Università degli Studi di Ferrara

PH.D. COURSE IN
ECONOMICS AND MANAGEMENT OF
INNOVATION AND SUSTAINABILITY

In cooperation with Università degli studi di Parma

CYCLE XXXVI

Coordinator Prof. Emidia Vagnoni

Methodological solutions based on Combined Permutation Tests with application to sustainability problems

Scientific/Disciplinary Sector (SDS) SECS-S/01

Candidate	**Supervisor**
Dott. Borghesi Michela	Prof. Bonnini Stefano

YEARS 2020 – 2023

Contents

Abstract

This thesis concerns the proposal of effective statistical methods for hypothesis testing problems, in particular for environmental sustainability problems. The proposed methodological solution belongs to the family of Combined Permutation Tests (CPTs). This method consists of carrying out a set of partial tests, and then combining the p-values of the partial tests in order to determine an overall combined test statistic. The decomposition of the general testing problem into partial sub-problems is at the basis of the CPTs. Through this statistical methodology, it is possible to take into account potential confounding factors, through the stratification procedure. Another advantage of the approach proposed extends to the case of multidimensional variables and to variables that follow the Bernoulli distribution. The good performances of this statistical approach are provided through some Monte Carlo simulation studies, specifically created for each methodological problem. From an application point of view, the first part is dedicated to the application of CPTs to solve complex tests of hypotheses, in particular non-directional hypotheses (such as the V-shaped or inverted V-shaped relationship). Hence, the characteristics of Italian Small and Medium Enterprises such as age, size, or business sector, and their propensity or reluctance to undertake activities focused on Circular Economy are considered. The second part concerns the application of CPTs to evaluate time series related to ESG titles. Through this statistical approach, it is possible to consider the comparison between two time series. The basic procedure concerns the decomposition of the general testing problem into a partial problem for all the time points of the time series. From an application point of view, two groups of titles (ESG and non-ESG) are compared in terms of returns relative variations in five years. The last part provides an application of combined permutation tests to multivariate regression analysis. Such methodology is suitable to test the significance of the estimates of the regression coefficients of all the equations jointly considered. Concerning the application, the characteristics of Italian SMEs such as age and size, together with variables that represent the "circularity" of these companies are the explanatory variables of the model to predict the economic performances of companies (i.e. Net Working Capital, Revenues, ROA,...). The work results are reflected in many similar problems presented

and studied in the literature.

Introduction

In this PhD thesis, the underlying concept concerns the proposal of effective statistical methods for hypothesis testing problems. In particular, these methodologies are applied to environmental sustainability problems.

It is crucial to monitor both material consumption and environmental impacts, and it is essential to establish targets for the effective implementation of a Circular Economy (CE) paradigm. The monitoring outcomes and objectives should not be significant to policymakers, but they should also encourage industry and the general public to take action (Haupt and Hellweg, 2019). Circular Economy is seen as a crucial model for industries to achieve sustainable development. Due to its importance, CE has been widely researched. Many authors have assessed the effect of a CE on the natural environment, while others have examined its impact on a broader view of sustainability. This includes not only environmental impact but also social and economic aspects. In some literature, there are four common factors driving research about CE: the active debate about CE's role in reaching sustainability, the interdisciplinary background of scholars, the development agendas supported by CE and the importance of financial support. There are many similarities between CE and sustainability, including integrating non-economic aspects into development, the need for cooperation among different stakeholders, using regulations and incentives as core implementation tools, and business model innovation for industry transformation. Circularity is often seen as a path towards sustainability. According to some researchers, circular economy is considered a necessity for sustainability or a trade-off in literature. Meanwhile, other authors have identified sustainability as a goal that can be achieved through a circular economy model (Hysa et al., 2020).

There are several criteria for evaluating sustainability disclosure, with social responsibility, economic sustainability, and environmental sustainability being key dimensions. As sustainability disclosure has progressed, so has awareness of ESG, community initiatives and global reporting initiatives (Bosi et al., 2022). The growing focus on environmental sustainability has led to a push towards adopting circular economy principles. However, this transition requires businesses to redesign their systems and modify their processes. By embracing disruptive technologies, companies can experience several benefits such as improved resource utilization and productivity. As a result, organizations are increasingly pressured to adopt and maintain digital supply chains to remain competitive in global markets (Dwivedi and Paul, 2022).

The thesis is outlined in three main chapters in which various aspects of environmental sustainability are explored deeply. In particular, we will talk about environmental sustainability regarding Circular Economy/circularity (as mentioned before), ESG titles and finally circularity versus company performances.

Taking the ESG aspect into consideration, it is well known that ESG scores are increasingly used to evaluate sustainability in organizations, encompassing environmental, social, and governance factors. Furthermore, it has become increasingly evident from recent literature that ESG scores are not an accurate measure of sustainability in terms of sustainable development. The main reason for this inadequacy is that ESG scores are not designed to measure sustainability concepts like temporality, impact, resource management, and interconnectivity. ESG scores are based on the idea that external firms would develop a methodology using scientific consensus, assumptions, and other external metrics. These different metrics would then be combined to provide an evaluation of businesses based on each pillar of E, S, and G. The ESG scores are an aggregation of measures from each pillar, used to propose a metric that represents how well an organization performs on the issue that ESG scores represent. The primary objective of ESG scores is to identify companies that perform better than others in terms of sustainability. However, scholars use ESG scores to compare the sustainability of organizations against one another, along with other metrics, to make assumptions about their sustainability. In a world where climate change, social challenges, and environmental issues are becoming increasingly dominant, it is not possible to rely only on ESG scores to evaluate the sustainability of an organization (Clément et al., 2022). According to a panel data regression conducted by Godínez-Reyes et al. (2022), generating sustainable value in firms has a positive effect on their corporate efficiency. The study found that Corporate Governance (G) is the most important factor in generating value. Interestingly, the study revealed that social performance (S) is the least valued dimension for the generation of sustainable value, both in theory and in practice. Despite social performance being the dimension with the greatest number of indicators in most internationally used models to assess firms' sustainability, it is often the least valued when deciding business sustainability strategies. The research findings suggest that socially responsible companies that create sustainable value are more profitable, making it a sound business strategy which large companies are increasingly adopting (Godínez-Reyes et al., 2022).

The third theme concerns the aspect of the economic performance of a firm. In some studies, a positive relationship was found between sustainability performance and financial performance, while a negative relationship was found in other studies. Furthermore, some studies did not find a significant relationship between these two aspects. Many investors and customers appreciate and respect firms that fully embrace corporate social responsibility (Park and Lee, 2018). At the same time, implementing sustainable practices can improve financial and corporate performance (Algarni et al., 2022). In the study of Kılıç et al. (2022), the authors used four proxies for financial performance to examine the effect of sustainability performance on financial performance indicators in two countries. The proxies were: Return on Asset (ROA), Return on Equity (ROE), Return on Sales (ROS), and market-to-book value ratio (MV/BV). The analysis revealed that sustainability performance has a significant impact on the return on assets (ROA). These findings are consistent with similar studies, which also found a positive correlation between sustainability performance and ROA. These results demonstrate that the relationship between sustainability performance and financial performance could exist between companies (Kılıç et al., 2022; Marota et al., 2023).

Obviously, in all three topics of this thesis, the literature has found a need to deepen the study of these aspects. Therefore one of the objectives of this work is precisely to provide an empirical contribution to the existing literature but also to offer a valid methodological solution capable of managing these types of problems related to environmental sustainability from a statistical point of view.

The methodological approach studied and presented in this PhD thesis concerns the family of combined permutation tests (CPTs) (Pesarin, 2001; Pesarin and Salmaso, 2010). This method consists of carrying out a set of partial sub-problems of the general test of the hypothesis under study, and then combining the p-values of the partial tests to determine a unique combined test statistic reasonable for the general problem. The CPT methodology is used in various application problems (Alibrandi et al., 2022; Giacalone and Alibrandi, 2011), for categorical data (Bonnini, 2014), big data (Bonnini and Melak Assegie, 2022), regression analysis (Giacalone and Alibrandi, 2015) and other problems. CPTs are part of the family of non-parametric statistics. This type of non-parametric approach has been studied a lot in the literature such as concerning ranking problems (Alvo and Cabilio, 1995; Philip et al., 2002), one-way and two-way factorial designs (Gao et al., 2008; Gao and Alvo, 2008) and also in non-directional alternatives such as the umbrella alternative (Alvo, 2008).

The dissertation is structured in three main chapters which correspond to three papers as illustrated below. Chapter 1 concerns the application of the proposed methodology to complex problems related to a Circular Economy application. The part related to the application of the statistical method proposed to compare time series is presented in Chapter 2. Chapter 3 presents an application linked to the relationship between the implementation of CE activities and the economic/financial performances of firms. Finally, there are the overall conclusions that summarize the main results and the future implications.

Chapter 1

Combined permutation tests for complex problems with applications to Circular Economy

This chapter is dedicated to the application of combined permutation tests (CPTs) to solve complex tests of hypotheses, in particular non-directional hypotheses. Through this statistical methodology, it is possible to take into account potential confounding factors, through the stratification procedure. In fact, the decomposition of the general testing problem into partial sub-problems is at the basis of the CPTs. Furthermore, another advantage of the approach proposed in this chapter extends to the cases of multidimensional variables (very common in a great variety of applied problems) and to variables that follow the Bernoulli distribution.

From an application point of view, the characteristics of Italian Small and Medium Enterprises such as age, size, or business sector, and their propensity or reluctance to undertake activities focused on Circular Economy are taken into consideration. Hence, the main research question of this chapter can be formulated as: *How do some characteristics of SMEs affect their propensity or reluctance to circularity?*

1.1 Introduction

Nowadays, the term Circular Economy (CE) is spreading very rapidly. This term first appeared in the 1960s and since then, there has been a rapid increase in its use in all its applications. Already Boulding (1966) presented the idea of a circular circuit of materials. Later, Stahel and Reday-Mulvey (1981) outlined the vision of a circular economy and its impact on job creation, resource savings and waste reduction. The concept of Circular Economy (CE) is based on the pillars of the

3Rs: Reduce, Reuse and Recycle. One of the objectives of this decade is precisely to promote sustainable development in order to reverse the inescapable course of man-made climate change. Thanks to the prudent use of resources and the recovery of materials, it is possible to limit and reduce waste from production processes. One of the practical approaches which hence makes the CE crucial for this global goal concerns the revolution of the economic system to replace the "end-of-life" concept with the reduction and reuse of materials (Austin and Rahman,2022). The foundation of this construct is that a linear economic model (that is of the type take-make-dispose) is no longer sustainable, especially with regard to the purchasing of materials. On the other hand, we have a circular economic system, which is considered one of the most effective solutions to environmental sustainability problems (Cristoni and Tonelli, 2018; Lieder and Rashid, 2016). For the effective implementation of CE, a re-design of products and materials for reuse in the production cycle is essential. In this regard, product and process innovations play a key role in achieving sustainable development goals. So the circular economy turns out to be a pivotal principle to allow companies to achieve a good level of innovation to realize these goals. For this reason, statistical inference, as well as sample surveys, are fundamental to investigating the position of companies in relation to circular economy practices. More and more companies have recently adopted the concept of a transition to the circular economy (CE) (Bressanelli et al., 2017; Elia et al., 2017; Winans et al., 2017). Since 2015, Italy has also taken part in this practice, particularly as regards the separate collection of waste (Ghisellini et al., 2020).

From a business perspective, the circular economy has spread particularly in larger companies and therefore has difficulty spreading in Small and Medium Enterprises (SMEs) (Ormazabal et al., 2018; Williamson et al., 2006; Garrido-Prada et al., 2021). What most likely influenced this fact is that, unlike large companies, SMEs have moderate financial flows, less business practice and expertise, and a lack of adequate technology as well as less support from public institutions (Rizos et al., 2016; Malik et al., 2022; Saidani et al., 2019; Schot and Steinmueller, 2018). With respect to SME attitudes toward environmental practices, there is a widely held view in the literature. In particular, the reluctance to circularity may be due to the economic organization because the switch from a linear to a circular view means reorganizing the entire company (Kaimal et al., 2020). The other important point that can determine the success or not of companies to achieve the ambition of being circular concerns consumers. For example, Singhal et al. (2019) claims that consumers are reluctant to buy recycled products for reasons of quality, efficiency, security and maintenance.

Looking specifically at the characteristics of enterprises, in literature it is widely believed that age and size of the company are the main important factors that can influence their circularity. Indeed, the size of the enterprise is a key factor in determining the extent of CE practices (Gennari, 2022). On one side, the scarcity of resources and the increased concerns of smaller companies could influence their behaviour in a negative way (Neubaum et al., 2004). On the other side, authors such as Hockerts and Wüstenhagen (2010) argue that smaller companies would be more likely to be involved in environmental practices. In the middle, there is the point of view of Hoogendoorn and Guerra (2015). They support the thesis that mid-sized SMEs are more likely to implement environmental practices linked to circular economy, suggesting an inverted "U-shaped" relationship between firm size and circularity.

Another point of view is that the relationship between the size of the company and the magnitude of the attitude towards environmental sustainability strategies and circularity is monotonic increasing (Brammer and Pavelin, 2008; Bassi and Dias, 2020; McIntyre and Ortiz, 2016; Brondoni, 2020; Tagliaferro, 2020). As a result, literature and empirical studies on the relationship between firm size and CE propensity or reluctance have not yet determined whether the size effect is increasing, decreasing or not monotonous. Thus, such a non-monotonic effect of company size on circularity is possible and represents an interesting hypothesis to test in order to deepen this relationship. Another important factor affecting the trend towards a circular economy is the age of the firm. (Bassi and Dias, 2019; Ghisetti and Montresor, 2020; Yadav et al., 2018). It is well known that older enterprises are generally in a better position to participate in CE activities than younger enterprises.

Since in the literature it is necessary to deepen these aspects of reluctance (or propension) to the practices of Circular Economy, we want to contribute in this sense especially from the statistical point of view by addressing techniques for the resolution of complex tests of hypotheses. Such assumptions are the non-monotonous relationships between factors as mentioned above. These hypotheses assume a growing and then decreasing relationship or vice versa. Another crucial aspect is the multidimensionality that characterizes CE, in fact, the propensity (or reluctance) to be circular involves more aspects such as water saving, the reuse of waste in the production process, and so on. Moreover, if we want to analyze in what terms the characteristics of companies (such as size or age) affect circularity, it is essential to take into account possible confounding factors. Confounding factors (or from a statistical point of view, stratification factors) are variables capable of producing an apparent association (or non-association) between the phenomena to be studied. If they are not taken into account, there is a risk of obtaining false results or, in general, non-reliable inferential results.

1.2 Combined permutation tests for V-shaped or inverted V-shaped alternatives

1.2.1 Introduction

In Section 1.1 it was highlighted that the hypotheses to be tested regarding the reluctance or propensity to the Circular Economy are not always monotonic (i.e. increasing or decreasing stochastic ordering). Indeed, some articles in the literature consider the possibility of a non-monotonic "V-shaped" or "inverted V-shaped" relationship (or similarly "U-shaped" or "inverted U-shaped" relationship). For this reason, this section focuses on the non-monotonous relationship between the size of the company (generally divided into micro, small and medium enterprises) and the propensity (or reluctance) to adopt CE practices.

From a statistical point of view, the aim is to achieve inferential methods that are suitable for this type of complex hypothesis. The difficulty concerns a testing problem when the relationship between the response and the given factor (i.e. treatment) follows a V-shaped pattern, so decreases up to a peak point and then increases after the peak point. The case similar to the "V-shape" is the "U-shape" case, in which the only difference is that the peak point is not unique and there is a constant trend of the effect between the decreasing and increasing phase. Since this non-monotonic stochastic ordering represents the alternative hypothesis of the testing problem, the null hypothesis is equality in distribution of the compared populations. Wolfe (2006), Millen and Wolfe (2005) and Mack and Wolfe (1981) employed non-parametric solutions based on rank approach. Hence, in these testing problems, the test proposed is based on the Jonckheere-Terpstra statistic and furthermore, the alternative hypothesis is not completely specified (Mack and Wolfe, 1981; Pan, 1996). Regarding the peak point, a standard solution based on the maximum of the test statistics is used when this particular point is not indicated in H_1 (Mack and Wolfe, 1981; Hettmansperger and Norton, 1987; Chen and Wolfe, 1990; Shi, 1988; Hartlaub and Wolfe, 1999). Generally speaking, these tests are applicable only for specific α significance levels. Instead, authors such as Manly (1997) and Neuhäuser et al. (2003) started to think about solutions based on the permutation approach. Finally, Basso and Salmaso (2011) proposed an exact permutation test that is applicable for any α value.

The proposed solution concerns the application of a nonparametric method within the family of combined permutation tests (CPTs). In particular, it involves a permutation test based on multiple pairwise comparisons where the test rise to be exact, unbiased and powerful (Pesarin, 2001, Bonnini et al., 2014a; Pesarin and Salmaso, 2010; Bonnini et al., 2014b; Basso et al., 2009; Basso and Salmaso, 2011; Finos et al., 2007). Moreover the proposed method assumes that the overall problem could be broken down into sub-problems, and a parametric technique is not always suitable (Bonnini et al., 2014a). From an application standpoint, the case of Italian SMEs is analyzed. The number of innovations introduced for enforcing CE practices is considered and the three groups under comparison are the three categories of SMEs according to size: micro, small and medium enterprises. To our knowledge, the few empirical studies that look at the relationship between company size and innovation intensity for CE involve inappropriate methods: they ignore the existence of confounding factors, they are not robust with respect to the underlying distribution or they are not suitable for small sample sizes or for multivariate problems. The proposed method manages all these challenges and again, in case of significance of the overall test, it lets to attribute such significance to one or more partial tests.

1.2.2 Statistical problem

The goal consists of testing complex V or U-shaped hypotheses between a numerical response variable (such as the reluctance of SMEs to CE practices) and an ordinal factor with more than two levels (such as the size of the firm). The random variable representing the response for the j-th population is denoted by X_j, with $j = 1, \ldots, C$ and $C > 2$. In the null hypothesis H_0, there is a null treatment effect, so equality in distribution of the C random variables. On the other hand, the representation of the i-th observation in the j-th sample in the fixed effects model, can be considered as the realization of the random variable X_{ij} as, $X_{ij} = \mu + \delta j + \varepsilon_{ij}$, with $i = 1, \ldots, n_j$ and $j = 1, \ldots, C$. Hence, μ represents the overall mean and $\delta_1, \ldots, \delta_C$ represent the C specific treatment effects.

In the alternative hypothesis H_1 (Bonnini et al., 2014a; Basso and Salmaso, 2011), there is a non-increasing stochastic relationship up to a peak point (that corresponds to the factor level p) and then a non-decreasing stochastic relationship after the peak point (see Equation 1.1).

$$\begin{cases} H_0 : X_1 \overset{d}{=} X_2 \overset{d}{=} \ldots \overset{d}{=} X_C \\ H_1 : X_1 \overset{d}{\geq} \ldots \overset{d}{\geq} X_{p-1} \overset{d}{\geq} X_p \overset{d}{\leq} X_{p+1} \overset{d}{\leq} \ldots \overset{d}{\leq} X_C \end{cases} . \tag{1.1}$$

In other words, in terms of cumulative distribution functions, assuming that $F_j(x) = P(X_j \leq x)$ with $x \in \mathbb{R}$ and $j = 1, \ldots, C$, the null hypotheses could be represented as $F_1(x) = F_2(x) = \ldots = F_C(x)$ and the alternative hypothesis is $F_1(x) \leq \ldots \leq F_p(x) \geq \ldots \geq F_C(x)$.

The alternative hypothesis of non-monotonic relationship, i.e. the V-shaped case, could be represented as the intersection of two stochastic ordering alternatives (Pesarin and Salmaso, 2010) as follows:

$$H_1 : H_{1p}^{\searrow} \cap H_{1p}^{\nearrow}. \tag{1.2}$$

Besides to what has just been explained for the case of the V-shaped hypothesis, the U-shaped relation can be considered an extension of the V-shaped one, in which the peak point could be not unique and the alternative hypothesis is the following:

$$H_1 : X_1 \overset{d}{\geq} \ldots \overset{d}{\geq} X_{p-1} \overset{d}{=} X_p \overset{d}{=} X_{p+1} \overset{d}{\leq} \ldots \overset{d}{\leq} X_C \tag{1.3}$$

From now on, since the solution for the U-shaped case is easily obtainable from that of the V-shaped case, we will consider only the V-shaped problem.

Actually, to represent the inverted V-shaped hypothesis, it is sufficient to change the direction of the inequalities of the alternative hypothesis in Equation 1.1. We obtain (Chen and Wolfe, 1990; Hartlaub and Wolfe, 1999; Casella and Berger, 2002):

$$H_1 : X_1 \overset{d}{\leq} \ldots \overset{d}{\leq} X_{p-1} \overset{d}{\leq} X_p \overset{d}{\geq} X_{p+1} \overset{d}{\geq} \ldots \overset{d}{\geq} X_C \tag{1.4}$$

and also the inverted U-shaped case is obtained as described above.

Finally, it is worth noting that given $p \in \{1, 2, \ldots, k\}$, the monotonic stochastic ordering can be considered as a case limit of the general problem defined above. If $p = 1$ we have an increasing stochastic ordering, and when $p = k$ we are in the case of decreasing stochastic ordering.

1.2.3 Methodological solution

The proposed solution to the complex testing problem described in Section 1.2.2 can be found as a member of the family of combined permutation tests (Pesarin, 2001; Pesarin and Salmaso, 2010). This method consists of carrying out a set of partial tests, and then combining the p-values of the partial tests in order to determine a combined test statistic suitable for the general problem. The CPT methodology is used in various application problems (Alibrandi et al., 2022; Giacalone and Alibrandi, 2011), for categorical data (Bonnini, 2014), big data (Bonnini and Melak Assegie, 2022), regression analysis (Giacalone and Alibrandi, 2015) and many other problems.

The solution that comes closest to the proposal is that of Basso and Salmaso (2011). Their method takes into account only the stochastic ordering and is based on a **m**ultiple **t**est based on **p**ooling (*mtp*): the first two-sample test compares the first sample and the remaining ones pooled together; the second two-sample test compares the first two samples pooled together and the remaining ones pooled together; etc. For this section, this procedure will be called *mtp*.

Following the technique of decomposing the global problem into sub-problems, a possible solution for the V-shaped problem consists in the application of the *mtp* method to each of the two sub-problems, i.e. the decreasing and the increasing stochastic ordering. Then an additional combination of the two tests provides a final test statistic, whose p-value solves the overall testing problem (the null hypothesis is rejected when the p-value is less than or equal to the significance level α). The most common functions used to combine the partial tests are:

- Fisher combination function: $T_F = -2\sum_i \ln(\lambda_i)$,
- Tippett combination function:$T_T = \max_i(1-\lambda_i)$,

where λ_i is the p-value function for the i-th partial test.

We proposed a methodological solution based on the combination of partial tests for pairwise comparisons between consecutive samples. The two main components of the hypotheses system of the overall testing problem are:

$$\begin{cases} H_{0p}^{\searrow} : \cap_{s=1}^{p-1} H_{s,s+1}^{0} \\ H_{1p}^{\searrow} : \cup_{s=1}^{p-1} H_{s,s+1}^{-} \end{cases} \tag{1.5}$$

and

$$\begin{cases} H_{0p}^{\nearrow} : \cap_{s=p}^{C-1} H_{s,s+1}^{0} \\ H_{1p}^{\nearrow} : \cup_{s=p}^{C-1} H_{s,s+1}^{+} \end{cases} \tag{1.6}$$

that correspond to the non-increasing and to the non-decreasing part of the relationship.

Since the exchangeability is satisfied in the partial null hypothesis, to test the decreasing or increasing stochastic dominance, a permutation test for two independent samples is carried out and suitable test statistics are $T_{s,s+1}^{-} = \bar{X}_s^* - \bar{X}_{s+1}^*$ and $T_{s,s+1}^{+} = \bar{X}_{s+1}^* - \bar{X}_s^*$ respectively. Without loss of generality, the (partial and overall) null hypotheses are rejected for large values of the test statistics. The latter procedure, presented as an alternative to *mtp*, will be called *mpc*: **m**ultiple **p**airwise **c**omparison.

1.2.4 Simulation study

This subsection aims to compare the performance of the permutation tests presented previously, in terms of power behaviour, through a Monte Carlo simulation study. The significance level considered was $\alpha = 0.05$, then 1000 datasets were generated and (for each simulation) 1000 random permutations of the dataset were carried out. In order to study the consistency of the tests, different sample sizes were considered. Furthermore, the cases of $C = 3$ and $C = 5$ balanced samples were taken into account, i.e. $n_1 = \ldots = n_C = n$. In H_1, when $C = 3$, the peak point was set at $p = 2$; when $C = 5$ the three cases $p = 2$, $p = 3$ and $p = 4$ were investigated.

The data were generated from Normal distributions, so $X_j \sim \mathcal{N}(\mu_j, 1)$ with $j = 1, \ldots, C$ and $\mu_j = 0 \ \forall j$ under the null hypothesis. Since the mean of the population corresponding to the peak point is equal to zero under H_1, the mean values of the other populations are determined by the rule $\mu_{p+k} = |k| \cdot \delta$ with $k \in \{-2, -1, 0, 1, 2\}$. The parameter δ represents the shift between consecutive populations. For example if $C = 3$ and $p = 2$ you get that $\mu_1 = \delta$, $\mu_2 = 0$ and $\mu_3 = \delta$, etc. Some values of δ were taken into account: 0, 0.25, 0.50 and 0.70. Hence, the rejection rates of the three tests were compared. The three tests were: *mtp* (multiple test based on pooling, i.e. the main competitor of our method), *mpc F* (multiple pairwise comparison with the Fisher combination) and *mpc T* (multiple pairwise comparison with the Tippett combination).

Table 1.1: Rejection rates under H_0 with different sample sizes in case of $C = 3$ samples.

n	mtp	mpc F	mpc T
10	0.057	0.088	0.057
20	0.057	0.064	0.043
30	0.059	0.071	0.045
40	0.060	0.077	0.047
50	0.061	0.071	0.046
60	0.062	0.065	0.044
70	0.064	0.078	0.051
80	0.065	0.091	0.057
90	0.059	0.077	0.048
100	0.052	0.063	0.038

Table 1.1 reports the rejection rates in the case of $C = 3$ under the null hypothesis. The tests *mpc F* and *mtp* seems to be anti-conservative because they don't respect the nominal α level. On the other hand, *mpc T* is well approximated for almost any value of n and when its rejection rate is greater than α, the difference is not much evident.

In Table 1.2 it can be seen that the power is always greater than under H_0, hence the tests are unbiased. Furthermore, the power increases with n and with δ in all the three cases. Consequently, the consistency is verified. The power of the solution based on pairwise comparisons is greater than the power of the competitor (*mtp*). The most powerful test is *mpc F*, but this result must be considered without ignoring the anti-conservative behavior under H_0.

Table 1.3 reports the rejection rates under H_0 in the case of $C = 5$. It can be remarked that the conclusions obtained in the case of $C = 3$ are confirmed but with a visible improvement of *mpc F*, whose rejection rates are always lower than

Table 1.2: Rejection rates under H_1 as a function of the sample sizes and for different values of δ in case of $C = 3$ samples.

n	δ	mtp	mpc F	mpc T	δ	mtp	mpc F	mpc T
10	0.25	0.097	0.212	0.142	0.5	0.203	0.418	0.277
20		0.132	0.278	0.195		0.381	0.640	0.491
30		0.182	0.349	0.251		0.528	0.756	0.638
40		0.231	0.420	0.307		0.674	0.872	0.784
50		0.267	0.470	0.350		0.762	0.915	0.842
60		0.303	0.520	0.392		0.850	0.957	0.900
70		0.353	0.583	0.448		0.890	0.970	0.927
80		0.402	0.646	0.503		0.929	0.982	0.953
90		0.440	0.688	0.542		0.951	0.989	0.967
100		0.477	0.730	0.581		0.972	0.996	0.981
n	δ	mtp	mpc F	mpc T				
10	0.7	0.349	0.592	0.449				
20		0.650	0.852	0.724				
30		0.789	0.920	0.839				
40		0.927	0.987	0.954				
50		0.957	0.994	0.972				
60		0.986	1.000	0.990				
70		0.992	1.000	0.995				
80		0.998	1.000	0.999				
90		0.999	1.000	1.000				
100		1.000	1.000	1.000				

α.

