Received: 26 November 2022 | Revised: 2 August 2023 | Accepted: 24 September 2023

DOI: 10.1002/asmb.2827

RESEARCH ARTICLE

WILEY

Simultaneous marginal homogeneity versus directional alternatives for multivariate binary data with application to circular economy assessments

Stefano Bonnini[1] | **Michela Borghesi**[1] | **Massimiliano Giacalone**[2]

[1]Department of Economics and Management, University of Ferrara, Ferrara, Italy

[2]Department of Economics, University of Campania "Luigi Vanvitelli", Capua (CE), Italy

Correspondence
Massimiliano Giacalone, Department of Economics, University of Campania "Luigi Vanvitelli", Corso Gran Priorato di Malta, 81043 Capua (CE), Italy.
Email: massimiliano.giacalone@unicampania.it

Funding information
Ministero dell'Università e della Ricerca

Abstract

Commodity price volatility is a major source of instability in those countries that are primarily commodity-dependent and has a negative impact, especially on economic growth. With this premise, commodities represent an effective financial exchange tool that nowadays finds relevance in being involved in the processes inherent to environmental sustainability. This work focus on raw materials and their demand, connected with the need for a transition towards the Circular Economy, as part of a strategy to address commodity supply disruptions. It presupposes changes in the mentality and behavior of companies in the various economic sectors. A crucial issue debated in the literature concerns whether or not the size of the companies favors their attitude towards Circular Economy. We propose a nonparametric method to test the effect of firm size on their propensity to undertake Circular Economy activities. Considering k of such activities, this propensity is a multidimensional concept and it can be represented by a k-dimensional vector of proportions. Each element of the vector represents the share of companies of the population under study that implement a specific Circular Economy activity. The main difficulty of such a multivariate testing problem, together with the multidimensional nature of the dichotomous response, is the one-sided type alternative, which is a stochastic dominance for multidimensional binary variables. A Monte Carlo simulation study proves the good power behavior of the proposed solution, based on a nonparametric approach. Case studies related to Italian small and medium enterprises in some strategic sectors are also addressed.

KEYWORDS

binary data, circular economy, multivariate analysis, permutation testing, simultaneous marginal homogeneity

Stefano Bonnini, Michela Borghesi, and Massimiliano Giacalone contributed equally to this work.

1 INTRODUCTION

In recent years, a growing number of companies have embraced the idea of a transition to the circular economy (CE).[1-3] The main reason is that a linear economy model (of the type take-make-dispose) is no longer sustainable, both from the point of view of production and with regard to the procurement of materials. CE is therefore considered one of the most effective solutions to environmental sustainability problems, also at the policy level.[4,5] Among the areas most involved in this transition, we have European Union[6] and China,[7] which have established an action program to support the transition to CE goals until 2050. Since 2015, Italy has also been contributing to this practice, especially in terms of separate collection of waste.[8]

This work focuses on raw materials and their demand, connected with the need for a transition towards the CE model, which presupposes changes in mentality and behavior. Considering the limited resources of the earth, the linear economy model proved ineffective. The continuous growth of the world population affects the unsustainable consumption of natural resources, which leads to an increase in the cost of raw materials. For this reason, the CE is an approach in which waste can be used as a raw material for the same or another industry, reducing the use of natural resources and the cost of raw materials.[9] With this premise, commodities represent an effective financial exchange tool that today finds relevance within the processes inherent to CE. CE represents an opportunity to increase resilience and overcome the risks due to uncertain future commodity supply and cost volatility.

Given that CE can be considered very important as a part of a commodity risk management strategy, in particular as a solution to commodity supply disruption, the paper focuses on the need to assess the actual feasibility of the CE goals, connected to the propensity and willingness of companies to implement CE activities. Such evaluation could be carried out through a sample survey, involving a representative sample of companies, and the application of appropriate inferential statistical techniques for data analysis. In other words, an empirical study that sheds light on companies' culture of innovation and sustainability, is necessary. To investigate the behaviors and attitudes of enterprises towards CE and therefore their propensity to follow CE-compatible business models, we focus on the main important factor affecting the propensity of companies to undertake activities related to this new economic approach based on recovery, recycling and reuse: the firm size. Specifically, we deal with a two-sample test for comparing two different types of companies (according to size) and compare their propensities towards CE. A company operates in CE if it implements one or more activities aimed at reusing production waste and protecting the environment. Hence, the response variable of the problem is multivariate and we are in the presence of binary data.

The business activities and choices compatible with the CE, as support to commodity risk management, are numerous and heterogeneous. Therefore, one way to assess the attitude towards the CE of a company is to administer a questionnaire with a list of k CE activities and, for each item, ask the company to indicate whether it undertakes or not the corresponding CE activity. The data collected from an empirical survey on such a topic, which involves n companies, regarding these k questions, consists of a $n \times k$ matrix of values equal to 0 or 1. The observed value x_{ij} is equal to 1 if the i-th company undertakes the j-th CE activity and it is equal to 0 otherwise, with $i = 1, \dots, n$ and $j = 1, \dots, k$. In other words, from the statistical point of view, we are in presence of a multidimensional response whose components are dichotomous variables.

In order to test whether the propensity towards CE of the first group of companies is less than the propensity towards CE of the second group, we should focus on the comparison between the proportion of circular companies in the first and in the second group, for each of the k CE activities taken into account. Definitely, we should compare the two k-dimensional vectors of sample proportions in order to make inference on the two k-dimensional vectors of population proportions. The inferential problem is not trivial due to the multidimensional response and to the directional alternative hypothesis. The formal definition and the detailed description of the hypotheses under study are included in the next section.

The proposed solution for the testing problem is nonparametric because based on the permutation approach. Its nonparametric nature makes it robust with respect to the assumptions about the distribution underlying the data. For this reason, it is preferable to a possible parametric solution. In fact, in general, permutation tests are much more powerful when the distributional assumptions of parametric tests do not hold but they have good power performance (similar to that of parametric tests) also when such assumptions are satisfied.[10] The proposed permutation test is also particularly appropriate for a multivariate problem such as the one under consideration. This is due to the fact that it allows taking into account the dependence between variables without knowing the underlying distribution and without modeling the dependence structure. Another advantage of the proposed method is that it is applicable also in the presence of small samples and large number of variables. In the literature, solutions to multivariate problems for stochastic dominance and stochastic ordering alternative hypotheses have been proposed. Most of them are suitable for numeric data, some others are valid for categorical and binary variables. Only few works are dedicated to the case of binary data for one-tailed hypotheses. In particular, we mention the interesting work of Davidov and Peddada.[11] A literature review on this topic is included in the next section. We wish to give a scientific contribution in this specific framework, with the study of a nonparametric method for multivariate binary data and directional alternatives based on the permutation approach. Our proposal is applicable under the condition of exchangeability and, in many situations, preferable to the parametric methods. In particular, we face the problem with solutions that belong to the family of combined permutation tests. This family of permutation tests was proposed by Pesarin.[10] An innovative methodological aspect of the paper concerns the detailed simulation study and, in particular, the comparison of the power behavior of the two main combined permutation tests applied to binary data. An important finding concerns the detection of which member of such a family of tests is the most powerful depending on the proportion of true partial alternative hypotheses. The illustrated method is valid under certain conditions, in particular in the presence of exchangeability of data under the null hypothesis. In the absence of exchangeability, the method is not applicable unlike other solutions not based on the permutation approach.

Definitely, we deal with an extension of the one-tailed two-sample test on proportions in a multivariate sense. The motivating example relates to a sample survey carried out in Italy to investigate the propensity towards CE of Italian small and medium enterprises (SMEs). In this survey, circular attitudes and behaviors are represented by six different CE activities such as reuse of waste, the introduction of innovations or investments in R&D related to the CE practices, etc. The goal is to contribute to the debate on the effect of company's size on the propensity towards CE. Some authors believe that the propensity increases with the size but in the literature there is not enough empirical evidence in favor of such a hypothesis and there is not a common and generally accepted position on the topic by the experts in the field. In the motivating example, the hypothesis under investigation is that the propensity towards CE of small enterprises is less than that of medium enterprises.

Hence, the paper provides various scientific contributions. First, it describes a valid methodological solution to a complex testing problem by proposing a flexible and powerful nonparametric method and studying its main properties and its performance. Second, it proposes a suitable approach to commodity risk management, based on carrying out sample surveys and using suitable inferential statistical techniques for monitoring and assessments regarding CE feasibility. This is crucial to support decision-making processes by managers and policymakers. Furthermore, it contributes to the debate on the comparative assessment of the propensity to CE, as a possible solution to commodity supply disruptions, of small and medium enterprises.

Section 2 focuses on the definition of the statistical problem. The description of the methodological proposal is presented in Section 3. Section 4 is dedicated to a simulation study. The results of the application of the proposed method to case studies concerning Italian SMEs, are shown and discussed in Section 5 and the conclusions are reported in Section 6.

