Volume 37
Number 2

ITALIAN JOURNAL OF APPLIED STATISTICS

Statistica Applicata

ISSN: 1125-1964

Statistica Applicata • Italian Journal of Applied Statistics

Volume 37 • Number 2 • 2025

FIRENZE UNIVERSITY PRESS

Statistica Applicata – ITALIAN JOURNAL OF APPLIED STATISTICS

A.S.A CONTACTS

Principal Contact

Francesco Palumbo (Editor in Chief)
editor@sa-ijas.org

Support Contact

Domenico Vistocco (Editorial Manager)
ijas@sa-ijas.org

JOURNAL WEBPAGE
https://oajournals.fupress.net/index.php/ijas

Statistica Applicata – Italian Journal of Applied Statistics is a four-monthly journal published by the Associazione per la Statistica Applicata (A.S.A.), Largo Gemelli 1 – 20123 Milano, Italy (phone + 39 02 72342904). Advertising:Cleup sc, via G. Belzoni, 118/3 – 35128 Padova, Italy (phone +39 049 8753496 – Fax +39 049 9865390), email: info@cleup.it.
Rules for manuscript submission: https://www.sa-ijas.org/ojs/index.php/sa-ijas/about/submissions
Subscription: yearly €103.30; single copy €40.00; A.S.A. associates €60.00; supporting institutions: €350.00. Advertisement lower than 70%. Postal subscription Group IV, Milan. Forum licence n. 782/89. CLEUP SC on behalf of ASA, 7 March 2023.

Statistica Applicata – Italian Journal of Applied Statistics is associated to the following Italian and international journals:

QTQM – Quality Technology & Quantitative Management (http/web.it.nctu.ew/~qtqm/)

SINERGIE – Italian Journal of Management

Pubbblished: April 2025

Published by Firenze University Press
Powered by Firenze University Press

Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

Summary

Statistica Applicata
ITALIAN JOURNAL OF APPLIED STATISTICS

Vol. 37, Number 2

doi.org/10.26398/IJAS.267

A COMPARATIVE STUDY OF GOODNESS-OF-FIT TESTS FOR THE GUMBEL DISTRIBUTION

Hadi Alizadeh Noughabi
Department of Statistics, University of Birjand, Birjand, Iran
E-mail: alizadehhadi@birjand.ac.ir

Abstract. *The Gumbel distribution is one of the most used models to carry out risk analysis in extreme events, in reliability tests, and in life expectancy experiments. In this article, we extend the general statistics for goodness-of-fit tests proposed by Noughabi (2019), specifically focusing on the Gumbel distribution. Our approach utilizes a new estimate of Kullback-Leibler information to develop a goodness-of-fit test. The properties of the test statistic are presented, and the unknown parameters of the Gumbel distribution are estimated by the maximum likelihood method. Critical points of the proposed test statistic are obtained through Monte Carlo simulation. A simulation study is conducted to evaluate the power of the test and compare its performance with existing tests. Finally, two real data examples are presented and analyzed.*

Keywords. *Gumbel distribution, Kullback-Leibler information, Goodness-of-fit tests, Test power, Monte Carlo simulation.*

1. INTRODUCTION

The Gumbel distribution is a popular, asymmetric, extreme value distribution (EVD), used to model maximums and minimums. For example, the EVD Type I has been used to predict earthquakes, floods, and other natural disasters, as well as modeling operational risk in risk management and the life of products that quickly wear out after a certain age.

Various applications based on the Gumbel distribution assumption are widely addressed in different fields of science. (e.g., Kotz and Nadarajah, 2000; Koutsoyiannis, 2003; Aryal and Tsokos, 2009; Yolanda et al., 2019; Eledum and Mohammed 2022; Osatohanmwen et al. (2022); and Krishna and Goel (2023)).

However, misspecification of the Gumbel distribution can have serious consequences, particularly when modeling extreme events. Incorrectly assuming a Gumbel distribution could lead to:

- Underestimation of risk: For instance, in risk management, using a Gumbel distribution when another skewed distribution is more appropriate could result in underestimating the likelihood of extreme events, leading to inadequate risk mitigation strategies.
- Inaccurate predictions: When modeling phenomena like natural disasters, using the wrong distribution could produce inaccurate predictions, impacting disaster preparedness and response efforts.

Therefore, finding a powerful goodness-of-fit test for the Gumbel distribution is crucial to ensure accurate model selection and reliable analysis. This is especially important when dealing with extreme events and other critical applications where misspecification can have significant consequences.

In this article, we investigate different goodness of fit tests for the Gumbel distribution based on the empirical distribution function.

Assuming that $X_1,...,X_n$ is the sample from a distribution F, we wish to assess whether the unknown $F(x)$ can be satisfactorily approximated by a Gumbel model $G(x)$. Goodness-of-fit (GOF) tests are designed to measure how well a proposed model fits the observed sample data. There are various classes of GOF tests, each based on different principles and measures of fit. One prominent class consists of tests based on the distance between the empirical and hypothesized distribution functions. These tests, such as the Cramer-von Mises (W^2), Kolmogorov-Smirnov (D), Kuiper (V), Watson (U^2), and Anderson-Darling (A^2), assess how well the hypothesized distribution function aligns with the empirical distribution function derived from the observed data. For this study, we focus on this class of GOF tests because:

- They are widely used and well-established.
- They provide a direct measure of the discrepancy between the proposed model and the observed data.
- They have robust theoretical properties and have been extensively studied in the literature.

For more details about these tests, see D'Agostino and Stephens (1986), Lemeshko et al. (2007), and Lemeshko and Gorbunova (2013).

The Kullback-Leibler (KL) discrimination has been widely studied in the literature as a central index for measuring quantitative similarity between two probability distributions. The KL discrimination of f from g is defined by

$$D(f,g)=\int f(x)\log\frac{f(x)}{g(x)}dx.$$

Note that $D(f,g)=0$ if and only if $f(x)=g(x)$ with probability 1.

