-dimensional normality assumption on H, when a difference in shift occurs. On the other hand, when the disturbances in scales appear, the Morgan–Pitman test [9,15] is optimal under such a parametric model. Relaxation of the gaussianity assumption often

E-mail address: wylupek@math.uni.wroc.pl.

https://doi.org/10.1016/j.jmva.2023.105229

causes an application of the Wilcoxon signed-rank test or the sign test in statistical practice. To explain this phenomenon it is important to note that under exchangeability of (X, Y), the problem (1) can be reduced to verifying (symmetry)

$$\mathcal{H}_0 : D(z) + D(-z) = 1, \quad \text{for all } z \in \mathbb{R},$$

$$\mathcal{H}_1 : D(z) + D(-z) \neq 1, \quad \text{for some } z \in \mathbb{R}, \tag{2}$$

where D is the cumulative distribution function of the differences Z_is. For a direct calculation, see, for instance, Section 1.8.2 of [14]. Such a formulation justifies the common use of the Wilcoxon signed-rank test or the sign test for inference. Unfortunately, those two solutions are sensitive to the only one specific direction of disturbances from $\mathcal{H}$, i.e., to the so-called "shift in location" alternatives. Specifically, their optimality can be shown under the underlying logistic or Laplace distribution of D, respectively. Cf. Section 4.5, [6].

A general nonparametric approach to testing $\mathcal{H}$ against $\mathcal{A}$, while still verifying $\mathcal{H}_0$ against $\mathcal{H}_1$, can be found in [24]. Also, see [20,25] for the related contributions. Importantly, without exchangeability of the data at hand, such a test can be highly biased. Cf. Table 1, Section 4.2, as well as Tables S19 and S20 in Supplementary Material. Consequently, the empirical likelihood ratio solution proposed there should be used with caution.

To overcome the lack of distribution-freeness of the classical t-test under $\mathcal{H}$, Konietschke and Pauly [7] investigate its bootstrapped and permuted variants. They test

$$H_0 : \mu_X = \mu_Y \quad \text{against} \quad H_1 : \mu_X \neq \mu_Y,$$

where $\mu_X = EX_1$ and $\mu_Y = EY_1$.

The fully nonparametric solutions to the problem (1) include the Brunner et al. [2] Wald-type and ANOVA-type tests, the Martínez-Camblor [8] test based on the kernel density estimator, the data-driven test of Ghattas et al. [5], the Cramér–von Mises and characteristic function tests of Quessy and Éthier [16], as well as the Bayesian based procedure of Pereira et al. [13].

This paper seeks to meet the following objectives:

(i) to elaborate a new nonparametric omnibus test for verifying $\mathcal{H}$ against $\mathcal{A}$,

(ii) to investigate its asymptotics by deriving the limiting null distribution of the test statistic, as well as proving the consistency of the related test under the general alternatives,

(iii) to examine the finite sample behaviour of the solutions to the problem (1).

The paper is organized as follows. Section 2 presents the new test statistic and studies its asymptotics under the null model. Its wild bootstrap counterpart together with its limiting behaviour both under the null and alternative hypothesis are introduced in Section 3. Section 4 demonstrates the outcome of an extensive simulation study. The conclusions constitute Section 5. Appendix A contains a justification of the basic convergence results, while Appendix B provides the proofs for the main outcomes. The R code, which allow one to easily repeat the conducted numerical experiment and its description, as well as the results of an additional simulation study are attached as Supplementary Material.

2. A new test statistic

In this section, we introduce four new statistics that allow one to assert in the problem (1). The first one (formula (6)) is a sup-type functional of the two-sample paired data empirical process. The remaining three (formulas (7), (8), and (9)) are the sup-type functionals of the weighted two-sample paired data empirical process considered in the literature for the first time. The test based on the last defined statistic (formula (9)) seeks to verify $\mathcal{H}$ against $\mathcal{A}$.

First, we introduce an auxiliary notation. Let $\hat{F}(x) = (1/n)\sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$ and $\hat{G}(y) = (1/n)\sum_{i=1}^{n} \mathbb{1}(Y_i \leq y)$, where $\mathbb{1}$ stands for the indicator of the set $\cdot$. Let B be a two-dimensional Brownian bridge with the covariance structure

$$\mathrm{cov}(B(u_1, v_1), B(u_2, v_2)) = C(u_1 \wedge u_2, v_1 \wedge v_2) - C(u_1, v_1)C(u_2, v_2), \tag{3}$$

$u_1, u_2, v_1, v_2 \in [0, 1]$, where $u \wedge v = \min\{u, v\}$. Put $B_1(u) = B(u, 1)$ and $B_2(v) = B(1, v)$. Obviously, B_1 and B_2 are the one-dimensional Brownian bridges.

A construction of the test is based on the following weak convergence result

$$\sqrt{n}\left(\hat{F} - F, \hat{G} - G\right) \xrightarrow{\mathcal{D}} (B_1 \circ F, B_2 \circ G), \tag{4}$$

in $\ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R})$, where the limiting process is a two-dimensional (F, G)-Brownian bridge with the covariance structure (a general element of the 2×2 covariance matrix)

$$\begin{aligned}\mathrm{cov}(B(F(x_1), G(y_1)), B(F(x_2), G(y_2))) &= C(F(x_1) \wedge F(x_2), G(y_1) \wedge G(y_2)) - C(F(x_1), G(y_1))C(F(x_2), G(y_2))\\ &= H(x_1 \wedge x_2, y_1 \wedge y_2) - H(x_1, y_1)H(x_2, y_2),\end{aligned}$$

$x_1, x_2, y_1, y_2 \in \mathbb{R}$. Under the null hypothesis $\mathcal{H}$, (4) and the continuous mapping theorem imply the convergence

$$\mathbb{E}_n := \sqrt{n}(\hat{F} - \hat{G}) \xrightarrow{\mathcal{D}} B^0, \qquad B^0 \overset{\mathcal{D}}{=} B_1 \circ F - B_2 \circ G, \tag{5}$$

with the covariance structure (see Proposition 7 in Appendix A)

$$\text{cov}(B^0(x), B^0(y)) = F(x \wedge y) - C(F(x), G(y)) - C(F(y), G(x)) + G(x \wedge y).$$

The above allow us to define a new Kolmogorov–Smirnov type statistic

$$KS = \sup_{x\in\mathbb{R}} |\mathbb{E}_n(x)| = \sqrt{n} \sup_{x\in\mathbb{R}} |\hat{F}(x) - \hat{G}(x)|. \tag{6}$$

Proposition 1. *Under $\mathcal{H}$,*

$$KS \xrightarrow{\mathcal{D}} \sup_{x\in\mathbb{R}} |B^0(x)|.$$

Proof. The outcome is a consequence of the convergence result (5) and the continuous mapping Theorem 18.11(i), [21]. □

Unfortunately, both the finite sample and the asymptotic distribution of the KS statistic strongly depend on the copula C and the construction of the related unconditional test, rejecting the null hypothesis $\mathcal{H}$ in favour of the alternative $\mathcal{A}$ for its large values, is not possible as regards the control of the Type I error under the whole set of distributions from $\mathcal{H}$. This can be remedied by applying a conditional counterpart based on the wild bootstrap technique. However, first, we approach the problem in a slightly different manner. Let $x \in \mathbb{R}$ be such that $\mathbb{E}_n(x) \neq 0$. By (11), Appendix A, the asymptotic variance of the random variable $\mathbb{E}_n(x) = \sqrt{n}\{\hat{F}(x) - \hat{G}(x)\}$ is $F(x) + G(x) - 2H(x, x) - [F(x) - G(x)]^2$. Then, the Slutsky lemma implies under $\mathcal{H}$ the convergence

$$\frac{\sqrt{n}\{\hat{F}(x) - \hat{G}(x)\}}{\sqrt{\hat{F}(x) + \hat{G}(x) - 2\hat{H}(x, x) - [\hat{F}(x) - \hat{G}(x)]^2}} \xrightarrow{\mathcal{D}} N(0, 1),$$

where $\hat{H}(x, y) = (1/n)\sum_{i=1}^n \mathbb{1}(X_i \le x, Y_i \le y)$ provided that $F(x) + G(x) - 2H(x, x) \neq 0$. Therefore, the