2 | TESTING PROBLEM

The null hypothesis of the problem (i.e., the equality of the two vectors of population proportions) is usually referred to in the literature as *simultaneous marginal homogeneity*.[12] For the case of binary data with two-sided alternative hypotheses, Chuang-Stein and Mohberg[13] proposed a test based on a multivariate extension of the Wald type statistic. For the same problem, Agresti and Klingenberg[12] presented a solution based on the permutation approach. This nonparametric test is proved to be powerful and flexible in general, but it is particularly useful in the presence of small n and large k, because in these conditions a parametric approach becomes much less competitive if not impossible. To overcome the impossibility of the chi-square test to take into account the possible ordinal nature of data, Agresti[14] proposed a specific test for problems with ordered categorical data. To test *simultaneous marginal homogeneity* versus stochastic ordering alternatives, various solutions have been proposed.[15-17] For clustered matched paired data, the method of Yan, Sun and Harging[18] and that of Deng and Carriere[19] are available. A Bayesian nonparametric approach to model spatially distributed multivariate binary data, in the presence of a particular correlation structure was proposed by Kang et al.[20]

Extending the interest to the general case of tests for stochastic order, some nonparametric solutions, distinct from the methodology of combined permutation tests, have been proposed.[11,21-23] Davidov[21] deals with stochastic dominance problems and presents a solution based on a family of linear rank tests, investigating the relationship between such a solution and the methods based on the combination of p-values. Davidov and Peddada[22] consider the two-sample linear stochastic order for multivariate discrete or continuous variables and introduce a nonparametric inferential approach that generalizes Roy's largest root test. For testing multivariate stochastic ordering, a nonparametric solution is proposed by Davidov and Peddada.[23] To test a multivariate stochastic order in the presence of binary data, an inferential nonparametric approach based on the use of the estimators of the vector of marginal probabilities and of the variance/covariance matrix, in the test statistic, is presented by Davidov and Peddada.[11] Such a test statistic asymptotically follows a chi-bar distribution but the weights of the combination depend on unknown parameters that must be estimated. To overcome this problem, the authors propose to use bootstrap methods. This solution is proved to be very powerful and preferable to some permutation tests. As said, we face the problem of simultaneous marginal homogeneity against one-tailed alternative hypotheses for multivariate binary variables, hence we are interested on such specific testing problem. As far as we know, a deep and extended simulation study to assess the power performance of combined permutation tests for such a problem is missing.

Let us assume that the value $x_{ij} \in \{0,1\}$ of the j-th variable (e.g., j-th CE activity) observed on the i-th statistical unit (e.g., firm) is a realization of the random variable X_{ij} with $i = 1, \dots, n$ and $j = 1, \dots, k$. Such a random variable follows the Bernoulli distribution with parameter θ_{sj}, where $s \in \{1,2\}$ denotes the population to which the i-th unit belongs. Formally, for every j:

$$X_{ij} \sim \begin{cases} \mathscr{B}e(\theta_{1j}) & \text{if } i \text{ is in sample 1} \\ \mathscr{B}e(\theta_{2j}) & \text{if } i \text{ is in sample 2.} \end{cases}$$

Without loss of generality, let $\{1, 2, \dots, n_1\}$ denote the units in sample 1 and $\{n_1+1, n_1+2, \dots, n_1+n_2 = n\}$ denote the units in sample 2. The probabilistic assumption mentioned above means that $X_{1j}, X_{2j}, \dots, X_{n_1 j}$ are identically distributed according to $\mathscr{B}e(\theta_{1j})$ and $X_{n_1+1,j}, X_{n_1+2,j}, \dots, X_{nj}$ are identically distributed according to $\mathscr{B}e(\theta_{2j})$, with $j = 1, 2, \dots, k$.

Actually, θ_{sj} can be seen as the proportion of firms of the s-th population that undertakes the j-th CE activity. Let us assume that Y_{sj} denotes a Bernoulli random variable with parameter θ_{sj}, that is, $Y_{sj} \sim \mathscr{B}e(\theta_{sj})$. The parameter θ_{sj} can also be defined as the probability of "success" of the Bernoulli random variable Y_{sj}, that is, the probability that a firm randomly selected from the s-th population undertakes the j-th CE activity. In other words, $P(X_{ij} = x) = P(Y_{sj} = x) = \theta_{sj}^{x}(1-\theta_{sj})^{1-x}$, with $x \in \{1,0\}$ and s denoting the population (and the sample) of i.

As a consequence, the (row) vector of values $(x_{i1}, x_{i2}, \dots, x_{ik})$ observed on the i-th unit is realization of the multidimensional random variable $(X_{i1}, X_{i2}, \dots, X_{ik})$ or, equivalently, of the random variable $(Y_{s1}, Y_{s2}, \dots, Y_{sk})$ which follows a k-variate Bernoulli distribution with vector of success probabilities $\boldsymbol{\theta_s} = (\theta_{s1}, \theta_{s2}, \dots, \theta_{sk})^T$, with $s = 1, 2$ denoting the population of i. It is worth noting that the vector of parameters $\boldsymbol{\theta_s}$ does not exhaustively characterize the distribution of the multivariate variable $(Y_{s1}, Y_{s2}, \dots, Y_{sk})$. Each component θ_{sj} determines the central tendency and the variability of the marginal distribution of the j-th component, given that $E[Y_{sj}] = \theta_{sj}$ and $Var[Y_{sj}] = \theta_{sj}(1-\theta_{sj})$. However, there is no specification on the dependence between the k components of the multivariate response, and the vector of parameters $\boldsymbol{\theta_s}$ is not informative about this dependence. This point is very important, because an advantage of the proposed solution, which is based on the nonparametric approach of the Combined Permutation Test (*CPT*), is the flexibility due to the fact that it is not necessary to make assumptions on the dependence between components nor to estimate nuisance parameters that represent such dependence. Unless the strong and implausible condition of independence between the k components is assumed, the dependence must be taken into account. In the *CPT* family of tests, such dependence is implicitly considered, even if unknown and not explicitly defined, through the permutations of the rows of the dataset to determine the null distribution of the test statistics.

The null hypothesis of the problem consists of the equality of the θ parameters of the two populations for each of the k components. In the alternative hypothesis, for at least one component, the strict inequality holds (i.e., for one or more CE activities, the proportion of firms that undertake such activities is lower in the first population). Hence, the system of hypotheses of the problem is the following:

$$\begin{cases} H_0: & \boldsymbol{\theta_1} = \boldsymbol{\theta_2} \\ H_1: & \boldsymbol{\theta_1} \leq \boldsymbol{\theta_2}, \end{cases}$$

where the alternative hypothesis means that, for each component, either the equality or the strict inequality "$<$" holds, and the strict inequality is satisfied at least for one component. Equivalently,

$$\begin{cases} H_0: & \theta_{1j} = \theta_{2j} \forall j \in \{1, \dots, k\} \\ H_1: & \exists j \in \{1, \dots, k\} \text{ s.t. } \theta_{1j} < \theta_{2j}. \end{cases}$$

It is worth noting that the null hypothesis is not equivalent to the equality in distribution of the two k-variate responses $\mathbf{Y_1} = (Y_{11}, Y_{12}, \dots, Y_{1k})$ and $\mathbf{Y_2} = (Y_{21}, Y_{22}, \dots, Y_{2k})$ but only to the equality of the k marginal distributions. In other words, given the joint probability mass function $f_s(\mathbf{y}) = f_s(y_1, y_2, \dots, y_k) = P(Y_{s1} = y_1, Y_{s2} = y_2, \dots, Y_{sk} = y_k)$ with $s = 1, 2$ and $\mathbf{y} \in \{0, 1\}^k$, the null hypothesis is not equivalent to $f_1(\mathbf{y}) = f_2(\mathbf{y})$. In fact, as said above, the hypothesis does not involve the joint k-variate distribution of the multivariate response but only the marginal distributions of the k components. This is the reason why such a null hypothesis is often indicated as *simultaneous marginal homogeneity*. Rather, given that $\theta_{1j} < \theta_{2j} \Leftrightarrow Y_{1j} \overset{d}{<} Y_{2j}$, the alternative hypothesis can be considered as a stochastic dominance for multivariate binary variables.

3 | PERMUTATION TEST

The solution proposed is based on the idea that the problem presented in the previous section may be considered a multiple test. In fact, the general problem can be broken down into k different partial tests. As mentioned in Section 2, in the j-th partial test, we are in presence of the following system of hypotheses, with $j = 1, \dots, k$:

$$\begin{cases} H_{0,j} : \theta_{1j} = \theta_{2j} \\ H_{1,j} : \theta_{1j} < \theta_{2j} \end{cases}$$

and the general problem can be represented as

$$\begin{cases} H_0 : \bigcap_{j=1}^{k} H_{0,j} \\ H_1 : \bigcup_{j=1}^{k} H_{1,j}, \end{cases}$$

where the symbols of intersection and union mean that "all" the partial hypotheses hold and "at least one" partial hypothesis hold respectively.

The general theoretical framework of the *CPT* methodology was formalized by Pesarin in 2001[10] and then deepened and developed in several works published in the last two decades. Among the most important monographs dedicated to the topic, we mention the contributions of Pesarin and Salmaso in 2010[24] and Bonnini et al. in 2014.[25] Such works have contributed to popularizing the methodology, in terms of theory, applications and software. The empirical studies in which the *CPT* methodology has been successfully applied are several and heterogeneous.[26-29] The flexibility and robustness of the approach contribute to make it suitable for big data,[30] categorical data,[17,31,32] regression analyses,[33-35] ranking problems,[36] non monotonic alternatives[37] and many other methodological frameworks.