Table 1.4, Table 1.5 and Table 1.6 show the power under H_1 in the case of 5 samples and with different peak points. One difference with respect to the case $C = 3$ concerns the lower power of *mpc T*. Another important result is that the power increases as the number of samples C increases. Furthermore, the power also varies with respect to the position of the peak point in the "V-shaped" relationship: when $p = 3$ (symmetric case) the tests are less powerful than in the cases where $p = 2$ or $p = 4$ (asymmetric cases).

1.2.5 Application to the reluctance to CE

The first case study concerns the hot topic of Circular Economy and in particular the *reluctance* towards this paradigm. The data refer to Italian Small and Medium Enterprises, in the sector of *Industry of wood and products in wood and cork (ex-*

Table 1.3: Rejection rates under H_0 with different sample sizes in case of $C = 5$ samples.

n	mtp	mpc F	mpc T
10	0.051	0.043	0.050
20	0.049	0.038	0.044
30	0.050	0.042	0.051
40	0.051	0.045	0.057
50	0.052	0.042	0.054
60	0.052	0.039	0.050
70	0.050	0.042	0.052
80	0.048	0.044	0.053
90	0.048	0.042	0.055
100	0.047	0.040	0.057

cluding furniture); manufacture of straw articles and woven materials. The dataset was collected through a sample survey about CE on Italian SMEs carried out by the Department of Economics and Management of the University of Ferrara, with the collaboration of a specialized company, via CATI method in January 2020. For the analysis, the focus was placed on old companies, i.e. firms with more than six years of activity. Indeed old firms seem to be much involved in this type of activities than young firms.

The response variable indicates the *reluctance* of the firm to CE. This variable was computed as the inverse of the number of innovations introduced in the years 2017-1018. Such innovations were aimed at the *Transfer of waste to other companies, which use it in their production cycle.* The factor under study denoted firm size and takes three levels: (1) for micro firms, i.e. firms with less than or equal to 9 employees; (2) for small firms, i.e. firms with more than 9 employees but less than or equal to 49 employees; (3) for medium firms, i.e. firms with more than 49 employees).

The overall sample of the survey consisted of 546 Italian old firms but the companies operating in the economic sector *Industry of wood and products in wood and cork (excluding furniture); manufacture of straw articles and woven materials* (according to the classification of economic activities ATECO 2007) consisted of only 36 companies. Of these companies, 2 were micro, 32 small and 2 medium, hence we are in presence of very small samples, which also justifies the previously mentioned non-parametric approach.

The aim of this problem is to test the V-shaped hypothesis between reluctance ad firm size, in other words, that small enterprises are less reluctant towards CE than micro and medium enterprises at the significance level α equal to 0.10.

Table 1.4: Rejection rates under H_1 as a function of the sample sizes and for different values of δ in case of $C = 5$ samples and peak point in 2.

n	δ	mtp	mpc F	mpc T	δ	mtp	mpc F	mpc T
10	0.25	0.424	0.484	0.336	0.5	0.962	0.972	0.874
20		0.739	0.762	0.575		1.000	1.000	0.994
30		0.852	0.866	0.738		1.000	1.000	0.997
40		0.964	0.969	0.901		1.000	1.000	1.000
50		0.981	0.984	0.944		1.000	1.000	1.000
60		0.997	0.999	0.987		1.000	1.000	1.000
70		0.999	1.000	0.992		1.000	1.000	1.000
80		1.000	1.000	0.997		1.000	1.000	1.000
90		1.000	1.000	0.998		1.000	1.000	1.000
100		1.000	1.000	0.998		1.000	1.000	1.000
n	δ	mtp	mpc F	mpc T				
10	0.7	1.000	1.000	0.990				
20		1.000	1.000	1.000				
30		1.000	1.000	1.000				
40		1.000	1.000	1.000				
50		1.000	1.000	1.000				
60		1.000	1.000	1.000				
70		1.000	1.000	1.000				
80		1.000	1.000	1.000				
90		1.000	1.000	1.000				
100		1.000	1.000	1.000				

The evidence that a V-shape relationship is possible is shown in Figure 1.1. To test the significance of the V-shaped effect of size, the test *mpc F*, based on the combination function of Fisher, is applied. We obtain a general p-value equal to **0.0579**, indicating overall significance, i.e. empirical evidence in favor of the alternative hypothesis of "V-shaped" relationship.

1.2.6 Application to the propension to CE

In this subsection, the other case of complex hypotheses explained in Section 1.2.3, namely the "inverted V-shaped" relationship is considered. In addition, stratification factors and the extension concerning a multivariate response variable are also introduced in this case study.

Here the effect of firm size on the intensity of innovation in Circular Economy of Italian SMEs is studied, considering also the presence of confounding factors

Table 1.5: Rejection rates under H_1 as a function of the sample sizes and for different values of δ in case of $C = 5$ samples and peak point in 3.

n	δ	mtp	mpc F	mpc T	δ	mtp	mpc F	mpc T
10	0.25	0.239	0.256	0.179	0.5	0.642	0.678	0.367
20		0.414	0.439	0.263		0.926	0.933	0.650
30		0.540	0.572	0.331		0.962	0.965	0.793
40		0.665	0.704	0.398		0.998	0.996	0.935
50		0.753	0.783	0.484		0.999	0.998	0.966
60		0.840	0.862	0.569		1.000	1.000	0.997
70		0.887	0.901	0.629		1.000	1.000	0.998
80		0.933	0.939	0.689		1.000	1.000	0.999
90		0.951	0.956	0.750		1.000	1.000	1.000
100		0.968	0.973	0.810		1.000	1.000	1.000
n	δ	mtp	mpc F	mpc T				
10	0.7	0.904	0.915	0.612				
20		0.997	1.000	0.927				
30		0.999	1.000	0.964				
40		1.000	1.000	1.000				
50		1.000	1.000	1.000				
60		1.000	1.000	1.000				
70		1.000	1.000	1.000				
80		1.000	1.000	1.000				
90		1.000	1.000	1.000				
100		1.000	1.000	1.000				

such as the age of the company and the sector of economic activity. Often, in the literature, these confounding factors are ignored. The proposed methodology takes into account the confounding effects by stratificating the sample of companies according to the confounding factors.

Considering two age classes (young companies, i.e. with less than or equal to 6 years of activity, and old companies, with more than 6 years of activity) and m economic sectors, there are $2 \times m$ strata. Within each stratum there are companies that are homogeneous with respect to sector and age (confounders). Hence, any difference in the distribution of the responses must be attributed only to the factor under investigation, that is size. In this case study, the size (factor) has three levels (micro, small and medium), so there are two pairwise comparisons (micro vs small and small vs medium). Looking at Figure 1.2, it can be seen the logic of combination and stratification of the complex problem under study: there are 2 age classes, m sectors, q variables (because of the q-variate response variable) and

Table 1.6: Rejection rates under H_1 as a function of the sample sizes and for different values of δ in case of $C = 5$ samples and peak point in 4.

n	δ	mtp	mpc F	mpc T	δ	mtp	mpc F	mpc T
10	0.25	0.443	0.497	0.324	0.5	0.954	0.963	0.857
20		0.743	0.771	0.595		1.000	0.999	0.993
30		0.849	0.870	0.745		1.000	1.000	0.997
40		0.954	0.968	0.895		1.000	1.000	1.000
50		0.974	0.982	0.935		1.000	1.000	1.000
60		0.993	0.995	0.974		1.000	1.000	1.000
70		0.997	0.998	0.985		1.000	1.000	1.000
80		1.000	1.000	0.996		1.000	1.000	1.000
90		1.000	1.000	0.998		1.000	1.000	1.000
100		1.000	1.000	1.000		1.000	1.000	1.000
n	δ	mtp	mpc F	mpc T				
10	0.7	0.999	1.000	0.990				
20		1.000	1.000	1.000				
30		1.000	1.000	1.000				
40		1.000	1.000	1.000				
50		1.000	1.000	1.000				
60		1.000	1.000	1.000				
70		1.000	1.000	1.000				
80		1.000	1.000	1.000				
90		1.000	1.000	1.000				
100		1.000	1.000	1.000				

2 pairwise comparisons. Hence the number of partial tests is $2 \times m \times q \times 2$. Then, a first combination with respect to the pairwise comparisons is carried out. After that, the remaining partial tests are $2 \times m \times q$ and they are combined subsequently.

The original dataset considered in this case study is the same of Subsection 1.2.5, but here the overall sample was taken into account. The firms interviewed were 3946, and the response variable is multidimensional. As a matter of fact, the response variables include ten counting variables indicating the number of innovations connected to ten different CE practices adopted in the years 2017-18. The variables were the following:

1. red-wat: number of innovations for reduction of water use in the production process,

2. red-mat: number of innovations for reduction in the use of materials,

3. use-ren-energy: number of innovations for use of energy generated from re-

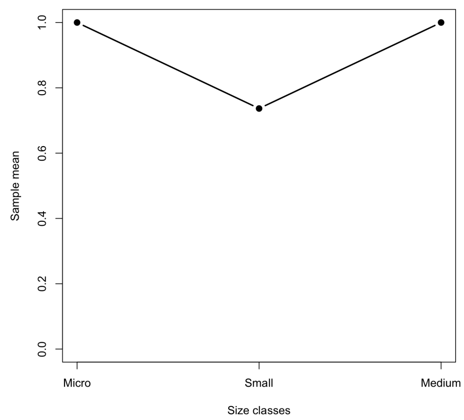


Figure 1.1: Main effect of firm size on CE reluctance of Italian SMEs in the period 2017-2018 for the sector: *Industry of wood and products in wood and cork (excluding furniture); manufacture of straw articles and woven materials.*

newable sources,

4. red-electr: number of innovations for reduction in the use of electricity,

5. red-waste: number of innovations for reduction of emitted waste,

6. reuse-waste: number of innovations for reuse of waste in its production cycle,

7. transf-waste: number of innovations for transfer of their waste to other companies, which use them in their production cycle,

8. change-design-min: number of innovations for change in product design to minimize the use of raw materials,

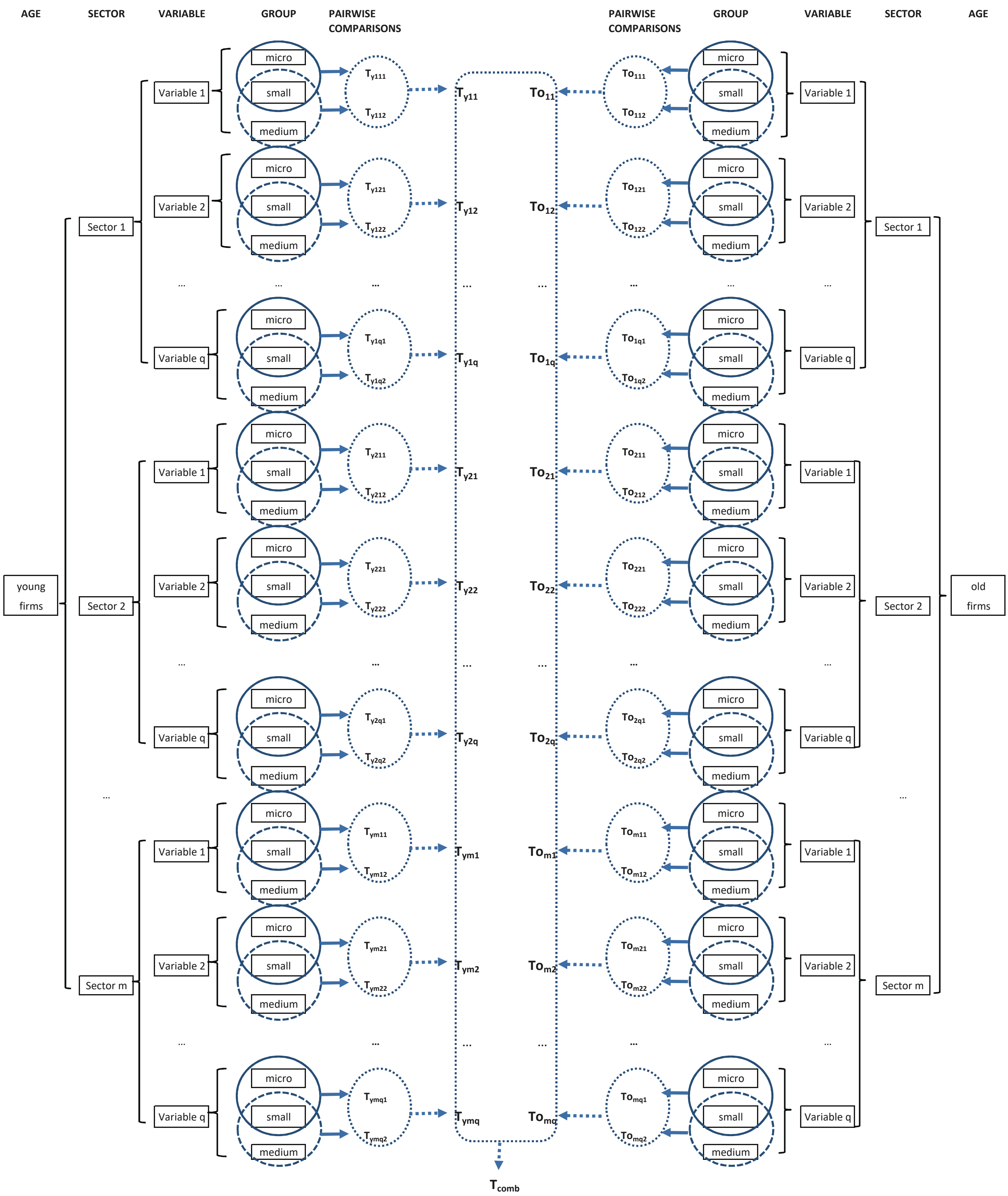
9. change-design-max: number of innovations for change in product design to maximize their recyclability,

10. change-prod-process: number of innovations for change in the production process to reduce greenhouse gas emissions.

Moreover, in order to overcome the confounding effect of the economic sector, the sectors defined according to the ATECO 2022 classification were aggregated into eight categories. To conduct this operation, Dimitar (2017) provided an example of aggregation (details are in the Appendix 3.5), so the eight economic sectors were the following:

- food-bev: food products, beverages and tobacco products,

Figure 1.2: Combination strategy of partial tests with respect to pairwise comparisons, response variables, and confounding factors, i.e. age and economic sector of companies.



- text-wear: textiles and wearing apparel,
- wood-print: wood, paper, printing and reproduction,
- chem-plast: chemical-pharmaceutical, plastic and refined petroleum products, coke,
- metals: basic metals and fabricated metal products,
- comp-machinery: computer, electronic and optical products,
- motors: motor vehicles, trailers, semitrailers and other transport equipment,
- other: other sectors.

As in Subsection 1.2.5 the factor under investigation is firm size, characterized by the number of employees. So the three groups under comparison are micro, small and medium firms, and the goal is to test the inverted "V-shaped relationship" between the intensity of the innovation for CE of Italian SMEs (represented by the ten variables explained before) and the firm size. In other words, testing if the intensity of innovation of small firms is greater than that of micro and medium firms. In Table 1.7 the sample sizes for each age group, for each sector and for each dimension are reported. As can be seen, in many strata we are in the presence of small sample sizes, this reinforces the need to use a non-parametric methodology because it is distribution-free and take into account the dependence structure of the ten response variables, without using asymptotic properties or other limiting conditions.

Table 1.7: Sample sizes for each age group, for each sector and for each dimension.

	Young firms			Old firms		
Sectors	Micro	Small	Medium	Micro	Small	Medium
food and beverages products	7	50	6	51	280	25
textiles and wearing apparel	3	60	2	17	259	24
wood, paper, printing	5	29	1	39	272	17
chemical-pharmaceutical, plastic	2	26	4	10	216	32
metals and metal products	10	116	10	68	759	63
computer, electronic	1	45	4	22	365	49
motors	1	16	6	4	47	6
other	12	135	6	72	620	72

In Figure 1.3 and Figure 1.4, the main effect plots of the firm's size with respect to the ten variables and for all eight sectors are represented. It can be noticed that, from a descriptive point of view, the "inverted V-shaped" relationship is more evident, especially for young companies and only in specific sectors of activity.

At this point, the non-parametric test described in Section 1.2.3, with the combination function of Fisher, was carried out. The overall p-value was equal to **0.0001**, which is significant at the level $\alpha = 0.10$. This implies the rejection of the null hypothesis of equality in distribution in favor of the alternative "inverted V-shaped" relationship. One of the advantages of using CPTs is that the overall problem can be broken down into subproblems. Furthermore, when the global test is significant (as in this case), it is possible to attribute such significance to one or more partial tests (in this specific problem there are 160 partial tests). For this purpose, all the partial p-values were adjusted with the Bonferroni-Holm method to control the familywise error rate of the test and prevent the type I error rate from exceeding α (Pesarin and Salmaso, 2010; Giacalone et al., 2018). In Table 1.8 and in Table 1.9 the adjusted p-values are reported. The significance can be attributed more to the sector of wood, paper, printing and reproduction, and to the sector of chemical-pharmaceutical, plastic and refined petroleum products. Given that it is commonly accepted in the literature that these industries are among the most active in the CE, this outcome is not surprising (Aggestam and Giurca, 2022; Stavenhagen, 2020). Furthermore, as said before looking to the main effect plots, the "inverted V-shaped" effect of the size on the intensity of innovation is more diffuse in young firms. On the other hand, types of innovation for which we have significance are:

- reduction of emitted waste,
- reuse of waste in the production cycle of the firm,
- transfer of their waste to other companies, which use it in their production cycle,
- change in product design to minimize the use of raw materials,
- change in product design to maximize their recyclability.

1.3 CPTs for multivariate Bernoulli distributions

1.3.1 Introduction

In this Section, to present the propensity of companies to CE, by taking into account the multidimensional nature of such propensity, we will introduce the

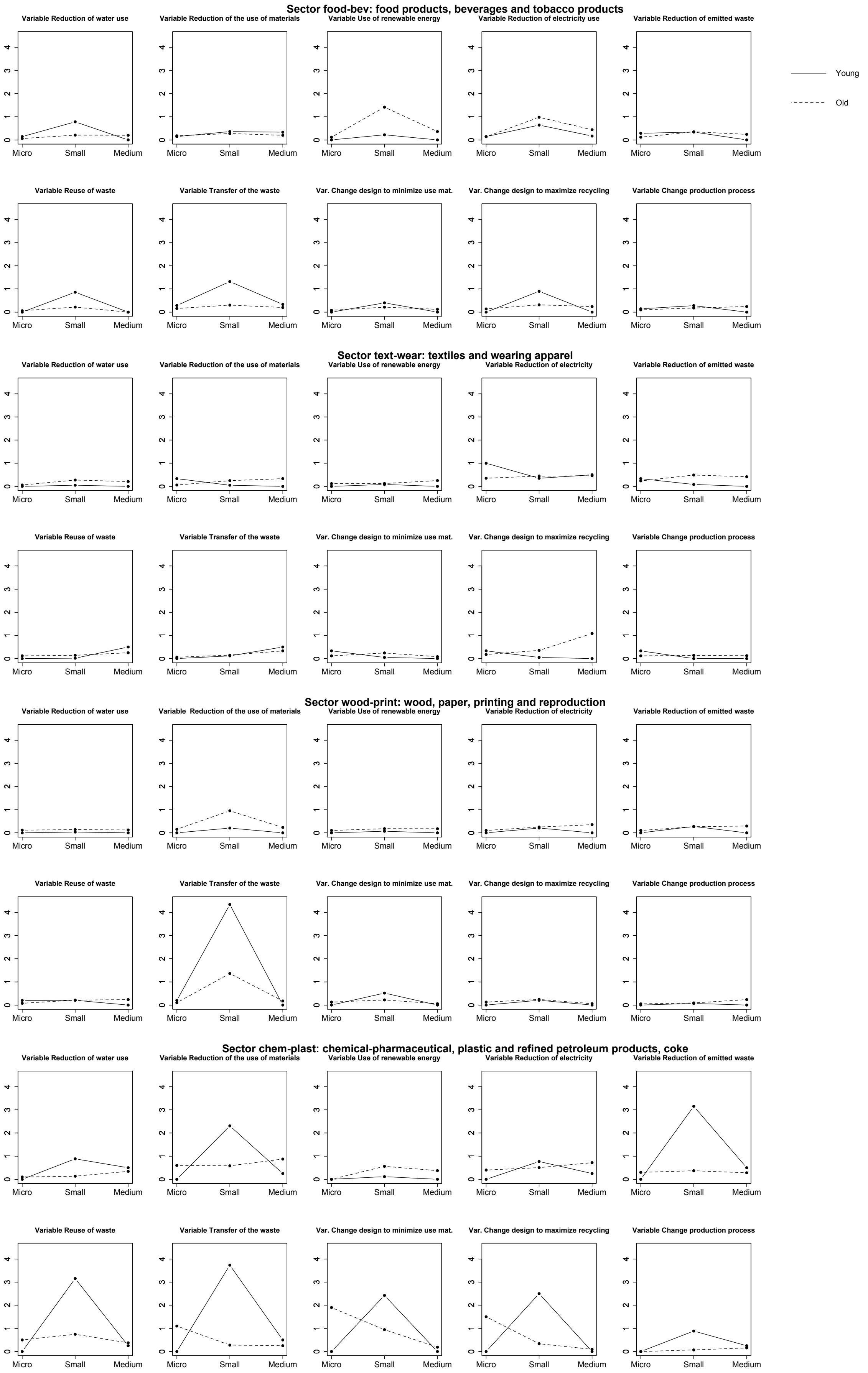


Figure 1.3: Main effect of firm size on all response variables, by age and for the sectors: food-bev, text-wear, wood-print and chem-plast.

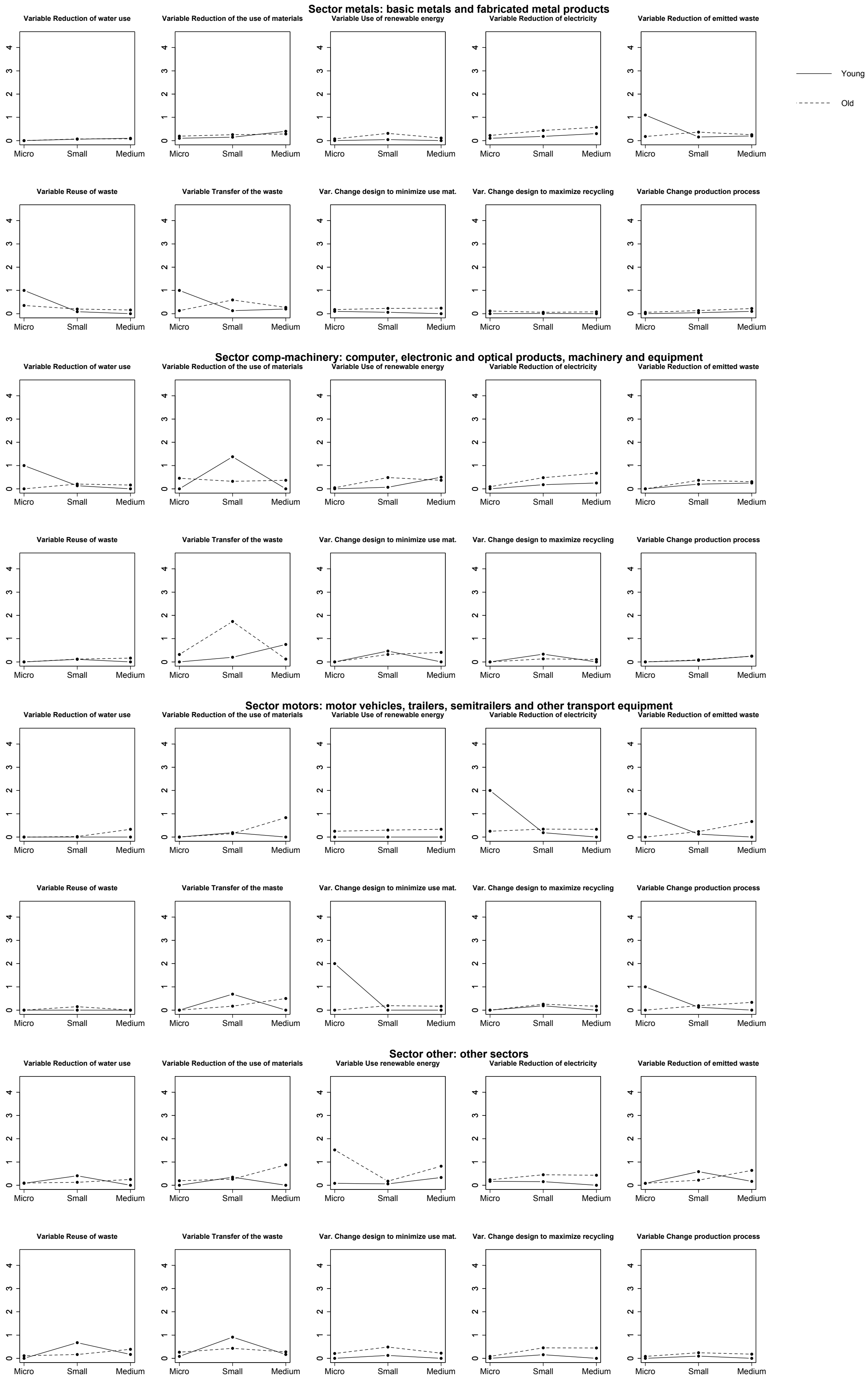


Figure 1.4: Main effect of firm size on all response variables, by age and for the sectors: metals, comp-machinery, motors and other.

concept of multivariate Bernoulli distribution. Specifically, the purpose is to compare, through a two-sample test, two different types of companies in terms of their propensity to circularity. In the literature, a circular company is a company that implements one or more activities correlated to recycling or to environmental sustainability. As for the case study of Subsection 1.2.6, the response variable is multivariate. The difference is that now the components of the multivariate response do not concern the number of innovations introduced in a certain activity but they consist of binary variables (of the type "presence" or "absence" of a given practice). In fact, one method of evaluating a firm's attitude toward CE is to administer a questionnaire with a list of k CE activities and ask the company to declare whether it engages in the appropriate CE activity or not. A $n \times k$ matrix containing values that are 0 or 1 represents the information gathered from an empirical investigation on this subject that includes n firms and asks these k questions. 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 $i = 1, \ldots, n$ and $j = 1, \ldots, k$. We should compare the proportion of circular companies in the two groups, for each of the k CE activities taken into account, in order to determine whether the propensity toward CE of the first group of companies is less than the propensity toward CE of the second group. Of course, in order to draw conclusions about the two k-dimensional vectors of population proportions, it is essential to compare the two k-dimensional vectors of sample proportions. Due to the multidimensional response and the directional alternative hypothesis, the inferential problem is complex. The proposed solution is based on the permutation approach, hence it is nonparametric. Its nonparametric nature makes it flexible with regard to the underlying distribution assumptions. This makes it more robust than a parametric solution. Permutation tests are powerful also when the assumptions of parametric tests are true, but they are generally much more performant when the distributional assumptions of parametric tests are violated (Pesarin, 2001). The proposed permutation test is also very suitable for a multivariate situation like the one under consideration. This is because it allows to take into account the interdependence between variables without assuming an underlying distribution or modeling the interdependence structure. The proposed approach also has the benefit of being effective when there are several variables and small sample sizes.

In the literature, a valid solution to the multivariate test for binary data with one-tailed hypotheses is that of Davidov and Peddada (2011). This inferential nonparametric approach is based on the computation of the estimator of the vector of marginal probabilities and on the bootstrap null distribution of the test statistic. Conversely, our idea is to face the problem of simultaneous marginal homogeneity against one-tailed alternative hypotheses for multivariate binary variables. Hence, the proposal can be used if exchangeability holds. In particular, we propose a solution that belongs to the family of combined permutation tests, suitable for a multivariate extension of the one-tailed two-sample test on proportions. The inspiring example is based on a sample study conducted in Italy to examine the small and medium-sized firms (SMEs) inclination toward CE. Six distinct CE activities, including trash reuse, the adoption of innovations, investments in R&D pertaining to CE practices, etc., are included in this poll to represent circular attitudes and behaviors. The objective is to add results to the discussion on how firm size affects a company's propensity toward CE.