$$wKS_x = \frac{\sqrt{n}\{\hat{F}(x) - \hat{G}(x)\}}{\sqrt{\hat{F}(x) + \hat{G}(x) - 2\hat{H}(x, x) - [\hat{F}(x) - \hat{G}(x)]^2}}$$

statistic could be a nice asymptotically distribution-free tool for testing $F = G$ in the point x. However, verifying $\mathcal{H}$ on the basis of such a statistic can be very inefficient. Therefore, in order to detect all possible deviations from the null model, we define a weighted Kolmogorov–Smirnov type statistic

$$wKS = \sup_{A\le x\le B} \left| \frac{\mathbb{E}_n(x)}{\sqrt{\mathbb{w}_n(x)}} \right| = \sup_{A\le x\le B} \left| \frac{\sqrt{n}\{\hat{F}(x) - \hat{G}(x)\}}{\sqrt{\hat{F}(x) + \hat{G}(x) - 2\hat{H}(x, x) - [\hat{F}(x) - \hat{G}(x)]^2}} \right|, \tag{7}$$

where $A = \min\{X_{1:n}, Y_{1:n}\}$, $B = \max\{X_{n:n}, Y_{n:n}\}$, while $X_{i:n}$ is the ith order statistic in the sample $X_1, \ldots, X_n$ and $Y_{i:n}$ is the ith order statistic in the sample $Y_1, \ldots, Y_n$, $i \in \{1, \ldots, n\}$. We reject $\mathcal{H}$ in favour of $\mathcal{A}$ for large values of wKS.

Let (a_1, b_1) and (a_2, b_2), $a_1, a_2, b_1, b_2 \in \overline{\mathbb{R}}$, be the supports of F and G, respectively. Put $a = \min\{a_1, a_2\}$ and $b = \max\{b_1, b_2\}$. Since A tends to a, in probability, and B tends to b, in probability, $\mathbb{w}_n(x)$ goes to 0, in probability, when n grows and x approaches a or b. As a result, the wKS statistic can take arbitrary large values under the null model. Since we do not know any existing outcome related to the asymptotic behaviour of the weighted two-sample paired data empirical process $\mathbb{V}_n(\cdot) := \mathbb{E}_n(\cdot)/\sqrt{\mathbb{w}_n(\cdot)}$, whereas deriving it is beyond the scope of this paper and can be a subject of a separate study, we propose to bound the weight $\mathbb{w}_n(\cdot)$ defining the supremum over a large compact subset of the support (a, b). For this purpose, let $I = (1/2)F + (1/2)G$, $\varepsilon = 0.0001$, $a_I = I^{-1}(\varepsilon)$, and $b_I = I^{-1}(1 - \varepsilon)$. We reject $\mathcal{H}$ in favour of $\mathcal{A}$ for large values of

$$WKS = \sup_{a_I\le x\le b_I} |\mathbb{V}_n(x)| = \sup_{a_I\le x\le b_I} \left| \frac{\sqrt{n}\{\hat{F}(x) - \hat{G}(x)\}}{\sqrt{\hat{F}(x) + \hat{G}(x) - 2\hat{H}(x, x) - [\hat{F}(x) - \hat{G}(x)]^2}} \right|. \tag{8}$$

Such a modification permits us to keep the weight function bounded, derive the asymptotic distribution of that statistic under $\mathcal{H}$, and prove the consistency of the related test under a broad spectrum of alternatives. Due to the choice of $\varepsilon = 0.0001$, for a sample size under $n = 5000$, WKS works as wKS, whereas above this size, it is blind for certain very subtle differences in tails. For more comments, see Section S1 in Supplementary Material.

Assumption 1. Let the copula C for the pair (X_1, Y_1) have maximal correlation strictly less than one.

Assumption 1 is necessary to assure positiveness of $F(x)+G(x)-2H(x,x)-[F(x)-G(x)]^2$ and its estimate for any $x \in [a_I, b_I]$. Such a restriction is related to the Fréchet–Hoeffding upper bound expressed by the copula $M(u,v)=\min\{u,v\}$, which corresponds to the condition $P(U=V)=1$, where $U=F(X_1)$ and $V=G(Y_1)$. Cf. Section 2.5 in [10]. Therefore, it holds under the null hypothesis, where the random variables are almost surely equal. Cf. the Type I errors for the Mardia copula in Section 4.2, Table 1, and in Supplementary Material, Sections S3 and S4, Tables S19 and S20, respectively.

Proposition 2. *Under $\mathcal{H}$ and Assumption 1,*

$$WKS \xrightarrow{\mathcal{D}} \sup_{a_I \le x \le b_I} \left| \frac{B^0(x)}{\sqrt{F(x)+G(x)-2H(x,x)}} \right|.$$

The proof of Proposition 2 is provided in Appendix B.

Under $\mathcal{H}$, both the finite sample and the asymptotic distribution of the WKS statistic still depend on the dependence structure expressed by the copula C. Therefore, the usage of the wild bootstrap for the inference is a necessity; see Section 3 below.

To combine the advantages of the first and the final defined statistics, we propose to reject $\mathcal{H}$ for large values of

$$M = \max\{KS, WKS\}, \tag{9}$$

where

$$KS = \sqrt{n} \sup_{x\in\mathbb{R}} |\hat{F}(x)-\hat{G}(x)|,$$

$$WKS = \sup_{a_I \le x \le b_I} \left| \frac{\sqrt{n}\{\hat{F}(x)-\hat{G}(x)\}}{\sqrt{\hat{F}(x)+\hat{G}(x)-2\hat{H}(x,x)-[\hat{F}(x)-\hat{G}(x)]^2}} \right|,$$

while $I=(1/2)F+(1/2)G$, $\varepsilon=0.0001$, $a_I=I^{-1}(\varepsilon)$, $b_I=I^{-1}(1-\varepsilon)$, and I^{-1} is the generalized inverse.

On the one hand, the use of the KS statistic guarantees consistency of the test based on M. On the other hand, the employment of the WKS statistic ensures high power of that test, see Section 4.3 and Section S4 in Supplementary Material. As a result, the M statistic combines the strengths of those two statistics.

Proposition 3. *Under $\mathcal{H}$ and Assumption 1,*

$$M \xrightarrow{\mathcal{D}} \max\left\{ \sup_{x\in\mathbb{R}} |B^0(x)|, \sup_{a_I \le x \le b_I} \left| \frac{B^0(x)}{\sqrt{F(x)+G(x)-2H(x,x)}} \right| \right\}. \tag{10}$$

Proof. Propositions 1 and 2 and the continuous mapping Theorem 18.11(i), [21], imply (10). □

3. The wild bootstrap technique

Both under finite sample size and asymptotically, four introduced test statistics (formulas (6), (7), (8), and (9)) are not distribution-free under $\mathcal{H}$ because their distributions depend on the copula C. Therefore, to infer on them, we need to introduce their conditional counterparts. A technique that allows one to do that is the wild bootstrap. It is worth mentioning that the idea comes from [26], where heteroscedastic linear regression was investigated and residuals were multiplied by the weights. Since usually, the results are different to the original residuals, the word "wild" has been stuck to the "bootstrap". Cf. Section 7, [26]. The wild bootstrap has been profoundly elaborated and investigated in many different contexts ever since. At present, it is a popular resampling technique inevitable in many challenging statistical problems.