A solution for the j-th partial problem, $H_{0,j}$ versus $H_{1,j}$, is provided by the permutation test for the proportions of two independent samples. A reasonable test statistic is

$$T_j = \hat{\theta}_{2j} - \hat{\theta}_{1j} = \sum_{i=n_1+1}^{n} \frac{X_{ij}}{n_2} - \sum_{i=1}^{n_1} \frac{X_{ij}}{n_1}.$$

We note that the test statistic is not standardized because we don't need to assume an exact or approximate standard normal distribution or other known probability law for the test statistic under H_0. For each partial test, the null hypothesis is rejected for large values of the test statistic. Therefore, the p-value of the j-th test is $P(T_j \geq t_{j,obs} | H_0)$, that is, the probability that, when the null hypothesis is true, the test statistic takes values at least equal to the observed one. By definition, the significance level function of the j-th partial test is $L_j(t) = P(T_j \geq t | H_{0,j})$, therefore the p-value can also be defined as $L_j(t_{j,obs})$.

Under the null hypothesis, the condition of exchangeability is satisfied because all the possible assignments of the n observed data to the samples (i.e., all the possible data permutations) have the same probability of occurrence. Therefore, the null permutation distribution of the test statistic can be obtained by considering all the possible dataset permutations and, for each permuted dataset, the corresponding value of the test statistic. Actually, for computational reasons, instead of considering the exact distribution based on all the possible permutations, a random generation of B permutations is carried out, where B is not large (usually greater than or equal to 1000) with respect to the cardinality of the permutation space. This approach, called the Conditional Monte Carlo method, is very common in practice. It allows obtaining a null distribution of the test statistic very similar to the exact permutation distribution with a high degree of approximation according to the Glivenko-Cantelli theorem. The significance level function of the j-th partial permutation test can be estimated as follows:

$$L_j^P(t) = \frac{\sum_{b=1}^{B} I_{\left(-\infty, t_{j,b}^P\right]}(t) + 0.5}{B+1},$$

where $t_{j,b}^P$ is the value of T_j corresponding to the b-th dataset permutation and $I_A(t)$ is the indicator function of the set A.

The proposed solution for the general multivariate testing problem consists in the combination of the significance level functions of the k partial tests. Without loss of generality, assume that the general null hypothesis is rejected for large values of the combined test statistic. Let $(t_1, t_2, \ldots, t_k)$ be a generic set of values taken by the k-variate random variable $(T_1, T_2, \ldots, T_k)$ and $(l_1, l_2, \ldots, l_k)$ the corresponding values of the significance level functions, that is, $l_j = L_j(t_j)$ with $j = 1, 2, \ldots, k$. The test statistic for the general testing problem is

$$T_\psi(t_1, t_2, \ldots, t_k) = \psi(l_1, l_2, \ldots, l_k),$$

where $\psi : \mathbb{R}^k \to \mathbb{R}$ is a function that satisfies some mild and simple conditions. The combining function must be non-increasing in each argument, it tends to its supremum when one argument tends to zero and provides a critical value for the combined test strictly less than the supremum. The class of combining functions that satisfy such conditions is very wide and each combining function corresponds to a different combined test with specific properties and performance.

Let $t_{\psi,obs}$ denote the value of the combined test statistic corresponding to the observed non-permuted dataset and $L_\psi^P(t)$ the *CPT* significance level function corresponding to the combination ψ. The p-value of the general testing problem, according to the permutation distribution, is $L_\psi^P(t_{\psi,obs})$. Within the wide family of combined permutation tests, we will consider the following members:

1. CPT based on the Fisher combination $T_F = -\sum_{j=1}^{k} \log(l_j)$
2. CPT based on the Tippett combination $T_T = \max_{j=1,\ldots,k}(1 - l_j)$.

Even if the testing problem concerns the marginal distributions of the multivariate response, in order to have a reliable and powerful solution, the dependence between the k components must be taken into account. In the reduction of the dimensionality of the multivariate test statistic (from k to 1), through the application of the nonparametric combination, the dependence between the partial tests is not ignored even if not explicitly modelled. Such important property is assured by the use of a suitable combining function and by the dataset permutations. In fact, the data permutations of the different columns are not independent because permutations concern the rows of the data matrix, that is, the individual response profiles.

We would like to point out some differences between the proposed permutation test and the nonparametric test introduced by Davidov and Peddada.[11] First of all, the very interesting nonparametric competitor uses the bootstrap approach to overcome the problem of the weights necessary to compute the test statistic, which depend on unknown variances that must be estimated. Our proposal, based on the permutation approach, is clearly applicable under the condition of exchangeability and, in many situations, as said, preferable to the parametric methods. Far from being the absolute best solution, the illustrated permutation method is valid under certain conditions, in particular, as said, in the presence of exchangeability of data under the null hypothesis. One peculiarity of our work is the extended simulation study in terms of different sample sizes, in both balanced and unbalanced designs, also considering different imbalance levels. This aspect is important also because the proposed permutation approach is powerful for large sample sizes (due to the consistency of the test) but performant and applicable also for small sample sizes. Moreover, we will prove that the higher the number of components, the greater the power of the test, regardless of the sample sizes. Hence, such a permutation test is applicable also when the number of variables is larger than the sample sizes. Unlike the method of Davidov and Peddada, the proposed permutation test doesn't need the estimation of any nuisance parameter, in particular the scaled variance of the estimator of the vector of marginal probabilities. Furthermore, it is performant and can be applied also in situations of sparse data (with small probabilities of success) without the need of a ridge regression (by adding an arbitrary constant) to guarantee the invertibility of the estimated matrix. In the absence of exchangeability, the method is not applicable, unlike other solutions not based on the permutation approach such as that of Davidov and Peddada.

4 | MONTE CARLO SIMULATION STUDY

A Monte Carlo simulation study was carried out in order to investigate the power behavior of the proposed permutation test. The application of the test, the data random generations and the comparative performance assessments were carried out through new *R* scripts created by the authors. For the computation of *p*-values, according to the null permutation distribution of the test statistics, 1000 random permutations were carried out. For each setting, 1000 datasets (replicates) were randomly generated and the power was estimated as a proportion of replicates in which the null hypothesis is rejected (rejection rate). Sample data were randomly generated from normal *k*-variate distributions and then transformed into binary values according to the probabilities of success of the two populations. The number of components of the multivariate response considered in the simulation settings are $k = 5$ and $k = 10$.

Let $\mathbf{0_k}$ denote the vector of k elements equal to zero, $\mathbf{1_{k\times k}}$ be the square matrix of order k of which all elements are equal to one and $\mathbf{I_k}$ the identity matrix of order k, that is,

$$\mathbf{0_k} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \mathbf{1_{k\times k}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \text{ and } \mathbf{I_k} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Formally, the simulated data for each unit, that is, for each individual response profile, were obtained by randomly generating a vector of k values from $Z \sim \mathcal{N}_k(\mathbf{0_k}, \mathbf{\Sigma})$, a normal k-variate random variable whose marginal components are standard normal, with $\mathbf{\Sigma} = \rho\mathbf{1_{k\times k}} + (1-\rho)\mathbf{I_k}$, that is,

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}.$$

The binary values for both the samples were obtained by transforming the normal ones. For the first population we have $E[Y_1] = \boldsymbol{\theta_1} = \theta_1\mathbf{1_k}$, where $\mathbf{1_k} = (1, 1, \dots, 1)^T$ is the vector of k ones. For the second population, by assuming that under H_1 a proportion p of the k partial null hypotheses is false and a proportion $1 - p$ of the k partial null hypotheses is true, we set $E[Y_2] = \boldsymbol{\theta_2} = \boldsymbol{\theta_1} + \delta\left(\mathbf{1}^T_{\mathbf{pk}}, \mathbf{0}^T_{\mathbf{(1-p)k}}\right)^T$. Hence, for all the n units, a k-dimensional vector of normal values was generated by the same identical random variable Z but the n_1 vectors of normal values of sample 1 were transformed according to $\boldsymbol{\theta_1}$ and the n_2 vectors of normal values of sample 2 were transformed according to $\boldsymbol{\theta_2}$. Each component of Y_s, with $s = 1, 2$, was generated n_s times as a function of the corresponding component of Z, according to the following rule:

$$Y_{sj} = \begin{cases} 1 & \text{if} \quad Z_j \le z_{1-\theta_{sj}} \\ 0 & \quad\text{otherwise,} \end{cases}$$

with $s = 1, 2$ and $j = 1, \dots, k$, where $z_{1-\theta_{sj}}$ is the quantile of the standard normal cumulative distribution function that leaves a probability of $1 - \theta_{sj}$ on the right, that is, $z_{1-\theta_{sj}} = \Phi^{-1}(\theta_{sj})$. The n k-dimensional realizations were generated independently of each other. Both balanced and unbalanced sample cases were simulated. Therefore, the setting parameters varied in the simulations are the following scalars:

1. n_1 and n_2: sample sizes,
2. $u = \frac{n_2}{n_1}$: imbalance ratio,
3. ρ: Pearson correlation parameter that represents the dependence between the k components of the multivariate response,
4. θ_1: probability of success for each component in population 1,
5. δ: positive "shift" parameter for the probability of success that represents the "distance" between the two populations for the components (partial tests) under H_1,
6. k: number of components of the multivariate response,
7. p: proportion of the k components (partial tests) for which the alternative hypothesis is true, that is, proportion of true partial alternative hypotheses (among the k of the multiple test).