Recently, Alizadeh Noughabi (2019) proposed a new estimate of the Kullback-Leibler discrimination and then constructed a test statistic for testing the validity of a model. His test statistic is

$$DA_{mn}=-\frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{2m}\left(G(X_{(i+m)};\hat{\theta})-G(X_{(i-m)};\hat{\theta})\right)\right\},$$

where G is the distribution function of g, m is a positive integer, $m\le n/2$, and $X_{(1)}\le X_{(2)}\le ...\le X_{(n)}$ are the order statistics and $X_{(i)}=X_{(1)}$ if $i<1$, $X_{(i)}=X_{(n)}$ if $i>n$. Here, θ is a model parameter which is usually unknown, and $\hat{\theta}$ is a reasonable equivariant estimate of θ.

Alizadeh Noughabi (2019) showed that the test statistic is non-negative just like the Kullback-Leibler divergence, i.e., $DA_{mn}\ge 0$. Also, the test based on DA_{mn} is consistent. Then, He proposed tests for normal, exponential, Laplace and Weibull distributions and compared the power of these tests with the other existing tests and showed that his test has a good power against different alternatives. In this paper, we apply the Alizadeh Noughabi's test statistic and introduce a goodness of fit test for the Gumbel distribution.

In section 2, we express some properties of the Gumbel distribution and then propose a goodness of fit test statistic for the Gumbel distribution based on an estimate of Kullback-Leibler divergence. In Section 3, the critical points and the power values of the proposed test are computed by Monte Carlo simulations and then compared with some known competing tests. Section 4 contains two real examples for illustrative purpose. The last section contains a brief conclusion.

2. The GUMBEL DISTRIBUTION AND TEST STATISTIC

This section begins by presenting key properties of the Gumbel distribution. We then extend the general statistics for goodness-of-fit tests proposed by Aizadeth Noughabi (2019), tailoring this framework to specifically address the Gumbel distribution.

2.1 THE GUMBEL DISTRIBUTION

The probability density function of the Gumbel distribution has the following form.

$$g(x;\mu,\sigma)=\frac{1}{\sigma}\exp\left\{\left(-\frac{x-\mu}{\sigma}\right)-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\},\qquad -\infty<x<\infty,\ \ \mu\in\mathbb{R},\ \ \sigma>0,$$

where μ and σ are the location and scale parameters, respectively. The cumulative distribution function can be obtained as

$$G(x;\mu,\sigma)=\exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right).$$

The mean and variance of the distribution are

$$E(X)=\mu+\sigma\gamma \quad \text{and} \quad Var(X)=\frac{\pi^2\sigma^2}{6},$$

where γ is the Euler constant.

If $Z=(X-\mu)/\sigma$, then Z is called the standard Gumbel random variable with the following density.

$$g(z)=e^{-(z+e^{-z})},\qquad -\infty<z<\infty.$$

Suppose that $X_1,X_2,...,X_n$ are a random sample from a Gumbel distribution. The maximum likelihood estimates for the Gumbel distribution are the solution to the following simultaneous equations

$$\overline{x}-\frac{\sum_{i=1}^{n}x_i\exp(-x_i/\hat{\sigma})}{\sum_{i=1}^{n}\exp(-x_i/\hat{\sigma})}-\hat{\sigma}=0,$$

$$-\hat{\sigma}\log\left(\frac{1}{n}\sum_{i=1}^{n}\exp(-x_i/\hat{\sigma})\right)-\hat{\mu}=0.$$

It is clear that the MLEs of the parameters cannot be obtained explicitly. Therefore, these equations need to be solved numerically and this is typically accomplished by using statistical software packages. We will use the MLEs to computation of the proposed test statistic.

2.2 THE PROPOSED TEST STATISTIC

Given a random sample $X_1,...,X_n$ from a continuous probability distribution F with a density function $f(x)$, the hypothesis of interest is

$$H_0: f(x) = g(x;\mu,\sigma) = \frac{1}{\sigma}\exp\left\{\left(-\frac{x-\mu}{\sigma}\right) - \exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \qquad \textit{for some } (\mu,\sigma)\in\Theta,$$

where μ and σ are specified or unspecified and $\Theta = \mathbb{R} \times \mathbb{R}^+$. The alternative to H_0 is

$$H_1: f(x) \neq g(x;\mu,\sigma), \qquad \textit{for any } (\mu,\sigma).$$

We extend the following test statistic for test of the Gumbel distribution.

$$DA_{mn} = -\frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{2m}\left(G(X_{(i+m)};\hat{\mu},\hat{\sigma}) - G(X_{(i-m)};\hat{\mu},\hat{\sigma})\right)\right\},$$

where G is the Gumbel distribution function and $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates of the unknown parameters.

We reject the null hypothesis for large values of the test statistic. According to Alizadeh Noughabi (2019), the test statistic is non-negative, i.e., $DA_{mn} \geq 0$, and also the test based on DA_{mn} is consistent.

Remark 1. Clearly, the proposed test statistic is invariant to transformations of location-scale and also the parameter space is transitive. Therefore, the distribution of the proposed test statistic DA_{mn} does not depend on the unknown parameters μ and σ. We will use this property to obtain the critical values of the test statistic.

3. CRITICAL POINTS AND POWER STUDY

At the significance level α, we reject H_0 if the value of the test statistic is greater than $C(\alpha)$, where the critical value $C(\alpha)$ is obtained by the $(1-\alpha)-$quantile of the distribution of the test statistic under the null hypothesis H_0.

Since deriving the exact distribution of the proposed test statistic is complicated, we study the null distribution of the proposed test statistic via Monte Carlo simulations using 100,000 runs for each sample size.

We use the following steps to determine the critical values of the proposed test statistics:

1) Generate a sample $X_1,...,X_n$ with size n from the standard Gumbel distribution;
2) Calculate the proposed statistics based on the sample $X_1,...,X_n$;
3) Repeat Steps 1–2 a large number of times and then determine the $(1-\alpha)$th quantile of the test statistics.

The obtained critical values for the proposed test statistics and sample sizes $5 \le n \le 50$ are presented in Table 1.