Given the sample $(X_1,Y_1),\dots,(X_n,Y_n)$, we independently generate the i.i.d. random variables $\xi_1,\dots,\xi_n$ such that $E\xi_1=0$, $E\xi_1^2=1$, and $\int_0^{+\infty}\sqrt{P(|\xi_1|>x)}dx < +\infty$. Then, we define the wild bootstrap counterparts of the processes: $\mathbb{E}_n$, $\mathbb{w}_n$, and $\mathbb{V}_n$ as

$$\mathbb{E}_n^*(x) = (1/\sqrt{n})\sum_{i=1}^n \xi_i[\mathbb{1}(X_i\le x)-\mathbb{1}(Y_i\le x)],$$

$$\mathbb{w}_n^*(x) = (1/n)\sum_{i=1}^n \frac{|\xi_i|}{E|\xi_i|}[\mathbb{1}(X_i\le x)+\mathbb{1}(Y_i\le x)-2\cdot\mathbb{1}(X_i\le x, Y_i\le x)] - \left\{(1/n)\sum_{i=1}^n \frac{|\xi_i|}{E|\xi_i|}[\mathbb{1}(X_i\le x)-\mathbb{1}(Y_i\le x)]\right\}^2,$$

$$\mathbb{V}_n^*(x) = \frac{\mathbb{E}_n^*(x)}{\sqrt{\mathbb{w}_n^*(x)}},$$

and the related statistics:

$$KS^* = \sup_{x\in\mathbb{R}}|\mathbb{E}_n^*(x)|, \quad WKS^* = \sup_{a_I\le x\le b_I}\left|\mathbb{V}_n^*(x)\right|, \quad M^* = \max\{KS^*, WKS^*\}.$$

We do not introduce the wild bootstrap version of the wKS statistic because it lacks a theoretical justification. This is due to an explosion of its value when x tends to infinity.

Proposition 4. *Under $\mathcal{H}$, conditionally, given $(X_1, Y_1), \ldots, (X_n, Y_n)$,*

$$KS^* \xrightarrow{\mathcal{D}} \sup_{x\in\mathbb{R}} |B^0(x)|.$$

Proposition 5. *Under $\mathcal{H}$ and Assumption 1, conditionally, given $(X_1, Y_1), \ldots, (X_n, Y_n)$,*

$$WKS^* \xrightarrow{\mathcal{D}} \sup_{a_I \le x \le b_I} \left| \frac{B^0(x)}{\sqrt{F(x)+G(x)-2H(x,x)}} \right|.$$

Proposition 6. *Under $\mathcal{H}$ and Assumption 1, conditionally, given $(X_1, Y_1), \ldots, (X_n, Y_n)$,*

$$M^* \xrightarrow{\mathcal{D}} \max\left\{ \sup_{x\in\mathbb{R}} |B^0(x)|, \sup_{a_I \le x \le b_I} \left| \frac{B^0(x)}{\sqrt{F(x)+G(x)-2H(x,x)}} \right| \right\}.$$

The proofs of Propositions 4–6 are postponed to Appendix B. Propositions 4–6 are the wild bootstrap counterparts of Propositions 1–3. They allow us to define asymptotically valid level α tests based on KS, WKS, and M. Specifically, let Q be one of the statistics: KS, WKS, M. Given the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, generating, independently, `nbr` multiplier samples, $\xi_1^{(1)}, \ldots, \xi_n^{(1)}; \ldots; \xi_1^{(\mathrm{nbr})}, \ldots, \xi_n^{(\mathrm{nbr})}$, we obtain `nbr` values of the Q^* statistic, say, $Q^{*(1)}, \ldots, Q^{*(\mathrm{nbr})}$. Let $q_Q^*(\alpha)$ be the $(1-\alpha)$-quantile of the distribution of $Q^{*(1)}, \ldots, Q^{*(\mathrm{nbr})}$. The wild bootstrap test is defined as $\Phi^*_{Q,n,\alpha} = \mathbb{1}(Q > q_Q^*(\alpha))$.

The asymptotic behaviour of the $\Phi^*_{Q,n,\alpha}$ test under $\mathcal{A}$ is summarized by Lemma 1.

Lemma 1. *Set $\alpha \in (0, 1)$.*

(i) *The $\Phi^*_{KS,n,\alpha}$ test is consistent under any alternative from $\mathcal{A}$.*

(ii) *Under Assumption 1, the $\Phi^*_{WKS,n,\alpha}$ test is consistent under any alternative from $\mathcal{A}$ such that $F(x) \neq G(x)$ for $x \in [a_I, b_I]$.*

(iii) *Under Assumption 1, the $\Phi^*_{M,n,\alpha}$ test is consistent under any alternative from $\mathcal{A}$.*

The above outcome shows the universal consistency of the conditional KS and M tests.

In the next section, we investigate the behaviour of these two tests under finite sample size and a large spectrum of scenarios in comparison to the other solutions of the testing problem. The results of an additional numerical experiment can be found in Supplementary Material. We can see how the w_n weight influences the power curve of the $\Phi^*_{M,n,\alpha}$ test. As a result, we recommend that solution for practical applications only.

4. Simulation study

Although the problem is classical and has a long history, it does not have many solutions. Therefore, we include into the numerical experiment all the significant contributions. These are presented in Section 4.1. Since dependence of X and Y can cause some troubles with the control of the Type I error, we investigate the tests in this regard in Section 4.2. Section 4.3 demonstrates the behaviour of the tests under a wide spectrum of alternatives. All computations have been carried out in `R` [17] under the seed 1 using the packages: `Matrix` [1], `nparLD` [11], `matrixcalc` [12], `MASS` [23], `copula` [27]. The significance level $\alpha = 0.05$. The sample size is $n = 50$. In the simulations, we estimate the values of the power functions of the tests using 1000 Monte Carlo runs. Dependence of the test on the asymptotic, the finite sample or the bootstrap approach for making a decision in any of the 1000 MC iterations is stressed in Table 1 below. (See, the column “Method” there.) Additional simulation results under $n = 20$ and $n = 200$ can be found in Supplementary Material, Sections S3 and S4, respectively. The accompanying files: Power_Functions.R, Generators_Alternative.R, and Generators_Null_Model.R contain the R codes, which allow one to repeat the conducted simulation study. Their description is provided in Supplementary Material, Section S5.

4.1. Competitive solutions

We include into the simulation study the classical Wilcoxon signed-rank W test statistic, the Wald-type statistic of Brunner et al. [2] denoted as BMP, the solution of Martínez-Camblor [8] abbreviated as M–C, the data-driven test of Ghattas et al. [5], $GPRY$ for short, the Cramér–von Mises test of Quessy and Éthier [16] denoted as QE, the empirical likelihood statistic of Vexler et al. [24], VGH for brevity, as well as the Konietschke and Pauly [7] wild bootstrap KP test, and two new tests based on the KS and M statistics with the multiplier $\xi_1, \ldots, \xi_n$ being the i.i.d. standard normal variables.

Table 1
Empirical Type I errors under different scenarios. $n = 50$, $\alpha = 0.05$, $\varepsilon = 0.0001$, 1000 MC runs, 1000 wild bootstrap runs. Errors multiplied by 100.

	$FGM(\theta)$			$Clayton(\theta)$			$Mardia(\theta)$			t_θ			
Test\\θ	−0.5	0	0.5	−0.5	0.5	1	0.5	0.7	0.9	1	3	5	Method
W	5.6	5.7	5.4	6.3	5.5	4.7	4.2	4.4	0.3	5.6	4.7	5.5	Asymptotic
BMP	5.3	6.0	5.3	5.5	6.3	6.2	4.9	5.3	5.2	5.2	5.8	5.1	Asymptotic
M-C	3.3	2.2	1.1	3.5	0.7	0.5	0.9	0.2	0.0	5.7	5.3	4.5	Smoothed bootstrap
GPRY	4.8	5.3	5.2	5.1	5.5	4.8	5.6	4.4	3.1	5.4	5.7	5.8	Asymptotic approx.
QE	7.3	7.0	6.4	7.4	5.8	4.9	4.3	5.0	4.0	1.3	4.5	6.1	Bayesian bootstrap
VGH	6.2	5.0	5.3	5.0	5.1	4.7	8.6	79.2	100	5.7	4.6	5.4	Finite sample
KP	4.4	5.2	5.2	4.8	5.4	6.2	4.7	5.0	5.9	5.4	5.7	5.7	Wild bootstrap
KS	6.2	5.6	5.3	6.1	5.0	4.9	4.4	4.1	3.5	5.3	5.4	4.7	Wild bootstrap
M	12.3	12.1	13.8	10.8	15.1	13.9	10.6	12.2	14.0	11.3	10.9	13.6	Wild bootstrap

Table 2
Empirical Type I errors of the $\Phi^*_{M,n,\alpha,c}$ test against c under different scenarios. $n = 50$, $\alpha = 0.05$, $\varepsilon = 0.0001$, 1000 MC runs, 1000 wild bootstrap runs. Errors multiplied by 100.