The nominal significance level chosen for the simulations, under all the settings considered in the study, is $\alpha = 0.05$. In this section we report the output of the simulations for some settings of the considered parameters. The results corresponding to other settings, are provided in the Appendix. Hence, since the probability of rejection of the null hypothesis when it is true should not exceed α, the rejection rates of the proposed permutation test should be not greater than 0.05 when data are simulated under the null hypothesis. The results of the simulation study under H_0, when the number of components of the multivariate Bernoulli distribution is $k = 5$ and $k = 10$, are reported in Figure 1. In particular, in these simulations, data were generated by assuming a correlation level $\rho = 0.3$, moderate but quite realistic and common in real applications. The probability of success is set at $\theta_1 = 0.3$. Such a value corresponds to a situation of sparse data or quite rare events, that is we expect to observe mostly zeros and rarely ones. These are usually unfavorable conditions for the performance of the testing problems on proportions like the one under study. The rejection rates represented in the graphs of Figure 1, are plotted as a function of the total sample size $n = n_1 + n_2$, in the balanced problem, that is, when $u = 1$ and consequently $n_2 = u \cdot n_1 = n_1$. The size of each sample varies from a minimum of 20 to a maximum of 100. Obviously, the proportion p of the k components (partial tests) for which the alternative hypothesis is true is 0 because data are simulated under the null hypothesis. For both the *CPT*s defined above, that is, those based on the Fisher and on the Tippett combination, the rejection rates are very close to 0.05 and usually less than 0.05. Therefore the tests are well approximated and, under the null hypothesis, they respect the nominal α level. In Table A1 of the Appendix, the detailed rejection rates plotted in Figure 1 are reported and the additional case of $n = 1000$ is included, in order to take into consideration the case of very big sample sizes. Also in this specific situation, the power of the two tests, in both cases $k = 5$ and $k = 10$, is very close to the nominal α level. Hence, it is confirmed that the tests are well approximated.

Figures 2,3,4,5,6,7, and 8 show the power behavior under the alternative hypothesis in different settings. From Figures 2 to 7 the balanced design ($u = 1$) is considered. In particular, in Figure 2, the rejection rates of the tests based on the Fisher and on the Tippett combination can be compared.

In these settings, the values of ρ and θ_1 are the same of the simulations under H_0 considered above. The number of components is $k = 10\%$ and 20% of them (i.e., two of them because $p = 0.2$) are under H_1. For these two variables, the probability of success is 0.35, hence slightly greater than θ_1 because $\delta = 0.05$. From Figure 2, it is evident that the rejection rates are increasing function of the sample size and the tests are consistent. Even if we are very close to the null hypothesis (low values of p and δ), the power of the tests are always greater than α, therefore the tests are unbiased. Moreover, the *CPT* based on the Fisher combination is more powerful than the *CPT* based on the Tippett combination, uniformly with respect to the sample size. In Table A2 of the Appendix, the results for very large sample sizes ($n \geq 1000$)

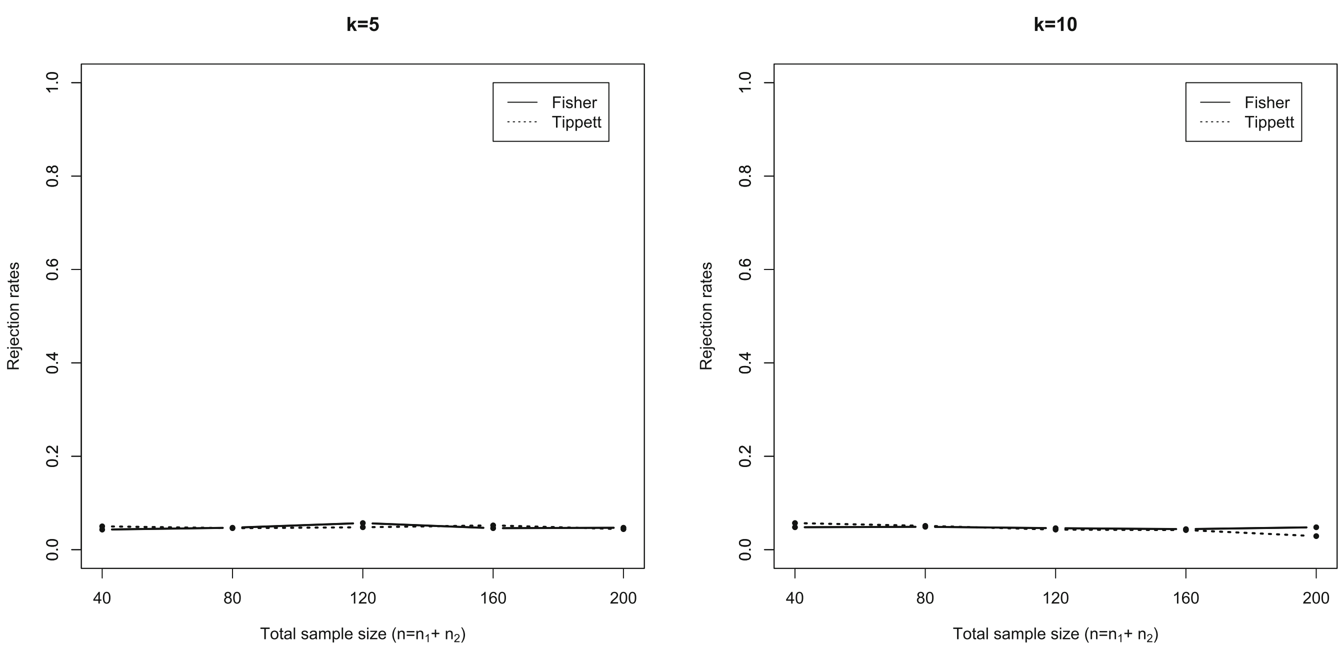

FIGURE 1 Rejection rates of CPTs (combination of Fisher and Tippett) versus sample size under the null hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $u = 1$, $k = 5$ and 10.

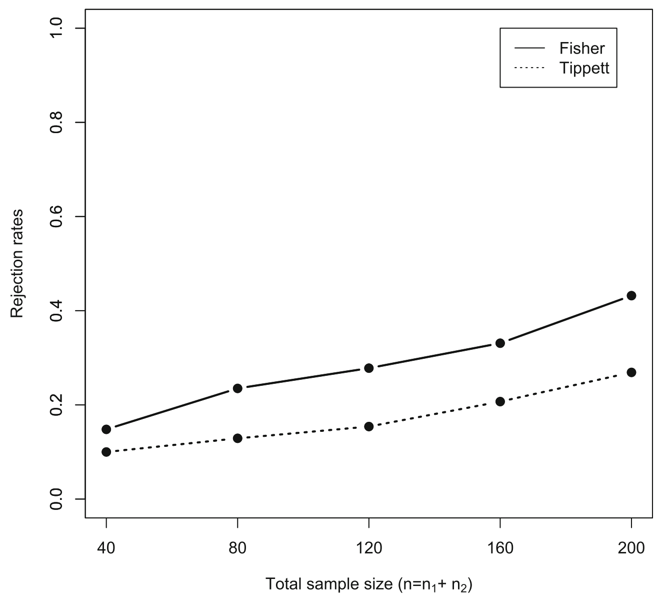

FIGURE 2 Rejection rates of CPTs (combination of Fisher and Tippett) versus sample size under the alternative hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $\delta = 0.05$, $k = 10$, $p = 0.2$, and $u = 1$.

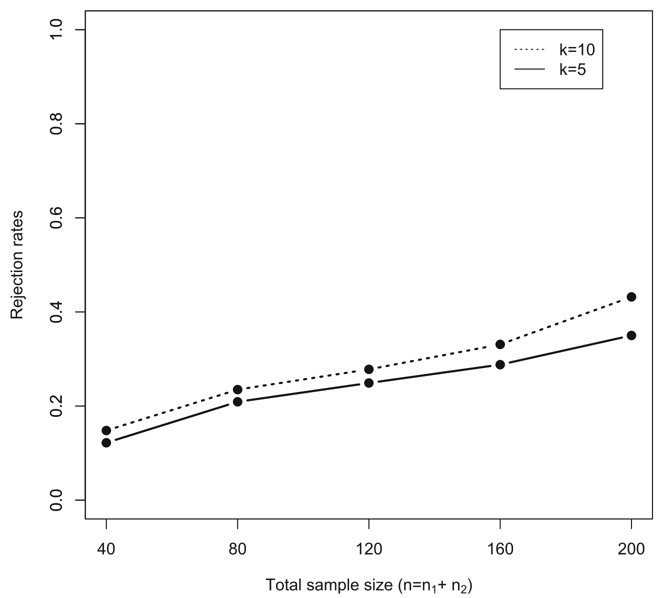

FIGURE 3 Rejection rates of CPT with combination of Fisher versus sample size under the alternative hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $\delta = 0.05$, $p = 0.2$, $u = 1$, $k = 5$ and 10.

are added in order to assess the convergence rate to 1 of the power. We found that the power of the CPT based on the Fisher combination tends to 1 when $n \geq 2400$, whilst the power of the CPT based on the Tippett combination reach the value 1 when $n = 4000$ (lower convergence rate). Given that the Fisher combination seems to be preferable, the performance of the corresponding combined test was deepened by analyzing the power as a function of various setting parameters.