1.3.2 Statistical problem

Simultaneous marginal homogeneity is the term used most frequently in the literature to refer to the null hypothesis of the problem, which is the equality of the two vectors of population proportions (Agresti and Klingenberg, 2005). Chuang-Stein and Mohberg (2019) provided a test for the scenario of binary data with two-sided alternative hypotheses based on a multivariate version of the Wald-type statistic. Agresti and Klingenberg (2005) proposed an alternative based on the permutation method for the same issue. This nonparametric test has been shown to be effective and adaptable in general, but it is especially advantageous when very small n and big k are present because under these circumstances a parametric technique becomes significantly less effective and in some cases impossible. Agresti (1983) presented a test for problems with ordered categorical data to overcome the Chi-square test's inability to consider the potential ordinal nature of the data. Numerous strategies have been put forth to test simultaneous marginal homogeneity versus stochastic ordering alternatives (Gao and Kuriki, 2006; Klingenberg et al., 2008; Bonnini et al., 2014b).

Assuming that $x_{ij} \in \{0,1\}$ is the value of the j-th response variable, observed on the i-th firm (e.g. statistical unit). It is a realization of the random variable X_{ij} with $i = 1,\ldots,n$ and $j = 1,\ldots,k$, which 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:

$$X_{ij} \sim \begin{cases} \mathcal{B}(\theta_{1j}) \text{ if } i \text{ is in sample 1} \\ \mathcal{B}(\theta_{2j}) \text{ if } i \text{ is in sample 2 .} \end{cases} \tag{1.7}$$

Let $\{1,2,\ldots,n_1\}$ denote the units in sample 1 and $\{n_1+1, n_1+2,\ldots,n_1+n_2 = n\}$ the units in sample 2. The assumption presented in Equation 1.7 means that $X_{1j}, X_{2j}, ..., X_{n_1 j}$ are identically distributed according to $\mathcal{B}(\theta_{1j})$ and $X_{n_1+1,j}, X_{n_1+2,j},$
$, ..., X_{nj}$ are identically distributed according to $\mathcal{B}(\theta_{2j})$, with $j = 1,\ldots,k$.

θ_{sj} represents the proportion of firms of the s-th population that undertakes the j-th CE activity. Hence it can be defined as the probability of "success" of the Bernoulli random variable. Each component θ_{sj} determines the central tendency and the variability of the marginal distribution of the j-th component, but there is no specification on the dependence between the k components of the multivariate response. This is crucial to understand because one benefit of the suggested solution, which is based on the nonparametric Combined Permutation Test (CPT) method, is the flexibility provided by the fact that no assumption about the dependence between the components is required, and it is not necessary to estimate nuisance parameters that represent such dependence. The dependency must be considered unless the strong and improbable condition of independence between the k components is assumed. In the CPT family of tests, it is implicitly taken into account through the permutations of the dataset rows, useful to determine the test statistics' null distribution. The equality of the parameters of the two populations for each of the k components is the problem's null hypothesis. The strict inequality is maintained under the alternative hypothesis for at least one component (i.e., for one or more CE activities, the fraction of companies engaging in such activities is smaller in the first population). Thus, the system of hypotheses of the problem is the following:

$$\begin{cases} H_0 : \theta_{1j} = \theta_{2j} \forall j \in \{1, \ldots, k\} \\ H_1 : \exists j \in \{1, \ldots, k\} \text{ s. t. } \theta_{1j} < \theta_{2j}. \end{cases} \tag{1.8}$$

1.3.3 Methodological solution

The proposed methodological solution concerns the use of combined permutation tests, hence it is based on the idea that the test raised in the previous section may be seen as a multiple testing procedure. The global problem may actually be divided into k separate partial tests. The general problem, rewriting Equation 1.8, can be written as follows:

$$\begin{cases} H_0 : \cap_{j=1}^{k} H_{0,j} \\ H_1 : \cup_{j=1}^{k} H_{1,j} \end{cases} . \tag{1.9}$$

Pesarin (2001) defined the overarching theoretical foundation of the CPT methodology, and subsequent studies published in the previous two decades refined and extended it. We mention Pesarin and Salmaso (2010) and Bonnini et al. (2014a) contributions as two of the most significant works in this area. Such publications have helped the technique become more widely used in terms of theory, applications, and software. There are numerous and various empirical investigations where the CPT methodology has been effectively used (Giacalone and Alibrandi, 2011; Giacalone et al., 2018). The method is appropriate for big data (Bonnini and Melak Assegie, 2022), categorical data (Arboretti Giancristofaro and Bonnini, 2009; Bonnini, 2014), regression analyses (Giacalone and Alibrandi, 2015), ranking problems (Corain et al., 2017), and many other methodological frameworks because of its flexibility and robustness.

In the j-th partial problem, a suitable test statistic could be $T_j = \hat{\theta}_{2j} - \hat{\theta}_{1j} = \sum_{i=n_1+1}^{n} \frac{X_{ij}}{n_2} - \sum_{i=1}^{n_1} \frac{X_{ij}}{n_1}$. For each partial test, the null hypothesis is rejected for large values of the test statistic.

Under H_0, all potential assignments of the n observed data to the samples have the same probability of occurring, hence the criterion of exchangeability is met under the null hypothesis. Therefore, by taking into account all potential dataset permutations and the consequent value of the test statistic for each permuted dataset, it is feasible to determine the null permutation distribution of the test statistic. Indeed, for computational reasons, a random generation of B permutations is performed rather than taking into account the exact distribution based on all possible permutations, where B is not huge (often larger than or equal to 1000) with respect to the cardinality of the permutation space. Such Conditional Monte Carlo technique is widely used in practice. The significance level function of the j-th partial permutation test can be estimated as:

$$L_j^P(t) = \frac{\sum_{b=1}^{B} I_{(-\infty, t_{j,b}^P]}(t) + 0.5}{B+1}, \tag{1.10}$$

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

The combination of the significance level functions of the k partial tests is the test statistic for the general multivariate testing problem. Assume that for large values of the combined test statistic, the overall null hypothesis is rejected without losing generality. The partial tests are combined with a similar procedure explained in Subsection 1.2.3. Hence, when one argument tends to zero, the combining function tends to its supremum and produces a critical value for the combined test that is strictly smaller than the supremum. The combining function has not increasing relationship with each argument. There is a large variety of combining functions that fulfill these requirements, and each combining function corresponds to a unique combined test with unique characteristics and capabilities. Within the wide family of combined permutation tests, CPTs based on Fisher and on Tippett combination are considered.

It is worth remembering that, another important strength of this solution concerns the dependence between the k components. In fact, even though it is not explicitly described, the dependency between the partial tests is taken into account by the implementation of the nonparametric combination and the permutation strategy.

1.3.4 Simulation study

To examine the power behavior of the proposed permutation test, a Monte Carlo simulation study was carried out. 1000 random permutations were performed in order to compute the p-values in accordance with the test statistics' null permutation distribution. The power was calculated as the proportion of simulations in which the null hypothesis is rejected (rejection rate) for each setting. After being randomly created from a Normal k-variate distribution, sample data were converted into binary values in accordance with the success probability of the two populations. In the parameters of the simulation, $k = 5$ and $k = 10$ components of the multivariate response are taken into account.

The formal method for obtaining the simulated data for each unit, or for each unique response profile, was to generate a vector of k values from $Z \sim \mathcal{N}_k(\mathbf{0_k}, \mathbf{\Sigma})$, a normal k-variate random variable with null mean whose marginal components are standard Normal with

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}. \tag{1.11}$$

The binary values for both the samples were obtained by transforming the normal ones. Consequently, the following scalars were used as the setup parameters that changed in the simulations:

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

The significance level chosen for all the simulations is $\alpha = 0.05$. In Table 1.10 the results of the simulation study under H_0 are represented and in this case, the rejection rates of the permutation test should be not greater than α. The rejection rate is very close to 0.05 and, on average, less than 0.05, for both the CPTs defined above based on Fisher and Tippett's combination. Therefore the tests are well approximated and, under the null hypothesis, they respect the nominal α level.

The powers under the alternative hypothesis are shown in Table 1.11 and onwards. In these tables, it is clear that the rejection rates are increasing function of the sample size, hence the tests are consistent. The performance of the corresponding combined test has been further investigated by analyzing the power as a function of different parameter settings given that Fisher's combination seems to be better suited.

In Table 1.12, the rejection rates of the test based on the Fisher combination with different numbers of variables k are shown. The values show that the higher the number of variables the greater the power. Furthermore, it is interesting that the convergence rate (that is the number n for which the power tends to the maximum value 1) decreases as the number k increases.

The rejection rate of CPTs from the Fisher combination has been assessed with different values of p, in order to measure the impact of the proportion of true partial alternatives hypotheses. The proportion of variables in the alternative hypothesis p seems to have no influence on the test's power unlike what occurs with regard to the number of variables k.

The effect of shift parameter δ on the power behaviour of a CPT combination with Fisher has been studied. As expected, the power increases with δ. In fact, when $\delta = 0.45$, the power is uniformly equal to 1 with respect to the sample size (see Table 1.14).

Moreover, sparse and non-sparse data were compared changing the values of θ_1 as in Table 1.15. Here it is found that the power is a non-monotonic function of θ_1, but it is a convex function of it: the power reaches the minimum when $\theta_1 = 0.5$. For population 1 the variance of the underlying Bernoulli distribution is equal to $\theta_1(1-\theta_1)$ and for population 2 it is equal to $(\theta_1+\delta)(1-\theta_1-\delta)$. Those fit to the parabolic functions with the maximum in $\theta_1 = 0.5$ for population 1 and maximum in $\theta_1 = 0.5 - \delta = 0.45$ for population 2. The thesis follows since the maximum of the variance corresponds to the minimum of the power ceteris paribus.

Table 1.16 shows how the correlation ρ and therefore the degree of dependence between the variables can affect the ability of the test to detect the true hypothesis when this hypothesis is H_1. Rejection rates for each sample size shall be reduced when the correlation between Gaussian variables used in randomly generated simulation data is high. The stronger the correlation between variables, the smaller the marginal information contribution of each partial test to the general testing problem. In addition, it would appear that ρ has an effect on the slope of a power function. As a result, when ρ is high, the slope is very low, hence the increase rate of the power with respect to the sample size is less than when ρ takes small values. To put it differently, a strong correlation implies that the convergence of the power to one when n diverges, due to the consistency of the test, is slower.

Finally, the relationship between the power of the test and u (imbalance ratio), is reported in Table 1.17. The power is maximum when $u = 1$ (balanced design) and, as the imbalance becomes more pronounced the power decreases. Moreover, it doesn't count if the imbalance is in favor of n_1 or n_2.

1.3.5 Application

Case studies related to circular economy are examined in this section. The data relates to SMEs (Small and Medium Enterprises) in Italy that operate in many different sectors. Old enterprises, or those that have been in business for more than six years, were considered because, as previously said, they are more engaged and inclined toward CE activities than new firms (Bassi and Dias, 2019), are more prevalent, and represent the benchmark population of Italian SMEs. The dataset is the same of the previous section, concerning a sample survey on CE that was conducted in January 2020 by interviewing a sample of Italian SMEs. We wish to provide a further contribution to the empirical literature on the impact of firms' size on the propensity to CE.

The comparison of small and medium-sized companies aims to determine if the former have a lower inclination to engage in CE activities than the latter. The percentage of businesses adopting a particular practice can be used to characterize an enterprise population's predisposition towards a particular CE practice. The following practices were taken into account in this study:

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.

The equality of the two proportions of CE firms (in the populations of small and medium companies) for each of the six activities under consideration is the null hypothesis. 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 CE activities in the list are each represented by a dichotomous variable that takes one when a company adopts the practice and zero otherwise. As a consequence, we have a two-sample multivariate testing problem on proportions with a directional alternative hypothesis as well as a multivariate binary response variable with six components. The factor, or "treatment", for the problem is the size of the company as determined by the number of employees. The two-factor levels are small and medium enterprises (categorized by the number of employees as in Subsection 1.2.5).

There are primarily two justifications for conducting three within-sector comparisons. First, it is important to compare companies that are similar, apart from size, because the industry in which a company works is clearly a confounding factor. The second reason is the requirement to conduct sector-specific analyses, taking into account independently one critical sector for the Italian economic system as well as any policy ramifications. The significance level of the test is $\alpha = 0.10$. The p-values of the partial tests were adjusted with the MinP method (Pesarin and Salmaso, 2010) to control the familywise error rate and prevent the type I error rate from exceeding α in cases when the general test showed significance, allowing us to assign the general significance to certain specific variables.

According to the ATECO 2007 classification of economic activity, we paid particular attention to the sector *manufacture of leather and similar item*s. Studies examining CE practices in such a sector are rare in the literature. One of the key causes is that a lack of funding prevents CE activities from being carried out successfully. Big enterprises often have the finest financial foundation to execute the reforms required to fulfill the goals of CE (Moktadir et al., 2020). The sample included 15 medium-sized companies and 118 small companies. Table 1.18 displays the sample percentages of companies among the groups of small and medium companies that carry out the specific mentioned CE activities. The proportions range from 0.051 (for small companies that reduced waste during the production cycle and engage in R&D to reduce the production's impact on the environment) to 0.467 (for medium companies that reduce emitted waste). Every activity has a lower percentage of small companies than medium-sized companies. The combined p-value of the test is equal to **0.008**, hence the alternative hypothesis that small companies have a lower tendency to circularity than medium-sized companies is accepted in place of the general null hypothesis. The significance of the global test must be attributed, according to the adjusted p-values, to the differences in the proportions of companies that reduce their use of raw materials, their emitted waste, their production cycle waste, and their investment in R&D to reduce their impact on the environment. The proportional differences for the other factors are not meaningful.

1.4 Conclusions

A powerful solution to a complex problem such as the test for V- or inverted V-shaped alternatives (and similarly for U- or inverted U-shaped relationship) is provided by the combined permutation test methodology. Monte Carlo simulations have shown that both *mtp* and *mcp* solutions are reliable, objective and powerful. The test based on the multiple comparisons procedure and the combination of Fisher is more powerful, although it seems to be anti-conservative in the null hypothesis in the 3-sample problem. There are few empirical studies in the literature on factors affecting SMEs' reluctance to be "circular", but there are no statistical analyses to test complex hypotheses such as V shaped and U shaped alternatives. The proposed solution, also because it is distribution free and robust with respect to the normality or other strong assumptions related to the distribution of data, constitutes a valid tool for such studies.

We have contributed to the empirical literature on the V-shaped impact of firm size on CE reluctance in Italian SMEs, demonstrating that analysis should be carried out as a separate problems for each sector: specifically, the V-shaped relationship holds in the industry of wood and products in wood and cork and manufacture of paper and paper product. Hence, the tendency to oppose innovation aimed at the implementation of circular economy activity and company size is characterized by a V-shaped relationship.

Regarding the inverted V-shaped relationship, empirical evidence shows that the intensity of innovation for SMEs is greater for small firms. In particular, this conclusion holds with respect to innovations for the reduction of emitted waste and its reuse in the production cycle, for the transfer of the waste to other companies that use it in their own production cycle, for changing the product design to minimize the use of raw materials and maximize their recyclability. This finding applies less widely, and does not apply at all for the reduction of water consumption, in respect of a production process change to limit greenhouse gas emissions. In addition, younger companies are more exposed to the inverted V-shaped effect of size. In view of the results being sector dependent, the stratification by sector is appropriate even if not common in the empirical literature. In fact, firm size matters mainly in the sector of wood, paper, printing and reproduction, and in the sector of chemistry, plastic, refined petroleum products and coke.

Moreover, 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, the 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 simulations prove that the tests have a good performance with regard to power. It is, in particular, found that the tests are well approximated, unbiased, consistent and powerful. An important finding involves the validity for small samples. This solution is based on the permutation approach, so it is distribution-free and does not assume a specific asymptotic null distribution of the test statistic, so it is applicable and powerful for small sample sizes. In addition, the fact that this method applies to each sample size irrespective of the number of variables makes it an important advantage. Similarly, the power of the test is increased when there are more variables regardless of the percentage of tests under the alternative hypothesis and in the same sample sizes. However, the power of the test is not affected by the proportion of variables under H_1, regardless of the number of variables. In addition, even if we experienced a small loss of power in this situation, the proposed permutation test would apply and be performed as well when there are small sample sizes.

The use of this method to test the impact of firm size on the propensity towards CE in a multivariate framework for a specific strategic economic sector has been shown to be useful in a case study related to the recent survey carried out in Italy. In the textile industries sector, there was a lower degree of circularity for small businesses compared to medium ones. In fact, the proposed permutation test is a powerful and robust methodological solution for the two sample tests 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 example, this type of questionnaire could be used to define the circular nature of a company by means of one or more specific business practices. According to the information at hand, there are no competitors in both quantitative and qualitative tests. In addition, tools and good practices to manage commodity risk for managers and policymakers that consist of a multivariate statistical approach derived from suitable questionnaires and an appropriate inferential method, for assessments about CE feasibility are proposed. Finally, we provided a scientific contribution to the empirical literature on CE, in particular as regards the impact of firm size on firms' tendency towards circularity.

Table 1.8: p-values of the partial tests for the sectors: food-bev, text-wear, wood-print, chem-plast (where N.S. means that the p-values are not significant).

Variables	Young firms		Old firms	
Sector food-bev: food products, beverages and tobacco products				
	p-values	adj. p-values	p-values	adj. p-values
red-wat	N.S.	N.S.	0.0411	N.S.
red-mat	N.S.	N.S.	N.S.	N.S.
use-ren-energy	N.S.	N.S.	0.0001	0.0038
red-electr	N.S.	N.S.	0.0001	0.0038
red-waste	N.S.	N.S.	0.0193	N.S.
reuse-waste	0.0001	0.0038	N.S.	0.0421
transf-waste	0.0001	0.0038	N.S.	N.S.
change-design-min	N.S.	N.S.	0.0905	N.S.
change-design-max	0.0001	0.0038	N.S.	N.S.
chance-prod-process	N.S.	N.S.	N.S.	N.S.
text-wear: textiles and wearing apparel				
red-wat	N.S.	N.S.	N.S.	N.S.
red-mat	N.S.	N.S.	N.S.	N.S.
use-ren-energy	0.0001	0.0038	N.S.	N.S.
red-electr	N.S.	N.S.	N.S.	N.S.
red-waste	0.0001	0.0038	N.S.	N.S.
reuse-waste	N.S.	N.S.	N.S.	N.S.
transf-waste	N.S.	N.S.	N.S.	N.S.
change-design-min	N.S.	N.S.	N.S.	N.S.
change-design-max	N.S.	N.S.	N.S.	N.S.
change-prod-process	N.S.	N.S.	N.S.	N.S.
wood-print: wood, paper, printing and reproduction				
red-wat	N.S.	N.S.	N.S.	N.S.
red-mat	0.0001	0.0038	0.0003	N.S.
use-ren-energy	0.0001	0.0038	N.S.	N.S.
red-electr	0.0001	0.0038	N.S.	N.S.
red-waste	0.0001	0.0038	N.S.	N.S.
reuse-waste	0.0001	0.0038	N.S.	N.S.
transf-waste	0.0001	0.0038	0.0001	0.0038
change-design-min	0.0001	0.0038	0.0004	0.0079
change-design-max	0.0001	0.0038	0.0005	0.0119
change-prod-process	0.0001	0.0038	N.S.	N.S.
chem-plast: chemical-pharmaceutical, plastic and refined petroleum products				
red-wat	N.S.	N.S.	N.S.	N.S.
red-mat	0.0001	0.0038	N.S	N.S.
use-ren-energy	N.S.	N.S.	0.0018	0.0271
red-electr	N.S.	N.S.	N.S.	N.S.
red-waste	0.0001	0.0038	N.S.	N.S.
reuse-waste	0.0001	0.0038	0.0011	0.0382
transf-waste	0.0001	0.0038	N.S.	N.S.
change-design-min	0.0001	0.0038	0.0006	0.0442
change-design-max	0.0001	0.0038	0.0007	0.0382
change-prof-process	N.S.	N.S.	N.S.	N.S.

Table 1.9: p-values of the partial tests for the sectors: metals, comp-machinery, motors, other (where N.S. means that the p-values are not significant).

Variables	Young firms		Old firms	
Sector metals: food products, beverages and tobacco products				
	p-values	adj. p-values	p-values	adj. p-values
red-wat	N.S.	N.S.	0.0368	N.S.
red-mat	N.S.	N.S.	N.S.	N.S.
use-ren-energy	N.S.	N.S.	0.0001	0.00380
red-electr	N.S.	N.S.	N.S.	N.S.
red-waste	N.S.	N.S.	0.0001	0.0038
reuse-waste	0.0001	0.0038	N.S	N.S
transf-waste	N.S.	N.S.	0.0001	0.0038
change-design-min	N.S.	N.S.	N.S.	N.S.
change-design-max	N.S.	N.S.	N.S.	N.S.
change-prod-process	N.S.	N.S.	N.S.	N.S.
comp-machinery: computer, electronic and optical products, machinery				
red-wat	N.S.	N.S.	N.S.	N.S.
red-mat	0.0001	0.0038	N.S.	N.S.
use-ren-energy	N.S.	N.S.	N.S.	N.S.
red-electr	N.S.	N.S.	N.S.	N.S.
red-waste	N.S.	N.S.	0.0343	N.S.
reuse-waste	N.S.	N.S.	N.S.	N.S.
transf-waste	N.S.	N.S.	0.0003	0.0302
change-design-min	0.0001	0.0038	0.0795	N.S.
change-design-max	0.0001	0.0038	N.S.	N.S.
change-prod-process	N.S.	N.S.	N.S.	N.S.
motors: motor vehicles, trailers, semitrailers and other transport equipment				
red-wat	N.S.	N.S.	N.S.	N.S.
red-mat	0.0001	0.0038	N.S.	N.S.
use-ren-energy	N.S.	N.S.	N.S.	N.S.
red-electr	0.0001	0.0038	N.S.	N.S.
red-waste	0.0001	0.0038	N.S.	N.S.
reuse-waste	N.S.	N.S.	0.0001	0.0038
transf-waste	0.0001	0.0038	N.S.	N.S.
change-design-min	N.S.	N.S.	N.S.	N.S .
change-design-max	0.0001	0.0038	0.0001	0.0038
change-prod-process	0.0001	0.0038	N.S.	N.S.
other: other sectors				
red-wat	N.S.	N.S.	N.S.	N.S.
red-mat	0.0512	N.S.	N.S.	N.S.
use-ren-energy	N.S.	N.S.	N.S.	N.S.
red-electr	N.S.	N.S.	N.S.	N.S.
red-waste	0.0001	0.0038	N.S.	N.S.
reuse-waste	0.0001	0.0038	N.S.	N.S.
transf-waste	0.0001	0.0038	0.0252	N.S.
change-design-min	N.S.	N.S.	0.0007	0.0163
change-design-max	N.S.	N.S.	N.S.	N.S.
change-prod-process	N.S.	N.S.	N.S.	N.S.

Table 1.10: Rejection rates of CPTs vs 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		0.47	0.044	200		0.048	0.029
1000		0.053	0.048	1000		0.050	0.052

Table 1.11: Rejection rates of CPTs vs 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 1.12: Rejection rates of CPT based on Fisher's combination vs 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 1.13: Rejection rates of CPT with combination of Fisher vs 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.

n	p	power	p	power
40	0.1	0.150	0.2	0.148
80		0.208		0.235
120		0.291		0.278
160		0.333		0.331
200		0.405		0.432

Table 1.14: Rejection rates of CPT with combination of Fisher vs 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		0.977		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 1.15: Rejection rates of CPT with combination of Fisher vs 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 1.16: Rejection rates of CPT with combination of Fisher vs 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

Table 1.17: Rejection rates of CPT with combination of Fisher vs sample size under the alternative hypothesis, with $k = 10$, $p = 0.2$, $\delta = 0.05$, $\theta_1 = 0.3$, $\rho = 0.3$, $u = 0.1, 0.5, 1, 2, 10$.

n	u	power	u	power	u	power	u	power	u	power
40	0.1	0.106	0.5	0.140	1	0.163	2	0.128	10	0.093
80		0.128		0.204		0.228		0.186		0.124
120		0.151		0.245		0.315		0.292		0.151
160		0.78		0.309		0.344		0.319		0.170
200		0.222		0.376		0.376		0.403		0.171

Table 1.18: 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. Sector: *manufacture of leather and similar items*(significant in bold).

Practices	Prop. small ent.	Prop. medium ent.	Prop .difference small-medium	p-value	Adj-p-value
1. Reduction use of raw materials	0.059	0.333	-0.274	0.002	**0.010**
2. Reduction of emitted waste	0.127	0.467	-0.340	0.002	**0.009**
3. Reduction of waste in production cycle	0.051	0.200	-0.149	0.038	**0.092**
4. Transfer waste to other companies	0.076	0.133	-0.057	0.231	0.328
5. Change design to max recyclability	0.059	0.133	-0.074	0.170	0.328
6. R&D reducing environmental impact	0.051	0.200	-0.149	0.038	**0.092**

Chapter 2

Nonparametric test for time series comparison. Theory and application to ESG titles

The second chapter is dedicated to the application of combined permutation tests (CPTs) to evaluate time series related to ESG titles. Through this statistical methodology, it is possible to consider the comparison between two time series. The basic procedure concerns the decomposition of the general testing problem into partial sub-problems, in particular one problem for all the time points of the time series. From an application point of view, two groups of titles (ESG and non-ESG) are compared in terms of returns relative variation in five years. The aim is to understand the possible existing better performance of one group over the other. Hence, the main research question of this chapter can be formulated as: *Is there a difference between the time series of ESG and non-ESG titles in terms of returns relative variation?*

2.1 Introduction

A large amount of alternatives are available for financial assets and it is important to analyze the distribution of appropriate variables, such as returns, in order to make investment decisions. Generally speaking, the central tendency and the dispersion around the central values, typically measured by mean and variance, are the relevant information on the distribution of these numeric variables. Therefore, estimates of expected returns, deviations and covariates are needed for the implementation of a new portfolio theory (Elton et al., 2013). Let us consider the problem of comparing the performance of two types of investment, say A ad B, according to sample data. In other words two samples of financial titles (A and B) are compared in order to assess which is the (financially) more convenient. Hence, the data under consideration consists of time series of a response variable that represents the financial performance observed on two samples of titles (selected from the population of A-type investments and from the population of B-type investments) and the problem can be defined as a multivariate two-sample test. Specifically, the multidimensionality of the response refers to the number of time points of the series, usually greater than the sample sizes in the analysis of time series. The focus could be on the central tendency of the response, on the variability, or other aspects of interest. For example, location problems are taken into account because the typical comparison between A and B titles relates to expectations of the financial performance but also extensions to other aspects of the distribution such as variability, uncertainty, risks, skewness or others, are possible and can be easily implemented.

The motivating example concerns the comparative assessment of the financial performance of ESG and non-ESG titles. ESG, in line with a typical financial management objective taking account of environmental, social and governance elements, is an acronym that means to deal with the aspects which are covered by all Responsible Investment (RI) activities. As more and more research points out that it has an impact on both risk and returns, many authors in the literature agree that ESG titles have become increasingly popular with a number of investors. Nowadays, focusing only on the financial returns and fundamentals of a particular sector or company has become an understatement exercise and savings investors are increasingly paying attention to ESG factors in their investment decisions. Therefore, it is no longer only with regard to the operations of Institutional Investors but also with respect to investment advisors that ESG criteria have become a key factor for Financial Analysis (Trahan and Jantz, 2023). As a result, from the point of view of responsibility in terms of environmental, social and corporate governance, ESG ratings express a synthetic judgment to validate the sustainability of an issuer, title or fund. However, as yet there is no convergence between the views expressed by experts on the Economic Synergies of ESG Investments and empirical studies to date do not provide an accurate common assessment of their validity. Instead, in some articles in the literature, it is increasingly widely believed that investing in ESG titles is not beneficial in terms of financial returns (Wimmer, 2013).

The degree of success in achieving objectives can be assessed on the basis of Financial Performance. In particular, with regard to academia, today there is a growing attention on the measurement of financial performance. Most strategic business analyses apply accounting or market-based measures to operationalize corporate financial performance: a typical accounting measure is the company's net income, whilst a market measure is the company's market value at the end of the fiscal year (Pham et al., 2021; Riahi-Belkaoui and Picur, 1993).