	$FGM(\theta)$			$Clayton(\theta)$			$Mardia(\theta)$			t_θ		
c\\θ	−0.5	0	0.5	−0.5	0.5	1	0.5	0.7	0.9	1	3	5
1.00	12.3	12.1	13.8	10.8	15.1	13.9	10.6	12.2	14.0	11.3	10.9	13.6
0.95	10.3	10.5	11.3	9.3	13.6	12.1	9.0	9.3	11.3	10.1	9.4	11.5
0.90	7.5	7.6	7.4	8.0	9.1	7.7	6.8	7.5	8.1	8.2	6.9	8.6
0.85	6.4	6.3	5.8	5.9	7.9	6.1	5.4	5.8	5.7	6.4	6.1	7.4
0.80	4.9	4.8	4.4	4.8	5.3	4.4	4.0	4.1	4.1	5.2	4.0	5.1
0.75	4.0	3.9	3.4	3.3	4.0	2.9	2.5	3.2	3.3	3.9	2.6	3.8

4.2. Type I error control

In Table 1, we present the estimated Type I errors of the tests under four different dependence models. $FGM(\theta)$ is the Farlie–Gumbel–Morgenstern copula with the parameter $\theta \in [-1, 1]$. See, Exercise 3.23, p. 87, [10]. $Clayton(\theta)$ is the Clayton copula with the parameter $\theta \in [-1, \infty)$. Cf. Exercise 4.16, p. 134, and Exercise 4.14, p. 138, [10]. $Mardia(\theta)$ stands for the Mardia copula with the parameter $\theta \in [-1, 1]$. Cf. [10] p. 15, formula (2.2.9). t_θ denotes the bivariate t-Student distribution with $\theta > 0$ degrees of freedom. See [4], p. 112, formula (2).

With the exception of slight deviations from the nominal level α of the Quessy and Éthier [16] solution using standard exponential multiplier weights scaled by the sample mean, the only problematic case for the competitive solutions is the Mardia copula, where the Type I error of the Vexler et al. [24] test tends to 1 when θ approaches 1. This is an example of the influence of the strong dependence between X and Y, raised in Assumption 1, which takes place in that case under $\theta = 1$ and touches the empirical likelihood *VGH* solution.

Remark 1. Though the Vexler et al. [24] test is recommended as a paired data problem solution, actually, it is the solution to the problem (2) of verifying the symmetry, which is linked to the problem (1) via the exchangeability assumption. Since that test is widely advertised to the practitioners [20,25], we investigate its behaviour below.

Although the wild bootstrap assures an asymptotic validity of the $\Phi^*_{M,n,\alpha}$ test, Table 1 clearly shows that under finite sample size, its Type I error exceeds the nominal level α. Therefore, to keep the level well, we need to widen the distribution of the WKS^* statistic multiplying w^*_n by a constant $c < 1$. For this purpose, let $\mathrm{w}^*_{nc} = c\,\mathrm{w}^*_n$, $WKS^*_c = WKS^*/\sqrt{c}$, $M^*_c = \max\{KS^*, WKS^*_c\}$, and $\Phi^*_{M,n,\alpha,c} = \mathbb{1}(M > q^*_{M_c}(\alpha))$. Table 2 demonstrates the Type I errors of the $\Phi^*_{M,n,\alpha,c}$ test against c.

In what follows, we apply the $\Phi^*_{M,n,\alpha,c}$ test with $c = 0.8$. A recommendation for the selection of the correction c for other sample sizes is given in Supplementary Material, Section S2.

4.3. Power comparison

In this section, we present the behaviour of the tests under the alternative. In Fig. 1, we depict their estimated power functions in nine cases covering various differences between marginal distributions of H, i.e., F and G, under interesting dependence structures.

Description of the alternatives
Smooth alternatives defining differences in shift, scale, and dispersion.

- Example 1: $(X, Y) \sim N_2(\mu, \Sigma)$, where $\mu = (0, \mu)$ and $\Sigma = \begin{pmatrix} 1 & -0.8 \\ -0.8 & 1 \end{pmatrix}$.

- Example 2: $(X, Y) \sim N_2(\mu, \Sigma)$, where $\mu = (0, 0)$ and $\Sigma = \begin{pmatrix} 1 & 0.95 \\ 0.95 & \sigma^2 \end{pmatrix}$.
- Example 3: $(X, Y) \sim 0.6N_2(\mu_1, \Sigma_1) + 0.4N_2(\mu_2, \Sigma_2)$,
  where $\mu_1 = (1, 1)$ and $\Sigma_1 = \begin{pmatrix} 1 & -0.8 \\ -0.8 & 1 \end{pmatrix}$, while $\mu_2 = (2, 2)$ and $\Sigma_2 = \begin{pmatrix} 4 & -3.6 \\ -3.6 & \sigma^2 \end{pmatrix}$.

Regression models exploiting square, cubic, and biquadratic function dependence.

- Example 4: $Y = \beta X^2 + \varepsilon$, where $X \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$, X and ε independent.
- Example 5: $Y = -\beta X^3 + \varepsilon$, where $X \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$, X and ε independent.
- Example 6: $Y = \beta X^4 + \varepsilon$, where $X \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$, X and ε independent.

Long-tailed alternatives expressing subtle differences between margins.

- Example 7: $X = \theta V + \sqrt{V}Z_1$, $Y = -\theta V + \sqrt{V}Z_2$,
  where $V \sim Ig(5/2, 5/2)$ and $(Z_1, Z_2) \sim N_2(\mu, \Sigma)$ with $\mu = (0, 0)$ and $\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$.
- Example 8: $X = Z_1 + 0.15V_2$, $Y = V_1 + 0.15V_2$,
  where $Z_1 \sim N(0, \sqrt{2})$, $V_1 \sim S(a, 0)$, $V_2 \sim S(a, 0)$, Z_1, V_1, and V_2 independent.
- Example 9: $X = V_1 + 0.15V_3$, $Y = V_2 + 0.15V_3$,
  where $V_1 \sim S(0.4, 0.1)$, $V_2 \sim S(a, a)$, $V_3 \sim S(a, a)$, V_1, V_2, and V_3 independent.

Examples 1–3 come from [13]. Cf. scenarios I, II, and IV, [13]. Examples 4–6 are patterned after the models considered in [19]. Example 7 defines the skew t_5 distribution introduced in Section 5.1, [4]. Examples 8–9 demonstrate two-dimensional Stable distributions. For a generator of the $S(a, b)$ law, see Section 2, [3].