In Figure 3, the rejection rates of the test based on the Fisher combination under the setting of Figure 2 are compared with those of the same setting in which $k = 5$ instead of $k = 10$. The compared plots prove that the higher the number of variables k, *ceteris paribus*, the greater the power. This conclusion is confirmed in Table A3 of the Appendix, where the simulations were extended to the cases $k = 15$, $k = 20$ and $k = 25$. An important finding concerns the increasing of the convergence rate with the increase of k. In fact, when $k = 5$, the rejection rate of the test tends to 1 when $n = 3400$; when $k = 25$, the power reaches the maximum when $n = 2000$, much lower value.

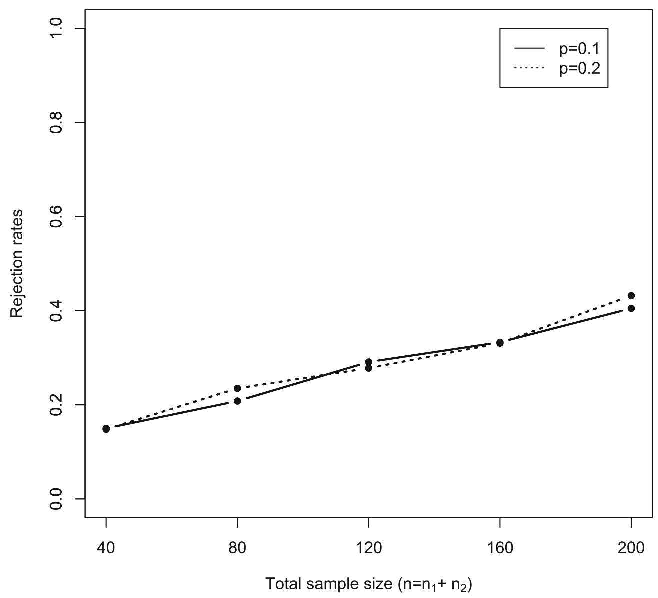

FIGURE 4 Rejection rates of CPT with combination of Fisher versus sample size under the alternative hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $\delta = 0.05$, $k = 10$, $u = 1$, $p = 0.2$ and 0.1.

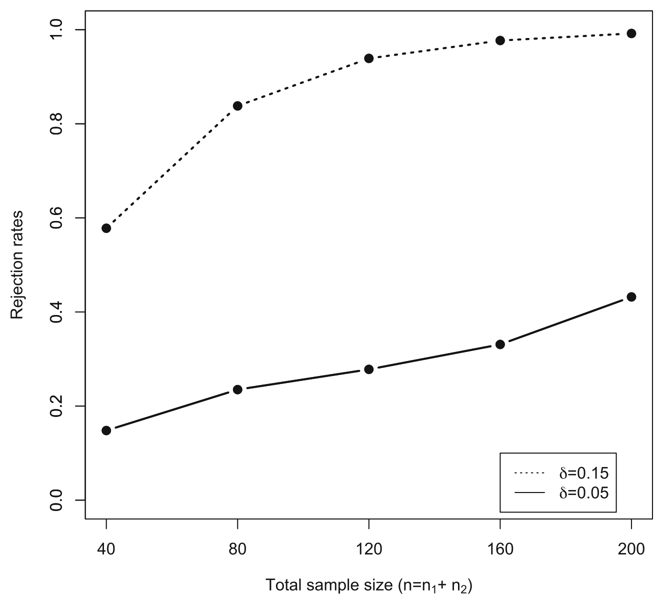

FIGURE 5 Rejection rates of CPT with combination of Fisher versus sample size under the alternative hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $k = 10$, $u = 1$, $p = 0.2$, $\delta = 0.05$ and 0.15.

In order to assess the effect of the proportion p of true partial alternative hypotheses, the rejection rates of the *CPT* based on the Fisher combination were evaluated under the setting of Figure 2, varying p from 0.2 to 0.1 (2 out of 10 and 1 out of 10 true partial alternative hypotheses respectively). In Figure 4 the power functions are compared. Unlike what happens for the number of variables k (for the same p), the proportion p of variables under the alternative hypothesis (for the same total number of variables k) does not seem to affect the power of the test.

Similarly to what done with k and p, a comparative evaluation was carried out to assess the possible effect of the shift parameter δ on the power behavior of the *CPT* with Fisher combination. The power related to the setting considered in Figure 2 was compared with the power regarding the setting obtained from the former by changing δ from 0.05 to 0.15. When $\delta = 0.15$, the power is much greater than when $\delta = 0.05$ (see Figure 5). The simulations were extended to the cases $\delta = 0.25$, $\delta = 0.35$ and $\delta = 0.45$ (see Table A4 in the Appendix). As expected, the power increases with δ. In fact,

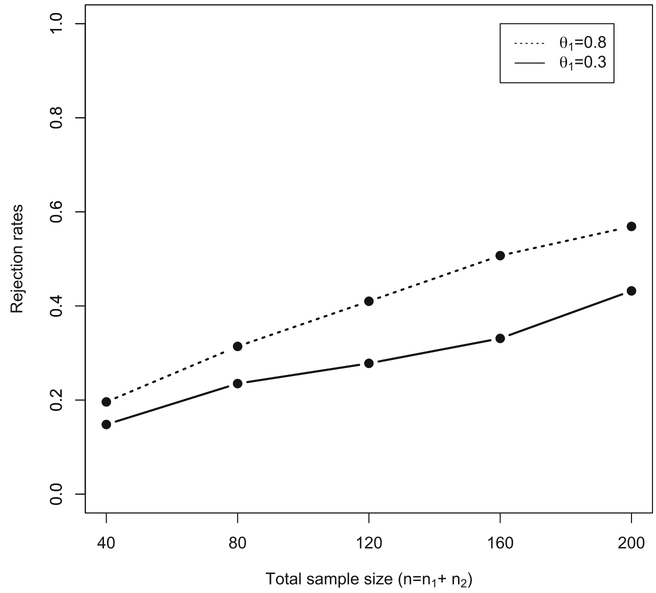

FIGURE 6 Rejection rates of CPT with combination of Fisher versus sample size under the alternative hypothesis, with $\rho = 0.3$, $k = 10$, $u = 1$, $p = 0.2$, $\delta = 0.05$, $\theta_1 = 0.3$ and 0.8.

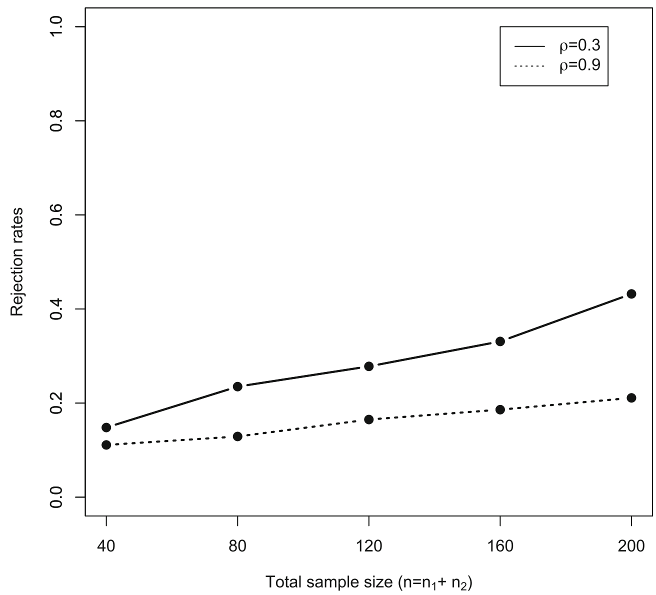

FIGURE 7 Rejection rates of CPT with combination of Fisher versus sample size under the alternative hypothesis, with $k = 10$, $u = 1$, $p = 0.2$, $\delta = 0.05$, $\theta_1 = 0.3$, $\rho = 0.3$ and 0.9.

when $\delta = 0.45$, the power is uniformly equal to 1 with respect to the sample size, within the considered range of values for n (40–1000).