Table 1. Critical values of the proposed test statistic for $\alpha = 0.05$

	m									
n	1	2	3	4	5	6	7	8	9	10
5	1.0889	0.6657								
10	0.7842	0.5222	0.4558	0.4560	0.5025					
15	0.6535	0.4320	0.3820	0.3648	0.3673	0.3930	0.4299			
20	0.5743	0.3763	0.3266	0.3115	0.3127	0.3189	0.3350	0.3605	0.3904	0.4191
25	0.5262	0.3397	0.2908	0.2765	0.2742	0.2797	0.2876	0.3009	0.3174	0.3402
30	0.4962	0.3115	0.2629	0.2477	0.2449	0.2490	0.2553	0.2651	0.2755	0.2895
40	0.4579	0.2774	0.2275	0.2103	0.2056	0.2065	0.2116	0.2182	0.2255	0.2339
50	0.4298	0.2557	0.2056	0.1870	0.1799	0.1792	0.1817	0.1859	0.1917	0.1986

Based on Remark 1, we can use any value of the parameters to obtain the critical values because the distribution of the test statistic does not depend on the unknown parameters μ and σ. Here, we considered $\mu = 0$ and $\sigma = 1$.

The power values of the proposed test against various alternatives are computed by Monte Carlo simulations. We compare the power values of the proposed test with the existing tests. In our power comparisons, we consider the well-known tests which are applied in practice and statistical software. The test statistics of these tests are briefly described as follows. For more details about these tests, on can see D'Agostino and Stephens (1986).

Let $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$ are the order statistics based on the random sample $X_1,...,X_n$.

1. The Cramer-von Mises statistic (1931): A quadratic statistic based on the integrated squared difference between the empirical and hypothesized cumulative distribution functions (CDFs).

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - G(X_{(i)}; \hat{\mu}, \hat{\sigma}) \right)^2 .$$

2. The Watson statistic (1961): A quadratic statistic similar to the Cramer-von Mises test but with a modified weighting function to account for the circularity of the data.

$$U^2 = CH - n\left(\bar{P} - 0.5\right)^2 ,$$

where $\bar{P}$ is the mean of $G(X_{(i)}; \hat{\mu}, \hat{\sigma}),\ i = 1,...,n$.

3. The Kolmogorov-Smirnov statistic (1933): A supremum statistic based on the maximum absolute difference between the empirical and hypothesized CDFs.

$$D = \max(D^+, D^-) .$$

where

$$D^{+}=\max_{1\le i\le n}\left\{\frac{i}{n}-G(X_{(i)};\hat{\mu},\hat{\sigma})\right\};\ D^{-}=\max_{1\le i\le n}\left\{G(X_{(i)};\hat{\mu},\hat{\sigma})-\frac{i-1}{n}\right\}.$$

4. The Kuiper statistic (1960): A supremum statistic similar to the Kolmogorov-Smirnov test but accounts for the cyclical nature of the data.
$$V=D^{+}+D^{-}.$$

5. The Anderson-Darling statistic (1952): A quadratic statistic that gives more weight to the tails of the distribution, making it particularly sensitive to deviations in the tails.
$$A^{2}=-n-\frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{\log G(X_{(i)};\hat{\mu},\hat{\sigma})+\log\left[1-G(X_{(n-i+1)};\hat{\mu},\hat{\sigma})\right]\right\}.$$

where G is the Gumbel distribution function.

The following alternatives are considered in power comparison. The considered alternatives can divide into two groups, symmetric alternatives and asymmetric alternatives.

Group I: Symmetric alternatives:

- the standard normal distribution, denoted by $N(0,1)$,
- the Student's t distribution with 10 degrees of freedom, denoted by $t(10)$,
- the Student's t distribution with 3 degrees of freedom, denoted by $t(3)$,
- the standard logistic distribution, denoted by $L(0,1)$,
- the standard Laplace distribution, denoted by $Laplace$,
- the standard Cauchy distribution, denoted by $C(0,1)$,
- the uniform distribution, denoted by $U(0,1)$,
- the beta distribution, denoted by $Beta(2,2)$,

Group II: asymmetric alternatives:

- the exponential, $Exp(1)$,
- the gamma, $\Gamma(0.5,1)$ and $\Gamma(2,1)$,
- the lognormal, $LN(0,1)$,
- the Weibull, $W(0.5,1)$ and $W(2,1)$,
- the inverse Gaussian, $IG(1,0.5)$, $IG(1,1)$ and $IG(1,2)$,
- the skew normal distribution, $SN(0,1,0.5)$, $SN(0,1,2)$ and $SN(0,1,3)$,
- the skew Laplace distribution, $SL(0,1,0.5)$, $SL(0,1,2)$ and $SL(0,1,3)$.

Under above alternatives the power values of the tests are obtained by means of Monte Carlo simulations. Under each alternative 100,000 samples of size 10, 20, 30 and 50 are generated and the test statistics are calculated. Then power value of the corresponding test is computed by the frequency of the event ‘‘the statistic is in the critical region’’. The power values of the tests at significance level $\alpha = 0.05$ are presented in Tables 2 and 3. For each sample size and alternative, the bold type in these tables indicates the tests achieving the maximal power.

Table 2. Empirical powers of the tests against symmetric distribution at significance level 5%.