In Example 1, where the pure difference in shift occurs, the *W*, *BMP*, *QE*, and *KP* tests are optimal. Therefore, their power is the greatest. The *M-C*, *KS*, and *M* tests, which also emphasize detection of other disturbances from the null model lose little. The loss of the *GPRY* test to the best one is dictated by its sensitivity to other directions, not only the first one. In Example 2, the scale is changed and the tests mostly dedicated to detection of the shift disturbances break down completely. There are the Wilcoxon, Brunner et al. [2], Vexler et al. [24], as well as Konietschke and Pauly [7] tests. The moderate power of the *KS* and *GPRY* solutions reflects their ability to detect changes in the second direction, that is, the changes in scale. It can be seen that the ability to detect such disturbances is the greatest in the case of the Martínez-Camblor [8] test, because the difference expressed in terms of the densities can be easily recognized by that solution. The *QE* and *M* tests do not lose too much because the second direction is still well detected by those procedures. Example 3 pertains to local changes in dispersion. Therefore, initially, the order of the tests is identical to that in Example 2; however, later, the possibility to detect such an alternative by the Martínez-Camblor [8] test is questionable. The sensitivity of the *GPRY* and *KS* tests in this direction is smaller than previously because such differences are more subtle than in the second case. The very low power of the *W*, *BMP*, *VGH*, and *KP* tests could be predicted because no difference in shift occurs there. In Examples 4 and 6, the square and biquadratic regression models generate differences in shift. As a result, the *KS*, *M-C*, *W*, and *BMP* tests work well, while the *QE*, *VGH*, *KP*, and *M* tests work very well. On the other hand, under Example 5, the cubic regression model generates no difference in shift. As a result, the Wilcoxon, Brunner et al. [2], Vexler et al. [24], as well as Konietschke and Pauly [7] tests break down completely. The behaviour of the *GPRY* and *KS* tests is pour. The *M* test works well because the weighted difference between the cdfs is large enough. The power of the *M-C* test is the highest because the integrated minimum of the densities is highly significant. Nevertheless, a kind of a saturation of that solution can be observed. In Example 7, where the difference in shift occurs and the model covers long tails, the Martínez-Camblor [8] solution breaks down because the disparity between the distributions expressed in terms of the densities is small. The power of the Ghattas et al. [5] test is weak. The *KS*, *M*, and *QE* tests work well, while the *BMP*, *VGH*, *W*, and *KP* solutions are the best. This is because the most sufficient disturbances come from the shift and the long tails also play a role. Under two-dimensional stable distribution, Example 8, when the parameter a is altered, there is no difference in shift. Therefore, the *KP*, *W*, *VGH*, and *BMP* solutions break down completely. The *QE*, *KS*, and *GPRY* tests are weak because such subtle disparities between the distributions are barely detected by the measures on which the related tests statistics are built. The *M-C* and *M* tests are the leading procedures. Once more, measuring the differences between the distributions in terms of values of those statistics is profitable. In Example 9, both parameters of the Stable distribution are changed and subtle differences occur in the middle of the pooled distribution given by $I = (1/2)F + (1/2)G$. Such a situation is well detected by the *GPRY*, *KS*, and *M* tests.

To summarize, the Wilcoxon signed-rank test, the Brunner et al. [2] test, the Vexler et al. [24] test, as well as the Konietschke and Pauly [7] test work very well when differences in shift occur and break down completely under more subtle changes in the distributions. The Ghattas et al. [5] solution usually works poorly. The more standard deviations from the null hypothesis, the better behaviour of the Quessy and Éthier [16] procedure. The Martínez-Camblor [8] solution usually works very well. However, it breaks down under certain circumstances losing much to the leading procedure. The *KS* statistic is not suggested for inference and its investigation only allows for better comparisons. It works very well only when the largest difference between F and G occurs in the middle of the pooled distribution. The new *M* test always has high power and is the only stable solution, i.e., it does not break down at all.

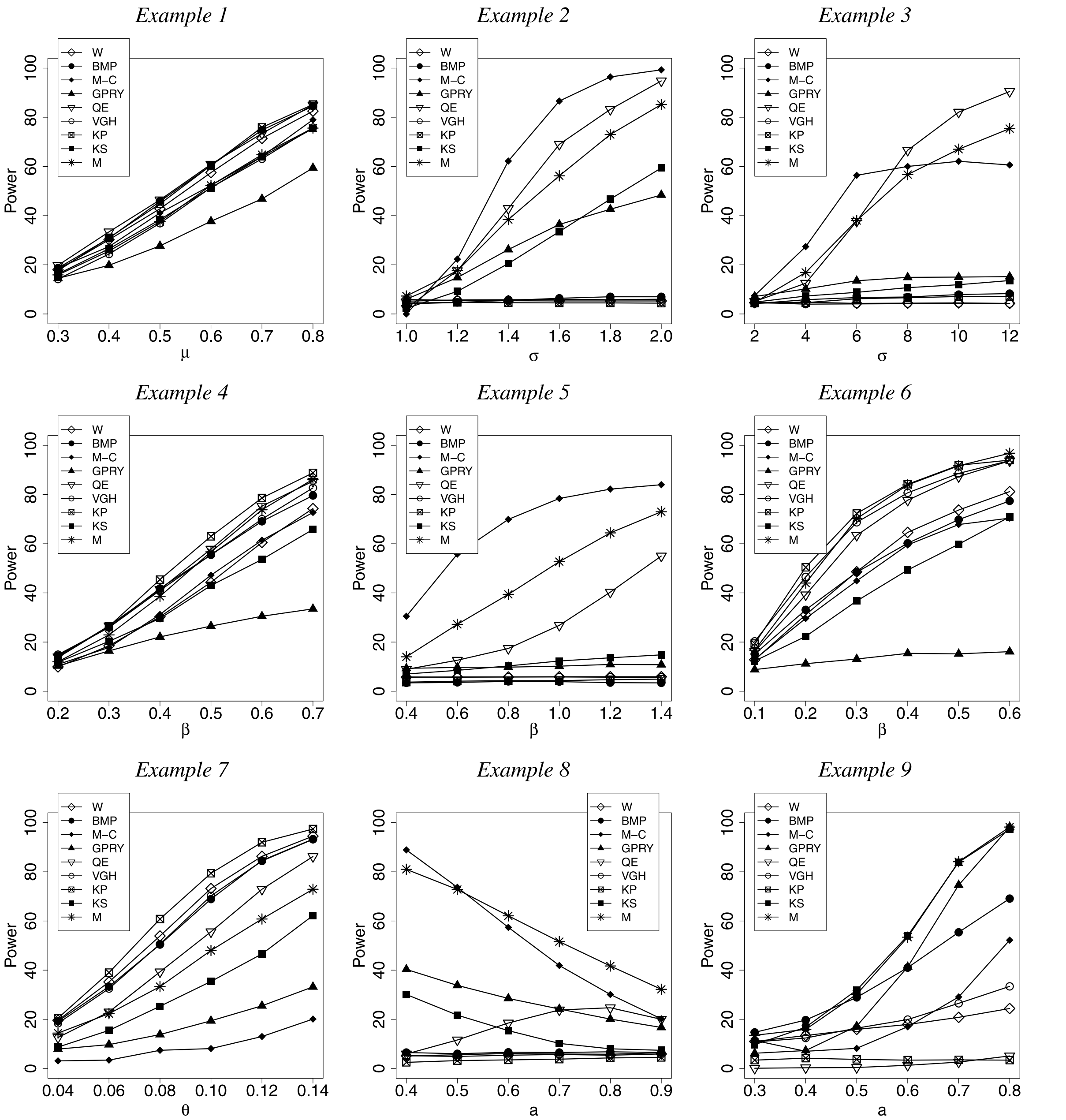

Fig. 1. Empirical powers against a parameter. $\alpha = 0.05$, $n = 50$, $\varepsilon = 0.0001$, $c = 0.8$. Based on 1000 MC runs and 1000 wild bootstrap runs. Values multiplied by 100.

Comment on the behaviour of the *BNP* solution under Examples 1–9.

Recently, Pereira et al. [13] have introduced a Bayesian nonparametric testing procedure for the problem (1). Since the paper is new and has been published in a recognized journal in the field, we wanted to include such a solution into the simulation study. It turns out that under smooth alternatives, Examples 1–6, that solution works very well and its power is comparable to that of the best test (Examples 1, 2, 4, 6) or is higher (Examples 3, 5). However, under long-tailed alternatives, Example 7, that solution breaks down, i.e., a similar behaviour to the *MC* solution can be observed. Furthermore, under Examples 8–9, that procedure is incomputable, that is, the posterior probability for $\mathcal{A}$ cannot be calculated, using the R package `BNPpairedtest` introduced by Pereira et al. [13] due to an error that occurs. For instance, under *Example 8* with $a = 0.4$, there is an error in the 'solve.default(var.cov)' command because the system is computationally singular: 1.49841e−17. As a result, that test could not undergo insightful finite-sample investigation and cannot be safely applied to the general non-restrictive problem (1).

5. Conclusions

The paper has been aimed at elaborating a new nonparametric omnibus test for verifying $\mathcal{H}$ against $\mathcal{A}$, investigating its asymptotics by deriving the limiting null distribution of the test statistic and proving the consistency of the related test under the general alternatives, as well as examining the finite sample behaviour of the solutions to the problem (1).

The first issue has been addressed by combining the Kolmogorov–Smirnov type statistic with a sup-type functional of the weighted two-sample paired data empirical process. The second objective is met by employing empirical processes' theory. The conducted extensive simulation study reveals properties of the investigated solutions of the considered testing problem under small, moderate, and large sample sizes. We examine the behaviour of the power functions of the tests both under the null model, as well as under the alternative. It turns out that the new test is a promising procedure which competes well with the best solutions of the two-sample problem for paired data.