The case of sparse data, $\theta_1 = 0.3$, was also compared to the case of non-sparse data $\theta_1 = 0.8$, all the other parameters remaining unchanged with respect to the setting of Figure 2. As reported in Figure 6, the rejection rates when $\theta_1 = 0.8$ are uniformly greater. Therefore, even if the test performance is good in both cases, we can confirm that the situation of sparse data is more difficult to deal with because the probability that the test will make the correct decision by rejecting the null hypothesis when it is false is lower. Indeed, the power is non monotonic function of θ_1. Specifically, it is a convex function of θ_1 and it reaches the minimum when θ_1 is approximately 0.5. In fact, the variance of the underlying Bernoulli distribution, for population 1 and population 2, is equal to $\theta_1(1 - \theta_1)$ and $(\theta_1 + \delta)(1 - \theta_1 - \delta)$ respectively. Those correspond to the parabolic functions with maximum at $\theta_1 = 0.5$ and $\theta_1 = 0.5 - \delta = 0.45$ respectively. Hence, since the

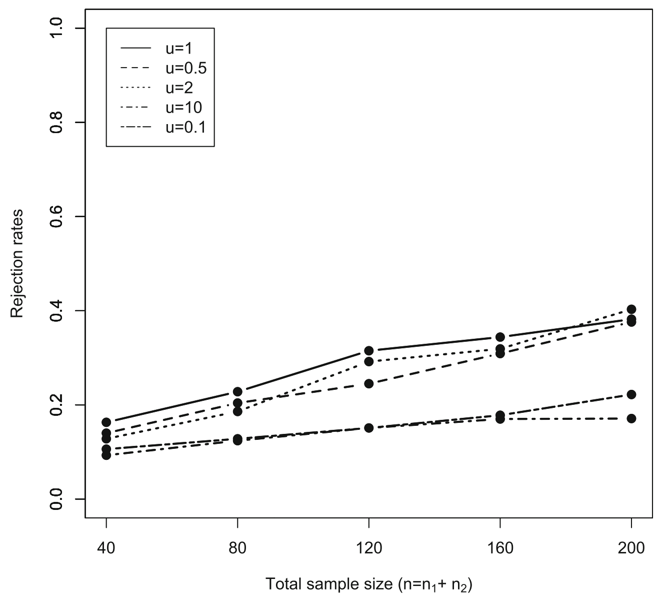

FIGURE 8 Rejection rates of CPT with combination of Fisher versus sample size under the alternative hypothesis, with $k = 10$, $p = 0.2$, $\delta = 0.05$, $\theta_1 = 0.3$, $\rho = 0.3$ and various values of u.

maximum of the variance corresponds to the minimum of the power, the thesis follows. The simulations, extended to the cases $\theta_1 = 0.1$, $\theta_1 = 0.5$, $\theta_1 = 0.7$, and $\theta_1 = 0.9$, prove such property (see Table A5 in the Appendix).

In the plots of Figure 7 it is instead possible to evaluate how the correlation ρ, and therefore the level of dependence between the variables, can affect the capacity of the test to recognize the true hypothesis when such hypothesis is the alternative one. When the correlation between the gaussian variables used to randomly generate the simulated data is strong ($\rho = 0.9$), the rejection rates are lower for every sample size. The stronger the correlation between variables, the lower the marginal informative contribution of each variable (i.e., of each partial test) to the general testing problem. Moreover, it seems that also the slope of the power function is affected by ρ. In fact, when ρ is high, the slope is low, hence the increase rate of the power with respect to the sample size is less than when ρ takes small values. In other words, a strong correlation implies that the convergence of the power to one when n diverges, due to the consistency of the test, is slower. This is confirmed by Table A6 in the Appendix, where we added $\rho = 0.1$, $\rho = 0.5$, and $\rho = 0.7$ to the considered settings.

Finally, Figure 8 shows the relationship between the power of the test and the imbalance ratio u. The power is maximum in the balanced design ($u = 1$). As the imbalance becomes more pronounced, that is, the ratio moves away from 1 with the same total sample size n, the power decreases, even if when the imbalance is very accentuated, the test keeps a good power. In general, the effect of u is symmetric with respect to the balanced case that is, the power behavior, when the ratio is u and when the ratio is $1/u$ with the same n, tends to be equivalent. In other words, if the imbalance is in favor of n_1 or n_2 it doesn't matter. For example, the rejection rates when $u = 2$ are lower than those of the balanced case and they are greater than when $u = 10$. However, the rejection rates when $u = 2$ and when $u = 0.5$ are very similar. The same conclusion holds when $u = 10$ and when $u = 0.1$.

5 | APPLICATIONS

This section deals with case studies concerning the Circular Economy. The data refer to Italian SMEs (Small and Medium Enterprises) operating in different sectors. We focused our attention on old firms, that is, firms with more than six years of activity, because more involved and well disposed towards CE activities than young firms,[38] and because they are much more numerous and characterizing the reference population of Italian SMEs. The dataset is original and related to a sample survey carried out in January 2020, by interviewing a random sample of Italian SMEs on the topic of CE. In particular, we want to contribute to the empirical literature about the effect of company's size on the propensity towards CE both in application and methodological terms.

The goal is to compare *small* and *medium* enterprises to test whether the propensity to undertake CE activities of the former is less than that of the latter. The propensity towards CE, related to a specific CE practice, of one population of enterprises can be defined as the proportion of companies of the population adopting that specific practice. The null hypothesis consists of the equality of the two proportions of CE enterprises (in the populations of small and medium enterprises) for all the six considered activities. The alternative hypothesis is that, for at least one activity, the proportion of CE companies in the population of small enterprises is less than the proportion of CE companies in the population of medium enterprises. The practices investigated in this study are the following:

1. reduction of the use of raw materials,
2. reduction of emitted waste (per unit of produced output),
3. reduction of waste in the production cycle,
4. conferral of waste to other companies to be used in the production cycle,
5. change in the design of products to maximize their recyclability,
6. investments in R&D aimed at reducing the environmental impact of production.

Each CE activity in the list corresponds to a dichotomous variable that takes one when a firm adopts such a practice and zero otherwise. Hence, we are in the presence of a multivariate binary response variable with six components and a two-sample multivariate testing problem on proportions with directional alternative hypothesis. The factor, or symbolic "treatment", of the problem consists in firms' size measured in terms of number of employees. The two levels of the factor are:

1. small-enterprise (to be distinguished from the micro-enterprise): a SME with more than nine but less than or equal to 49 employees,
2. medium enterprise: a SME with more than 49 employees.

Specifically, we focused on the following three different economic sectors: (1) textile industries and manufacture of clothing items (leather and fur), (2) manufacture of leather and similar items, (3) manufacture of rubber and plastic materials. Each sector corresponds to a different case study. The reasons to carry out three within-sector studies are mainly two. First of all, the need to compare similar companies, except for the size, given that the sector of economic activity in which the firm operates is an obvious confounding factor. The second reason lies in the need to make sector-specific assessments, considering each of these three strategic sectors for the Italian economic system separately, also in the perspective of possible policy implications. The significance level of the test is $\alpha = 0.10$. In case of significance of the general test, in order to attribute such significance to some specific variables, the p-values of the partial tests were adjusted with the MinP method[24] to control the family wise error rate and prevent the type I error rate from exceeding α.

5.1 | First case study

In the first case study, we took into account the sector *textile industries and manufacture of clothing items (leather and fur)*, according to the ATECO 2007 classification of economic activities. In 2019, the European Commission has identified the textile, clothing and fabrics sector as a priority product category within CE. A more circular textile system could contribute to the achievement of both EU and global sustainability objectives and consequently contribute to the attainment of a number of climate, environment and waste goals.[39]

The samples consisted in 259 small and 24 medium enterprises. In Table 1 the sample proportions of "circular" companies for each group of enterprises and each CE practice are reported. The proportions range from 0.069 (small enterprises that reduce waste in the production cycle) to 0.333 (medium enterprises that reduce the use of raw materials and the emitted waste). For all six activities, the proportion of CE firms among small enterprises is less than the proportion of CE firms among medium enterprises. The p-value of the CPT with Fisher combination is equal to 0.060, which is less than α and provides empirical evidence in favor of the hypothesis that the propensity to CE of small enterprises is less than that of medium enterprises. According to the adjusted p-values, the significance of the global test should be attributed to the difference of the proportions of companies that reduce the use of raw materials. For all the other variables, the differences in the proportions are not significant.

5.2 | Second case study

In the second case study, we examined the sector *manufacture of leather and similar items*, according to the ATECO 2007 classification of economic activities. In the literature, studies exploring CE practices in such a sector are scarce. One

TABLE 1 Sample proportions of CE companies in the groups of small and medium enterprises and *p*-values of the partial one-sided two-sample test for proportion comparison.

Practice	Proportion Small ent.	Proportion Medium ent.	Prop. diff. Small-medium	*p*-value	Adj. *p*-value
Reduction use of raw materials	0.143	0.333	−0.190	0.013	**0.066**
Reduction of emitted waste	0.158	0.333	−0.175	0.024	0.107
Reduction of waste in prod. cycle	0.069	0.167	−0.098	0.065	0.199
Waste to other companies	0.104	0.208	−0.104	0.078	0.234
Change design to max recyclab.	0.104	0.167	−0.063	0.178	0.245
R&D reducing envir. impact	0.089	0.167	−0.078	0.120	0.245

Note: Sector: *textile industries and manufacture of clothing items (leather and fur).*

TABLE 2 Sample proportions of CE companies in the groups of small and medium enterprises and *p*-values of the partial one-sided two-sample test for proportion comparison.