altern.	n	W^2	D	V	U	A^2	DA_{mn}
$N(0,1)$	10	0.1090	0.0929	0.1002	0.1088	0.1008	**0.1859**
	20	0.2092	0.1620	0.1821	0.2039	0.2187	**0.3899**
	30	0.3026	0.2289	0.2592	0.2895	0.3340	**0.5252**
	50	0.4965	0.3694	0.4263	0.4717	0.5551	**0.7037**
$t(10)$	10	0.1330	0.1113	0.1217	0.1319	0.1254	**0.2138**
	20	0.2690	0.2094	0.2401	0.2643	0.2801	**0.4378**
	30	0.3907	0.3033	0.3467	0.3819	0.4202	**0.5719**
	50	0.6057	0.4771	0.5442	0.5912	0.6485	**0.7511**
$t(3)$	10	0.2352	0.2000	0.2167	0.2334	0.2301	**0.2893**
	20	0.4572	0.3848	0.4312	0.4570	0.4653	**0.5322**
	30	0.6220	0.5391	0.5908	0.6218	0.6376	**0.6795**
	50	0.8252	0.7505	0.7990	0.8262	0.8392	**0.8503**
$C(0,1)$	10	**0.5971**	0.5617	0.5759	0.5934	0.5931	0.4526
	20	**0.8708**	0.8350	0.8560	0.8701	0.8703	0.7591
	30	0.9612	0.9442	0.9546	0.9611	**0.9617**	0.8837
	50	0.9971	0.9946	0.9960	0.9971	**0.9973**	0.8094
$L(0,1)$	10	0.1439	0.1197	0.1300	0.1427	0.1367	**0.2258**
	20	0.2978	0.2344	0.2681	0.2939	0.3098	**0.4564**
	30	0.4317	0.3382	0.3856	0.4242	0.4586	**0.5941**
	50	0.6512	0.5297	0.5945	0.6409	0.6879	**0.7739**
Laplace	10	0.2243	0.1900	0.2037	0.2228	0.2139	**0.2819**
	20	0.4676	0.3922	0.4344	0.4691	0.4670	**0.5599**
	30	0.6456	0.5592	0.6061	0.6464	0.6498	**0.7313**
	50	0.8602	0.7902	0.8289	0.8616	0.8623	**0.9080**
$U(0,1)$	10	0.1244	0.0995	0.1233	0.1295	0.1177	**0.1683**
	20	0.2454	0.1822	0.2270	0.2483	0.2654	**0.2888**
	30	0.3787	0.2732	0.3418	0.3745	**0.4422**	0.3855
	50	0.6553	0.4776	0.5991	0.6415	**0.7616**	0.5662
$Beta(2,2)$	10	0.0903	0.0787	0.0854	0.0920	0.0805	**0.1522**
	20	0.1578	0.1315	0.1367	0.1531	0.1595	**0.2975**
	30	0.2319	0.1854	0.1912	0.2169	0.2544	**0.4102**
	50	0.4071	0.3036	0.3345	0.3724	0.4739	**0.5816**

Table 3. Empirical powers of the tests against asymmetric distribution at significance level 5%.

altern.	n	W^2	D	V	U	A^2	DA_{mn}
$Exp(1)$	10	0.1585	0.1396	0.1280	0.1482	**0.1919**	0.1304
	20	0.3047	0.2439	0.2255	0.2733	**0.3769**	0.3324
	30	0.4446	0.3515	0.3296	0.3963	0.5506	**0.5608**
	50	0.7020	0.5596	0.5651	0.6428	0.8132	**0.8637**
$\Gamma(0.5,1)$	10	0.4350	0.3769	0.3593	0.4114	**0.5075**	0.3986
	20	0.7764	0.6624	0.6878	0.7361	**0.8513**	0.7915
	30	0.9285	0.8448	0.8875	0.9019	**0.9680**	0.8931
	50	0.9955	0.9770	0.9925	0.9915	**0.9993**	0.9175
$\Gamma(2,1)$	10	0.0630	0.0616	0.0593	0.0617	**0.0712**	0.0680
	20	0.0853	0.0752	0.0737	0.0804	**0.0992**	0.0951
	30	0.1106	0.0960	0.0911	0.1042	0.1334	**0.1334**
	50	0.1653	0.1318	0.1310	0.1534	0.2053	**0.2245**
$LN(0,1)$	10	0.2850	0.2621	0.2317	0.2626	**0.3332**	0.1298
	20	0.5219	0.4481	0.3995	0.4647	**0.5962**	0.3214
	30	0.7043	0.6130	0.5594	0.6349	**0.7803**	0.5208
	50	0.9068	0.8313	0.8022	0.8547	**0.9496**	0.7455
$W(0.5,1)$	10	0.6997	0.6363	0.6264	0.6745	**0.7586**	0.5051
	20	0.9600	0.9144	0.9344	0.9453	**0.9790**	0.6488
	30	0.9965	0.9857	0.9936	0.9942	**0.9989**	0.5775
	50	1.0000	0.9997	1.0000	1.0000	**1.0000**	0.4524
$W(2,1)$	10	0.0484	0.0468	0.0510	0.0504	0.0444	**0.0622**
	20	0.0559	0.0548	0.0583	0.0582	0.0509	**0.0799**
	30	0.0620	0.0613	0.0626	0.0632	0.0582	**0.0998**
	50	0.0808	0.0751	0.0737	0.0807	0.0768	**0.1489**
$IG(1,0.5)$	10	0.4135	0.3741	0.3390	0.3848	**0.4731**	0.1995
	20	0.7340	0.6469	0.6140	0.6795	**0.7997**	0.5164
	30	0.8957	0.8224	0.8055	0.8535	**0.9368**	0.6944
	50	0.9890	0.9648	0.9654	0.9774	**0.9960**	0.7622
$IG(1,1)$	10	0.2314	0.2116	0.1835	0.2121	**0.2760**	0.1171
	20	0.4313	0.3670	0.3196	0.3789	**0.5062**	0.2920
	30	0.5162	0.5203	0.4590	0.5434	**0.7024**	0.4963
	50	0.8494	0.7469	0.7020	0.7829	**0.9115**	0.7889
$IG(1,2)$	10	0.1102	0.1047	0.0919	0.1017	**0.1316**	0.0731
	20	0.1841	0.1617	0.1333	0.1594	**0.2246**	0.1262
	30	0.2567	0.2178	0.1761	0.2161	**0.3186**	0.1963
	50	0.4218	0.3375	0.2800	0.3542	**0.5105**	0.3539
$SN(0,1,0.5)$	10	**0.1057**	0.0900	0.0972	0.1053	0.0978	0.0684
	20	0.2003	0.1559	0.1740	0.1955	**0.2092**	0.1275
	30	0.2895	0.2208	0.2492	0.2773	**0.3219**	0.2126
	50	0.4760	0.3544	0.4074	0.4524	**0.5337**	0.4019

Table 3. Continued.