Since the considered testing problem is the $LD - F1$ design in longitudinal data analysis for $t = 2$, extensions to $t > 2$ and other more complex models can be examined in future research.

Acknowledgements

The author is grateful to the Editor, Associate Editor, and Reviewers for the comments which led to the improvement in the presentation. He would like to express appreciation to P. Martínez-Camblor for sharing the R code, L. Gutiérrez for sending the R package BNPpairedtest, and J.-F. Quessy for comments on his multiplier. The research has been supported by the grant 2017/25/B/ST1/00500 from the National Science Centre, Poland. Calculations have been carried out at Wrocław Centre for Networking and Supercomputing (http://www.wcss.wroc.pl) under grant No. 199. M. Jedrzejewski's aid in ensuring optimal use of the Centre's resources is acknowledged with much gratitude, likewise.

Appendix A. Justification for the convergence (4) and (5)

Let $\ell^\infty(\mathbb{R})$ be the space of all functions $f : \mathbb{R} \mapsto \mathbb{R}$ such that $\|f\|_{\mathbb{R}} = \sup_{x\in\mathbb{R}} |f(x)| < +\infty$. We will also work in the product space $\ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R})$ with a proper product sup-norm $\|f \times g\|_{\mathbb{R}\times\mathbb{R}} = \|f\|_{\mathbb{R}} \vee \|g\|_{\mathbb{R}}$, where $x \vee y = \max\{x, y\}$.

Mind that B is the two-dimensional Brownian bridge with the covariance structure (3), while $B_1(\cdot) = B(\cdot, 1)$ and $B_2(\cdot) = B(1, \cdot)$ are the one-dimensional Brownian bridges.

Proposition 7. *Under* $\mathcal{H}$*,*

$$(i) \qquad \sqrt{n}\left(\hat{F} - F, \hat{G} - G\right) \xrightarrow{\mathcal{D}} (B_1 \circ F, B_2 \circ G) \qquad \textit{in} \quad \ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R}),$$

where the limiting process is the two-dimensional (F, G)*-Brownian bridge with* 2×2 *covariance matrix whose any element can be defined, in general, as follows:*

$$cov(B(F(x_1), G(y_1)), B(F(x_2), G(y_2))) = H(x_1 \wedge x_2, y_1 \wedge y_2) - H(x_1, y_1)H(x_2, y_2), \quad x_1, x_2, y_1, y_2 \in \mathbb{R};$$

$$(ii) \qquad \sqrt{n}(\hat{F} - \hat{G}) \xrightarrow{\mathcal{D}} B^0 \qquad in \quad \ell^\infty(\mathbb{R}),$$

$$where \quad B^0 \overset{\mathcal{D}}{=} B_1 \circ F - B_2 \circ G, \quad with \quad cov(B^0(x), B^0(y)) = F(x \wedge y) - H(x, y) - H(y, x) + G(x \wedge y), \quad x, y \in \mathbb{R}.$$

Proof. Consider the universally Donsker, see Example 2.5.4, [22], class of functions

$$\mathcal{F} = \{f : \mathbb{R} \mapsto \{0, 1\}, \text{ where } f(x_1) = \mathbb{1}(x_1 \le y_1), y_1 \in \mathbb{R}\}.$$

The ecdfs $\hat{F}$ and $\hat{G}$ can be identified with the empirical distributions of the observations indexed by the functions from the class $\mathcal{F}$. Then, the main empirical process theorem (cf. Theorems 2.5.2 and 2.5.6, [22]) implies

$$\sqrt{n}\left(\hat{F} - F\right) \xrightarrow{\mathcal{D}} B_1 \circ F \quad \text{in } \ell^\infty(\mathbb{R}) \quad \text{and} \quad \sqrt{n}\left(\hat{G} - G\right) \xrightarrow{\mathcal{D}} B_2 \circ G \quad \text{in } \ell^\infty(\mathbb{R}).$$

By Lemma 1.3.8, [22], $\sqrt{n}(\hat{F} - F)$ and $\sqrt{n}(\hat{G} - G)$ are asymptotically measurable and tight. Lemmas 1.4.3 and 1.4.4, [22], imply the same for the pair $(\sqrt{n}(\hat{F} - F), \sqrt{n}(\hat{G} - G))$. Then, Prohorov's theorem (1.3.9 (ii), [22]) assures that it has a weakly converging subsequence. A natural candidate for the limit is the two-dimensional (F, G)-Brownian bridge with the general element of the 2×2 covariance matrix given by

$$\begin{aligned} cov(B(F(x_1), G(y_1)), B(F(x_2), G(y_2))) &= C(F(x_1) \wedge F(x_2), G(y_1) \wedge G(y_2)) - C(F(x_1), G(y_1))C(F(x_2), G(y_2)) \\ &= H(x_1 \wedge x_2, y_1 \wedge y_2) - H(x_1, y_1)H(x_2, y_2), \quad x_1, x_2, y_1, y_2 \in \mathbb{R}. \end{aligned}$$

Furthermore, a possible distinct subsequence converges to the same limiting process by its nature. As a result, the assertion (i) follows. To prove (ii), it suffices to note that the above and the continuous mapping Theorem 18.11(i), [21], yield

$$\sqrt{n}(\hat{F}-\hat{G}) \xrightarrow{\mathcal{D}} B^0, \quad \text{where} \quad B^0 \overset{\mathcal{D}}{=} B_1 \circ F - B_2 \circ G, \quad \text{with}$$

$$\begin{aligned}\operatorname{cov}(B^0(x), B^0(y)) &= \operatorname{cov}(B_1(F(x)) - B_2(G(x)), B_1(F(y)) - B_2(G(y)))\\ &= F(x)\wedge F(y) - F(x)F(y) - C(F(x), G(y)) + F(x)G(y)\\ &\quad - C(F(y), G(x)) + G(x)F(y) + G(x)\wedge G(y) - G(x)G(y)\\ &= F(x\wedge y) - C(F(x), G(y)) - C(F(y), G(x)) + G(x\wedge y)\\ &= F(x\wedge y) - H(x,y) - H(y,x) + G(x\wedge y), \quad x, y \in \mathbb{R}. \quad \square \end{aligned} \tag{11}$$

Appendix B. Proofs

Here, we provide the proof of Proposition 2 exploiting the convergence result (4), the continuous mapping Theorem 18.11(i), [21], and the functional Delta-Method (Theorem 3.9.4, [22]). To prove Propositions 4–6 and Lemma 1, we employ the Multiplier Central Limit Theorem 2.9.6, [22], and the functional Delta-Method for the bootstrap (Theorem 3.9.11, [22]).

First, note that the process $\mathbb{E}_n/\sqrt{\mathbb{W}_n}$ can be considered as a function of $(\hat{F}, \hat{G})$. Specifically,

$$(\hat{F}, \hat{G}) \overset{\psi_1}{\longmapsto} (\hat{F}, \hat{G}, \hat{H}) \overset{\psi_2}{\longmapsto} (\hat{F}-\hat{G};\ \hat{F}+\hat{G}-2\hat{H}-[\hat{F}-\hat{G}]^2) \overset{\psi_3}{\longmapsto} (\mathbb{E}_n/\sqrt{\mathbb{W}_n})/\sqrt{n}. \tag{12}$$

In the next step, we show Hadamard differentiability of the map $\psi = \psi_3 \circ \psi_2 \circ \psi_1$. For this purpose, recall that $I = (1/2)F + (1/2)G$, $\varepsilon = 0.0001$, $a_I = I^{-1}(\varepsilon)$, and $b_I = I^{-1}(1-\varepsilon)$. Let $D[a_I, b_I]$ be the Banach space of all càdlàg functions $f : [a_I, b_I] \mapsto \mathbb{R}$ equipped with the uniform norm $\|f\|_{[a_I,b_I]} = \sup_{x\in[a_I,b_I]} |f(x)|$. Let $B0_\varepsilon[a_I, b_I]$ stand for the set of all càdlàg functions bounded below by $\varepsilon > 0$. We equip the product spaces with a proper product norms. See, for instance, [21], p. 257, Example 18.8.