Particularly, the number of time points with respect to the number of observed companies is a key factor affecting the complexity of the problem. It is equivalent to consider a multivariate two-sample test with a number of response variables, much larger than the sample sizes. Operationally, a vector of data that corresponds to the time series studied for each statistical unit of financial titles was actually observed and analyzed. In such a case, there is no valid parametric solution (Pesarin and Salmaso, 2010). The stochastic dominance alternative hypothesis and the high dependence between response variables (due to the time series autocorrelation), represent the main challenges of this problem. The family of combined permutation tests (CPTs) (Pesarin, 2001) provides flexible and robust solutions for such a complex problem, particularly for problems with a number of variables much larger than the sample size. Pesarin (2001) proposed the use of this nonparametric approach, which has evolved over years as described in several methodological publications, and today can be used effectively for a large range of difficult test problems (Arboretti et al., 2014; Bonnini et al., 2014a; Arboretti et al., 2016).

There are a wide range of different permutation tests within the CPT family and each member, which is characterized by a particular type of combination, displays an individual power behavior with regard to its characteristics and complexity. By analyzing these from a new point of view, the properties of the most common CPTs were analyzed. In addition, a solution to the problem of comparing groups of time series is provided. Finally, we contribute to the empirical literature about the financial performance of ESG titles, by applying such a solution to real financial data.

2.2 The statistical problem of comparing groups of time series

A descriptive method for comparing time series is to use a similarity measure, which includes metric and non-metric methods, to compare two groups of time series and return a value that encodes how similar the two groups are. However, the use of metrics is not always possible. Some non-metric similarity measures provide a different perspective on comparing time series. In fact, they are able to process data that metric methods cannot, and provide results more meaningful than metric methods.

Ramsay (1982), through the functional data analysis, provided a significant methodological contribution to problems related to the comparisons of groups of time-dependent data. His solution is in contrast to the approach commonly used based on a random selection of a sample of data at a limited number of time points, with consequent loss of much of the original available information about the evolution over time of the variables under study. Ramsay et al. (1996) introduced a functional analysis of variance to evaluate the statistical relevance of differences between trajectories over time for groups of functions. Nevertheless, there are two main weaknesses in Ramsey's and other methods based on functional data analysis that relate to the multiple comparisons on which they are based: the extensive need to adjust p-values to control the type I error rate for the multiple comparisons and the computationally demanding procedures that require high processing times. A functional principal component analysis (Ryan et al., 2006; Harrison et al., 2007) has been identified as a valid option for reducing the number of comparisons, but it does not have widespread adoption in practice due to its complexity. Pataky et al. (2015a, 2015b) performed a method based on the combined application of Random Field Theory (Adler and Taylor, 2007) and Statistical Parametric Mapping (Friston et al., 1994). With regard to the departure of normality of the data, this approach is completely parametric and therefore not robust. Niiler (2020) presents an approach with locally weighted scatter-plot smoothing as a within-group statistic and serial Welch t-tests for between-group inferential comparisons. From an explorative (non-inferential) perspective, there is a stream of literature concerning clustering methods. In this context, the comparison of periods may be carried out by means of similarity measures. For the purposes of comparing two time periods and assessing whether they are alike, these similarity measures shall be carried out in either metric or nonmetric methods. But it is not always possible to apply metrics. Some nonmetric similarity measures have the advantage of providing a different perspective on time series comparisons and giving more relevant results as compared with metric methods. However, sometimes, the application of these methods directly to the original time series is computationally complex and therefore shorter representations are created and the pairwise distances between time series can only be estimated. The discrete Fourier transform, the discrete wavelet transformation, a linear aggregate approximation or an overall symbolic approximation are mentioned among these methods (Vlachos et al., 2002; Veltkamp, 2001). As a consequence of this, it is possible to compare dynamic models and therefore multidimensional multivariate series as long as they have the same number of variables. However, where both dynamic models have a number of component variables other than those referred to above, that approach does not apply (Sutcliffe, 2006).

Tapinos and Mendes (2013) proposed the semi-metric time series (Semi Metric Ensemble Time Series or SMETS), to compare multivariate time series with different numbers of components. Even in the presence of large data sets, this algorithm can be used as a computational tool for most common applications.

SMETS uses all the information contained in both time series, which makes this exploratory method unique and preferable to other classic methods such as Euclidean distance, dynamic time bias and Frobenius extended norm whose use is limited to applications where time series are of equal size. In addition to these cases, it is apparent from the literature that there are no other problems or solutions related to comparing groups of time series. Consequently, to the best of our knowledge, our problem and the proposed solution are of great originality and innovation. The problem is a test of hypothesis in which two groups of time series are compared. Each time series is univariate but can be seen as determination of a multidimensional random variable with a certain unknown multivariate distribution, trajectory over time of a specific response observed on a given statistical unit. The null hypothesis is that the underlying distribution of the two groups of time series is equal. The alternative hypothesis can be of multivariate stochastic dominance (one-sided alternative) or inequality of the two multivariate distributions (two-sided alternative). We will consider the one-sided test because more typical in real world problems and more difficult to solve. In other words, two populations are compared not in terms of point observations but with regard to a multiplicity of time series. For example, it might be useful to test whether the financial performance of ESG and non-ESG titles is the same according to a variable observed over time versus the hypothesis that the performance of non-ESG titles is higher.

Formally speaking, we observe n time series of a variable X over T time points $\{x_{ut}; u = 1, \ldots, n; t = 1, \ldots, T\}$ where u denotes a sample statistical unit. Let us assume that x_{ut} is the realization of the random variable X_{jt} with $j = 1, 2$, and the u-th unit belongs to the j-th sample. Without losing generality, assume that the first n_1 statistical units (and dataset rows) belong to sample 1 (selected from population 1) and the subsequent n_2 statistical units belong to sample 2 (selected from population 2). Moreover, $\mu_{jt} = E[X_{jt}]$ and $\sigma_t^2 = Var[X_{jt}] = E[(X_{jt} - \mu_{jt})^2]$ are mean and variance of the probability distribution of X_{jt} respectively. The probability distribution of X_{jt} is completely unknown. The system of hypotheses of the test is:

$$\begin{cases} H_0 : \mu_{1t} = \mu_{2t} \forall t \in \{1, \ldots, T\} \\ H_1 : \mu_{1t} \geq \mu_{2t} \forall t \in \{1, \ldots, T\} \text{ and } \exists t \in \{1, \ldots, T\} \text{ such that } \mu_{1t} > \mu_{2t} \end{cases} . \tag{2.1}$$

Similarly, a problem of equality in distributions versus stochastic dominance can be defined as follows:

$$\begin{cases} H_0 : X_{1t} \stackrel{d}{=} X_{2t} \forall t \in \{1, \ldots, T\} \\ H_1 : X_{1t} \stackrel{d}{\geq} X_{2t} \forall t \in \{1, \ldots, T\} \text{ and } \exists t \in \{1, \ldots, T\} \text{ such that } X_{1t} \stackrel{d}{>} X_{2t} \end{cases} . \tag{2.2}$$

It is very likely that serial autocorrelation exists, but any assumption or estimation of its function is required in the proposed nonparametric solution. The proposed method takes into account and does not ignore the serial dependence structure, but does not need explicit modelling or estimation of it. It is worth nothing that, also under H_0, means and variance may differ over time. Hence stationarity should not be assumed for the application of the proposed method.

2.3 Non-parametric solution

The CPT method is suitable for problems that can be broken down into (dependent) partial tests. For instance, the testing problem of time series comparisons presented above may be decomposed into T partial tests, one for each time point. In brief, we have T partial null hypotheses and T partial alternative hypotheses. Hence the overall problem can be defined as

$$\begin{cases} H_0 : \cap_{t=1}^{T} H_{0t} \\ H_1 : \cup_{t=1}^{T} H_{1t} \end{cases}, \tag{2.3}$$

where $H_{0t} : (\mu_{1t} = \mu_{2t})$ and $H_{1t} : (\mu_{1t} > \mu_{2t})$. This, among other advantages, allows us to define the problem and interpret the results more easily. In the case of marginal or separate inferences, partial tests would be useful (Westfall and Young, 1993; Basso et al., 2009). But jointly considered, they provide information on the overall hypothesis, which represents the main goal of the problem.

The CPT's flexibility and effectiveness are based on its nonparametric nature. Suche method does not require the underlying probability law to be known and the dependence structure of the response variables to be modeled since it is distribution free. However, in order to determine the null distribution of the test statistics, the dependence between partial tests is implicitly taken into account by permuting the rows of the data set. This is a very important advantage, in particular when it is generally too difficult to define and model dependence relationships, e.g. with continuous responses when the actual relationship is not linear (Joe, 1997). Permutation tests generally have much greater powers than parametric counterparts, if the probability conditions applicable to parametric competitors are not met. Nevertheless, the power of permutation tests is slightly lower than that of parametric tests if these conditions are satisfied (Pesarin 1990, 1992, 2001).

On one hand partial tests may help to identify the general problem and at the same time they can provide insufficient information for each specific sub-hypothesis. The p-value adjustments are necessary in order to avoid the inflation of the type I error rate of the overall problem when we wish to attribute the overall significance to one ore more partial tests. In accordance with the theory of CPT, a proper combination of the p-values of the partial tests must be carried out.

Without loss of generality, the test statistics of the partial tests can be defined in such a way that the null hypothesis is rejected for large values. Let T_t and $L_t(z) = P(T_t \geq z|\mathbf{X})$ denote the test statistic and the significance level function of the t-th partial test respectively, with $z \in \mathbb{R}$ and $\mathbf{X}$ denoting the observed dataset. According to CPT theory, we need a suitable combining function for the p-values, with the purpose of solving the general testing problem. This combining function is denoted by $\psi : (0,1)^T \rightarrow \mathbb{R}$. Through the function ψ, the dimensionality of the test statistic is reduced from T to 1, obtaining a (combined) univariate statistic T_ψ informative on the overall problem. Hence, the p-value of the overall test can be computed from the null distribution of T_ψ. For any vector of values $[z_1, \ldots, z_T]$ that can be taken by the T-variate statistic $[T_1, \ldots, T_T]$, the corresponding value of the combined test statistic can be obtained as $T_\psi = \psi(l_1, \ldots, l_T)$ where $l_t = L_t(z_t)$ with $t = 1, \ldots, T$.

Considering the t-th partial test regarding H_{0t} versus H_{1t}, a suitable partial test statistic is $T_t = \sum_{u=1}^{n_1} X_{ut}$, that represents the sum of the first n_1 values of the t-th column of the dataset, i.e. the sum of the values of sample 1. It is worth nothing that this test statistic is permutationally equivalent to the mean of sample 1 and to the difference between the mean of sample 1 and the mean of sample 2. Hence, the observed value of the test statistic of the t-th partial problem is $T_{t,\text{obs}} = \sum_{u=1}^{n_1} x_{ut}$.

Exchangeability (under the null hypothesis) is satisfied and the null permutation distribution of each partial test can be obtained by permuting the rows of $\mathbf{X}$ or, equivalently, re-assigning the (row) vectors of values observed on the statistical units to the two samples. For computational reasons, it is common practice to take into account an approximation, by selecting a random sample of B permutations, according to the so-called Conditional Monte Carlo (CMC) approach. CMC's estimation method provides an unbiased and consistent estimate of the significance level function and of the p-values of permutation tests (Pesarin, 2001).

Let $x^*_{ut,b}$ denote the element of the t-th column in the u-th row of the b-th permuted dataset. The b-th permutation value of the t-th partial test statistic is $T^*_{t,b} = \sum_{u=1}^{n_1} x^*_{ut,b}$. The estimate of the significance level function of the t-th partial test, according to the CMC approach is the same as Equation 1.10:

$$\hat{L}_t(z) = \frac{\sum_{b=1}^{B} I_{(-\infty, T^*_{t,b}]}(z) + 0.5}{B+1} \tag{2.4}$$

where $I_A(z)$ is the indicator function of the set A, i.e. $I_A(z) = 1$ if $z \in A$ and $I_A(z) = 0$ if $z \notin A$. In order to simplify the notation, the significance level function computed from the observed and the b-th permutation values of the t-th partial test statistic may be indicated as $\hat{L}_t(T_{t,\text{obs}}) = l_{t,0}$ and $\hat{L}_t(T^*_{t,b}) = l_{t,b}$ respectively, where $l_{t,0}$ is the marginal p-value of the t-th partial test with $t = 1, \ldots, T$.

Thus, the observed and b-th permutation value of the combined test statistic are $T_{\psi,\text{obs}} = \psi(l_{1,0}, \ldots, l_{T,0})$ and $T^*_{\psi,b} = \psi(l_{1,b}, \ldots, l_{T,b})$ respectively. So the estimated significance level function of the combined test can be written as

$$\hat{L}_{\psi}(z) = \frac{\sum_{b=1}^{B} I_{(-\infty, T^*_{\psi,b}]}(z) + 0.5}{B+1}, \tag{2.5}$$

hence, $\hat{L}_{\psi}(T_{\psi,\text{obs}})$ is the final p-value of the overall test.

In general, the parametric methods used for the solution of univariate and multivariate problems, as a main limitation, have the need to introduce very restrictive assumptions, often unjustified if not impossible to justify, unrealistic, not always clear, difficult to interpret, formulated ad hoc in order to make inference. Moreover, it is not always possible to obtain results by approximation and in most cases the assumptions necessary for the application of these methods (normality, homoscedasticity, independence and an equal distribution of a volatile stochastic component are not met) (Bonnini et al., 2014a).

The combination function ψ shall meet at most the properties set out below (Pesarin, 2001; Goutis et al., 1996):

1. monotonicity: ψ is non-increasing function of each p-value, i.e. if $l_t > l_t^{'}$ ceteris paribus then $\psi(\ldots, l_t, \ldots) \leq \psi(\ldots, l_t^{'}, \ldots)$

2. supremum: when at least one p-value tends to (attains) zero, ψ tends to (attains) its supremum ψ_U (possibly not finite). Formally: $\exists t \in \{1, \ldots, T\} | l_t \to 0 \implies \psi(\ldots, l_t, \ldots) \to \psi_U$

3. finite critical value: for every significance level α, the corresponding critical value of the combined test $t_{\psi,\alpha}$ is finite and strictly less than the supremum. Formally: $\forall \alpha \in (0, 1)$, $t_{\psi,\alpha} < \psi_U$.

Assuming as usual that each row of the dataset corresponds to a statistical unit, and considering for example a test for independent samples, the procedure of the nonparametric combination works as follows:

1. calculate the vector of observed values of the test statistics T, $T_{\text{obs}} = [T_1(X), \ldots, T_k(X)]' = [T_{1(0)}, \ldots, T_{k(0)}]$.

2. consider a permutation of the dataset rows, i.e. a reallocation of group units, and calculate the corresponding values of the test statistics, $T^*_{(1)} = [T_1(X^*_{(1)}), \ldots, T_k(X^*_{(1)})]'$.

3. perform B independent repetitions of step (2) and obtain $T^*_{(b)} = [T^*_{1(b)}, \ldots, T^*_{k(b)}]'$, with $b = 1, \ldots, B$.

4. for each t time point, compute an estimate of the significance level function $\Pr\{T^*_t \geq z\} : \hat{L}_{\psi}(z)$ as in Equation 2.5.

5. for each b, compute $l^*_{t(b)} = \hat{L}_t(T^*_{t(b)})$, $b = 1, \dots, B$ and $l_{t(0)} = \hat{L}_t(T_{t(0)})$, $t = \dots, T$.

6. for each b, compute the combined values $T^*_{\psi(b)} = \psi(l^*_{1(b)}, \dots, l^*_{T(b)})$ and $T_{(0)} = \psi(l_{1(0)}, \dots, l_{T(0)})$ using a suitable combining function ψ.

7. compute the estimated p-value of the test.

The condition of measurability of the combining function ψ is always met since this property can be exploited for any continuous function under permutation. Given that the CPT is not combination invariant, each combination function is a different permutation test and the problem of finding the uniformly most powerful test arises. The main advantage, compared to standard parametric methods, is that the multivariate distribution of the test statistic need not to be known or estimated, and in particular the dependency structure between the variables need not to be explicitly specified. The dependence is implicitly taken into account through the permutation strategy and the application of the combination function ψ. The most common combination functions suggested in the literature are presented below:

- Fisher combining function,

$$T_F = -2 \sum_{t=1}^{T} \log(l_t),$$

- Liptak combining function,

$$T_L = \sum_{t=1}^{T} \phi^{-1}(1 - l_t),$$

- Tippett combining function,

$$T_T = \max_{t \in \{1, \dots, T\}} (1 - l_t).$$

The Tippett combination provides powerful tests when one or a few but not all of the alternative partial hypotheses are true; the Liptak function has a more powerful behavior when all alternative partial hypotheses are jointly true; Fisher's solution is intermediate between the others (Bonnini et al., 2014a). The parametric methods used for the solution of univariate and multivariate problems have some limitation and, in addition, the assumptions on which they are based (normality, homoskedasticity, independence and identical distribution of the erratic stochastic component) are usually rarely satisfied and, even if satisfied, the results are often obtained by approximation.

It can be noticed that CPT solutions, as well as all the permutation tests, are conditional to a set of sufficient statistics. Therefore, it is not necessary to model the dependence between the components of the multivariate response or between the partial tests. Thus, unlike parametric approach, the permutation method is very good in terms of power behavior under very general and mild conditions. The combination takes into account the joint permutation distribution of the partial test statistics, conditional on the observed data. Consequently, the latent dependence relationships that characterizes the population in question are nonparametric. Therefore, even if the null permutation distributions of the partial test statistics correspond to their marginal distributions, the combination approach and the permutation strategy permit the joint analysis of the partial tests through the appropriate determination of the joint (permutation) distribution of the multivariate test statistic.

2.3.1 Generalization of the combination functions

A general formula may be applied to the combining functions which have been introduced in the previous section, where each combination can be regarded as a special case. In particular, the first two are written in summation form as follows:

- Fisher combination function $T_F = -2\sum_{t=1}^{T}\log(l_t)$,
- Liptak combination function $T_L = \sum_{t=1}^{T}\phi^{-1}(1-l_t)$, where ϕ is the distribution function of a standard normal.

On the other hand, it is also possible to write the Tippett combination function as a summation, using the p-norm (see Appendix 3.5 for details), as follows:

$$T_T = \lim_{p\to+\infty}\left(\sum_t |1-l_t|^p\right)^{\frac{1}{p}}. \tag{2.6}$$

In general, therefore, the combining function can be seen as the p-norm of a certain function of the p-values l, i.e., formally, if l is the vector of the T p-values (or of the significance level function), then

$$\psi(l_1,\ldots,l_T) = ||h(l)||_p = \left[\sum_t |h(l_t)|^p\right]^{\frac{1}{p}}. \tag{2.7}$$

Therefore, with respect to what has been said previously regarding the p-norm, the following cases are obtained:

- Fisher: for $p = 1$ and $h(l) = -2\ln(l)$,
- Liptak: for $p = 1$ and $h(l) = \phi^{-1}(1 - l) + k$ with $k \in \mathbb{R}^+$,
- Tippett: for $p \to +\infty$ and $h(l) = 1 - l$

and k is a suitable constant such that $h(l_t)$ assumes non-negative values. This function $h(l)$ transforms each single p-value to obtain, depending on the case, either a relevant increase of the corresponding addendum $h(l_t)$ in the calculation of the value of the test statistic (reward effect) or a relevant decrease of $h(l_t)$ (penalty effect).

In Tippett's rule, we have the limiting case in which there is no penalty for partial tests with high p-value and only the most significant partial test (with lower p-value) is considered. Therefore the significance of just one partial p-value might be enough in order to make the overall combined test significant, because the other partial p-values do not contribute to the value of the combined statistic, and their possible large values do not compensate for the significance of the smallest partial p-value.

In Fisher's combination, due to the convexity of the curve of the $h(\cdot)$ transformation, when a single p-value tends to zero (from the right, because the p-value is a probability and assumes values between 0 and 1), there is a reward effect because the corresponding summand $h(l_t)$ tends to grow much more than proportionally as l_t varies according to the graph of the logarithm function. All partial p-values contribute to the calculation of the value of the test statistic but in cases of high partial p-values, even if tending to one, there is no penalty effect, while each small partial p-value determines a great increase in the combined statistic due to the reward effect.

Finally, in Liptak's function, the curve of $h(l_t)$ is firstly convex (for $l_t < 0.5$) and then concave (for $l_t > 0.5$). As a p-value tends to zero, the growth speed of the transformed value increases generating a reward effect and the symmetry of the graph with respect to the point of coordinates $(0.5, 0)$ is such that as a p-value tends to one, the the speed of decrease of the transformed value increases in a similar way generating a penalty effect. Therefore, only when all or at least many partial tests are significant, and none is non significant, the statistic based on Liptak's combination assumes very high values such that the combined test is significant.

Generally speaking, if the test based on the Liptak combination is preferable because it is more powerful in H_1 when all the partial tests are significant, the test based on the Tippett combination is more effective when only one or several partial tests are significant. It can therefore be considered as the case limit of the reward effect, since it considers only the maximum complement of one of the

p-values to be the minimum p-value, thus ignoring the information provided by all the other p-values.

Table 2.1: Characteristics of the main combining functions considered as a p-norm of the $h(\cdot)$ transformation of the vector of p-values: $\psi_{p,h}(l_1,\ldots,l_T) = ||h(l)||_p$.

Combination	p	$h(l)$	Reward effect	Penalty effect
Tippett	∞	$1-l$	yes	no
Fisher	1	$-2\log(l)$	yes	no
Liptak	1	$\phi^{-1}(1-l)+k$	yes	yes

In terms of the parameter p and the function $h(\cdot)$, the characterization of the main combined permutation tests, viewed as p-norms of the transformation $h(\cdot)$ of the vector of partial p-values, are reported in Table 2.1.

2.4 Simulation study

The method can be applied to any multivariate two-sample test on location, regardless of whether the number q of components of the multidimensional response corresponds to the number T of time points like in our case study. Data were simulated from a q-variate normal distribution with different settings. In addition, multivariate homoscedasticity is assumed. Different correlations were considered, with different number of components (q) and number of components under the alternative hypothesis (q_1). A plausible dependence structure, with observations at different time points correlated only if the time points are close, e.g. consecutive, and not correlated if the time points are far, was assumed. Hence, the random variables that represent the response at the time points v_1 and v_2 respectively, X_{jv_1} and X_{jv_2}, are considered correlated if $|v_1 - v_2| = 1$ and uncorrelated otherwise. All new and original R scripts were specifically developed in order to conduct this simulation study.

The two random variables that represent the multidimensional response are $[X_{j1},\ldots,X_{jq}]' \sim \mathcal{N}_q(\boldsymbol{\mu}_{\mathbf{j}}, \boldsymbol{\Sigma})$ with $j = 1,2$. Furthermore, the vectors of means for the two populations are: $\boldsymbol{\mu}_{\mathbf{1}} = [0,0,\ldots,0]' = \mathbf{0}_{\boldsymbol{q}}$ and $\mu_2 = [\delta,\ldots,\delta,0,\ldots,0]' = \delta[\mathbf{1}'_{\boldsymbol{q_1}}, \mathbf{0}'_{\boldsymbol{q-q_1}}]'$ where δ is the shift parameter (i.e. the "distance" between population 1 and population 2 in terms of central tendency), $\mathbf{0}_{\boldsymbol{q}}$ and $\mathbf{0}_{\boldsymbol{q-q_1}}$ are the zero vectors of size q and $q-q_1$ respectively and $\mathbf{1}_{\boldsymbol{q_1}}$ is the all-ones vector of size q_1. Finally, the correlation matrix Σ, according to what said above, is

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & 0 & 0 & 0 & \dots & 0 & 0 \\ \rho & 1 & \rho & 0 & 0 & \dots & 0 & 0 \\ 0 & \rho & 1 & \rho & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & \rho \\ 0 & 0 & 0 & 0 & 0 & \dots & \rho & 1 \end{pmatrix}.$$

In order to estimate the power, 1000 datasets have been randomly generated for each setting. Each dataset contains n q-dimensional observations in both samples, so $n_1 = n_2 = n$. 1000 CMC resamplings have been taken into account for the estimation of the p-values. The significance level was fixed at $\alpha = 0.05$ and the comparative study was held by computing and comparing the rejection rates of the CPTs based on the combination functions of Fisher, Liptak and Tippett.

The rejection rate of the four permutation tests under the null hypothesis, with $q = 10$ components of a multivariate response, is shown in Table 2.2. Since we're dealing with a null hypothesis, $\delta = 0$ or similarly $q_1 = 0$. The simulations were carried out with different sample sizes, from 10 to 500, and with different correlation levels, i.e. uncorrelation ($\rho = 0$), weak correlation ($\rho = 0.3$), intermediate correlation ($\rho = 0.6$) and strong correlation ($\rho = 0.9$).

The rejection rates under H_0 are expected to be lower than 0.05, the significance level α of the test. Actually such rates in Table 2.2 are less than or very close to α. Table 2.3 shows the rejection rates under the null hypothesis for different numbers of variables q, ranging from 20 to 100, for both small and large sample sizes, $n = 20$ and $n = 200$, respectively, and in the case of null and weak correlation. This result confirms, again, that these three tests are well approximated and the level of significance α is respected when the null hypothesis is true. It can bee seen that, the test based on the Tippett combination, despite having a good behavior in general, seems to be more unstable than the other tests, because it is sometimes anticonservative, especially for small sample sizes (small n) and high number of variables (large q).

Simulations were performed under H_1 with different conditions, to assess the power behaviour of these three permutation tests. Rejection rates were reported as a function of the sample size n, for different correlation values, with the shift parameter δ equal to 0.2, $q = 10$ variables, all in the alternative hypothesis ($q_1 = q$), in Table 2.4. The power of all the tests is increasing function of the sample size, tends to one when n diverges and has a decreasing relationship with ρ. The higher the correlation ceteris paribus, the lower the single marginal informative contribution of each partial test, the lower the power. Since in these simulations $q_1 = q$ and 100% of partial alternative hypotheses are true, the test based on

Table 2.2: Rejection rates under H_0 as a function of sample size n and correlation ρ, with $q = 10$ and $\alpha = 0.05$.

n	ρ	Fisher	Liptak	Tippett	ρ	Fisher	Liptak	Tippett
10	0	0.056	0.052	0.056	0.3	0.054	0.052	0.042
20		0.050	0.049	0.050		0.055	0.053	0.047
30		0.052	0.051	0.050		0.049	0.044	0.045
50		0.052	0.047	0.048		0.048	0.051	0.051
70		0.049	0.058	0.041		0.042	0.044	0.046
100		0.044	0.048	0.047		0.051	0.043	0.052
150		0.052	0.053	0.036		0.052	0.055	0.058
200		0.039	0.036	0.047		0.056	0.056	0.039
250		0.052	0.050	0.038		0.052	0.052	0.052
300		0.053	0.047	0.049		0.045	0.045	0.057
500		0.042	0.047	0.047		0.053	0.053	0.048
n	ρ	Fisher	Liptak	Tippett	ρ	Fisher	Liptak	Tippett
10	0.6	0.047	0.045	0.046	0.9	0.047	0.048	0.050
20		0.048	0.047	0.046		0.048	0.048	0.045
30		0.042	0.038	0.044		0.051	0.050	0.051
50		0.042	0.040	0.048		0.047	0.047	0.050
70		0.049	0.049	0.048		0.045	0.045	0.048
100		0.047	0.049	0.040		0.048	0.047	0.051
150		0.056	0.054	0.054		0.051	0.051	0.055
200		0.057	0.054	0.050		0.050	0.050	0.044
250		0.056	0.056	0.049		0.046	0.046	0.048
300		0.047	0.048	0.039		0.062	0.062	0.065
500		0.054	0.054	0.052		0.044	0.044	0.039

Liptak combining function is the most effective even if its performance is very similar and almost equivalent to the one of the test based on Fisher's combination.

For small and large sample sizes, $n = 20$ and $n = 200$ respectively, with $q_1 = q$, $\delta = 0.2$ and $\rho = 0.3$, the rejection rates are determined as a function of the number q of variables, as shown in Table 2.5. Since we are in the most favorable case to Liptak and Fisher combinations (i.e. $q_1 = q$), the two corresponding CPTs are the most powerful ones. The rejection rates of CPTs are supposed to converge to one for fixed sample sizes when the number of variables q and the global non-centrality induced by the CPT statistics diverge (Pesarin and Salmaso, 2010a). Therefore, an analysis of the rejection rates as a function of the component variables' percentage under the alternative hypothesis $100\frac{q_1}{q}$ % is carried out (see Table 2.6 and Table 2.7).