altern.	n	W^2	D	V	U	A^2	DA_{mn}
$SN(0,1,2)$	10	0.0644	0.0577	0.0631	**0.0649**	0.0587	0.0553
	20	0.0887	0.0757	0.0831	**0.0888**	0.0874	0.0685
	30	0.1118	0.0946	0.0998	0.1091	**0.1194**	0.0881
	50	0.1650	0.1284	0.1392	0.1574	**0.1849**	0.1286
$SN(0,1,3)$	10	0.0507	0.0487	0.0514	0.0518	0.0476	**0.0536**
	20	0.0565	0.0526	0.0583	0.0586	0.0532	**0.0598**
	30	0.0582	0.0558	0.0594	0.0588	0.0589	**0.0689**
	50	0.0679	0.0629	0.0660	0.0678	0.0703	**0.0817**
$SL(0,1,0.5)$	10	**0.4059**	0.3275	0.3714	0.4012	0.3927	0.1867
	20	0.7535	0.6448	0.7099	0.7473	**0.7576**	0.5536
	30	0.9087	0.8329	0.8785	0.9035	**0.9147**	0.7880
	50	0.9904	0.9703	0.9831	0.9892	**0.9919**	0.9609
$SL(0,1,2)$	10	0.1092	0.0991	0.1003	0.1071	**0.1116**	0.0469
	20	0.1936	0.1646	0.1821	0.1945	**0.1997**	0.0652
	30	0.2712	0.2292	0.2569	0.2737	**0.2839**	0.0953
	50	0.4254	0.3548	0.4055	0.4343	**0.4395**	0.1646
$SL(0,1,3)$	10	0.0978	0.0929	0.0863	0.0933	**0.1083**	0.0527
	20	0.1533	0.1359	0.1313	0.1458	**0.1679**	0.0620
	30	0.2034	0.1771	0.1736	0.1924	**0.2231**	0.0729
	50	0.3158	0.2602	0.2723	0.3044	**0.3358**	0.1012

The power of the proposed test statistic depends on the alternative distribution and the window size. It is not possible to have the best value of m which attains the maximum powers for all alternatives. Therefore, based on a broad Monte Carlo analysis, we determine the optimal m to be the values of m which attain good (not best) powers for symmetric of asymmetric alternative distributions. For a given n, the value of m can be obtained from heuristic formula $m=[n/2-1]$ and $m=[n/10]$, for symmetric of asymmetric alternatives, respectively. Here, [x] means the integer part of x. For example, when $n=20$, we recommend $m=2$ and $m=9$, against asymmetric and symmetric alternatives, respectively, as the optimal values which the proposed test attains good (not best) power values. We observe that the optimal m increases as n increases.

From Table 2, the symmetric alternatives, it is seen that the proposed test based on DA_{mn} statistic has the most power (with the exception of the case where Cauchy was the alternative). The differences of power values between the test DA_{mn} and the other tests are substantial. Therefore, against symmetric alternatives, the proposed test based on DA_{mn} statistic should be recommended in practice.

In Table 3, the asymmetric alternatives, it is evident that no single test can be said to perform the best against all alternatives. However, the test A^2 has the most power against mostly alternatives.

Our analysis indicates that the DA_{mn} and A^2 tests exhibit the highest power against their respective types of alternatives: DA_{mn} for symmetric and A^2 for asymmetric distributions. Overall, both tests demonstrate robust performance against a range of alternatives, making them reliable tools for practical applications.

4. APPLICATIONS

In this section, we examine two real-world data set to test the goodness-of-fit for the Gumbel distribution when a sample is available.

Example 1. The first real data set consists of 30 observations of time between failures for the repairable item. It was introduced by Murthy et al. (2004) and then applied by Hossam et al. (2022). The real data set is as follows.

1.43, 0.11, 0.71, 0.77, 2.63, 1.49, 3.46, 2.46, 0.59, 0.74, 1.23, 0.94, 4.36, 0.40, 1.74, 4.73, 2.23, 0.45, 0.70, 1.06, 1.46, 0.30, 1.82, 2.37, 0.63, 1.23, 1.24, 1.97, 1.86, 1.17.

In Figure 1, we present a graphical comparison of the observed data and the Gumbel distribution using an empirical distribution function (EDF) plot. Additionally, we provide a quantile-quantile (Q-Q) plot to visually assess the agreement between the two distributions.

The proposed procedure can be used to investigate whether the data come from a Gumbel distribution. The value of the considered test statistics is computed and also the critical value of each test at the significance level 0.05 is obtained by Monte Carlo simulation. Results are summarized in Table 4. Also, the values of estimated parameters are $\hat{\mu} = 1.06$ and $\hat{\sigma} = 0.77$.

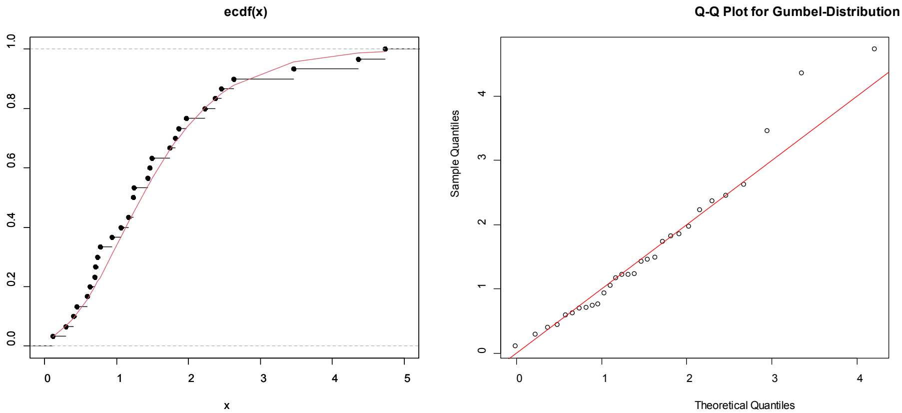

Figure 1. EDF plot and Q-Q plot of the observed data to the Gumbel distribution.

Table 4. The value of the test statistics and critical values at 5% level.

Test	Value of the test statistic	Critical value	Decision
W^2	0.0335	0.1226	Not Reject H_0
D	0.1023	0.1566	Not Reject H_0
V	0.1589	0.2641	Not Reject H_0
U^2	0.0295	0.1165	Not Reject H_0
A^2	0.2748	0.7461	Not Reject H_0
DA_{mn}	0.1062	0.2629	Not Reject H_0

Based on the considered tests, we can find that the values of these test statistics are smaller than the corresponding critical values and consequently the Gumbel hypothesis is not rejected at the significance level of 0.05. Therefore, based on our analysis, we do not have sufficient evidence to reject the Gumbel distribution as the underlying distribution of these data.