Lemma 2. *The maps, defined in* (12)*, i.e.,*
(i) $\psi_1 : (D[a_I, b_I])^2 \mapsto (D[a_I, b_I])^3$,
(ii) $\psi_2 : (D[a_I, b_I])^3 \mapsto (D[a_I, b_I])^2$,
(iii) $\psi_3 : D[a_I, b_I] \times B0_\varepsilon[a_I, b_I] \mapsto D[a_I, b_I]$,
(iv) $\psi : B0_\varepsilon[a_I, b_I] \times B0_\varepsilon[a_I, b_I] \mapsto D[a_I, b_I]$,
are Hadamard-differentiable.

Proof.
(i) The ψ_1 map is linear and continuous, hence Hadamard-differentiable.
(ii) We have $\psi_2 = (\psi_{21}, \psi_{22})$. The maps ψ_{21} and ψ_{22} are Hadamard-differentiable by checking.
(iii) The map ψ_3 is a composition of $\psi_{3\cdot 1}$, $\psi_{3\cdot 2}$, and $\psi_{3\cdot 3}$, where

$$(A, B) \overset{\psi_{3\cdot 1}}{\longmapsto} (A, B^{1/2}) \overset{\psi_{3\cdot 2}}{\longmapsto} (A, B^{-1/2}) \overset{\psi_{3\cdot 3}}{\longmapsto} AB^{-1/2}.$$

The maps $\psi_{3\cdot 1}$ and $\psi_{3\cdot 3}$ are Hadamard-differentiable by simple checking, while $\psi_{3\cdot 2}$ is Hadamard-differentiable because B is bounded below from zero on $[a_I, b_I]$. Finally, the chain rule (Lemma 3.9.3 in [22]) implies Hadamard differentiability of ψ_3.

(iv) An application of the chain rule once more, leads to Hadamard differentiability of the composition $\psi = \psi_3 \circ \psi_2 \circ \psi_1$. $\square$

Proof of Proposition 2. Since in $(\ell^\infty[a_I, b_I])^2$

$$\sqrt{n}(\hat{F}-F, \hat{G}-G) \xrightarrow{\mathcal{D}} (B_1 \circ F, B_2 \circ G),$$

where the limiting process is the two-dimensional (F, G)-Brownian bridge with 2×2 covariance matrix whose element can be defined, in general, as follows:

$$\begin{aligned}\operatorname{cov}(B(F(x_1), G(y_1)), B(F(x_2), G(y_2))) &= C(F(x_1)\wedge F(x_2), G(y_1)\wedge G(y_2)) - C(F(x_1), G(y_1))C(F(x_2), G(y_2))\\ &= H(x_1\wedge x_2, y_1\wedge y_2) - H(x_1, y_1)H(x_2, y_2), \quad x_1, x_2, y_1, y_2 \in [a_I, b_I],\end{aligned}$$

the functional Delta-Method (Theorem 3.9.4, [22]) in combination with Lemma 2, under $\mathcal{H}$ and Assumption 1, imply

$$\mathbb{E}_n/\sqrt{\mathbb{W}_n} = \sqrt{n}\,\psi(\hat{F}, \hat{G}) = \sqrt{n}\,[\psi(\hat{F}, \hat{G}) - \psi(F, G)] \xrightarrow{\mathcal{D}} \psi'(B_1 \circ F, B_2 \circ G) \quad \text{in} \quad \ell^\infty[a_I, b_I],$$

where ψ' is the Hadamard derivative of ψ. As a result,

$$\psi'(B_1 \circ F, B_2 \circ G)(x) \stackrel{\mathcal{D}}{=} B_w^0(x) \stackrel{\mathcal{D}}{=} \frac{B_1 \circ F(x) - B_2 \circ G(x)}{\sqrt{F(x) + G(x) - 2H(x,x)}} \stackrel{\mathcal{D}}{=} \frac{B^0(x)}{\sqrt{F(x) + G(x) - 2H(x,x)}},$$

where B_w^0 is a zero mean Gaussian process with the covariance structure

$$E[B_w^0(x)B_w^0(y)] = \frac{F(x \wedge y) - H(x,y) - H(y,x) + G(x \wedge y)}{\sqrt{F(x) + G(x) - 2H(x,x)}\sqrt{F(y) + G(y) - 2H(y,y)}}, \qquad x, y \in [a_I, b_I].$$

An application of the continuous mapping Theorem 18.11(i), [21], completes the proof. $\square$

Wild bootstrap's part: Recall that, $\xi_1, \ldots, \xi_n$ are, independent from the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, i.i.d. random variables such that $E\xi_1 = 0$, $E\xi_1^2 = 1$, $\int_0^{+\infty} \sqrt{P(|\xi_1| > x)}dx < +\infty$. The wild bootstrap counterparts of the processes: $\mathbb{E}_n$, $\mathbb{w}_n$, and $\mathbb{V}_n$ take the form

$$\mathbb{E}_n^*(x) = (1/\sqrt{n})\sum_{i=1}^n \xi_i[\mathbb{1}(X_i \le x) - \mathbb{1}(Y_i \le x)],$$

$$\mathbb{w}_n^*(x) = (1/n)\sum_{i=1}^n \frac{|\xi_i|}{E|\xi_i|}[\mathbb{1}(X_i \le x) + \mathbb{1}(Y_i \le x) - 2 \cdot \mathbb{1}(X_i \le x, Y_i \le x)] - \left\{(1/n)\sum_{i=1}^n \frac{|\xi_i|}{E|\xi_i|}[\mathbb{1}(X_i \le x) - \mathbb{1}(Y_i \le x)]\right\}^2,$$

$$\mathbb{V}_n^*(x) = \frac{\mathbb{E}_n^*(x)}{\sqrt{\mathbb{w}_n^*(x)}},$$

and the related statistics are

$$KS^* = \sup_{x \in \mathbb{R}} |\mathbb{E}_n^*(x)|, \quad WKS^* = \sup_{a_I \le x \le b_I} \left|\mathbb{V}_n^*(x)\right|, \quad M^* = \max\{KS^*, WKS^*\}.$$

Additionally, define $\hat{F}^*(x) = (1/n)\sum_{i=1}^n \xi_i\, \mathbb{1}(X_i \le x)$, $\hat{G}^*(y) = (1/n)\sum_{i=1}^n \xi_i\, \mathbb{1}(Y_i \le y)$, and $\hat{H}^*(x,y) = (1/n)\sum_{i=1}^n \xi_i\, \mathbb{1}(X_i \le x, Y_i \le y)$, as well as $F^*(x) = (1/n)\sum_{i=1}^n \xi_i\, F(x)$, $G^*(y) = (1/n)\sum_{i=1}^n \xi_i\, G(y)$, and $H^*(x,y) = (1/n)\sum_{i=1}^n \xi_i\, H(x,y)$, $x, y \in \mathbb{R}$.

Proof of Proposition 4. Consider the class of functions $\mathcal{G} = \{g : \mathbb{R}^2 \mapsto \{0,1\},$ where $g(x_1, y_1) = \mathbb{1}(x_1 \le x_2, y_1 \le y_2), x_2, y_2 \in \mathbb{R}\}$. Since the class is universally Donsker, see, Example 2.5.4, [22], the conditional Multiplier Central Limit Theorem 2.9.6, [22], entails the convergence

$$\sqrt{n}\left(\hat{H}^* - H^*\right) \xrightarrow{\mathcal{D}} B \circ (F, G),$$

conditionally, given $(X_1, Y_1), \ldots, (X_n, Y_n)$. Since $\mathbb{E}_n^*(x) = \sqrt{n}\{\hat{H}^*(x, \infty) - \hat{H}^*(\infty, x)\}$, a double application of the conditional continuous mapping Theorem 1.9.5, [22], completes the proof. $\square$

Proof of Proposition 5. Define the weight $w(x) = F(x) + G(x) - 2H(x,x)$, for $x \in [a_I, b_I]$. Under $\mathcal{H}$ and Assumption 1,

$$\sup_{a_I \le x \le b_I} \left|\frac{\mathbb{w}_n^*(x)}{w(x)}\right| \xrightarrow{\mathcal{P}} 1, \qquad \inf_{a_I \le x \le b_I} w(x) > 0.$$

Therefore,

$$WKS^* = \sup_{a_I \le x \le b_I} \left|\mathbb{V}_n^*(x)\right| = \sup_{a_I \le x \le b_I} \left|\frac{\mathbb{E}_n^*(x)}{\mathbb{w}_n^*(x)}\right| = \sup_{a_I \le x \le b_I} \left|\frac{\mathbb{E}_n^*(x)}{w(x)}\right| + O_P(1).$$

The argument as in the proof of Proposition 4 finishes the job. $\square$

Proof of Proposition 6. A combination of Propositions 4 and 5 together with the conditional continuous mapping theorem makes the proof complete. $\square$

Proof of Lemma 1. It suffices to show that $q_Q^*(\alpha) = O_P(1)$ and $Q \xrightarrow{\mathcal{P}} +\infty$ for $Q \in \{KS, WKS, M\}$. The first holds because Q^* has a proper asymptotic distribution under $\mathcal{A}$.