Practice	Proportion Small ent.	Proportion Medium ent.	Prop. diff. Small-medium	*p*-value	Adj. *p*-value
Reduction use of raw materials	0.059	0.333	−0.274	0.002	**0.010**
Reduction of emitted waste	0.127	0.467	−0.340	0.002	**0.009**
Reduction of waste in prod. cycle	0.051	0.200	−0.149	0.038	**0.092**
Waste to other companies	0.076	0.133	−0.057	0.231	0.328
Change design to max recyclab.	0.059	0.133	−0.074	0.170	0.328
R&D reducing envir. impact	0.051	0.200	−0.149	0.038	**0.092**

Note: Sector: *manufacture of leather and similar items.*

of the main reasons is that the lack of financial support does not help the successful implementation of CE activities. In general, big enterprises have the best financial basis to implement the reforms necessary to achieve the objectives of CE.[40]

The samples consisted in 118 small and 15 medium enterprises. In Table 2, sample proportions of firms that undertake the specific listed CE activities within the groups of small and medium enterprises are shown. The proportions range from 0.051 (small enterprises that reduce waste in the production cycle and invest in R&D with the aim of reducing the environmental impact of production) to 0.467 (medium enterprises that reduce the emitted waste). For every activity, the proportion of small enterprises is smaller than the proportion of medium enterprises. The combined *p*-value of the test is equal to 0.008, therefore again the general null hypothesis is rejected in favor of the alternative hypothesis that the propensity to CE of small enterprises is less than that of medium enterprises. According to the adjusted *p*-values, the significance of the global test must be attributed to the difference of proportions of companies that reduce the use of raw materials, reduce the emitted waste, reduce waste in the production cycle and invest in R&D for reducing the environmental impact of production. For the other variables, the differences in the proportions are not significant.

5.3 | Third case study

The third case study is dedicated to the sector *manufacture of rubber and plastic materials*, according to the ATECO 2007 classification of economic activities. This sample consisted in 166 small and 22 medium enterprises. The sample proportions of CE small enterprises and CE medium enterprises are included in Table 3. The proportions range from 0.090 (small enterprises that reduce waste in the production cycle and invest in R&D with the aim of reduce the environmental impact of production) to 0.364 (medium enterprises that reduce the use of raw materials). The differences between the sample proportions of CE small and medium enterprises are not negative for all the sectors. The combined

TABLE 3 Sample proportions of CE companies in the groups of small and medium enterprises and *p*-values of the partial one-sided two-sample test for proportion comparison.

Practice	Proportion Small ent.	Proportion Medium ent.	Prop. diff. Small-medium	*p*-value	Adj. *p*-value
Reduction use of raw materials	0.223	0.364	−0.141	0.085	0.315
Reduction of emitted waste	0.247	0.227	0.020	0.565	0.816
Reduction of waste in prod. cycle	0.235	0.318	−0.083	0.198	0.572
Waste to other companies	0.211	0.227	−0.016	0.415	0.816
Change design to max recyclab.	0.127	0.045	0.082	0.851	0.862
R&D reducing envir. impact	0.090	0.182	−0.092	0.115	0.385

Note: Sector: *manufacture of rubber and plastic materials.*

p-value of the general test is equal to 0.293, which is greater than α and indicates no significance. Given that there is not empirical evidence in favor of the alternative hypothesis, there is no need to assess the significance of partial tests.

6 | CONCLUSIONS

The proposed nonparametric test based on the *CPT* methodology is a valid solution for the multivariate extension of the two-sample test on proportions or, equivalently, two-sample test for multivariate Bernoulli distributions. In particular, we developed a solution for the challenging test on simultaneous marginal homogeneity versus one-sided alternative hypotheses. The Monte Carlo simulation study proved the good performance of the test in terms of power. In particular, the test was proved to be well approximated, unbiased, consistent and powerful. An important finding concerns the validity for small samples. Such a solution is based on the permutation approach, hence it is distribution free and doesn't assume a specific asymptotic null distribution of the test statistic, thus it is applicable (and powerful) also for small sample sizes. Another important advantage of such a method is that it is applicable to any sample size regardless of the number of variables. Moreover, the greater the number of variables (with the same proportion of variables under the alternative hypothesis and with the same sample sizes), the higher the power of the test. Nevertheless, the proportion of variables under H_1, regardless of the number of variables, does not affect the power of the test. Finally, the proposed permutation test is applicable and performant also in the presence of sparse data, even if in such a situation we observed a slight loss of power. The illustrated method is valid when exchangeability of data under the null hypothesis is satisfied. In the absence of exchangeability, the method is not applicable unlike other solutions not based on the permutation approach.

The application of the method to the case studies related to the recent survey on CE and SMEs carried out in Italy, showed its utility in testing the effect of firm's size on the propensity towards CE in a multivariate framework for specific strategic economic sectors. The lower "circularity" of small enterprises with respect to medium enterprises was found in two of these sectors (textile industries and manufacture of leather). Such evidence was not confirmed in the sector regarding manufacture of rubber and plastic materials.

Definitely, the proposed permutation test represents a powerful and robust methodological solution for the two-sample test for multivariate binary data with one-tailed alternative hypotheses. Such kind of data are very common when dealing with empirical studies connected to sample surveys with a questionnaire that includes a list of questions with only two possible answer alternatives (two-choice questions). For instance, this is typical of questionnaires aimed at defining the circularity of companies based on the implementation of one or more specific business practices. As far as we know, at the moment there are no competitors among both parametric and nonparametric tests. Furthermore, we propose tools and good practices, for managers and policymakers, for commodity risk management, that consists in a multivariate statistical approach based on a suitable questionnaire and an appropriate inferential method, for assessments about CE feasibility. Finally, we provide a scientific contribution on the empirical literature about CE, in particular about the effect of firm size on the tendency of enterprises to be circular. Such comparative assessments are important to foster the creation of a circular economy that allows for successfully dealing with commodity supply disruptions, in particular for small and medium enterprises.

AUTHOR CONTRIBUTIONS

The authors thank the Associate Editor and the anonymous Referees for the valuable comments and suggestions that have contributed to significantly improve the quality of the paper.

ACKNOWLEDGMENTS

The authors wish to thank the Italian Ministry of Education, University Research that funded the departmental development program (DEM – University of Ferrara) for the period 2018–2022, to promote excellence in education and research ("Dipartimenti di Eccellenza"). This paper represents an output of such research funding initiative.

CONFLICT OF INTEREST STATEMENT

All authors have no financial/commercial conflict of interest of any kind to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Massimiliano Giacalone https://orcid.org/0000-0002-4284-520X

REFERENCES

1. Bressanelli GM, Perona M, Saccani N. Reshaping the washing machine industry through the circular economy and product-service system business models. *Procedia CIRP*. 2017;64:43-48.
2. Elia V, Gnoni MG, Tornese F. Measuring circular economy strategies through index methods: a critical analysis. *J Clean Prod*. 2017;142(4):2741-2751.
3. Winans K, Kendall A, Deng H. The history and current applications of the circular economy concept. *Renew Sustain Energy Rev*. 2017;68:825-833.
4. Cristoni N, Tonelli M. Perception of firms partecipating in a circular economy. *Eur J Sustain Dev*. 2018;7(4):105-118.
5. Lieder M, Rashid A. Towards circular economy implementation: a comprehensive review in context of manufacturing industry. *J Clean Prod*. 2016;115:36-51.
6. Michelini G, Moraes RN, Cunha RN, Costa JMH, Ometto AR. From linear to circular economy: PSS conducting the transition. *Procedia CIRP*. 2017;64:2-6.
7. Zhu J, Fan C, Shi H, Shi L. Efforts for a circular economy in China. A comprehensive review of policies. *J Ind Ecol*. 2018;23(1):110-118.
8. Ghisellini P, Ulgiati S. Circular economy transition in Italy. Achievements, perspectives and constraints. *J Clean Prod*. 2020;243:118360. doi:10.1016/j.jclepro.2019.118360
9. Sreedevi SC, Salahudeen RM. Sustainability route for industry 4.0: the future of global circular economic transition. *Circular Economy-Recent Advances, New Perspectives and Applications*. IntechOpen; 2021.
10. Pesarin F. *Multivariate Permutation Tests with Applications in Biostatistics*. Wiley; 2001.
11. Davidov O, Peddada S. Order-restricted inference for multivariate binary data with application to toxicology. *J Am Stat Assoc*. 2011;106(496):1394-1404.
12. Agresti A, Klingenberg B. Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. *Appl Stat*. 2005;54(4):691-706.
13. Chuang-Stein C, Mohberg N. A unified approach to the analysis of safety data in clinical trials. In: Sogliero-Gilbert G, ed. *Drug Safety Assessments in Clinical Trials*. Dekker; 2019.
14. Agresti A. Testing marginal homogeneity for ordinal categorical variables. *Biometrics*. 1983;39:505-510.
15. Gao W, Kuriki S. Testing marginal homogeneity against stochastically ordered marginals for $r \times r$ contingency tables. *J Multivar Anal*. 2006;97:1330-1341.
16. Klingenberg B, Solari A, Salmaso L, Pesarin F. Testing marginal homogeneity against stochastic order in multivariate ordinal data. *Biometrics*. 2008;65:452-462.
17. Bonnini S, Prodi N, Salmaso L, Visentin C. Permutation approaches for stochastic ordering. *Commun Stat Theory Methods*. 2014;43(10):2227-2235.
18. Yang Z, Sun X, Hardin J. Testing marginal homogeneity in clustered matched-pair data. *J Stat Plan Inference*. 2011;141:1313-1318.
19. Deng B, Carriere KC. Testing simultaneous marginal homogeneity for clustered matched-pair multinomial data. *Int J Stat Prob*. 2018;7(5):86-94.
20. Kang J, Zhang N, Shi R. A Bayesian nonparametric model for spatially distributed multivariate binary data with application to a multidrug-resistant tuberculosis (MDR-TB) study. *Biometrics*. 2014;70(4):981-992.
21. Davidov O. Ordered inference, rank statistics and combining p-values: a new perspective. *Stat Methodol*. 2012;9(3):456-465.