Example 2. We consider the Covid-19 data set presented by Hassan et al. (2021). Covid-19 data belong to Italy of 111 days that are recorded from 1 April to 20 July 2020. This data formed of daily new deaths divided by daily new cases. It is available at https://covid19.who.int. The data set is as follows.

0.2070, 0.1520, 0.1628, 0.1666, 0.1417, 0.1221, 0.1767, 0.1987, 0.1408, 0.1456, 0.1443, 0.1319, 0.1053, 0.1789, 0.2032, 0.2167, 0.1387, 0.1646, 0.1375, 0.1421, 0.2012, 0.1957, 0.1297, 0.1754, 0.1390, 0.1761, 0.1119, 0.1915, 0.1827, 0.1548, 0.1522, 0.1369, 0.2495, 0.1253, 0.1597, 0.2195, 0.2555, 0.1956, 0.1831, 0.1791, 0.2057, 0.2406, 0.1227, 0.2196, 0.2641, 0.3067, 0.1749, 0.2148, 0.2195, 0.1993, 0.2421, 0.2430, 0.1994, 0.1779, 0.0942, 0.3067, 0.1965, 0.2003, 0.1180, 0.1686, 0.2668, 0.2113, 0.3371, 0.1730, 0.2212, 0.4972, 0.1641, 0.2667, 0.2690, 0.2321, 0.2792, 0.3515, 0.1398, 0.3436, 0.2254, 0.1302, 0.0864, 0.1619, 0.1311, 0.1994, 0.3176, 0.1856, 0.1071, 0.1041, 0.1593, 0.0537, 0.1149, 0.1176, 0.0457, 0.1264, 0.0476, 0.1620, 0.1154, 0.1493, 0.0673, 0.0894, 0.0365, 0.0385, 0.2190, 0.0777, 0.0561, 0.0435, 0.0372, 0.0385, 0.0769, 0.1491, 0.0802, 0.0870, 0.0476, 0.0562, 0.0138.

Figure 2 includes both an EDF plot and a Q-Q plot, visually comparing the observed data to the Gumbel distribution.

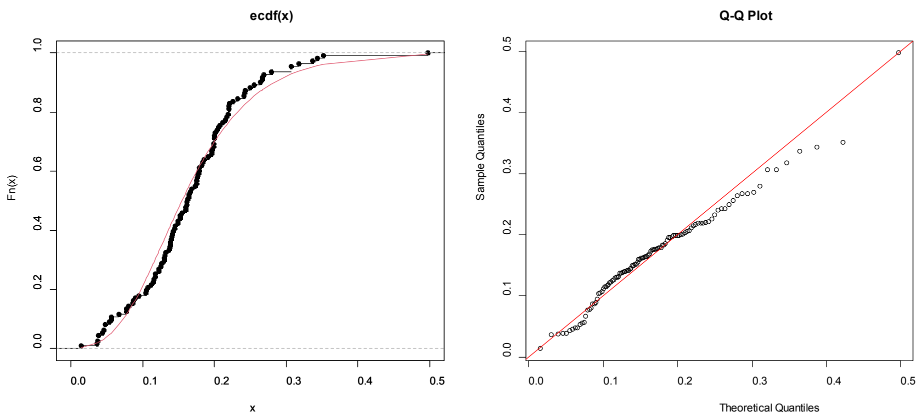

Figure 2. EDF plot and Q-Q plot of the observed data to the Gumbel distribution.

For this example, the values of estimated parameters are obtained as $\hat{\mu} = 0.13$ and $\hat{\sigma} = 0.07$. Applying the proposed procedure to this data set the value of the test statistic is obtained as 0.0967 and also the critical value of the test at the significance level 0.05 is obtained as 0.1146. The other procedures are also used to investigate whether this data come from a Gumbel distribution. The value of each test statistic is computed and also the critical value of each test is obtained by Monte Carlo simulation. Results are summarized in Table 5.

Table 5. The value of the test statistics and critical values at 5% level.

Test	Value of the test statistic	Critical value	Decision
W^2	0.1768	0.1235	Reject H_0
D	0.0816	0.0835	Not reject H_0
V	0.1401	0.1408	Not reject H_0
U^2	0.1632	0.1174	Reject H_0
A^2	1.1045	0.7552	Reject H_0
DA_{mn}	0.0967	0.1146	Not Reject H_0

Based on the tests D, V and DA_{mn}, we can find that the values of these test statistics are smaller than the corresponding critical values and consequently the Gumbel hypothesis is not rejected at the significance level of 0.05. Therefore, based on our analysis, we do not have sufficient evidence to reject the Gumbel distribution as the underlying distribution of these data. Based on the other tests, since the values of the test statistics are larger than the corresponding critical values, the Gumbel hypothesis is rejected at significance level 0.05.

Based on our simulations from Tables 2 and 3, we concluded that generally the proposed test DA_{mn} and A^2 are powerful against symmetric and asymmetric alternatives, respectively. Therefore, in this example, we prefer the proposed test A^2 over the other tests. Consequently, we choose this test and make a decision. From the results of Table 6, the test A^2 reject the null hypothesis and we can not conclude that these data follow a Gumbel distribution.

5. CONCLUSIONS

In this paper, we have extended a goodness-of-fit test for the Gumbel distribution based on an estimate of Kullback-Leibler information. We have examined the properties of the proposed test, computed critical values, and evaluated its power. While our findings demonstrate the test's effectiveness against symmetric alternatives, its true value lies in distinguishing the Gumbel distribution from other skewed distributions, particularly relevant in domains like extreme event modeling and survival analysis where misspecifying a skewed distribution as a Gumbel could lead to underestimation of risk or inaccurate predictions.

The current study focuses on complete data sets, but acknowledging the prevalence of type II censoring in survival analysis, future research should investigate the applicability of our proposed test in the presence of censoring. This extension would be particularly valuable for analyzing survival data and evaluating the fit of the Gumbel distribution in settings where complete data is not available.