(i) For the KS statistic, we have

$$\begin{aligned} KS &= \sup_{x \in \mathbb{R}} \left|\sqrt{n}\{\hat{F}(x) - \hat{G}(x)\}\right| \ge \left|\sqrt{n}\{\hat{F}(x_0) - \hat{G}(x_0)\}\right| \\ &= \left|\sqrt{n}\{\hat{F}(x_0) - F(x_0)\} - \sqrt{n}\{\hat{G}(x_0) - G(x_0)\} + \sqrt{n}\{F(x_0) - G(x_0)\}\right| = \left|O_P(1) + \sqrt{n}\{F(x_0) - G(x_0)\}\right| \xrightarrow{\mathcal{P}} +\infty, \end{aligned}$$

because $x_0 \in \mathbb{R}$ is such that $F(x_0) \ne G(x_0)$.

(ii) Under Assumption 1, let $x_0 \in [a_I, b_I]$ be such that $F(x_0) \neq G(x_0)$. We have

$$\begin{aligned}
WKS &= \sup_{a_I \le x \le b_I} \left| \frac{\sqrt{n}\{\hat{F}(x) - \hat{G}(x)\}}{\sqrt{\hat{F}(x) + \hat{G}(x) - 2\hat{H}(x,x) - [\hat{F}(x) - \hat{G}(x)]^2}} \right| \\
&\ge \left| \frac{\sqrt{n}\{\hat{F}(x_0) - \hat{G}(x_0)\}}{\sqrt{\hat{F}(x_0) + \hat{G}(x_0) - 2\hat{H}(x_0,x_0) - [\hat{F}(x_0) - \hat{G}(x_0)]^2}} \right| \\
&= \left| \frac{\sqrt{n}\{\hat{F}(x_0) - F(x_0)\} - \sqrt{n}\{\hat{G}(x_0) - G(x_0)\} + \sqrt{n}\{F(x_0) - G(x_0)\}}{\sqrt{\hat{F}(x_0) + \hat{G}(x_0) - 2\hat{H}(x_0,x_0) - [\hat{F}(x_0) - \hat{G}(x_0)]^2}} \right| \\
&= \left| \frac{O_P(1) + \sqrt{n}\{F(x_0) - G(x_0)\}}{\sqrt{O_P(1) + F(x_0) + G(x_0) - 2H(x_0,x_0) - [F(x_0) - G(x_0)]^2}} \right| \xrightarrow{\mathcal{P}} +\infty,
\end{aligned}$$

because the denominator is bounded below from zero for sufficiently large n.

(iii) The convergence $KS \xrightarrow{\mathcal{P}} +\infty$ and $WKS \xrightarrow{\mathcal{P}} +\infty$ implies $M \xrightarrow{\mathcal{P}} +\infty$. $\square$

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2023.105229.

We discuss there the outcomes of an additional simulation study and present a description of the R codes attached in 3 separate R files: Power_Functions.R, Generators_Alternative.R, and Generators_Null_Model.R.

References

[1] D. Bates, M. Maechler, M. Jagan, Matrix: sparse and dense matrix classes and methods, 2022, R, package version 1.5-3. https://CRAN.R-project.org/package=Matrix.
[2] E. Brunner, U. Munzel, M.L. Puri, Rank-score tests in factorial designs with repeated measures, J. Multivariate Anal. 70 (1999) 286–317.
[3] J.M. Chambers, C.L. Mallows, B.W. Stuck, A method for simulating stable random variables, J. Amer. Statist. Assoc. 71 (1976) 340–344.
[4] S. Demarta, A.J. McNeil, The t copula and related copulas, Internat. Statist. Rev. 73 (2005) 111–129.
[5] B. Ghattas, D. Pommeret, L. Reboul, A.F. Yao, Data driven smooth test for paired populations, J. Statist. Plann. Inference 141 (2011) 262–275.
[6] J. Hájek, Z. Šidák, P.K. Sen, Theory of Rank Tests, Academic Press, New York, 1999.
[7] F. Konietschke, M. Pauly, Bootstrapping and permuting paired t-test type statistics, Stat. Comput. 24 (2014) 283–296.
[8] P. Martínez-Camblor, Nonparametric k-sample test based on kernel density estimator for paired design, Comput. Statist. Data Anal. 54 (2010) 2035–2045.
[9] W.A. Morgan, A test for the significance of the difference between two variances in a sample from a normal bivariate population, Biometrika 31 (1939) 13–19.
[10] R.B. Nelsen, An Introduction to Copulas, 2nd edition, Springer, New York, 2006.
[11] K. Noguchi, Y.R. Gel, E. Brunner, F. Konietschke, nparLD : An R software package for the nonparametric analysis of longitudinal data in factorial experiments, J. Stat. Softw. 50 (2012) 1–23.
[12] F. Novomestky, matrixcalc: Collection of functions for matrix calculations, 2022, R package version 1.0-6. https://CRAN.R-project.org/package=matrixcalc.
[13] L.A. Pereira, D. Taylor-Rodríguez, L. Gutiérrez, A Bayesian nonparametric testing procedure for paired samples, Biometrics 76 (2020) 1133–1146.
[14] F. Pesarin, L. Salmaso, Permutation Tests for Complex Data: Theory, Applications and Software, Wiley, Chichester, 2010.
[15] E.J.G. Pitman, A note on normal correlation, Biometrika 31 (1939) 9–12.
[16] J.-F. Quessy, F. Éthier, Cramér-von Mises and characteristic function tests for the two and k-sample problems with dependent data, Comput. Statist. Data Anal. 56 (2012) 2097–2111.
[17] R Core Team, R : A language and environment for statistical computing, 2021, R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.
[18] Student, The probable error of a mean, Biometrika 6 (1908) 1–25.
[19] O. Thas, J.P. Ottoy, A nonparametric test for independence based on sample space partitions, Comm. Statist. Simulation Comput. 33 (2004) 711–728.
[20] W.-M. Tsai, A. Vexler, G. Gurevich, An extensive power evaluation of a novel two-sample density-based empirical likelihood ratio test for paired data with an application to a treatment study of attention-deficit/hyperactivity disorder and severe mood dysregulation, J. Appl. Stat. 40 (2013) 1189–1208.
[21] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, Cambridge, 2000.
[22] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes with Applications to Statistics, Springer, New York, 1996.
[23] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, Springer, New York, 2002.
[24] A. Vexler, G. Gurevich, A.D. Hutson, An exact density-based empirical likelihood ratio test for paired data, J. Statist. Plann. Inference 143 (2013) 334–345.
[25] A. Vexler, W.-M. Tsai, G. Gurevich, J. Yua, Two-sample density-based empirical likelihood ratio tests based on paired data, with application to a treatment study of attention-deficit/hyperactivity disorder and severe mood dysregulation, Stat. Med. 31 (2012) 1821–1837.
[26] C.F.J. Wu, Jackknife bootstrap and other resampling methods in regression analysis, Ann. Statist. 14 (1986) 1261–1295.
[27] J. Yan, Enjoy the joy of copulas: with a package copula, J. Stat. Softw. 21 (2007) 1–21.