Table 2.3: Rejection rates under H_0 as a function of sample size n, correlation ρ and number of variables q with $\alpha = 0.05$.

n	q	ρ	Fisher	Liptak	Tippett	ρ	Fisher	Liptak	Tippett
20	20	0	0.060	0.055	0.049	0.3	0.056	0.055	0.057
200	20		0.045	0.045	0.052		0.046	0.051	0.054
20	30		0.056	0.050	0.063		0.048	0.045	0.053
200	30		0.043	0.052	0.060		0.057	0.057	0.051
20	50		0.056	0.058	0.054		0.056	0.057	0.052
200	50		0.050	0.048	0.050		0.043	0.045	0.031
20	70		0.047	0.045	0.070		0.040	0.041	0.065
200	70		0.051	0.054	0.064		0.052	0.054	0.060
20	100		0.049	0.051	0.080		0.063	0.060	0.086
200	100		0.043	0.046	0.090		0.049	0.048	0.079

There is an increasing relationship between the power and the proportion of true alternative hypotheses. The Liptak-based combined test is proved to be the better choice, when all q variables are under H_1, although Fisher's combination shows almost equivalent results. In general, given this output, the Fisher combination test seems to be preferable when the percentage is higher than a threshold, and this threshold appears to increase with q. For instance, the case of small sample sizes is reported in Table 2.6, where the threshold is around 40% when $q = 10$, and it is more than 80% when $q = 100$. On the other hand, the case of large sample sizes is documented in Table 2.7, where the same conclusions hold.

2.5 Case study

The case study in this paragraph concerns an application related to ESG titles which, as mentioned, are becoming the center of many investors' attention. An increasing number of scientific works show that it has implications for both risk and return. The ESG rating (or sustainability rating) expresses a synthetic judgment that certifies the solidity of an issuer, a title or a fund from the point of view of the responsibility in the environmental, social and governance fields. Parameters such as carbon dioxide emissions, efficiency in the use of natural resources, attention to climate change and population growth refer to the environmental impact. On the other hand, the social sphere includes respect for human rights, working conditions and attention to equality and inclusion in the treatment of people. Finally, the governance aspect includes the presence of independent directors, diversity policies and top management remuneration linked to sustainability objectives (see Table 2.8).

Table 2.4: Rejection rates as a function of the sample size for different correlation levels: $q = 10$, $q_1 = q$, $\delta = 0.2$, $\alpha = 0.05$.

n	ρ	Fish.	Lipt.	Tipp.	ρ	Fish.	Lipt.	Tipp.	ρ	Fish.	Lipt.	Tipp.
10	0.3	0.184	0.182	0.154	0.6	0.143	0.142	0.123	0.9	0.102	0.105	0.103
20		0.270	0.270	0.197		0.180	0.180	0.154		0.154	0.154	0.152
30		0.357	0.359	0.242		0.268	0.261	0.247		0.215	0.218	0.211
50		0.490	0.503	0362		0.334	0.328	0.271		0.259	0.259	0.260
70		0.612	0.611	0.434		0.416	0.416	0.360		0.352	0.351	0.351
100		0.729	0.730	0.591		0.571	0.571	0.504		0.412	0.414	0.413
150		0.864	0.868	0.728		0.673	0.675	0.613		0.588	0.593	0.569
200		0.948	0.949	0.845		0.796	0.793	0.741		0.662	0.663	0.651
250		0.978	0.979	0.915		0.851	0.855	0.806		0.761	0.762	0.736
300		0.989	0.991	0.950		0.906	0.905	0.877		0.831	0.832	0.804
500		1.000	1.000	0.999		0.985	0.986	0.978		0.951	0.951	0.944

Table 2.5: Rejection rates as a function of the number of variables q for different sample sizes: $q_1 = q$, $\delta = 0.2$, $\rho = 0.3$, $\alpha = 0.05$.

q	$n_1 = n_2$	Fisher	Liptak	Tippett	$n_1 = n_2$	Fisher	Liptak	Tippett
10	20	0.279	0.286	0.199	200	0.940	0.946	0.849
20		0.258	0.259	0.200		0.959	0.956	0.870
30		0.250	0.263	0.178		0.961	0.965	0.877
40		0.315	0.311	0.231		0.968	0.967	0.901
50		0.258	0.258	0.171		0.969	0.969	0.855
70		0.294	0.293	0.230		0.967	0.971	0.909
100		0.287	0.287	0.268		0.966	0.967	0.933

Bender et al. (2017) argue that the potential for ESG to become part of a global equity portfolio is being stimulated by increasing data availability. ESG is a source of fresh and valuable information to investors, which increases the potential for return as well as risk (Ting et al., 2019). In support of these investments, Giudici and Bonaventura (2019) show that firms with higher ESG ratings achieve higher differential returns, with a standard deviation that is not significantly different from Khan et al. (2016). However, some works in literature argue that it is not always possible to have a greater return for titles with an excellent ESG score. There are indications of a significant discrepancy in ESG ratings between small and medium enterprises, given that firms with very high market capitalization and revenues consistently receive higher ESG scores than those with low levels of revenue. The predictive power of ESG indicators is inconsistent and literature has shown that, while some high ESG portfolios may be able to outperform the overall

Table 2.6: Rejection rates as a function of the percentage of partial tests under H_1 for different numbers of variables q: $n = 20$, $\delta = 0.2$, $\rho = 0.3$, $\alpha = 0.05$.

%	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.
10	10	0.055	0.056	0.055	20	0.058	0.053	0.080	30	0.065	0.056	0.079
20		0.071	0.074	0.081		0.076	0.073	0.085		0.084	0.077	0.089
30		0.105	0.108	0.111		0.092	0.086	0.094		0.087	0.083	0.094
40		0.112	0.103	0.107		0.136	0.125	0.129		0.128	0.129	0.135
50		0.139	0.139	0.142		0.145	0.143	0.136		0.145	0.135	0.153
70		0.175	0.169	0.143		0.216	0.193	0.198		0.190	0.181	0.158
100		0.279	0.286	0.199		0.258	0.259	0.200		0.250	0.263	0.178
%	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.
10	40	0.047	0.043	0.062	70	0.058	0.059	0.070	100	0.067	0.067	0.090
20		0.068	0.065	0.080		0.064	0.065	0.093		0.080	0.076	0.123
30		0.101	0.096	0.099		0.109	0.100	0.125		0.092	0.084	0.160
40		0.122	0.109	0.141		0.113	0.110	0.143		0.126	0.111	0.179
50		0.159	0.148	0.149		0.147	0.135	0.156		0.156	0.146	0.199
70		0.195	0.186	0.189		0.215	0.207	0.223		0.207	0.193	0.227
100		0.315	0.311	0.231		0.294	0.293	0.230		0.287	0.287	0.268

market, low ESG portfolios are also capable of doing so. It is therefore not possible to link ESG scores and financial performance with each other (Boffo and Patalano, 2020). Research and empirical evidence suggest significant benefits from ESG integration, but the myth of sustainability as a low-cost or low-return investment persists. It may be because exclusion screens, as an integration tool for ESGs, have continued to be used by many conventional investments. The investment practices do not allow additional value to be generated from sustainability initiatives in some companies and so, the returns that were expected by Kotsantonis et al. (2016), have not already been achieved through these funds.

Recent contributions argue that further studies on this subject are needed because the relationship between ESG score and brand value has changed over time, (see El Zein et al., 2019). In addition, it is noted in Auer and Schuhmacher (2016) that there are inconsistencies between the different geographic regions. In contrast to passive investments on the stock exchange, active selection of titles with high or low ESG ratings does not guarantee good results, regardless of geographical region, industry or ESG criterion. Lastly, these results continue to vary in terms of the percentage of portfolio termination used, time frame or impact on transaction costs.

In addition, financial performances are strongly influenced by the geographical and economic framework of an investment strategy that is underpinned by ESG

Table 2.7: Rejection rates as a function of the percentage of partial tests under H_1 for different numbers of variables q: $n = 200$, $\delta = 0.2$, $\rho = 0.3$, $\alpha = 0.05$.

%	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.
10	10	0.124	0.102	0.306	20	0.130	0.105	0.381	30	0.135	0.108	0.410
20		0.206	0.148	0.443		0.227	0.164	0.261		0.244	0.168	0.575
30		0.369	0.245	0.589		0.376	0.261	0.651		0.347	0.257	0.655
40		0.498	0.368	0.650		0.561	0.401	0.717		0.536	0.376	0.707
50		0.628	0.469	0.701		0.663	0.510	0.744		0.697	0.543	0.778
70		0.816	0.747	0.764		0.840	0.766	0.792		0.878	0.802	0.841
100		0.940	0.946	0.849		0.959	0.956	0.870		0.961	0.965	0.877
%	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.	q	Fish.	Lipt.	Tipp.
10	40	0.113	0.097	0.456	70	0.115	0.092	0.486	100	0.113	0.092	0.555
20		0.221	0.152	0.582		0.218	0.147	0.634		0.245	0.164	0.708
30		0.392	0.268	0.716		0.375	0.233	0.722		0.413	0.276	0.815
40		0.559	0.405	0.744		0.561	0.406	0.782		0.547	0.391	0.856
50		0.707	0.547	0.799		0.713	0.540	0.828		0.704	0.545	0.880
70		0.885	0.802	0.850		0.895	0.805	0.860		0.899	0.827	0.926
100		0.968	0.967	0.901		0.967	0.971	0.909		0.966	0.967	0.933

Table 2.8: ESG pillars and components (source Boffo and Patalano, 2020).

Environmental factors	Social Factors	Governance factors
Natural resource use	Workforce	Board independence
Carbon emissions	Human rights	Board diversity
Energy efficiency	Diversity	Shareholder rights
Pollution/waste	Supply chain	Management compensation
Environmental opportunities		Corporate ethics

criteria. In support of this hypothesis, Wimmer (2013) believes that value-oriented mutual fund investors seeking high ESG risk investments cannot count on a long-term continuation of high ESG scores and therefore must re-balance their wallet from time to time. The second consequence of the latter is that, for approximately two years following investment in a high-risk ESG fund, investors will have been able to count on persistently high ESG scores. While this period is relatively long compared to the persistence of financial performance reported in the literature, investors looking to see their wealth invested in companies with a high CSR (Corporate Social Responsibility) profile still need to monitor their investments over time to ensure that their mutual fund is still invested in companies with a high level of commitment to corporate social responsibility (Wimmer, 2013). Thus, there is a need to carry on additional investigations in order to strengthen the results achieved so far.

We analyzed data downloaded from the Yahoo Finance website. The time interval from 31/12/2013 to 31/12/2018 (a five-year period) was considered. All the titles in the S&P 500 list (Steadman and Perrone, 2019) updated to 31/12/2018 were selected. ESG titles were differentiated from non-ESG titles. For each title, the time series of the daily returns in the mentioned time period was downloaded. The closing return value has been taken into account in respect of each date. The Standard & Poor's 500 (S&P 500) is the most important North American stock index (Frino and Gallagher, 2001). This basket has become more important for investors in recent years, even though it was historically the first index developed by Dow Jones. In fact, it is a leading benchmark for stocks listed on the New York Stock Exchange and it underpins an incredibly broad range of derivatives products such as futures, options and certificates. The value of the S&P 500 is calculated automatically every 15 seconds based on the prices of the last contracts concluded during trading hours, i.e. from 09:30 (15:30 Italian time) to 16:00 (22:00 Italian time).) from Reuters America, a Thomson Reuters Corporation company.

Therefore, the total daily return of an overall 453 companies with 314 not-ESG and 139 being ESG is included in this dataset for a considered five-year period. The period is made up of 1258 observations as shown below. Some summary graphs on the trend of the five summary variables relating to the box-plot to compare ESG titles and non-ESG titles at a descriptive level can be seen in Figure 2.1.

Original R scripts were specifically created to carry out the statistical analysis.

Since the goal of the analysis is the comparison of the "evolution" over time of the returns of the two groups of time series, in a dynamic perspective, regardless of the starting values, we considered the relative variation of the returns. Specifically, for each company u and daily date t (with $t = 0, \ldots, T$ and $T = 1257$), we computed the relative variations $d_{ut} = \frac{r_{ut} - r_{u0}}{r_{u0}}$, where r_{ut} represents the return of the u-th company (financial title) in the day t. The relative differences between the two groups of titles are represented in Figure 2.2 by the time series of the sample.

From a descriptive point of view, it is clear that, except for the last few hundred dates in which a decreasing trend has been observed, the means of return relative variations between ESG and non-ESG titles tend to increase over time. There's no systematic dominance of one series over another, but the mean return relative variations of the non-ESG titles are higher than those of the ESG titles in some sub-periods.

Let D_{1t} denote the random variable from which the return relative variations observed at time t for the ESG titles are assumed to be generated. Similarly, D_{2t} represents the random variable of which the return relative variations observed at

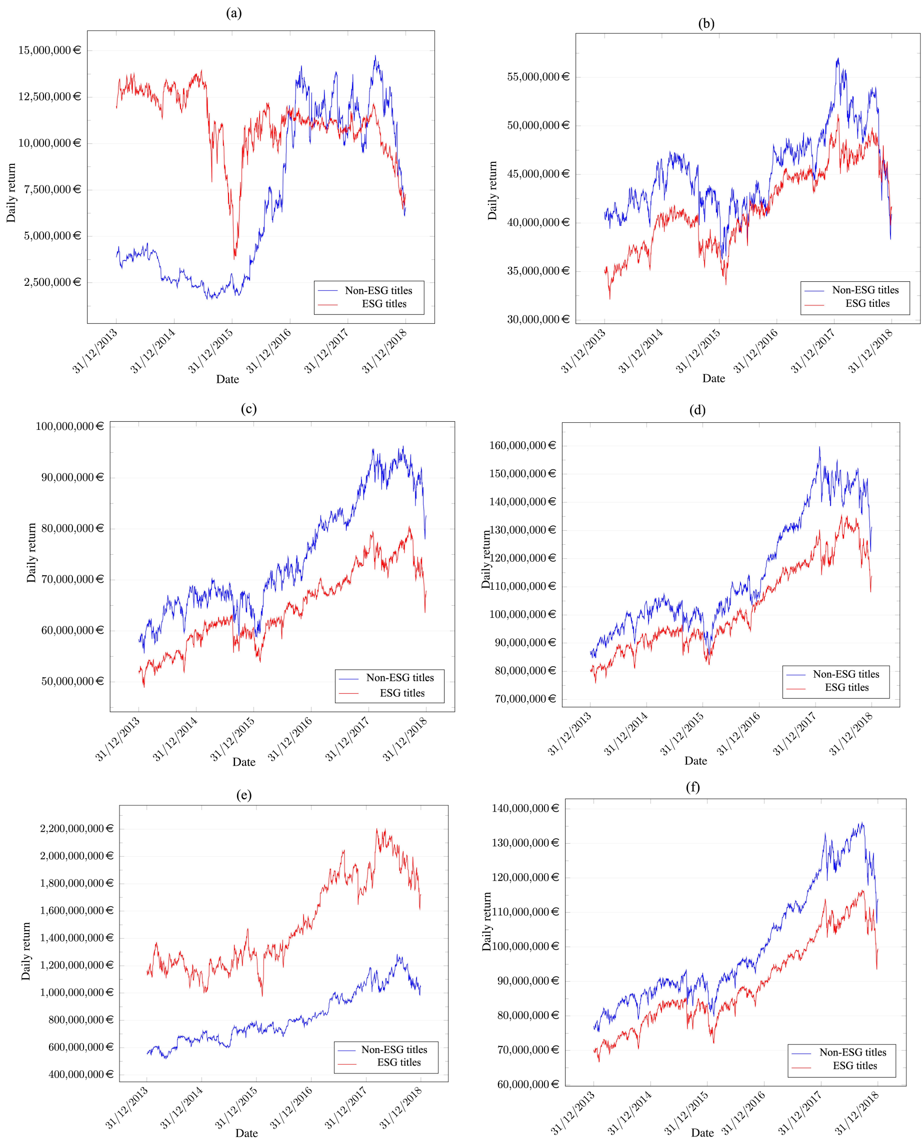


Figure 2.1: Descriptive sample statistics of the daily returns over time in the considered time period for the two types of titles. (**a**) minimum, (**b**) first quartile, (**c**) median, (**d**) third quartile, (**e**) maximum, (**f**) mean.

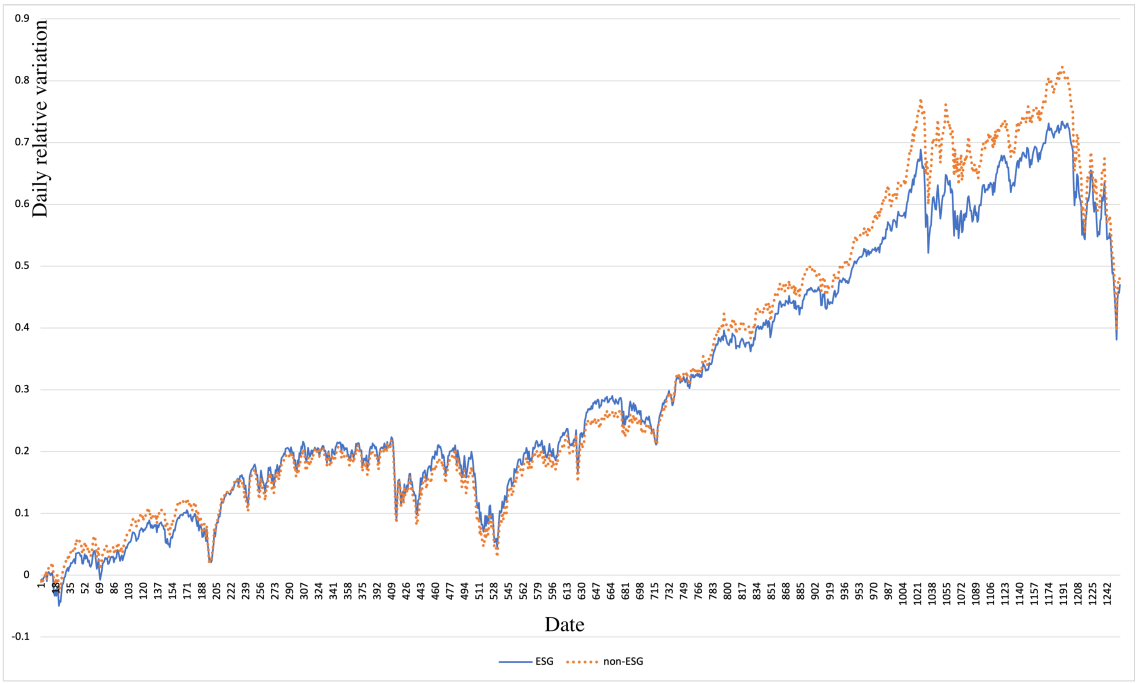


Figure 2.2: Time series of the sample means of the return relative variations of ESG and non-ESG financial titles.

time t for the non-ESG titles are realizations. The testing problem is the following:

$$\begin{cases} H_0 : \cap_{t=1}^{T}(D_{1t} \stackrel{d}{=} D_{2t}) \\ H_1 : \cup_{t=1}^{T}(D_{1t} \stackrel{d}{<} D_{2t}) \end{cases} . \tag{2.8}$$

The aim is to test the null hypothesis versus the alternative hypothesis in Equation 2.8 at the significance level $\alpha = 0.10$. In other words, the purpose is to see whether the non-ESG titles show better financial performance than the ESG titles in one or more time points, and therefore they are financially more advantageous in terms of relative variations.

At this point, Tippett's combination function was applied as it provides powerful tests when one or a few but not all of the alternative partial hypotheses are true. This was also demonstrated in Section 2.4 through the simulation study. The p-value of the test is 0.032. Since it is less than α, the null hypothesis of equality in distribution is rejected in favor of the alternative hypothesis of stochastic dominance. In this way, the hypothesis that non-ESG titles have greater convenience in terms of relative differences between returns is proven. Given that the difference in the sample means is significant in some of the 1257 considered time points, we can deepen the analysis to detect the specific partial tests that contribute to such a significance, i.e. the dates or time points where the sample means are significantly different. To control the Family Wise Error Rate (FWER) and avoid the overall type I error rate being greater than α because of the multiplicity of tests, a suitable method for the adjustment of the 1257 partial p-values is required. A closed testing method is appropriate and advisable because controls the FWER in a strong sense. In particular, we applied the minP method (Westfall and Young, 1992; Westfall and Young, 1989). In Table 2.9, the significant adjusted p-values and the corresponding time points are reported.

Table 2.9: Significant partial adjusted p-values and corresponding time points.

date	p-value
14	0.068
15	0.054
16	0.068
17	0.068
18	0.054
19	0.054
20	0.068
21	0.032
22	0.032
23	0.032
24	0.054
25	0.092
64	0.092

The difference between the time series of non-ESG titles and that of ESG titles, is significant only in 13 out of the 1257 time points. In particular, the first month and the beginning of the second month, i.e. January and February 2014, are covered by these 13 dates.

2.6 Conclusions

Combined permutation tests are an interesting, appropriate, powerful and therefore preferable solution, given the well-known limitations of parametric methods, for multivariate testing problems due to very restrictive assumptions, often unjustified, unrealistic and formulated ad hoc for the purpose of making inferences we focused on the two sample problem for continuous numeric variable, consistent to the two-sample test for time series comparisons.

We proved that the CPTs based on the three most popular combination functions are well approximated and satisfy unbiasedness, consistency, and the increasing relationship between power and proportion of true partial alternative hypotheses. They have a good power behavior, also (but not only) with small sample sizes and a large number of variables. These nonparametric tests represent a valid solution to testing problems on the comparison of groups of time series.

In this chapter, a two-sample test for the location of numeric variables has been considered, but the solution can be easily extended to other frameworks such as multisample tests, categorical variables, multispect comparisons, etc., based on the combination of p-values. For the combined permutation test, there is a significant advantage in comparison to standard parametric methods: the multivariate distribution of the test statistic does not need to be assumed or estimated. The dependence structures between variables are implicitly taken into account and they need not be explicitly known or modeled.

Finally, the application of the CPT based on the Tippett combination to the case study concerning the comparative evaluation of the financial performance of ESG and non-ESG titles, provides a contribution to the empirical literature on this topic. The empirical evidence supports the hypothesis that non-ESG titles are more financially convenient than ESG titles, at least in terms of return relative variation.

Chapter 3

Economic performances of "circular" SMEs: a nonparametric multivariate regression

This chapter is dedicated to the application of combined permutation tests (CPTs) to multivariate regression analysis. Such methodology is suitable to test the significance of the estimates of the regression coefficients of all the equations jointly considered. In fact, the decomposition of the general testing problem into partial sub-problems is at the basis of the CPTs, as reported in previous chapters. With regard to the application, the characteristics of Italian SMEs such as age and size, together with variables that represent the "circularity" of these companies are the explanatory variables of the model to predict the economic performances of companies (i.e. Net Working Capital, Revenues, ROA,...). The main research question of this chapter can be formulated as: *How do some characteristics and the "circularity" of SMEs affect their economic performances?*

3.1 Introduction

SMEs are the most widespread form of business throughout the world and constitute the majority of a nation's productive facilities. They are crucial not only for the survival of a country's economy, but also for the enormous amount of products (goods and services) they have the possibility to offer in an economic system (Berg et al., 2021). In depth, according to the European Commission (2023), they represent 99.8% of European companies and are also responsible for 65% of the employment in the European Union, hence embracing an important ethical value. Due to these non-negligible percentages, SMEs should carry the huge responsibility of protecting not only their image and reputation but also that of the country they belong to and be available to spread initiatives and behaviours. Indeed, because of their relatively smaller dimensions, SMEs can be very flexible in adapting to changes and in surviving shocks that are caused by both exogenous and endogenous factors. They also have the ability to develop closer connections with their customers and the social community in general, even more so when such SMEs are family businesses. This last statement suggests that small and medium enterprises are deemed as the most important production-related entities present in an economic system, able to create, enhance, and propagate networks and linkages with consumers (Abbasi et al., 2017).

On the other hand, SMEs also have weaknesses and vulnerabilities. First of all, many of those already established usually lack the knowledge and skills in financial management and accounting, and this might be very dangerous in that it compromises the possibility of success for these entities and, most of the time, their survival. Indeed, a good financial structure and management are deemed essential to monitor the capability of a company to cope with its short and long-term obligations, maintain solid financial and economic stability, boost the morale of employees and other stakeholders, balance equity and debt, and also keep a safety stock of liquidity for unexpected events or "simply" short-term debts. Such activities represent not only sustainable and circular behaviours. The lack of such practices is among the main and most relevant reasons why a non-negligible number of small and medium businesses fail after 5 years of activity (Fundera, 2020). Another fragility of small and medium firms concerns the difficulties they usually face when building their customer base and attracting customers: in fact, when it comes to competition, the larger the company, the easier to build and expand the group of trusted consumers. In particular, large firms are likely to dispose of a greater amount of funds and capital they can devote, for instance, to advertising, while small entities have a restricted number of options among which priorities are allocated, also because of limited access to finance. Small businesses face financial constraints that make sustainability and circularity difficult to achieve, as per discussions on the barriers to eco-innovation (Ghisetti, 2017).

In the long list of materials considered for recycling and reuse, *plastic* is among the most interesting ones. In fact, especially in recent years, many researchers have focused on the adoption of circular economy patterns by SMEs in the plastic industry. Maione et al. (2022) conducted a study on the circular practices of Italian plastic packaging companies, highlighting trends and areas for improvement. A qualitative analysis was conducted on the behaviors of 19 firms in the years 2011, 2015, and 2019. The study showed that circular practices have increased over the years, with a particular focus on recycling used packaging and improving connectivity between different actors in the value chain. It is becoming increasingly clear that circular and sustainable practices are essential. It is important to ensure that these practices are employed throughout all stages of production. This includes implementing innovative and resource-efficient business models, utilizing reuse and refill systems, adopting sustainable substitute materials, implementing effective waste management technologies, and enacting government policies that support sustainability (Lau et al., 2020) must be even more encouraged, spread and improved. It goes without saying that Circular Economy and sustainability present unique challenges for the discussed sector. Investigating the barriers to plastic circularity, some research has been conducted by Paletta et al. (2019), taking into account a set of Italian SMEs in the Emilia-Romagna region involved in plastic-converting operations. What is notable is that only a few companies are utilizing non-virgin plastics. This is due to various reasons such as the possible presence of impurities, which makes reusing plastic difficult. Additionally, fluctuating prices, the use of chemicals in previous productions, and the lack of legislative support for alternatives to virgin plastics are contributing factors.

Another important aspect is emphasized in the work of D'Angelo et al. (2023). In this study, the authors tested a model in which it seems that an inverted U-shaped effect is brought about by "Circular Economy Activities" on "Economic Performances". In particular, a firm's economic performance benefits from a limited number of activities. Consumers perceive the firm's environmental efforts positively and reward it by choosing it over competitors, resulting in increased sales. However, in some cases, investing in corporate social responsibility activities may not yield significant benefits if customers do not perceive the company's efforts to be aligned with their core business. For example, focusing on recycling, efficient use of resources and green logistics may not be enough to offset some negative impacts on the benefits of the company's efforts.

Since we are dealing with a highly debated topic in the literature, the objective is to contribute to the empirical literature regarding the relationship between the economic performance of SMEs and their "circularity".