Our findings underscore the potential of our proposed test in various domains. Future research should include a more comprehensive comparison of our test with existing

methods, particularly the Anderson-Darling test, to gain a clearer understanding of its advantages and limitations in both complete and censored data settings.
Finally, we have presented two real data sets to illustrate how the proposed test can be applied to items and removed from the life-test experiment items and removed from the life-test experiment assess the goodness-of-fit of the Gumbel distribution when a complete sample is available. This demonstrates the potential usefulness of our test in various domains, and further research will explore its applicability to censored data settings.

ACKNOWLEDGEMENTS

The author is grateful to the anonymous referees and the associate editor for providing useful comments on an earlier version of this manuscript.

REFERENCES

Alizadeh Noughabi, H. (2019). A new estimator of Kullback–Leibler information and its application in goodness of fit tests. *Journal of Statistical Computation and Simulation*, 89, 1914-1934.

Anderson, T.W. and Darling, D.A. (1954). A test of goodness of fit. *Journal of American Statistical Association*, 49, 765-769.

Aryal, G.R. and Tsokos, C.P. (2009). On the transmuted extreme value distribution with application, *Nonlinear Analysis*, 71, 1401-1407.

D'Agostino, R.B. and Stephens, M.A. (Eds.) (1986). *Goodness-of-Fit Techniques*, New York: Marcel Dekker.

Eledum, H. and Mohammed, A.M. (2022). A new generalization of the Gumbel distribution with climate application, *Information Sciences Letters*, 11, 278-286.

Hassan, A.S., Almetwally, E.M. and Ibrahim, G.M. (2021). Kumaraswamy inverted topple one distribution with applications to COVID-19 data. *CMC-Computers Materials and Continua*, 68, 337-358.

Hossam E., Abdulrahman, A.T., Gemeay A.M., Alshammari, N., Alshawarbeh E.and Mashaqbah, N.K. (2022). A novel extension of Gumbel distribution: Statistical inference with COVID-19 application, *Alexandria Engineering Journal*, 61, 8823-8842.

Kolmogorov, A.N. (1933), *Sulla determinazione empirica di une legge di distribuzione*. Giornale dell'Intituto Italiano degli Attuari, 4, 83-91.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*, World Scientific.

Koutsoyiannis, D. (2003). On the appropriateness of the Gumbel distribution for modelling extreme rainfall, In *Proceedings of the ESF LESC Exploratory Workshop held at Bologna,* 24-25.

Krishna, H. and Goel, R. (2023). Inferences based on correlated randomly censored Gumbel's Type-I bivariate exponential distribution, *Annals of Data Science,* https://doi.org/10.1007/s40745-023-00463-7.

Kuiper, N.H. (1960). Tests concerning random points on a circle, *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A*, 63, 38-47.

Lemeshko, B.Yu. and Gorbunova, A.A. (2013). Application and power of the nonparametric Kuiper, Watson, and Zhang tests of goodness-of-fit, *Measurement Techniques,* 56, 465-475.

Lemeshko, B.,Yu., Lemeshko, S.B. and Postovalov, S.N. (2007). The power of goodness of fit tests for close alternatives, *Measurement Techniques*, 50, 132-141.

Murthy, D.N.P., Xie, M. and Jiang, R., (2004), *Weibull models*, volume 505, John Wiley & Sons.

Osatohanmwen, P., Efe-Eyefia, E. and Oyegue, F.O. (2022), The exponentiated Gumbel–Weibull {logistic} distribution with application to Nigeria's COVID-19 infections data, *Annals of Data Science*, 9, 909–943.

von Mises, R. (1931). *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Leipzig and Vienna: Deuticke.

Watson, G.S. (1961). Goodness of fit tests on a circle, *Biometrika*, 48, 109-114.

Yolanda, M. Heleno, B. and Hector, W. (2019). Gumbel distribution with heavy tails and applications to environmental Data, *Mathematics and Computers in Simulation*, 157, 115-129.

doi.org/10.26398/IJAS.289

WHAT ELICITS EMPATHETIC PERFORMANCE IN DISCOURSE PRACTICES AMONG UNDERGRADUATE STUDENTS? A MULTIDISCIPLINARY PILOT STUDY

Rosaria Simone, Sole Alba Zollo[1]
Department of Political Science, University of Naples Federico II, Naples, Italy

Abstract *In the last few years scholars have started to study the process of operationalizing empathy in professional contexts such as investigative interviewing, medical schools, public administration and engineering. Since empathy is crucial also for politicians, political consultants, journalists and public relations representatives, the present research is meant as a pilot study to assess the empathetic performance of bachelor students at University of Napoli Federico II in Italy. On the basis of survey data collected to investigate respondents' behaviour faced to hypothetic scenarios, a statistical and linguistic study is pursued with Halliday's Transitivity Model and with the analysis of preference rankings to disclose discriminant elements in the elicitation of empathy in different scenarios.*

Keywords*: Empathy; Halliday's transitivity Model; Survey; Data Analysis; Consensus Ranking*

1. THE IMPORTANCE OF EMPATHY IN PROFESSIONAL SETTINGS

Traditionally, the term empathy was used to describe an unconscious reaction to an object involving our projection into it or our physical imitation of it. Then it lost its bodily connotation and started to be used to describe a psychological process. Recently it is related to the concept of sympathy since both describe how an individual can understand other people's feelings even if empathy is "the ability to understand another person's feelings, experience, etc." (https://www.oxfordlearnersdictionaries.com/) while sympathy is "the feeling of being sorry for somebody; showing that you understand and care about somebody's problems" (https://www.oxfordlearnersdictionaries.com/).

Over the years, many articles have been published on the concept of empathy,

[1] Corresponding author: Rosaria Simone (rosaria.simone@unina.it)
The authors discussed and conceived the paper jointly. Rosaria Simone is responsible for Sections 4 and 6 and Sole Alba Zollo is responsible for Sections 1,2,3 and 5.

offering different definitions, often conflicting, and developing a large variety of methodological approaches[2]. Apart from the different studies on clinical empathy (Greeno et al. 2017; Kim 2020), in particular on the importance of implementing empathy training in medical curriculum (Barnhill Bayne 2011), recently scholars have started to study the process of operationalizing empathy in professional contexts such as investigative interviewing, public administration and engineering. Research suggests that empathy can be helpful during police interviews because it can increase cooperation and help to obtain more detailed information from interviewees. For instance, by analysing a sample of police interviewers' self-reports, Baker-Eck et al. (2020) verified the application of empathy in interviews and definition or understanding of this concept. Moreover, given the lack of understanding of empathy in public administration and absence of method to include the practice of empathy in public services, Edlins (2021) explored the strengths and challenges of the practice of empathy in order to develop a model of empathy for public administration which is able to suggest good practices to improve relational interactions. Also for engineers it can be crucial to improve their degree of empathy especially when they manage project groups. So, according to Rasoal et al. (2012), it can be important also for engineering students to develop their empathic abilities by acquiring both theoretical and practical knowledge.