22. Davidov O, Peddada S. The linear stochastic order and directed inference for multivariate ordered distributions. *Ann Stat*. 2013;41(1):1-40.
23. Davidov O, Peddada S. Testing for the multivariate stochastic order among ordered experimental groups with application to dose–response studies. *Biometrics*. 2013;69(4):982-990.
24. Pesarin F, Salmaso L. *Permutation Tests for Complex Data. Theory, Applications and Software*. Wiley; 2010.
25. Bonnini S, Corain L, Marozzi M, Salmaso L. *Nonparametric Hypothesis Testing. Rank and Permutation Methods with Applications in R*. Wiley; 2014.
26. Giacalone M, Alibrandi A. A non parametric approach for the study of the controls in the production of agribusiness products. *Electron J Appl Stat*. 2011;4(2):235-244.
27. Giacalone M, Zirilli A, Moleti M, Alibrandi A. Does the iodized salt therapy of pregnant mothers increase the children IQ? Empirical evidence of a statistical study based on permutation tests. *Qual Quant*. 2018;52(3):1423-1435.
28. Giacalone M, Agata Z, Cozzucoli PC, Alibrandi A. Bonferroni-holm and permutation tests to compare health data: methodological and applicative issues. *BMC Med Res Methodol*. 2018;18(1):1-9.
29. Alibrandi A, Giacalone M, Zirilli A. Psychological stress in nurses assisting amyotrophic lateral sclerosis patients: a statistical analysis based on non-parametric combination test. *Mediterranean J Clin Psychol*. 2022;10(2):1-40.
30. Bonnini S, Melak Assegie G. Advances on permutation multivariate analysis of variance for big data. *Stat Trans New Series*. 2022;23(2):163-183.
31. Bonnini S. Testing for heterogeneity with categorical data: permutation solution vs. bootstrap method. *Commun Stat: Theory Methods*. 2014;43(4):906-917.
32. Arboretti Giancristofaro R, Bonnini S. Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives. *Stat Methods Appl J Italian Stat Soc*. 2009;18(2):221-236.
33. Bonnini S, Borghesi M. Relationship between mental health and socio-economic, demographic and environmental factors in the COVID-19 lockdown period-a multivariate regression analysis. *Mathematics*. 2022;10(18):3237.
34. Bonnini S, Cavallo G. A study on the satisfaction with distance learning of university students with disabilities: bivariate regression analysis using a multiple permutation test. *Stat Appl Italian J Appl Stat*. 2021;33(2):143-162.
35. Giacalone M, Alibrandi A. Overview and main advances in permutation tests for linear regression models. *J Math Syst Sci*. 2015;5(2):53-59.
36. Corain L, Arboretti R, Bonnini S. *Ranking of Multivariate Populations: A Permutation Approach with Applications*. CRC Press; 2017.
37. Bonnini S, Borghesi M, Giacalone M. Advances on multisample permutation tests for "V-shaped" and "U-shaped" alternatives with application to circular economy. *Ann Oper Res*. 2023;1-16.
38. Bassi F, Dias JG. The use of circular economy practices in SMEs across the EU. *Conserv Recycl*. 2019;146:523-533.
39. ETC and EEA experts. *Textiles and the Environment in a Circular Economy*. *Eionet Report*. ETC/WMGE; 2019:7.
40. Moktadir MA, Ahmadi HB, Sultana R, Zohra FT, Liou JJH, Rezaei J. Circular economy practices in the leather industry: a practical step towards sustainable development. *J Clean Prod*. 2020;251:119737. doi:10.1016/j.jclepro.2019.119737

How to cite this article: Bonnini S, Borghesi M, Giacalone M. Simultaneous marginal homogeneity versus directional alternatives for multivariate binary data with application to circular economy assessments. *Appl Stochastic Models Bus Ind*. 2024;40(2):389-407. doi: 10.1002/asmb.2827

APPENDIX

This section includes the results of the extended simulation study in tabular form.

TABLE A1 Rejection rates of CPTs versus sample size under the null hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $u = 1$, $k = 5$ and 10.

n	*k*	Fisher	Tippett	*n*	*k*	Fisher	Tippett
40	5	0.043	0.050	40	10	0.048	0.060
80		0.047	0.046	80		0.049	0.051
120		0.057	0.048	120		0.046	0.043
160		0.046	0.052	160		0.044	0.042
200		0.047	0.044	200		0.048	0.029
1000		0.053	0.048	1000		0.050	0.052

TABLE A2 Rejection rates of CPTs versus sample size under the alternative hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $\delta = 0.05$, $k = 10$, $p = 0.2$ and $u = 1$.

n	Fisher	Tippett	*n*	Fisher	Tippett
40	0.148	0.100	1000	0.934	0.760
80	0.235	0.129	2000	0.998	0.960
120	0.278	0.154	2200	0.998	0.975
160	0.331	0.207	2400	1.000	0.989
200	0.432	0.269	4000	1.000	1.000

TABLE A3 Rejection rates of CPT based on Fisher's combination versus sample size under the alternative hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $\delta = 0.05$, $p = 0.2$, $u = 1$, $k = 5, 10, 15, 20, 25$.

n	*k*	Power	*k*	Power	*k*	Power	*k*	Power	*k*	Power
40	5	0.122	10	0.148	15	0.167	20	0.156	25	0.161
80		0.209		0.235		0.279		0.260		0.243
120		0.249		0.278		0.305		0.337		0.345
160		0.288		0.331		0.343		0.371		0.400
200		0.350		0.432		0.443		0.490		0.464
1000		0.880		0.934		0.954		0.961		0.966
2000		0.988		0.998		0.999		1.000		1.000
2200		0.994		0.998		0.999		1.000		1.000
2400		0.995		1.000		1.000		1.000		1.000
3200		0.998		1.000		1.000		1.000		1.000
3400		1.000		1.000		1.000		1.000		1.000

TABLE A4 Rejection rates of CPT based on Fisher's combination versus sample size under the alternative hypothesis, with $\rho = 0.3$, $\theta_1 = 0.3$, $k = 10$, $u = 1$, $p = 0.2$, $\delta = 0.05, 0.15, 0.25, 0.35, 0.45$.

n	δ	Power	δ	Power	δ	Power	δ	Power	δ	Power
40	0.05	0.148	0.15	0.578	0.25	0.898	0.35	0.995	0.45	0.999
80		0.235		0.838		0.997		1.000		1.000
120		0.278		0.939		1.000		1.000		1.000
160		0.331		0.977		1.000		1.000		1.000
200		0.432		0.992		1.000		1.000		1.000
1000		0.934		1.000		1.000		1.000		1.000

TABLE A5 Rejection rates of CPT based on Fisher's combination versus sample size under the alternative hypothesis, with $\rho = 0.3$, $k = 10$, $u = 1$, $p = 0.2$, $\delta = 0.05$, $\theta_1 = 0.1, 0.3, 0.5, 0.7, 0.9$.

n	θ_1	Power	θ_1	Power	θ_1	Power	θ_1	Power	θ_1	Power
40	0.1	0.253	0.3	0.148	0.5	0.134	0.7	0.156	0.9	0.371
80		0.379		0.235		0.218		0.233		0.599
120		0.509		0.278		0.287		0.291		0.725
160		0.602		0.331		0.335		0.383		0.848
200		0.716		0.432		0.365		0.453		0.895
1000		0.998		0.934		0.916		0.951		1.000
2000		1.000		0.998		0.995		0.999		1.000
2200		1.000		0.998		0.998		0.999		1.000
2400		1.000		1.000		0.997		1.000		1.000

TABLE A6 Rejection rates of CPT based on Fisher's combination versus sample size under the alternative hypothesis, with $k = 10$, $u = 1$, $p = 0.2$, $\delta = 0.05$, $\theta_1 = 0.3$, $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$.

n	ρ	Power	ρ	Power	ρ	Power	ρ	Power	ρ	Power
40	0.1	0.191	0.3	0.148	0.5	0.140	0.7	0.112	0.9	0.111
80		0.310		0.235		0.174		0.161		0.129
120		0.394		0.278		0.235		0.222		0.165
160		0.500		0.331		0.297		0.241		0.186
200		0.590		0.432		0.313		0.258		0.211
1000		0.992		0.934		0.847		0.697		0.626
2000		1.000		0.998		0.984		0.945		0.878
2200		1.000		0.998		0.987		0.965		0.884
2400		1.000		1.000		0.997		0.975		0.917