From a statistical point of view, the concept of "economic performance" is multidimensional because it includes several partial aspects. Therefore, the dependent variable of the regression model is not univariate. As a consequence, we consider utilizing a multivariate regression model. A combined permutation test shall be performed in order to assess the effect of explanatory variables on the dependent variables (Pesarin, 2001). It combines all the tests on the significance of the estimates of the single regression coefficients. It is a non-parametric solution because it follows the permutation approach. Hence, it is more flexible and robust with respect to the underlying distribution than, for instance, Pillai's test (Harrar and Bathke, 2008) or other typical parametric methods. The general version of Pillai's test statistic is the trace of the product of the extra sum of squares and cross product (SSCP) matrix and the inverse of the error SSCP matrix of the reduced model. The extra SSCP matrix is the difference between the error SSCP of the reduced model (the model under the null hypothesis) and the error SSCP of the full model. One way to calculate Pillai's test statistic is based on the non-zero eigenvalues of the product between the extra SSCP matrix and the inverse of the error SSCP matrix of the full model. Under the null hypothesis, this test statistic approximately follows an F distribution, and the reliability of the inferential results depends on the level of approximation of such null distribution (Johnson and Wichern, 2007). The methodology applied in this Chapter is suitable for both normal and non-normal (multivariate) errors and, above all, it does not make assumptions about the dependence structure of the error terms. Permutation tests have been used in linear regression models since the 1930s (Fisher, 1935; Pitman, 1937a; Pitman, 1937b; Pitman, 1937c). To test the significance of the estimate of one regression coefficient in the presence of nuisance explanatory variables (with the need to control for the effects of such variables), several authors proposed solutions based on the idea that a suitable test statistic is the square of the correlation between the residuals of the regression of the response on the nuisance predictors and the residuals of the regression of the predictor of interest on the nuisance variables (Freedman and Lane, 1983; Kennedy, 1995; Anderson and Robinson, 2001). The approach based on the so-called permutation of regression residuals (PRR) consists of replacing the explanatory variable of interest in the model with the residuals of the linear regression of this explanatory variable on the other predictors. It was proposed by Potter (2005) for logistic regression (Potter, 2005), and then extended to a General Linear Model (Werft and Benner, 2010). In order to obtain the null distribution of the test statistic and compute the p-value, an alternative approach is the conditional permutation of the dependent variable (Kennedy, 1995). Mixed models have been effectively analyzed using permutation tests (Basso and Finos, 2012) and to Generalized Linear Models (Winkler et al., 2014; Goeman et al., 2011). For goodness-of-fit tests in linear regression models, some permutation solutions are based on partial sums or on cumulative sums of residuals (Stute et al., 1998; Hattab and Christensen, 2018; Blagus et al., 2019). Basso and Finos (2012) propose to use a non-parametric combination of dependent permutation tests to infer the effects of covariates. Another work wherein the overall test of the multivariate linear model is conceived as a multiple test and it was proposed by Solari et al. (2014). This solution is based on the different and effective theory of rotation tests, rather than on the permutation approach.

3.2 Statistical problem

Since the performance of companies, is multidimensional, specifically represented by twenty-one variables, a suitable model to investigate the relationship between economic performance and the explanatory variables concerning circular activities and characteristics of companies must be composed of twenty-one equations, one for each response. In other words, a multivariate regression analysis must be carried out.

In general, let us assume that the multivariate dependent variable is q-dimensional and that k predictors are considered in the study. The multivariate (specifically q-variate) linear model is:

$$Y_{ij} = \beta_{0j} + \sum_{v=1}^{k} \beta_{vj} x_{iv} + \varepsilon_{ij}, \tag{3.1}$$

with $i = 1, \ldots, n$ (companies) and $j = 1, \ldots, q$ (dependent variables). In the classic regression analysis, the error terms are supposed to be normally distributed with a null mean and constant variance within each equation, and uncorrelated (consequently independent because Gaussian) with respect to units. Formally, for the random variable ε_{ij}, the conditions $E[\varepsilon_{ij}] = 0$, $\mathrm{Var}[\varepsilon_{ij}] = \mathrm{Var}[Y_{ij}] = E[\varepsilon_{ij}^2] = \sigma_j^2$ and $\mathrm{Cov}[\varepsilon_{ij}\varepsilon_{ur}] = E[\varepsilon_{ij}\varepsilon_{ur}] = 0$ with $i \neq u = 1, \ldots, n$ and $j, r = 1, \ldots, q$. It is worth noting that $\mathrm{Cov}[\varepsilon_{ij}\varepsilon_{ir}] = E[\varepsilon_{ij}\varepsilon_{ir}]$ could be not equal to 0 because, for a given statistical unit, the errors referring to different responses could be correlated, thus the equations could be related.

The matrix representation of the model is:

$$\underset{n \times q}{Y} = \underset{n \times (k+1)(k+1) \times q}{XB} \underset{n \times q}{+E} \tag{3.2}$$

where the first column of X is the vector of ones corresponding to the q constants of the model. If we consider the generic i-th row of the random matrix E, denoted by $\varepsilon_{(i)}$, the model error's probabilistic assumptions can be represented as follows:

$$\varepsilon_{(i)} = (\varepsilon_{i1}, \ldots, \varepsilon_{iq}) \sim \mathcal{N}_q(0_q, \Sigma) \tag{3.3}$$

where the q-variate errors, in the classic multivariate linear regression model, are independent and identically (normally) distributed random variables, with null vector of means and a constant $q \times q$ covariance matrix Σ. In general, for inferential purposes, the assumptions of normality and independence can be relaxed. The use of a nonparametric approach makes model inference robust to non-normality of errors, without assuming a specific distribution family. The flexibility of a valid approach, without assuming independence in error terms, increases the possibility of inferential reliable results. For these main reasons, the permutation method was adopted to test the significance of regression coefficients. In this case, the assumption of independence is replaced by the milder condition of exchangeability of errors across units. This assumption is satisfied because, under the null hypothesis, all the regression coefficients are equal to zero, and the model in each of the q equations includes only the intercept. Consequently, when the null hypothesis H_0 is true, the conditional mean of Y given X does not depend on the predictors, and the null distribution of the test statistic can be obtained by permuting the residuals or, equivalently, reassigning the q-dimensional rows of Y to the $(k+1)$-dimensional rows of X (or vice-versa). In fact, the interest concerns testing the significance of all the regression coefficients, jointly considered, that is, testing the null hypothesis that no explanatory variable affects any dependent variable, versus the alternative hypothesis that at least one explanatory variable affects at least one dependent variable (in other words, the negation of the null hypothesis). In terms of regression coefficients, the hypothesis under test can be represented as follows:

$$\begin{cases} H_0: & \beta_{11} = \beta_{12} = \ldots = \beta_{kq} = 0 \\ H_1: & \bar{H}_0 \end{cases}. \tag{3.4}$$

The problem reported in Equation 3.4 is the classic Multivariate Analysis of Variance (MANOVA) of the linear regression model.

3.3 Methodological solution

The testing problem defined above can be considered as a multiple test composed of all the tests on the single regression coefficients (partial tests). The null and alternative hypotheses of the partial test concerning the coefficient related to the v-th independent variable and the j-th dependent variable can be represented as $H_0^{vj}: \beta_{vj} = 0$ and $H_1^{vj}: \beta_{vj} \neq 0$ respectively. Hence, the null and the alternative hypotheses of the overall problem are the following:

$$\begin{cases} H_0: \cap_{v=1}^{k} \cap_{j=1}^{q} H_0^{vj} \\ H_1: \cup_{v=1}^{k} \cup_{j=1}^{q} H_1^{vj} \end{cases} \tag{3.5}$$

where the symbol of "intersection" implies that, under the overall null hypothesis, all the partial null hypotheses are true, and the symbol of "union" means that under the overall alternative hypothesis, at least one partial alternative hypothesis is true.

The methodological solution applied to this problem consists of a combined permutation test. This method is useful for solving complex problems, such as multivariate problems or those requiring multivariate statistical tests. One significant benefit of this method compared to parametric methods is that there is no need to assume the multivariate distribution of the test statistic or estimate dependence parameters. Additionally, there is no need to know or model the dependence structure between variables (Bonnini et al., 2014a). Permutation tests are distribution-free, hence they are flexible and robust with respect to the departure from normality (Pesarin and Salmaso, 2010).

The main idea of a combined permutation test is to find a suitable test statistic to solve each partial test, and to combine the permutation p-values of the partial tests in order to solve the overall problem (Bonnini et al., 2014a). The absolute values of the regression coefficients' least square estimators are appropriate test statistics for the partial tests. The procedure consists of the following steps:

1. Computation of the vector of observed values of the test statistics $t_0 = (|b_{11}|, |b_{12}|, \dots, |b_{kq}|) = t(X)$;
2. B independent random permutations of the rows of the X matrix: $X_1^*, X_2^*, \dots, X_B^*$;
3. Computation of the values of the test statistic vector for the B dataset permutations t_b^* and the corresponding vector of p-values l_b^*, with $b = 1, 2, \dots, B$;
4. Computation of the value of the combined test statistic for each permutation and for the observed dataset through the combination of the partial p-values with a suitable function $\psi : [0, 1]^{kq} \to \mathbb{R}$, $t^*_{\text{comb},b} = \psi(l_b^*)$;
5. Computation of the p-value of the combined test according to the null permutation distribution.

The dependence of the partial tests is implicitly taken into account through the permutation of the rows of the X matrix. Assuming, without loss of generality, that the null (partial and overall) hypotheses are rejected for large values of the test statistics, a suitable combining function ψ should satisfy the following reasonable and mild conditions: (1) it is a monotonic non-increasing function of the p-values; (2) when one p-value tends to zero, it tends to the supremum, and when one p-value tends to one it tends to the infimum; (3) the acceptance region is limited. One of the most commonly used combining functions is that of Tippett:

$$t^*_{\text{comb},b} = \max_{v,j} \left(1 - l^*_{vj,b}\right). \tag{3.6}$$

In particular, Tippett's combination function provides powerful tests when one or few partial alternative hypotheses are true (Bonnini et al., 2014a). The parametric methods used to solve testing problems, both univariate and multivariate, are based on restrictive assumptions that are often unrealistic and lack empirical evidence, or asymptotic theories. In particular, the typical assumptions of the classic linear regression model, such as the normality and uncorrelation of the error terms, are rarely satisfied or only approximately satisfied. The permutation tests are preferable to the parametric tests when the parametric assumptions do not hold. However, they are performant also when these assumptions are valid, and in general, they represent a robust and flexible solution for complex testing problems. Since the permutation MANOVA solution for the linear regression model is defined as a multiple test, in the case of a rejection of the null hypothesis in favour of the alternative and in order to attribute the overall significance to specific partial tests (i.e., to specific coefficient estimates), the control of the family-wise error (FWE) is necessary (Westfall and Young, 1992a). In other words, to avoid the inflation of the type I error of the overall test, we must adjust the partial p-values. A suitable method is the Bonferroni-Holm rule. The typical parametric approach to test the goodness-of-fit of a model is based either on a general test on the whole model, or on a stepwise procedure based on the sequential application of t-tests on the significance of the single regression coefficients. The typical general test on the whole model is the F test in the univariate model, Pillai's trace, Wilk's lambda, or Roy's largest root in the multivariate case (Ates et al., 2019). For this reason, the use of the described nonparametric method, based on the combined permutation test and on the adjustment of partial p-values, is not only appropriate in cases of possible violations of the assumption of normality or other typical assumptions of the parametric tests, but also consistent with the definition of the MANOVA problem as the multiple test composed of $q \times k$ partial tests, represented in Equations (3.4)–(3.5). When the number of partial tests is much larger than the sample size, the parametric approach results in a loss of degrees of freedom and reduced power, which is an important limitation. In regression analysis, the parametric approach is not applicable when the number of explanatory variables exceeds the sample size. On the other hand, when the number of partial tests under the alternative hypothesis increases (e.g., when new predictors whose regression coefficients are not null are added in the model), the power of the combined permutation tests increases. Therefore, even if the sample size is smaller than the number of independent variables, the nonparametric procedure remains feasible and powerful (Pesarin and Salmaso, 2010a). When analyzing regression, it's important to address multicollinearity. To this aim, compute Variance Inflation Factors (VIFs), before carrying out the permutation MANOVA, especially when there are many predictors and eliminate from the model the predictors with high VIF (usually VIF greater than 5). There exists the possibility of developing the analysis in the context of Generalized Linear Models. Indeed, this approach is very flexible and compatible with the use of permutation tests. However, this idea was not pursued because the use of GLM is mainly adopted for non-normal errors or peculiar responses such as categorical, binary, mixed variables, etc. (Faraway, 2016). None of these reasons apply to our problem since the permutation approach is flexible and robust to errors that deviate from normal distribution. The analysis was performed using original R scripts, specifically created by the author, and packages from the CRAN network, as well as an online source file. The regression and diagnostic analysis for linear models was carried out with the basic commands *lm()*, *qqnorm()*, *qqline()*, *hist()* and *plot()* (packages stats and graphics). The Variance Inflation Factor (VIF) for the detection of multicollinearity in the matrix of regressors was performed with the command *vif()* available in the package *usdm*. For the implementation of the combined permutation tests, the following original commands were used (source file *cptlm.R*): *perm()* for creating the null permutation distribution of the multivariate test statistic by permuting the rows of the matrix of regressors; *slf()*: for the application of the significance level function for the computation of the p-values; *psi()*: for the nonparametric combination.

3.4 Application

The case study of this chapter concerns a multivariate regression analysis in order to understand whether SMEs adopting practices of circular economy are somehow able to receive some benefits at the economic level. The data were collected in a sample survey where 200 Italian SMEs operating in the plastics industry, were interviewed through a questionnaire designed by the Department of Economics and Management of the University of Ferrara in 2020. This survey is called "Questionario alle imprese Rev.1 Gennaio 2020 - Economia Circolare". The questionnaire, filled out by 4565 Italian enterprises operating in different sectors, had the purpose of collecting the following details: information about the firm and the leader/respondent; elements on the structure and characteristics of the company; insights concerning innovation and investments (i.e., R&D, sustainability, and industry 4.0); information on whether the enterprise implements circular practices, and of which kind; and particulars regarding the organization, formation/education/training, and industrial relationships of the responding firm. Questions concerning the introduced and performed innovations, investments, and practices are relative to the two-year periods 2017–2018 and 2019–2020.

At this point, it is evident that the Circular Economy's focus is on responsibly using what already exists, reducing the need for natural resource extraction, and revitalizing used products. This approach significantly helps to tackle the ecological crisis. However, companies need to adopt circular behaviours to thrive in competitive markets, generating economic returns and profits.

In this case study, exclusively SMEs operating in the plastics industry were selected, observing how they behave in terms of circularity in this contentious sector, famous for its polluting environmental impact. As said, this decision reduced the dataset to 200 units. The identification of such 200 companies belonging to the plastics industry has been possible after checking their ATECO code, which, in case the firm actually operates in the aforementioned sector, corresponds to *22.2 - Manufacture of plastic items.*

The variables of the dataset considered in this study (concerning innovations, sustainability, and circularity) were the following:

- n-patents: number of filed patents aimed at reducing the environmental impact
- n-inn-reduce-wat: number of innovations introduced to reduce water usage in the production process
- n-inn-red-mat: number of innovations introduced to reduce materials' usage
- n-inn-ren-eng: number of innovations introduced to use energy generated from renewable sources
- n-inn-red-eng: number of innovations introduced to reduce energy consumption;
- n-inn-red-waste: number of innovations introduced to reduce the generation of waste;
- n-inn-reuse-waste: number of innovations introduced to reuse waste in the production process;
- n-inn-trans-waste: number of innovations introduced to transfer internally generated waste to other companies, for reutilization in their production process;
- n-inn-min-mat: number of innovations introduced to change the product design, in order to minimize raw materials' usage (energy included);
- n-inn-max-recy: number of innovations introduced to change the product design. in order to maximize their recyclability;
- n-inn-change-prod-proc: number of innovations introduced to change the production process, in order to reduce greenhouse gas emissions;
- tot-expenses: total amount of expenses sustained to reduce pollution and for environmental protection;
- per-inv-red-waste17-18: percentage variation of investments in the years 2017-2018 for innovations aimed at reducing waste generation;

- per-inv-red-waste19-20: percentage variation of investments in the years 2019-2020 for innovations aimed at reducing waste generation;
- per-inv-reu-waste17-18: percentage variation of investments in the years 2017-2018 for innovations aimed at reusing waste in the production process;
- per-inv-reu-waste19-20: percentage variation of investments in the years 2019-2020 for innovations aimed at reusing waste in the production process;
- per-inv-trans-waste17-18: percentage variation of investments in the years 2017-2018 for innovations aimed at transferring internally generated waste to other companies, for reutilization in their production processes;
- per-inv-trans-waste19-20: percentage variation of investments in the years 2019-2020 for innovations aimed at transferring internally generated waste to other companies, for reutilization in their production processes;
- Operating expenses (OE-17-18): they refer to the expenditures sustained by the enterprise to increase its possibilities of survival over time, and embed, among the others, marketing expenses, and those related to R&D.

These particular points were selected because they showcase the essential aspects of interest regarding each entity's efforts towards circular behaviours. When viewed collectively, they present a comprehensive understanding of the overall direction taken by the analyzed organizations. It can be easily noted that the points investigating the "number of introduced innovations" have been selected because they are key in determining whether the company is circular. Indeed, such points summarize the set of actions and efforts on the basis of which a firm can be deemed circular. Because of these reasons, such factors have a central role in the regression analysis.

When conducting multivariate regression analysis, the goal is to identify a correlation between circular efforts, investments, and innovations and their possible economic and financial returns. Therefore, we considered eco-financial data for the analysis.

Some works in the literature suggest the most important economic and financial variables that are more likely to be positively influenced by circular behaviours. Yin et al. (2023) investigate the relationship between the adoption of circular actions and the firm's performance under different aspects, including the economic and financial ones. The study analyzed over 8,000 companies from various industries and sizes. Based on their research, the authors found evidence of a positive correlation between circular approaches and economic and financial variables. These approaches led to improved sales, increased market share, reduced warehouse investments, easier asset returns, better budget control, and ultimately improved financial performance for businesses. Although the degree of correlation is greatly influenced by certain variables, such as the country and industry type, the above-mentioned aspects are some of the most important financial and economic outcomes resulting from the adoption of a circular economy approach. Another work, conducted by Antonioli et al. (2022), represent a recent contribution to the existing literature in the same field of research, i.e., the linkage between practices of circular economy and eco-financial benefits. The paper at hand presents an analysis similar to the one in this case study. Using a similar dataset, the authors studied the same relationship but did not focus on any specific industrial sector, taking into account all respondent firms.

The researchers have chosen to examine two economic responses: sales revenue and total production costs, in the year 2019. These variables were selected because they are likely to be impacted in the short term by innovations introduced during the 2017-2018 period (Antonioli et al., 2022). The authors point out that SMEs do not seem to benefit from lower costs or higher revenues due to the implementation of circular procedures, leading to disappointing economic outcomes.

To analyze the economic outcomes of circularity, a comprehensive set of financial and economic variables were selected from AIDA. These variables have been used to assess the impact of circularity:

- Net working capital (NWC): it is obtained from the difference between current assets and current liabilities. A positive value is expressive of the company's ability to meet its short-term obligations;

- Current ratio (CR): it is calculated as current assets/current liabilities. Intuitively, if the ratio is greater than one, the company is in possession of enough liquidity to pay its current debts;

- Leverage (LEV): it is another measure of financial health, computed as net financial debt/shareholders' equity. Generally, the greater the ratio, the worse the financial condition of the company, since it means that the company prefers financing its business activities through debt instead of equity;

- Revenues from sales (REV): clearly resulting from the price of the goods/services sold multiplied by the number of units sold throughout the year;

- Return on sales (ROS): represents the average operating result per unit of revenue. It is computed as (operating income/sales revenues)*100;

- Return on assets (ROA): being computed as (net income/total assets)*100, it gives an indication of the total assets' weight and effectiveness in generating profit;

- Total production costs (TPC): a summary of the total fixed and variable costs incurred by the firm while operating its business activity.

It is important to note that all the previous variables were studied in the years 2019, 2020, and 2021. This decision was made to observe how economic effects change over consecutive years and to account for the introduction and sustenance of sustainable/circular innovations and expenses in the 2017-2018 biennium. It is not necessary to include periods prior to 2019 as these innovations will take some time before they yield positive results.

In addition, two more variables have been included in the explanatory ones, i.e., the number of employees (n-empl), and the age (age-2023). Such factors are meant to work, throughout the regression analysis, as *control variables* and, therefore included to better confine the possible linkage between the independent and dependent variables. In other words, the age and the number of employees might be two factors influencing the outcome of circular strategies (D'Angelo et al., 2023), and, by incorporating them into the model, their impact can be managed.

The dataset was composed of 200 companies (statistical units), 21 dependent variables (seven items, each considered for the years 2019-2020-2021), and 21 independent variables.

Since we have many explanatory variables in the dataset, the first step is to check for multicollinearity. Multicollinearity is a frequent phenomenon affecting multiple and multivariate regression analyses, and its likelihood of occurrence is generally higher with numerous explanatory variables. In detail, multicollinearity is present when such predictor variables influence each other and, thus, are linearly related to one another, potentially reducing the efficiency of the estimators of the regression coefficients. In cases of multicollinearity due to some independent variables, such variables must be excluded from the model. This process requires the identification of the "Variance Inflation Factor" (VIF), which can be obtained through an R command called *vif()*. After the computation of the VIF, the following explanatory variables were eliminated: n-inn-min-mat, n-inn-red-eng and perc-var-inv-reu-waste17-18. The remaining variables had a VIF less than 5 and this is considered acceptable (see Figure 3.1).

For the analysis, the final dataset was composed of 200 companies, 21 dependent variables and 18 independent variables (the remaining after the elimination of three of them because of the VIF value).

The application of the methodology presented in the previous section, provided a global p-value equal to 0.029, which is significant at the level $\alpha = 0.05$. In this specific case, there are 378 partial tests (21 dependent variables for 18 explanatory variables). To attribute the overall significance to one or more partial tests, the partial p-values were adjusted with the Bonferroni-Holm method.

In Tables 3.2, 3.3, 3.4 and 3.5 all the partial p-values related to the regression

Table 3.1: VIF of the remaining explanatory variables.

Variables	VIF
n-empl	1.60
age-2023	1.02
OE-17-18	1.25
n-patents	1.47
n-inn-reduce-waste	2.87
n-inn-red-mat	2.17
n-inn-ren-eng	2.06
n-inn-red-waste	4.31
n-inn-reuse-waste	1.91
n-inn-trans-waste	1.20
n-inn-max-recy	1.24
n-inn-change-prod-proc	3.28
tot-expenses	1.30
per-inv-red-waste17-18	2.49
per-inv-red-waste19-20	2.24
per-inv-reu-waste19-20	3.03
per-inv-trans-waste17-18	3.99
per-inv-trans-waste19-20	2.66

coefficients are reported. It can be observed that the coefficient estimate of the explanatory variable *number of employees* is significant and positive in the equation with the following dependent variables: NWC-2021, NWC-2020, NWC-2019, REV-2021, REV-2020, REV-2019, TPC-2021, TPC-2020 and TPC-2019. This means that the higher the dimension of the company, the higher the value of Net Working Capital, Revenues and Total Production Costs. Specifically, Net Working Capital represents the difference between current assets and current liabilities reported in the balance sheet. It is also a measure of management's ability to manage the company's current operational activities. Revenues represent the company's revenues. Finally, Total Production Costs summarize the total fixed and variable costs incurred by the firm while operating its business activity.

The histograms of the residuals are reported in Figures 3.1 and 3.2. As can be seen from the plots, the marginal distributions of the errors could be asymmetric. For this main reason, the non-parametric procedure explained before is needed. In fact, this method does not require that the data follow a specific distribution.

Table 3.2: Estimates and adjusted p-values of the partial permutation tests on the regression coefficients of the multivariate regression model. Dependent variables: NWC-2021, NWC-2020, NWC-2019, CR-2021, CR-2020, CR-2019 (significant in bold).

	NWC-2021		NWC-2020		NWC-2019	
	Coeff	Adj p-value	Coeff	Adj p-value	Coeff	Adj p-value
Intercept	7.281e+05		5.458e+05		4.146e+05	
n-empl	5.362e+04	**0.038**	4.523e+04	**0.038**	4.943e+04	**0.038**
age-2023	-1.526e+02	1.000	7.509e+01	1.000	1.347e+02	1.000
OE-17-18	6.549e-01	1.000	6.701e-01	1.000	7.274e-01	1.000
n-patents	1.284e+06	1.000	8.536e+05	1.000	1.148e+05	1.000
n-inn-reduce-waste	2.301e+05	1.000	3.319e+05	1.000	2.843e+05	1.000
n-inn-red-mat	1.165e+05	1.000	1.299e+05	1.000	9.537e+04	1.000
n-inn-ren-eng	2.170e+04	1.000	1.967e+04		2.428e+03	1.000
n-inn-red-waste	-1.253e+05	1.000	-4.375e+04	1.000	1.208e+05	1.000
n-inn-reuse-waste	3.187e+03	1.000	-5.740e+02	1.000	-1.105e+04	1.000
n-inn-trans-waste	-1.726e+05	1.000	-1.033e+05	1.000	-1.043e+05	1.000
n-inn-max-recy	6.848e+01	1.000	-6.566e+01	1.000	-2.107e+02	1.000
n-inn-change-prod-proc	-1.699e+05	1.000	-5.256e+05	1.000	-7.129e+05	1.000
tot-expenses	1.511e+00	1.000	1.163e+00	1.000	1.105e+00	1.000
per-inv-red-waste17-18	4.380e+04	1.000	6.203e+03	1.000	1.736e+03	1.000
per-inv-red-waste19-20	4.167e+03	1.000	4.968e+03	1.000	2.466e+02	1.000
per-inv-reu-waste19-20	-5.557e+03	1.000	-7.843e+02	1.000	8.414e+03	1.000
per-inv-trans-waste17-18	-2.784e+04	1.000	2.500e+03		1.083e+04	1.000
per-inv-trans-waste19-20	-3.425e+04	1.000	-4.129e+04	1.000	-4.380e+04	1.000
	CR-2021		CR-2020		CR-2019	
	Coeff	Adj p-value	Coeff	Adj p-value	Coeff	Adj p-value
Intercept	1.751e+00		1.855e+00		1.633e+00	
n-empl	1.470e-03	1.000	2.182e-03	1.000	1.579e-03	1.000
age-2023	3.192e-05	1.000	1.382e-04	1.000	7.697e-05	1.000
OE-17-18	-2.591e-07	1.000	-3.904e-07	1.000	-3.929e-07	1.000
n-patents	-2.123e-01	1.000	-2.505e-01	1.000	-1.530e-01	1.000
n-inn-reduce-waste	9.054e-03	1.000	4.662e-02	1.000	7.097e-02	1.000
n-inn-red-mat	-3.201e-02	1.000	-2.669e-02	1.000	-9.596e-03	1.000
n-inn-ren-eng	-8.955e-03	1.000	-1.619e-02	1.000	-1.395e-02	1.000
n-inn-red-waste	1.812e-01	1.000	1.719e-01	1.000	1.371e-01	1.000
n-inn-reuse-waste	2.509e-04	1.000	1.874e-03	1.000	1.184e-03	1.000
n-inn-trans-waste	-2.135e-02	1.000	-8.672e-03	1.000	-3.290e-03	1.000
n-inn-max-recy	-2.448e-04	1.000	-2.562e-04	1.000	-2.786e-04	1.000
n-inn-change-prod-proc	-4.209e-01	1.000	-3.373e-01	1.000	-3.455e-01	1.000
tot-expenses	1.807e-07	1.000	-8.581e-08	1.000	8.852e-08	1.000
per-inv-red-waste17-18	2.270e-02	1.000	2.157e-02	1.000	1.685e-02	1.000
per-inv-red-waste19-20	-3.417e-03	1.000	-3.388e-03	1.000	-3.975e-03	1.000
per-inv-reu-waste19-20	2.499e-03	1.000	-4.423e-03	1.000	-2.068e-03	1.000
per-inv-trans-waste17-18	-2.644e-02	1.000	-2.563e-02	1.000	-1.800e-02	1.000
per-inv-trans-waste19-20	7.050e-03	1.000	7.374e-03	1.000	1.891e-03	1.000

Table 3.3: Estimates and adjusted p-values of the partial permutation tests on the regression coefficients of the multivariate regression model. Dependent variables: LEV-2021, LEV-2020, LEV-2019, REV-2021, REV-2020, REV-2019 (significant in bold).

	LEV-2021		LEV-2020		LEV-2019	
	Coeff	Adj p-value	Coeff	Adj p-value	Coeff	Adj p-value
Intercept	5.039e+00		4.903e+00		7.942e+00	
n-empl	-2.598e-02	1.000	-3.481e-02	1.000	-3.779e-02	1.000
age-2023	-6.028e-05	1.000	-4.889e-04	1.000	-1.925e-03	1.000
OE-17-18	3.184e-06	1.000	3.324e-06	1.000	1.412e-06	1.000
n-patents	2.213e-01	1.000	9.694e-01	1.000	1.100e+00	1.000
n-inn-reduce-waste	1.641e-01	1.000	1.427e+00	1.000	1.945e+00	1.000
n-inn-red-mat	1.406e-01	1.000	5.917e-01	1.000	4.519e-01	1.000
n-inn-ren-eng	-2.168e-02	1.000	6.253e-02	1.000	5.634e-02	1.000
n-inn-red-waste	-1.563e-01	1.000	-8.983e-01	1.000	-1.155e+00	1.000
n-inn-reuse-waste	-6.224e-02	1.000	-8.887e-02	1.000	-3.491e-02	1.000
n-inn-trans-waste	-1.066e-01	1.000	1.949e-01	1.000	-2.003e-01	1.000
n-inn-max-recy	1.065e-05	1.000	2.327e-04	1.000	1.656e-04	1.000
n-inn-change-prod-proc	2.235e-01	1.000	-6.434e-01	1.000	5.764e-01	1.000
tot-expenses	-1.675e-06	1.000	-7.663e-08	1.000	-2.940e-06	1.000
per-inv-red-waste17-18	-3.831e-02	1.000	-3.389e-02	1.000	-1.430e-01	1.000
per-inv-red-waste19-20	-2.728e-02	1.000	-1.187e-02	1.000	-6.788e-02	1.000
per-inv-reu-waste19-20	3.785e-02	1.000	2.817e-02	1.000	7.766e-04	1.000
per-inv-trans-waste17-18	1.884e-03	1.000	-3.281e-02	1.000	-6.177e-02	1.000
per-inv-trans-waste19-20	7.787e-02	1.000	7.568e-02	1.000	3.343e-01	1.000
	REV-2021		REV-2020		REV-2019	
	Coeff	Adj p-value	Coeff	Adj p-value	Coeff	Adj p-value
Intercept	6.649e+05		4.506e+05		-5.526e+04	
n-empl	1.941e+05	**0.038**	1.555e+05	**0.038**	1.806e+05	**0.038**
age-2023	-1.495e+03	1.000	-1.102e+03	1.000	-7.131e+02	1.000
OE-17-18	2.501e+00	1.000	2.602e+00	1.000	3.139e+00	1.000
n-patents	2.864e+06	1.000	3.894e+05	1.000	3.387e+04	1.000
n-inn-reduce-waste	-2.007e+05	1.000	1.422e+05	1.000	-1.551e+04	1.000
n-inn-red-mat	3.934e+05	1.000	3.673e+05	1.000	2.132e+05	1.000
n-inn-ren-eng	1.097e+05	1.000	1.504e+04	1.000	-3.826e+04	1.000
n-inn-red-waste	-1.051e+06	1.000	-1.054e+05	1.000	3.433e+05	1.000
n-inn-reuse-waste	7.457e+04	1.000	-1.778e+04	1.000	-3.310e+03	1.000
n-inn-trans-waste	8.889e+04	1.000	-3.829e+04	1.000	-1.161e+05	1.000
n-inn-max-recy	9.285e+02	1.000	-1.460e+02	1.000	-5.092e+02	1.000
n-inn-change-prod-proc	1.719e+06	1.000	-1.013e+06	1.000	-9.258e+05	1.000
tot-expenses	-9.302e-01	1.000	2.387e+00	1.000	-4.120e-01	1.000
per-inv-red-waste17-18	2.355e+05	1.000	1.140e+05	1.000	9.160e+04	1.000
per-inv-red-waste19-20	-1.157e+03	1.000	-2.640e+03	1.000	3.124e+03	1.000
per-inv-reu-waste19-20	-1.330e+04	1.000	5.966e+04	1.000	2.530e+04	1.000
per-inv-trans-waste17-18	-2.272e+05	1.000	-1.949e+05	1.000	-1.076e+05	1.000
per-inv-trans-waste19-20	6.726e+04	1.000	4.309e+04	1.000	1.150e+04	1.000

Table 3.4: Estimates and adjusted p-values of the partial permutation tests on the regression coefficients of the multivariate regression model. Dependent variables: ROS-2021, ROS-2020, ROS-2019, ROA-2021, ROA-2020, ROA-2019 (significant in bold).

	ROS-2021		ROS-2020		ROS-2019	
	Coeff	Adj p-value	Coeff	Adj p-value	Coeff	Adj p-value
Intercept	4.033e+00		5.332e+00		5.300e+00	
n-empl	2.409e-02	1.000	5.326e-02	1.000	-4.743e-03	1.000
age-2023	-9.038e-03	1.000	-2.364e-02	1.000	-1.802e-03	1.000
OE-17-18	-3.084e-06	1.000	-7.932e-06	1.000	-1.907e-06	1.000
n-patents	-1.014e+00	1.000	-2.401e+00	1.000	-2.533e-01	1.000
n-inn-reduce-waste	-1.252e+00	1.000	-5.362e+00	1.000	-5.177e-01	1.000
n-inn-red-mat	-3.264e-01	1.000	-3.118e-01	1.000	4.939e-02	1.000
n-inn-ren-eng	-2.208e-02	1.000	1.009e-02	1.000	3.500e-02	1.000
n-inn-red-waste	5.935e-01	1.000	4.389e-01	1.000	-2.845e-01	1.000
n-inn-reuse-waste	5.833e-02	1.000	1.281e-01	1.000	3.321e-02	1.000
n-inn-trans-waste	6.771e-01	1.000	5.821e-01	1.000	2.695e-01	1.000
n-inn-max-recy	-4.447e-05	1.000	3.274e-04	1.000	-2.383e-04	1.000
n-inn-change-prod-proc	8.068e-01	1.000	6.772e+00	1.000	1.734e+00	1.000
tot-expenses	-7.028e-07	1.000	-9.907e-06	1.000	-1.882e-06	1.000
per-inv-red-waste17-18	2.268e-01	1.000	1.369e-01	1.000	6.503e-02	1.000
per-inv-red-waste19-20	6.824e-02	1.000	5.004e-04	1.000	4.821e-02	1.000
per-inv-reu-waste19-20	-5.941e-02	1.000	-3.700e-02	1.000	-9.795e-02	1.000
per-inv-trans-waste17-18	-4.249e-02	1.000	-2.080e-02	1.000	4.853e-03	1.000
per-inv-trans-waste19-20	-3.982e-02	1.000	8.871e-02	1.000	2.759e-02	1.000
	ROA-2021		ROA-2020		ROA-2019	
	Coeff	Adj p-value	Coeff	Adj p-value	Coeff	Adj p-value
Intercept	5.635e+00		4.283e+00		4.610e+00	
n-empl	-9.372e-03	1.000	1.555e-02	1.000	4.188e-03	1.000
age-2023	-1.435e-03	1.000	-2.757e-03	1.000	-5.595e-04	1.000
OE-17-18	-8.953e-07	1.000	-1.545e-06	1.000	-1.186e-06	1.000
n-patents	1.083e-01	1.000	-1.656e+00	1.000	-5.344e-01	1.000
n-inn-reduce-waste	-5.131e-01	1.000	-1.987e+00	1.000	-7.323e-01	1.000
n-inn-red-mat	-2.444e-01	1.000	1.810e-01	1.000	1.496e-02	1.000
n-inn-ren-eng	-6.394e-02	1.000	-9.859e-02	1.000	9.797e-03	1.000
n-inn-red-waste	4.902e-01	1.000	9.733e-01	1.000	-2.001e-01	1.000
n-inn-reuse-waste	2.066e-02	1.000	-3.169e-02	1.000	5.268e-02	1.000
n-inn-trans-waste	1.440e+00	1.000	4.059e-01	1.000	4.347e-01	1.000
n-inn-max-recy	-1.115e-04	1.000	-7.186e-04	1.000	-6.471e-06	1.000
n-inn-change-prod-proc	-4.252e-01	1.000	8.411e-01	1.000	1.753e+00	1.000
tot-expenses	2.190e-06	1.000	-1.443e-06	1.000	-3.066e-06	1.000
per-inv-red-waste17-18	1.301e-01	1.000	6.924e-02	1.000	8.986e-02	1.000
per-inv-red-waste19-20	1.179e-01	1.000	7.173e-02	1.000	1.205e-01	1.000
per-inv-reu-waste19-20	-9.187e-02	1.000	9.056e-03	1.000	-1.777e-01	1.000
per-inv-trans-waste17-18	-1.165e-01	1.000	-9.014e-02	1.000	7.669e-02	1.000
per-inv-trans-waste19-20	-4.051e-02	1.000	4.470e-04	1.000	-5.447e-03	1.000

Table 3.5: Estimates and adjusted p-values of the partial permutation tests on the regression coefficients of the multivariate regression model. Dependent variables: TPC-2021, TPC-2020, TPC-2019 (significant in bold).

	TPC-2021		TPC-2020		TPC-2019	
	Coeff	Adj p-value	Coeff	Adj p-value	Coeff	Adj p-value
Intercept	3.982e+05		3.480e+05		-4.246e+04	
n-empl	1.985e+05	**0.038**	1.507e+05	**0.038**	1.769e+05	**0.038**
age-2023	-1.300e+03	1.000	-7.293e+02	1.000	-4.834e+02	1.000
OE-17-18	2.869e+00	1.000	3.065e+00	1.000	3.577e+00	1.000
n-patents	2.607e+06	1.000	1.045e+06	1.000	-6.675e+04	1.000
n-inn-reduce-waste	-7.345e+04	1.000	1.968e+05	1.000	1.268e+05	1.000
n-inn-red-mat	3.948e+05	1.000	2.660e+05	1.000	1.570e+05	1.000
n-inn-ren-eng	1.083e+05	1.000	1.068e+04	1.000	-3.217e+04	1.000
n-inn-red-waste	-1.025e+06	1.000	-6.819e+04	1.000	2.832e+05	1.000
n-inn-reuse-waste	6.170e+04	1.000	-9.697e+03	1.000	6.873e+03	1.000
n-inn-trans-waste	5.991e+04	1.000	-5.472e+04	1.000	-1.876e+05	1.000
n-inn-max-recy	8.959e+02	1.000	-1.107e+02	1.000	-4.038e+02	1.000
n-inn-change-prod-proc	1.296e+06	1.000	-1.049e+06	1.000	-1.203e+06	1.000
tot-expenses	-1.576e+00	1.000	1.936e+00	1.000	1.079e-01	1.000
per-inv-red-waste17-18	2.108e+05	1.000	9.024e+04	1.000	9.102e+04	1.000
per-inv-red-waste19-20	-1.481e+04	1.000	-2.454e+03	1.000	-2.830e+03	1.000
per-inv-reu-waste19-20	3.491e+03	1.000	4.406e+04	1.000	3.465e+04	1.000
per-inv-trans-waste17-18	2.515e+05	1.000	-1.763e+05	1.000	-1.200e+05	1.000
per-inv-trans-waste19-20	1.104e+05	1.000	5.636e+04	1.000	1.796e+04	1.000

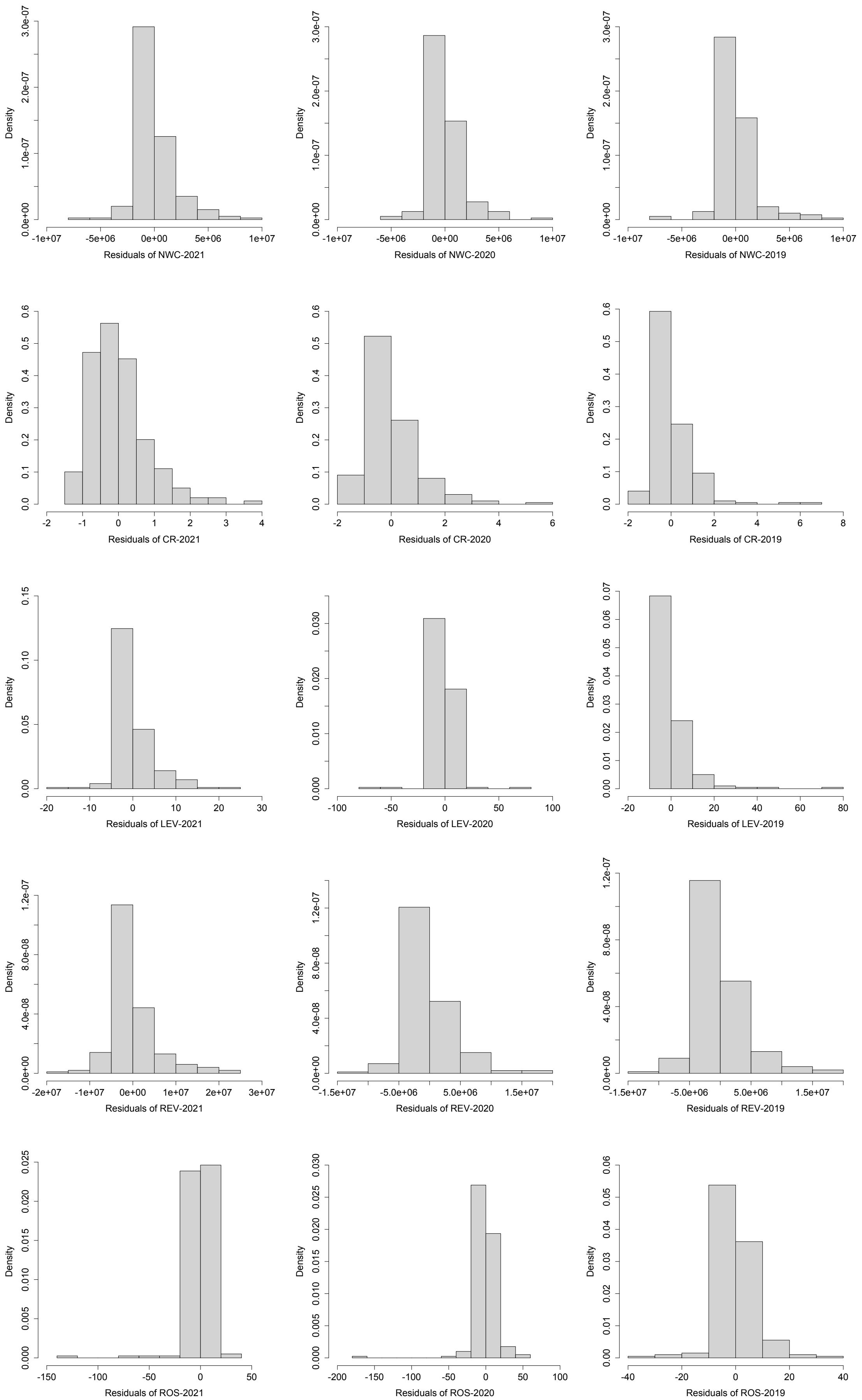


Figure 3.1: Histogram of the residuals of the dependent variables: NWC, CR, LEV, REV and ROS.

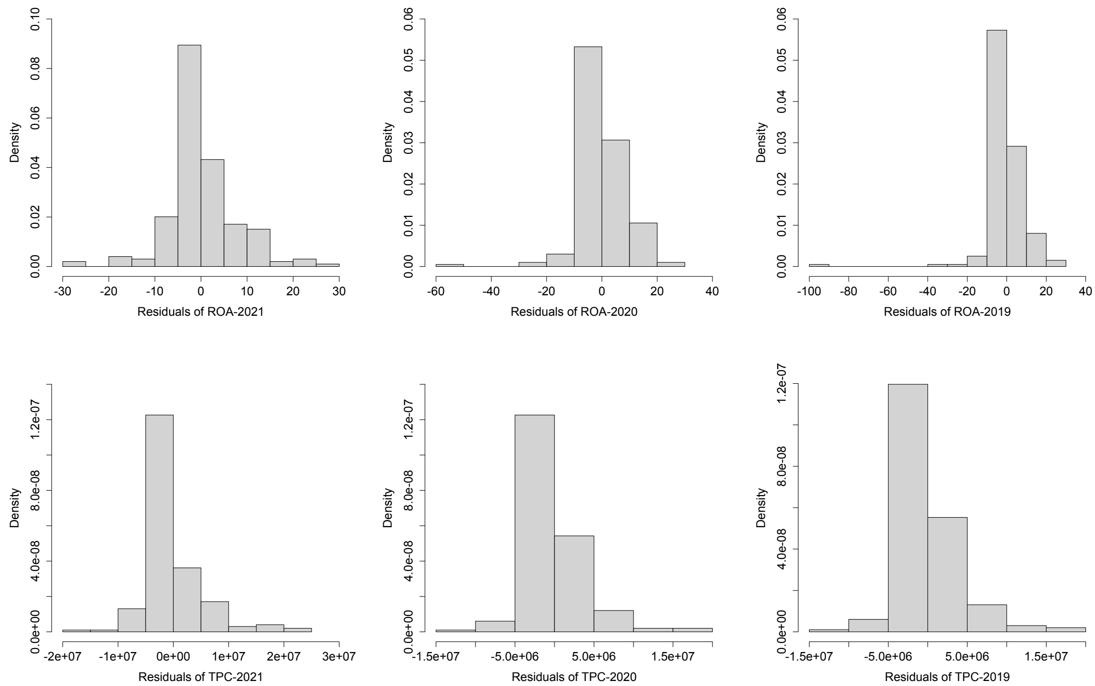


Figure 3.2: Histogram of the residuals of the dependent variables: ROA and TPC

3.5 Conclusions

The core purpose of this chapter was to test, whether Italian SMEs doing business in the plastics industry have economic and financial advantages from the adoption of practices relative to the Circular Economy framework, through a multivariate regression analysis with the application of a permutation test on the coefficients estimates. The emergence of an ecological crisis, characterized by pollution, CO2 emissions, an extreme accumulation of greenhouse gases, and global warming, is primarily caused by human societies and their economies. A circular socio-economic system can be a helpful solution to address this issue. Small and medium-sized enterprises (SMEs) are common in Italy and other economies worldwide. Despite their flexibility, social roots, and widespread presence, they encounter significant obstacles when trying to implement the aforementioned approach. The case study focuses on companies that prioritize social and environmental responsibility for sustainability, with the goal of contributing to the literature on the topic. For the purpose of enhancing the study, the analysis will focus on the plastics industry. As said, this industry has a significant impact on the current environmental challenges that our planet is facing. However, it is a controversial topic when discussing sustainability programs, mainly due to its appearance.

The multivariate non-parametric analysis carried out revealed that the possibility of gaining economic and financial benefits (measured in terms of Net Working Capital, Revenues and Total Production Costs respectively) in the three years, depends only on the number of employees of the company, i.e. on the size of the company. However, in the findings present in some of the reviewed literature, the advantages are accompanied by mirrored increasing trends in Total Production Costs, making it hard for the analyzed SMEs to fully appreciate and experience the positive outcomes.

These results do not fully confirm what is reported in the cited literature. This may be due to the fact that these works did not take into account company characteristics (such as age and/or size) or they did not stratify by sector of activity of the company.

In conclusion, this study may have a limitation due to the lack of available literature on the supporting tools of the circular economy in the plastics industry. This has made it difficult to provide comprehensive support and draw conclusive suggestions, resulting in some general recommendations.

To gain a comprehensive understanding of the current state of affairs and identify areas for progress, future research must explore various industries, business environments, and countries. It would be valuable to conduct follow-up analyses a few years after implementing new policies, incentives, and organizational strategies to evaluate their impact.

Final Remarks and Future Developments

In this PhD thesis, we proposed effective statistical methods for hypothesis testing problems. These approaches were applied to environmental sustainability problems.

The methodological approach developed in this thesis concerns Combined Permutation Tests (CPTs). These tests belong to the nonparametric tests family because they don't need some specific assumption as parametric tests do. The nonparametric nature of CPTs makes them flexible in relation to the underlying distribution assumptions. This makes the proposal more robust than a parametric solution and the only assumption required concerns the exchangeability of dataset rows (a very milder condition that is always satisfied). Another valid positive aspect of this method concerns the possibility of decomposing the overall testing problem into partial subproblems. Hence, this method consists of carrying out a set of partial tests, and then combining the p-values of the partial tests in order to determine a combined test statistic suitable for the general problem. Some Monte Carlo simulation studies were carried out in this thesis. We performed and simulated various settings of data and we confirmed that permutation tests are powerful also when the assumptions of parametric tests are true, but they are generally much more performing when the distributional assumptions of parametric tests are violated. The proposed permutation test is also very suitable for multivariate situations and in the case of complex non-monotonic hypotheses. That's because it allows us to take into account the interdependence between variables without assuming an underlying distribution or modelling the interdependence structure. The proposed approach also has the benefit of being effective when there are several variables, as well as small sample sizes and it allows to take into account possible confounding factors.

From an application point of view, Chapter 1 focuses on Small and Medium Enterprises (SMEs) and their relationship with Circular Economy activities. In general, many authors declare in the literature that SMEs are particularly involved in environmental sustainability activities (Ghisellini et al., 2020; Ghisetti and Montresor, 2020). The possible relationship between the size of the company and its propensity to undertake Circular Economy activities was inspired by Bassi and Dias (2019) e Bassi and Dias (2020). The complex non-monotonic relationship hypothesis was motivated by the work of Hoogendoorn and Guerra (2015). In this study, the authors obtained an inverted U-shaped relationship between company size and the adoption of greening processes. This type of hypothesis was the same one that we considered in the first subsection of the first chapter assuming a V-shaped relationship between company size and their reluctance to CE. The mentioned hypothesis concerned a V-shape relationship between company size (micro, small, medium) with more than six years of activity, focusing on the sector "Industry of wood and products in wood and cork (excluding furniture); manufacture of straw articles and woven materials". From a statistical point of view, it was a univariate test in which the response variable of the case study was the reluctance towards CE represented by the inverse of the number of innovations introduced by the company in the transfer of waste to other companies who would have reused it in their productive cycle. In fact, the focus in this first part was not so much the multivariate or stratified aspect, but rather the hypothesis of a non-monotonic relationship between the treatment (company size) and the response variable. The global test was significant and therefore the hypothesis was verified, confirming some of the existing results in the current literature (De Vass et al., 2023; Kromoser et al., 2022; Chen et al., 2022).

The same methodological approach was then applied in the second subsection of the first chapter. In this case, we demonstrated that the statistical procedure of breaking the non-monotonic complex hypothesis into several monotonic partial hypotheses is possible also in the opposite case (inverted V-shaped relationship). The statistical tests were multiple and multivariate. The test was stratified by age (both young and old), the treatment was the company size, the response variables were the number of innovations in 10 CE activities and all eight sectors were considered. From an application point of view, the relationship between the size of the company and the propension to adopt CE practices was taken into consideration. The global test was significant. Looking at the partial p-values it was possible to note that the significance can be attributed more to the sector of wood, paper, printing and reproduction, and to the sector of chemical-pharmaceutical, plastic and refined petroleum products. On the other hand, the variables that were found to be significant are: reduction of emitted waste, reuse of waste in the production cycle of the firm, transfer of their waste to other companies, which use it in their production cycle, change in product design to minimize the use of raw materials, change in product design to maximize their recyclability. These results have found positive confirmation with what has been found so far in the literature (Aggestam and Giurca, 2022; Stavenhagen, 2020). In conclusion, in both case studies of the first section of Chapter 1, we obtained significance through the application of CPTs to complex non-monotonic hypotheses, finding positive feedback from the literature.

In the second section of Chapter 1 we focused on the methodological aspect of proportions, considering a multivariate Bernoulli response variable. Consequently, in this application, we considered companies with more than 6 years of activity, six response variables represented by the proportions of companies that adopted the various activities related to CE and the sector of "manufacture of leather and similar items" was taken into account. The hypothesis of this test concerned that small companies were less inclined to adopt CE practices than medium companies. The global test was significant, confirming some results found in the literature, also regarding the involvement of the leather industry in Circular Economy models (Hu et al., 2011).

In Chapter 2 we studied the difference between the time series of ESG and non-ESG titles in terms of returns relative variation. In the literature there are discordant opinions regarding this topic, in fact, some authors report the clear prevalence, in terms of financial performance, of ESG titles over non-ESG titles. On the other hand, other authors are unable to see this prevalence. Our methodological approach breaks down the general problem into many partial problems, in this case one for each time point of the time series. This procedure allows us to take into account the dependence between each time point and the previous and the subsequent ones. From an application point of view, we concluded that non-ESG titles are more financially convenient than ESG titles in terms of returns relative variation. This conclusion is therefore in line with some literature research in which it is declared that the prevalence of ESG titles over non-ESG titles is not well defined and maybe there are other factors that can contribute in this sense (for example the methodology used to compose equity portfolios or the presence of specific ESG titles rather than others can lead to better economic benefits).

The last Chapter 3 was dedicated to the application of combined permutation tests (CPTs) to multivariate regression analysis. The application concerned whether the characteristics of Italian SMEs such as age and size, together with variables that represent the "circularity" of these companies, can predict the economic performances of companies (i.e. Net Working Capital, Revenues, ROA, ROS...). In literature, we found works in which this relationship had been highlighted. Because of this, the concept was deeply explored by applying the CPTs technique to the estimation of the coefficients in the multivariate regression analysis to verify the relationship that emerged in the literature. The results demonstrated a positive relationship between the number of company employees (i.e. firm size) and Net Working Capital, Revenues and Total Production Costs. This conclusion does not fully confirm what was found in the literature and this may be due to the fact that in our case study, we only considered companies in a specific sector (to avoid the confounding effect given by the sector itself) or because, through the permutation strategy, we verified the significance of each coefficient in every individual equation net of all the others.

In conclusion, the originality of the thesis lies in how the method of Combined Permutation Tests is applied to the specific problems faced in the thesis, i.e. in the combination strategy. The thesis provides an important scientific contribution by presenting a methodology for testing complex hypotheses that can be broken down into partial sub-hypotheses. According to the proposed method, to solve the overall testing problem, a suitable combination of the p-values of the partial tests should be carried out. The proposed methodological solution is valid for many statistical problems: complex non-monotonic hypotheses, classic monotone increasing or decreasing hypotheses (stochastic ordering), in the presence of several variables or small sample sizes, but also in simple, multiple and multivariate regression for goodness-of-fit tests. This is an innovative approach in inferential problems, typical of empirical studies on sustainability, which overcomes limitations of the classic methods usually applied to such problems. Another advantage of the method concerns the probabilistic assumptions. Indeed, it is applicable whether simple and mild conditions are satisfied. Such conditions are weaker than the assumptions of parametric tests. The original contribution consists of having explored the various areas of application of this methodology which is often used in two-sample comparisons. Furthermore, for the implementation of the method, for its application to real datasets, and for the Monte Carlo simulation studies to investigate its main properties, original R scripts have been created.

Suggestions for future research

The methodology developed in this thesis will be used in other applications. The idea of decomposing the general problem into sub-problems will have far-reaching consequences in many other contexts. Furthermore, what can be suggested for future research includes extending the simulations to other cases, for example simulating under other distributions other than Normal ones. Another aspect concerns the possibility of extending it to other applications relating to the Circular Economy field and beyond. In fact, the proposed methodology allows us to range from applications in different fields and sectors, such as in the medical, engineering, biological, socio-economic fields and so on. Finally, future studies may be dedicated to develop new, more powerful and performant combined permutation tests.

Bibliography

Abbasi, W.A., Wang, Z., Abbasi, D.A. (2017) Potential sources of financing for small and medium enterprises (SMEs) and role of government in supporting SMEs. *Journal of Small Business and Entrepreneurship Development*, 5(2), 39-47.

Adler, R.J., Taylor, J.E. (2007) *Random Fields and Geometry*, Springer, New York (NY).

Aggestam, F., Giurca, A. (2022). Implementing Circular-Bioeconomy Principles across Two Value Chains of the Wood-Based Sector: A Conceptual Approach. *Land*, 11(11), 2037.

Agresti, A. (1983) Testing Marginal Homogeneity for Ordinal Categorical Variables. *Biometrics*, 39, 505-510

Agresti, A., Klingenberg, B. (2005) Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. *Appl. Stat.*, 54(4), 691-706

Algarni, M.A., Ali, M., Albort-Morant, G., Leal-Rodríguez, A.L., Latan, H., Ali, I., Ullah, S. (2022) Make green, live clean! Linking adaptive capability and environmental behavior with financial performance through corporate sustainability performance. *J. Clean. Prod.*, 346, 131156.

Alibrandi, A., Giacalone, M., Zirilli, A. (2022). Psychological stress in nurses assisting Amyotrophic Lateral Sclerosis patients: A statistical analysis based on Non-Parametric Combination test. *Mediterranean Journal of Clinical Psychology* 10(2).

Alvo, M. (2008). Nonparametric tests of hypotheses for umbrella alternatives. *Canadian Journal of Statistics*, 36(1), 143-156.

Alvo, M., Cabilio, P. (1995). Rank correlation methods for missing data. *Canadian Journal of Statistics*, 23(4), 345-358.

Anderson, M.J., Robinson, J. (2001) Permutation tests for linear models. *Aust. N. Z. J. Stat.*, 43, 75–88.

Antonioli, D., Ghisetti, C., Mazzanti, M., Nicolli, F. (2022). Sustainable production: The economic returns of circular economy practices. *Business Strategy and the Environment*, 31(5), 2603-2617.

Arboretti, G.R., Bonnini, S., Corain. .L, Salmaso L (2016), Dependency and truncated forms of combinations in multivariate combination-based permutation tests and ordered categorical variables, *Journal of Statistical Computation and Simulation*, 86, 18, 3608-3619.

Arboretti, G.R., Bonnini, S., Corain, L. and Salmaso, L. (2014), A Permutation Approach for Ranking of Multivariate Populations, *Journal of Multivariate Analysis*, 132, 39-57.

Arboretti Giancristofaro, R., Bonnini, S. (2009) Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives. *Statistical Methods and Applications – Journal of the Italian Statistical Society*, 18(2), 221-236

Ates, C., Kaymaz, O., Kale, H.E., Tekindal, M.A. (2019) Comparison of Test Statistics of Nonnormal and Unbalanced Samples for Multivariate Analysis of Variance in terms of Type-I Error Rates. *Comput. Math. Methods Med.*, 2173638.

Auer, B.R., Schuhmacher, F., (2016), Do socially (ir)responsible investments pay? New evidence from international ESG data, *The Quarterly Review of Economics and Finance*, 59, 51-62.

Austin, A., and Ur Rahman, I. (2022). A triple helix of market failures: Financing the 3Rs of the circular economy in European SMEs. *J Clean Prod* 361, 132284.

Bassi, F., Dias, J. G. (2019). The use of circular economy practices in SMEs across the EU. *Conservation & Recycling*, 146, 523–533.

Bassi, F., Dias, J.G. (2020). Sustainable development of small- and medium-sized enterprises in the European Union: A taxonomy of circular economy practices. *Bus. Strat. Env.*, 29:2528–2541. https://doi.org/10.1002/ bse.2518

Basso, D., Finos, L. (2012) Exact Multivariate Permutation Tests for Fixed Effects in Mixed Models. *Commun. Stat. Theory*, 41, 2991–3001.

Basso, D., Pesarin, F., Salmaso, L., Solari, A. (2009). *Permutation tests for stochastic ordering and ANOVA: Theory and applications with R*. Springer.

Basso, D., Salmaso, L. (2011). A permutation test for umbrella alternatives. *Statistics and Computing*, 21, 45–54.

Bender, J., Bridges, T., He, C., Lester, A., and Sun, X. (2017), *A Blueprint for Integrating ESG into Equity Portfolios.*

Berg, J., Hilal, A., El, S., Horne, R. (2021). World employment and social outlook: Trends 2021. International Labour Organization. [pdf] Available at: link

Blagus, R., Peterlin, J., Stare, J. (2019). Goodness-of-fit Testing in Linear Regression Models. *arXiv preprint* arXiv:1911.07522.

Boffo, R., Patalano, R.(2020), *ESG Investing: Practices, Progress and Challenges*, OECD Press, Paris, www.oecd.org/finance/ESG-Investing-Practices-Progress-and-Challenges.pdf.

Bonnini, S. (2014). Testing for heterogeneity with categorical data: Permutation solution vs. bootstrap method. *Communication in Statistics: Theory and Methods*, 43(4), 906–917.

Bonnini, S., Corain, L., Marozzi, M., Salmaso, L. (2014). *Nonparametric hypothesis testing: Rank and permutation methods with application in R.* Wiley.

Bonnini, S., Melak Assegie, G. (2022) Advances on Permutation Multivariate Analysis of Variance for Big Data. *Statistics in Transitions – New Series*, 23(2), 163-183

Bonnini, S., Prodi, N., Salmaso, L., Visentin, C. (2014). Permutation approaches for stochastic ordering. *Communications in Statistics-Theory and Methods*, 43(10–12), 2227–2235.

Bosi, M. K., Lajuni, N., Wellfren, A. C., Lim, T. S. (2022). Sustainability reporting through environmental, social, and governance: A bibliometric review. *Sustainability*, 14(19), 12071.

Boulding, K. E. (1966). The economics of the coming spaceship earth. *Environmental quality issues in a growing economy.*

Brammer, R.G. and Pavelin, S. (2008). Factors influencing the quality of corporate environmental disclosure, *Bus. Strateg. Environ.* 17(2):120-136.

Bressanelli, G.M., Perona, M., Saccani, N. (2017) Reshaping the washing machine industry through the circular economy and product-service system business models. *Procedia CIRP*, 64, 43-48

Brondoni, S.M. (2020). Competitive circular ecology management. The Mitsubishi corporation case, *Symphonia Emerg Iss Manage* 1:14-27.

Casella, G., Berger, R.L. (2002). *Statistical Inference*. 2nd edn. Duxbury, Pacific Grove (CA).

Chen, Q., Feng, H., de Soto, B. G. (2022). Revamping construction supply chain processes with circular economy strategies: A systematic literature review. *Journal of cleaner production*, 335, 130240.

Chen, Y. I., Wolfe, D. D. (1990). A study of distribution-free tests for umbrella alternatives. *Biometrical Journal*, 32, 47–57.

Chuang-Stein, C., Mohberg, N. (2019) A unified approach to the analysis of safety data in clinical trials. In Sogliero-Gilbert, G. (eds.) *Drug Safety Assessments in Clinical Trials*. Dekker, New York

Clément, A., Robinot, É., Trespeuch, L. (2022). Improving ESG scores with sustainability concepts. *Sustainability*, 14(20), 13154.

Corain, L., Arboretti, R., Bonnini, S. (2017) *Ranking of Multivariate Populations: A Permutation Approach with Applications*. CRC Press, Taylor and Francis Group, Boca Raton (FL)

Cristoni, N., Tonelli, M. (2018) Perception of firms partecipating in a circular economy. *Eur. J. Sustain. Dev.*, 7 (4), 105e118

D'Angelo, V., Cappa, F., Peruffo, E. (2023) Walking the tightrope: Circular economy breadth and firm economic performance. Corporate Social Responsibility and Environmental Management, 1-14.

Davidov, O., Peddada, S. (2011) Order-restricted inference for multivariate binary data with application to toxicology. *Journal of the American Statistical Association*, 106(496), 1394-1404

De Vass, T., Nand, A. A., Bhattacharya, A., Prajogo, D., Croy, G., Sohal, A., Rotaru, K. (2023). Transitioning to a circular economy: lessons from the wood industry. *The International Journal of Logistics Management*, 34(3), 582-610.

Dimitar N. (2017). The impact of Foreign Direct Investments on employment: the case of the Macedonian manufacturing sector. *East. J. Eur. Stud.*, Volume 8, Issue 2, December 2017. https://www.researchgate.net/publication/322420660

Dwivedi, A., Paul, S. K. (2022). A framework for digital supply chains in the era of circular economy: Implications on environmental sustainability. *Business strategy and the environment*, 31(4), 1249-1274.

Elia, V., Gnoni, M.G., Tornese, F. (2017) Measuring circular economy strategies through index methods: a critical analysis. *J. Clean. Prod.*, 142 (4), 2741e2751

Elton, E. J., Gruber, M. J., Brown, S. J., Goetzmann, W. N. (2013), *Modern Portfolio Theory and investment analysis*, Wiley, 9th edition.

El Zein, S. A., Consolacion-Segura, C., and Huertas-Garcia, R. (2019), The Role of Sustainability in Brand Equity Value in the Financial Sector, *Sustainability*, 12(1), 1-19.

European Commission (2023) The Green Deal Industrial Plan: putting Europe's net-zero industry in the lead. Available at: link.

Faraway, J.J. (2016) *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 2nd ed.; CRC Press: Boca Raton, FL, USA.

Finos, L., Salmaso, A., Solari, A. (2007). Conditional inference under simultaneous stochastic ordering constraints. *Journal of Statistical Planning and Inference*, 137(8), 2633–2641.

Fisher, R.A. (1935) *Design of Experiments*; Oliver & Boyd: Edinburgh, UK.

Freedman, D., Lane, D. (1983) A nonstochastic interpretation of reported significance levels. *J. Bus. Econom. Statist.*, 1, 292–298.

Frino, A., Gallagher, D. R. (2001). Tracking S&P 500 index funds. *The Journal of Portfolio Management*, 28(1), 44-55.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C. (1994) Assessing the significance of focal activations using their spatial extent, *Human Brain Mapping*, 1, 210-220.

Fundera (2020) What Percentage of Small Businesses Fail? (And Other Need-to-Know Stats). Available at: link

Gao, X., Alvo, M. (2008). Nonparametric multiple comparison procedures for unbalanced two-way layouts. *Journal of Statistical Planning and Inference*, 138(12), 3674-3686.

Gao, X., Alvo, M., Chen, J., Li, G. (2008). Nonparametric multiple comparison procedures for unbalanced one-way factorial designs. *Journal of Statistical Planning and Inference*, 138(8), 2574-2591.

Gao, W., Kuriki, S. (2006) Testing Marginal Homogeneity against Stochastically Ordered Marginals for $r \times r$ Contingency Tables. *J. of Multivar. Anal.*, 97, 1330-1341

Garrido-Prada, P., Lenihan, H., Doran, J., Rammer, C., Perez-Alaniz, M. (2021). Driving the circular economy through public environmental and energy R&D: Evidence from SMEs in the European Union, *Ecol. Econ.* 182, 106884.

Gennari, F. (2022). The transition towards a circular economy. A framework for SMEs, *J. Manage. Gov.*, 1-35.

Ghisellini, P., Ulgiati, S. (2020) Circular economy transition in Italy. Achievements, perspectives and constraints. *J. Clean. Prod.*, 243, 118360 https://doi.org/10.1016/j.jclepro.2019.118360

Ghisetti, C. (2017) Demand-pull and environmental innovations: estimating the effects of innovative public procurement. *Technological Forecasting and Social Change*, 125, 178–187.

Ghisetti, C., Montresor, S. (2020). On the adoption of circular economy practices by small and medium size enterprises (SMEs): does "financing- as-usual" still matter?. *J. Evol. Econ.* 30, 559-586.

Giacalone, M., Agata, Z., Cozzucoli, P.C., Alibrandi, A. (2018). Bonferroni-Holm and permutation tests to compare health data: methodological and applicative issues. *BMC Medical Research Methodology* 18(1), 1-9

Giacalone, M., Alibrandi, A. (2011). A non parametric approach for the study of the controls in the production of agribusiness products. *Electronic Journal of Applied Statistical Analysis*, 4(2), 235–244.

Giacalone, M., Alibrandi, A. (2015). Overview and main advances in permutation tests for linear regression models. *Journal of Mathematics and System Science*, 5(2), 53–59.

Giudici, G., Bonaventura, M. (2019), *La relazione fra rating ESG e performance di mercato: uno studio sui titoli dell'indice Stoxx Europe 600*, Banor.

Godínez-Reyes, N. L., Gómez-Monge, R., Alfaro-Calderón, G. G., Calderón-Gutiérrez, A. (2022). Sustainable Value: An Empirical Research on Large Firms.

In *Soft Computing and Fuzzy Methodologies in Innovation Management and Sustainability* (pp. 197-208). Cham: Springer International Publishing.

Goeman, J.J., Van Houwelingen, H.C., Finos, L. (2011)Testing against a high-dimensional alternative in the generalized linear model: Asymptotic type I error control. *Biometrika*, 98, 381–390.

Goutis, C., Casella, G., Wells, M. T. (1996) Assessing evidence in multiple hypotheses. *Journal of the American Statistical Association*, 91(435), 1268-1277.

Harrar, S.W., Bathke, A.C. (2008) A non-parametric version of the Bartlett-Nanda-Pillai multivariate test. Asymptotics, approximations, and applications. *Am. J. Math. Manag. Sci.*, 28, 309–335.

Harrison, A.J., Ryan, W. and Hayes, K. (2007) Functional data analysis of joint coordination in the development of vertical jump performance, *Sports Biomechanics*, 6, 199-214.

Hartlaub, B. A., Wolfe, D. A. (1999). Distribution-free ranked-set sample procedures for umbrella alternatives in the m-sample setting. *Environmental and Ecological Statistics*, 6, 105–118.

Hattab, M.W., Christensen, R. (2018) Lack of fit tests based on sums of ordered residuals for linear models. *Aust. N. Z. J. Stat.*, 60, 230–257.

Haupt, M., Hellweg, S. (2019). Measuring the environmental sustainability of a circular economy. *Environmental and Sustainability Indicators*, 1, 100005.

Hettmansperger, T. P., Norton, R. M. (1987). Tests for patterned alternatives in k-sample problems. *Journal of the American Statistical Association*, 82, 292–299.

Hysa, E., Kruja, A., Rehman, N. U., Laurenti, R. (2020). Circular economy innovation and environmental sustainability impact on economic growth: An integrated model for sustainable development. *Sustainability*, 12(12), 4831.

Hockerts, K., Wüstenhagen, R. (2010). Greening Goliaths versus emerging Davids–theorizing about the role of incumbents and new entrants in sustainable entrepreneurship. Journal of Business Venturing, 125(5), 481–492.

Hoogendoorn, B., Guerra, D. (2015). What drives environmental practices of SMEs? *Small Business Economics*, 44, 759–781. https://doi.org/10.1007/s11187-014-9618-9.

Hu, J., Xiao, Z., Zhou, R., Deng, W., Wang, M., Ma, S. (2011). Ecological utilization of leather tannery waste with circular economy model. *Journal of Cleaner Production*, 19(2-3), 221-228.

Joe, H. (1997) Multivariate Models and Dependence Concepts, *Chapman & Hall*, London.

Johnson, R.A., Wichern, D.W. (2007) *Applied Multivariate Statistical Analysis*, 6th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA.

Kaimal, Manju M., Sajoy, P. B. (2020). Circular economy—a paradigm shift for sustainable development. *Perspectives on Business Management & Economics*, 1(1), 132–141.

Kennedy, P.E. (1995) Randomization tests in Econometrics. *J. Bus. Econom. Statist.*, 13, 85–94.

Khan, M., Serafeim, G., Yoon, A. (2016), Corporate Sustainability: First Evidence on Materiality, *The Accounting Review*, 91(6), 1697-1724.

Kılıç, M., Gurler, H. E., Kaya, A., Lee, C. W. (2022). The Impact of Sustainability Performance on Financial Performance: Does Firm Size Matter? Evidence from Turkey and South Korea. *Sustainability*, 14(24), 16695.

Klingenberg, B., Solari, A., Salmaso, L., Pesarin, F. (2008) Testing Marginal Homogeneity Against Stochastic Order in Multivariate Ordinal Data, *Biometrics*, 65, 452-462

Kotsantonis, S., Pinney, C., Serafeim, G. (2016). ESG integration in investment management: Myths and realities, *Journal of Applied Corporate Finance*, 28(2), 10-16.

Kromoser, B., Reichenbach, S., Hellmayr, R., Myna, R., Wimmer, R. (2022). Circular economy in wood construction–Additive manufacturing of fully recyclable walls made from renewables: Proof of concept and preliminary data. *Construction and Building Materials*, 344, 128219.

Lau, W. W. Y., Shiran, Y., Bailey, R. M., Cook, E., Stuchtey, M. R., Koskella, J., Velis, C. A., Godfrey, L., Boucher, J., Murphy, M. B., Thompson, R. C., Jankowska, E., Castillo, A. C., Pilditch, T. D., Dixon, B., Koerselman, L., Kosior, E., Favoino, E., Gutberlet, J., Baulch, S., Atreya, M. E., Fischer, D., He, K. K., Petit, M. M., Sumaila, U. R., Neil, E., Bernhofen, M. V., Lawrence, K., Palardy, J. E. (2020) Evaluating scenarios toward zero plastic pollution. *Science*, 369(6510), 1455-1461.

Lieder, M., Rashid, A. (2016) towards circular economy implementation: a comprehensive review in context of manufacturing industry. *J. Clean. Prod.*, 115, 36e51

Mack, G. A., Wolfe, D. A. (1981). K-sample rank tests for umbrella alternatives. *Journal of the American Statistical Association*, 76, 175–181.

Maione, C., Lapko, Y., Trucco, P. (2022) Towards a circular economy for the plastic packaging sector: Insights from the Italian case. *Sustainable Production and Consumption*, 34, 78-89.

Malik, A., Sharma, P., Sharma, P., Vinu, A., Karakoti, A., Kaur, K., Singh Gujral, H., Munjal, S., Laker, B. (2022). Circular economy adoption by SMEs in emerging markets: Towards a multilevel conceptual framework. *J. Bus. Res.* 142, 605–619. https://doi.org/10.1016/j.jbusres.2021.12.076

Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology* (2nd ed.). Chapman and Hall: London.

Marota, R., Suryadnyana, N. A., Sjam, J. M. E., Supriadi, T. (2023). Corporate Social Responsibility (CSR) Accounting Treatments on Financial Performance: Case Study in Manufacturing Public Companies. *Atestasi: Jurnal Ilmiah Akuntansi*, 6(1), 339-350.

McIntyre, K. and Ortiz, J.A. (2016). Multinational corporations and the circular economy: How Hewlett Packard scales innovation and technology in its global supply chain. In: Clift, R., Druckman, A. (eds.) *Taking stock of industrial ecology.* Springer.

Millen, B. A., Wolfe, D. A. (2005). A class of nonparametric tests for umbrella alternatives. *Journal of Statistical Research*, 39, 7–24.

Moktadir, M. A., Ahmadi, H. B., Sultana, R., Zohra, F. T., Liou, J. J. H., Rezaei, J. (2020) Circular economy practices in the leather industry: A practical step towards sustainable development. *J. Clean. Prod.*, 251, 119737. https://doi.org/10.1016/j.jclepro.2019.119737

Neubaum, D., Mitchell, M., Schminke, M. (2004). Firm newness, entrepreneurial orientation, and ethical climate. *Journal of Business Ethics*, 52(4), 335–347.

Neuhäuser, M., Leisler, B., Hothorn, L. A. (2003). A trend test for the analysis of multiple paternity. *The Journal of Agricultural, Biological and Environmental Statistics*, 8, 29–35.

Niiler, T. (2020). Comparing groups of time dependent data using locally weighted scatterplot smoothing alpha-adjusted serial T-tests, *Gait & Posture*, 76(9), 58-63.

Ormazabal, M., Prieto-Sandoval, V., Puga-Leal, R., and Jaca, C. (2018). Circular economy in Spanish SMEs: Challenges and opportunities. *J. Clean. Prod.* 185, 157–167.

Paletta, A., Filho, W.L., Balogun, A.L., Foschi, E., Bonoli, A. (2019) Barriers and challenges to plastics valorisation in the context of a circular economy: Case studies from Italy. *Journal of Cleaner Production*, 241, 118149.

Pan, G. (1996). Distribution-free tests for umbrella alternatives. *Communications in Statistics-Theory and Methods*, 25, 3185–3194.

Park, J.W., Lee, C.W. (2018) Performance of stock price with changes in SRI governance index. *Corp. Soc. Responsib. Environ. Manag.*, 25, 1121–1129.

Pataky, T.C., Vanrenterghem, J. and Robinson, M.A. (2015a), Two-way ANOVA for scalar trajectories, with experimental evidence of non-phasic interactions, *Journal of Biomechanics*, 48, 186-189.

Pataky, T.C., Vanrenterghem, J. and Robinson, M.A. (2015b), Zero- vs. one-dimensional, parametric vs. non-parametric, and confidence interval vs. hypothesis testing procedures in one-dimensional biomechanical trajectory analysis, *Journal of Biomechanics*, 48, 1277-1285.

Pesarin, F. (1990) *Permutation Testing of Multidimensional Hypotheses by Non-parametric Combination of Dependent Tests*, CLEUP, Padova.

Pesarin, F. (1992) *A resampling procedure for nonparametric combination of several dependent test*, Journal of the Italian Statistical Society.

Pesarin, F. (2001) *Multivariate permutation tests with applications in biostatistics.* Wiley.

Pesarin, F., Salmaso, L. (2010). *Permutation tests for complex data: Applications and software.* Wiley.

Pesarin, F. and Salmaso, L. (2010a) Finite-sample consistency of combination-based permutation tests with application to repeated measures designs, *Journal of Nonparametric Statistics*, 22(5), 669-684.

Pham, D., Do, T., Doan, T., Nguyen, T., Pham, T. (2021) The impact of sustainability practices on financial performance: empirical evidence from sweden, *Cogent Bus. Manag.* 8(1) 1912526.

Philip, L. H., Lam, K. F., Alvo, M. (2002). Nonparametric rank tests for independence in opinion surveys. *Austrian Journal of Statistics*, 31(4), 279-290.

Pitman, E.J.G. (1937) Significance tests which may be applied to samples from any populations. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 4, 119–130.

Pitman, E.J.G. (1937) Significance tests which may be applied to samples from any populations. II. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 4, 225–232.

Pitman, E.J.G. (1937) Significance tests which may be applied to samples from any populations. III. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 4, 322–335.

Potter, D.M. (2005) A permutation test for inference in logistic regression with small- and moderate-sized datasets. *Stat. Med.*, 24, 693–708.

Ramsay, O. (1982), When the data are functions, *Psychometrika*, 47, 379-396

Ramsay, J.O., Munhall, K.G., Gracco, V.L. and Ostry, D.J. (1996) Functional data analyses of lip motion, *Journal of Acoustical Society of America*, 99, 3718-3727.

Riahi-Belkaoui, A., Picur, R. D. (1993). An Analysis of the Use of Accounting and Market Measures of Performance, CEO Experience and Nature of the Deviation from the Analysts' Forecasts in CEO. *Managerial Finance*, 19(2), 20-32.

Rizos, V., Behrens, A., Van der Gaast, W., Hofman, E., Ioannou, A., Kafyeke, T., and Hirschnitz-Garbers, M. (2016). Implementation of circular economy business models by small and medium-sized enterprises (SMEs): Barriers and enablers. *Sustainability*, 8(11), 1212.

Ryan, W., Harrison, A. and Hayes, K. (2006) Functional data analysis of knee joint kinematics in the vertical jump, *Sports Biomechanics*, 5, 121-138.

Saidani, M., Yannou, B., Leroy, Y., Cluzel, F., Kendall, A. (2019). A taxonomy of circular economy indicators. *J. Clean. Prod.* 207, 542-559. doi: 10.1016/j.jclepro.2018.10.014.

Schot, J., Steinmueller, W. E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy*, 47(9), 1554–1567.

Shi, N. Z. (1988). Rank test statistics for umbrella alternatives. *Communications in Statistics-Theory and Methods*, 17, 2059–2073.

Singhal, D., Tripathy, S., Kumar Jena, S. (2019). Acceptance of remanufactured products in the circular economy: An empirical study in India. *Management Decision*. https://doi.org/10.1108/MD-06-2018-0686.

Solari, A., Finos, L., Goeman, J.J. (2014) Rotation-based multiple testing in the multivariate linear model. *Biometrics*, 70, 954–961.

Stahel, W. R., Reday-Mulvey, G. (1981). *Jobs for tomorrow: the potential for substituting manpower for energy*. Vantage Press.

Stavenhagen, M. (2020). Combine or combust?-Circular economy, digitalization and collaboration models for the new chemical Industry 4.0. *Journal of Business Chemistry*, 17(2), 70-82.

Steadman, R., Perrone, D. (2019), *The S&P 500 ESG Index: Integrating Enviromental, Social, and Governance Values into the Core*, S&P Dow Jones Indices.

Stute, W., Thies, S., Zhu, L. (1998) Model checks for regression: An innovation process approach. *Ann. Statist.*, 26, 1916–1934.

Sutcliffe, C.M.S. (2006), *Stock index futures*, Aldershot, England: Ashgate Publishing Ltd.

Tagliaferro, N. (2020). The circular economy at Enel X, *Symphonia Emerg Iss Manage* 1:101-116.

Tapinos, A., Mendes, P. (2013), A Method for Comparing Multivariate Time Series with Different Dimensions, *Plus One*, 8(2).

Ting, I.W.K., Azizan, N.A., Bhaskaran, R.K. and Sukumaran, S.K. (2019), Corporate Social Performance and Firm Performance: Comparative Study among Developed and Emerging Market Firms, *Sustainability*,12(1),26, https://doi.org/10.3390/su12010026.

Trahan, R., Jantz, B. (2023) What is esg? rethinking the "e" pillar, *Bus. Strategy. Environ.*

Veltkamp, R.C. (2001). Shape matching: Similarity measures and algorithms. In *Proceedings International Conference on Shape Modeling and Applications* (pp. 188-197). IEEE.

Vlachos, M., Kollios, G., Gunopulos, D. (2002) Discovering similar multidimensional trajectories, *Proceedings of the 18th International Conference on Data Engineering*, 673-684.

Werft, W., Benner, A. (2010) glperm: A Permutation of Regressor Residuals Test for Inference in Generalized Linear Models. *R J.*, 2, 39–43.

Westfall, P. H., Young, S. S. (1989), P-value adjustments for multiple tests in multivariate binomial models. *Journal of the American Statistical Association*, 84, 780-786.

Westfall, P.H., Young, S.S. (1992) Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment; Wiley-Interscience: New York, NY, USA.

Westfall, P.H., Young, S.S. (1992) On adjusting P-values for Multiplicity. *Biometrics*, 49, 941-945.

Westfall, P.H., Young, S.S. (1993) *Resampling-based multiple testing: Examples and methods for p-values adjustment*, Wiley, New York.

Williamson D., Lynch-wood G., Ramsay J. (2006). Drivers of Environmental SMEs Behaviour and the in Manufacturing Implications for CSR, *J. Bus. Ethics.*, 67(3), 317–330.

Wimmer, M. (2013) ESG-persistence in Socially Responsible Mutual Funds, *Journal of Management and Sustainability*, 3(1), 9.

Winans, K., Kendall, A., Deng, H. (2017) The history and current applications of the circular economy concept. *Renew. Sustain. Energy Rev.*, 68, 825e833

Winkler, A., Ridgway, G.R., Webster, M.A., Smith, S.M., Nicholas, T.E. (2014) Permutation inference for the general linear model. *NeuroImage*, 92, 381–397.

Wolfe, D. A. (2006). Nonparametric distribution-free procedures for order restricted alternatives. In M. Ahsanullah & M. Z. Raquab (Eds.), *Recent developments in order random variables.* New York: Nova Science.

Yadav N., Gupta K., Rani L., Rawat D. (2018). Drivers of Sustainability Practices and SMEs: A Systematic Literature Review, Eur J Sustain Dev, 7, 4, 531-544.

Yin, S., Jia, F., Chen, L., Wang, Q. (2023) Circular economy practices and sustainable performance: A meta-analysis. Resources, *Conservation & Recycling*, 190, 106838.

Appendix

ATECO classification

Stratification of sectors defined according to the ATECO 2022 classification:

- food-bev: food products, beverages and tobacco products (sectors 10 - 11 - 12 of the *Ateco* classification),
- text-wear: textiles and wearing apparel (sectors 13 - 14 of the *Ateco* classification),
- wood-print: wood, paper, printing and reproduction (sectors 16 - 17 - 18 of the *Ateco* classification),
- chem-plast: chemical-pharmaceutical, plastic and refined petroleum products, coke (sectors 19 - 20 - 21 - 22 of the *Ateco* classification),
- metals: basic metals and fabricated metal products (sectors 24 - 25 of the *Ateco* classification),
- comp-machinery: computer, electronic and optical products, machinery and equipment (sectors 26 - 28 of the *Ateco* classification),
- motors: motor vehicles, trailers, semitrailers and other transport equipment (sectors 29 - 30 of the *Ateco* classification),
- other: other (sectors8-15-23-27-31-32-33-43-46-47-73-95 of the *Ateco* classification).

The p-norm

By definition, a space X is called a vector space if it is an algebraic structure defined starting from a set of vectors, a field of scalars and two binary operations (sum between vectors and product of a vector by a scalar). Let X be a vector space on $\mathbb{R}$, then we say that $||\cdot||$ is a norm on X if

- $||x|| \geq 0, \forall x \in X$
- $||x|| = 0$ if and only if $x = 0$
- $||\lambda x|| = |\lambda| ||x||$ for every $\lambda \in \mathbb{R}$ and for every $x \in X$
- $||x + y|| \leq ||x|| + ||y||$ for every $x, y \in X$.

The p-norm of a function in a finite-dimensional space is written as

$$||x||_p = \left(\sum_t |x_t|^p \right)^{\frac{1}{p}} \quad \text{for } 1 \leq p \leq +\infty$$

as it is defined we have that it is a norm and when $p = +\infty$ the infinite norm is defined as the maximum of the vector because the limit for $p \to +\infty$ of the p-norm is the norm of the maximum:

$$||x||_\infty = \max_t |x_t|.$$