Since recent studies have shown that empathy is a key component not only for the doctor-patient relationship, the present study aims to analyse the level of empathy in discourse practices among political science undergraduate students through an exploratory study conducted in class between October-December 2019. Emulating the study carried out in Pounds et al. (2017), a survey in English language was administered to bachelor students, including mainly rating and ranking questions tailored to assess which response reaction to a given circumstance would be chosen by the respondent among listed options. On this basis, we discuss a preliminary investigation combining discourse analysis and statistical methods.

The paper is organized as follows: Section 2 recalls the foundation of Halliday's transitivity model; Section 3 reports a detailed description of the survey scheme and the results of the linguistic analysis of questions and available response options. Section 4 describes instead the results of the statistical analysis pursued to test differences in empathetic performances under the different

[2] Hall and Schwartz (2019) provide a review of empathy definitions and usages by examining and comparing two corpora of peer-reviewed journals published between 2001 and 2013, and in 2017.

scenarios and respondents' given gender. Specifically, first Quantile ANOVA (Mair and Wilcox 2020) is applied to test the significance of observed differences in the distribution of the designed empathetic score along its range; secondly, we resort to Kemeny's distance (Kemeny et al. 1962) to determine the so-called consensus ranking and the distance between observed rankings and the most empathetic one (D'Ambrosio 2021). Concluding remarks end the paper.

2. HALLIDAY'S TRANSITIVITY MODEL

According to Halliday's (1985) Systemic Functional Linguistics (SFL), language can have different functions, and among which it is primarily used to express peoples' outer and inner experience. He identifies three meta-functions of language: ideational, interpersonal, and textural. The ideational function is the use of language to communicate effectively, the interpersonal function is the use of language to create and maintain social relations, the textual function is the use of language to signify discourse. The system of transitivity concerns the ideational function. This system has been widely used by scholars to investigate literary and non-literary texts (written and spoken) from a discourse analysis perspective both qualitatively and quantitatively.

In traditional grammar transitivity is a term frequently used. It is a grammatical feature that indicates if a verb takes a direct object (transitive verb) or not (intransitive verb). Halliday introduces a new concept of transitivity where it is not a verbal phenomenon but a clausal one. Since the interpersonal function concerns the linguistic mechanisms of interaction among people such as speech acts, turn-taking and interruption, there is a connection between this meta-function and transitivity. In order to maintain social relationships, speakers not only express their opinions but also try to influence their interlocutors' viewpoints and behaviours. In fact, "a clause is the product of three simultaneous semantic processes. It is at one and the same time a representation of experience (ideational), an interactive exchange (interpersonal), and a message (textual)" (Halliday 1985: 53).

Given this broader definition of transitivity, three components of this process can be identified: the process itself, the participants in the process (animate or inanimate), and the circumstances associated with the process. The process is realized by verbs, which can be related with one or more participants and circumstances of time, space, manner, cause, etc.

As an effective tool for discourse analysis, Halliday (1985) presents a description of English transitivity. He identifies six major types of processes: material, mental, verbal, relational, behavioral, and existential (see Figure 1, left panel). Thanks to this model, the content of clauses can be more understandable as we will be able to identify the specific process.

Process type	Category meaning	Participants
Material:	'Doing'	Actor, Goal
Action	Doing	'Behaver'
Event	'Happening	
Behavioral:	'Behaving'	
Mental:	'Sensing'	
Perception	'Seeing'	Senser
Affection	'Feeling'	Phenomenon
Cognition	'Thinking'	
verbal:	'Saying'	Sayer, target
Relational:	'Being'	Token, value
Attribution	'Attributing'	Carrier, attribute
Identification	'Identifying'	Identified, identifier
Existential:	'Existing'	Existent

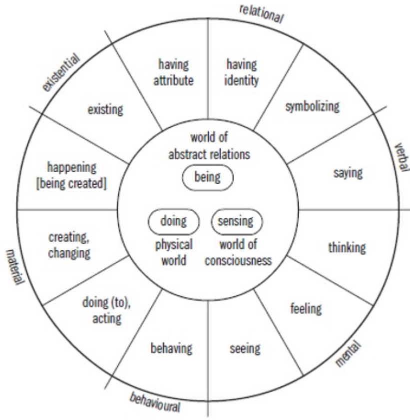

Figure 1: Left: Process types (Halliday 1985: 131). Right: Diagrammatic representation of process types (Halliday and Matthiessen 2004: 172)

There is a distinction between what we experience in the world around us and what we experience in our consciousness. Grammatically these two types of experience are expressed by material process clauses and mental process clauses. In addition, we have a third type of processes, the relational process clauses, since we are able to relate pieces of experience to others. There are also intermediate categories, that is the behavioural process (external manifestations of inner experience), the verbal process (symbolic relationships built in the world of consciousness and enacted in the form of language) and the existential process (phenomena of being or happening). Halliday and Matthiessen (2004) stress out that the process types are fuzzy categories and there is a circular continuity visually represented in Figure 1 (right panel).

SFL has been widely used by scholars to examine linguistic phenomena for many years. This paper adopts the transitivity theory of SFL to analyse the questionnaires used for our pilot study in order to identify the main recurrent processes in it in order to explain their functions of constructing an empathic message. Linguistic choices have significance, and transitivity plays a key role in meaning making. Transitivity can be a powerful instrument to analyse a text, focusing on agency and action. In particular, by seeing if responsibility is implicit or not, backgrounded or foregrounded, it can help discuss on the effectiveness of the questionnaires and how they can influence participants' responses.

Material processes are processes of doing and happening. They express tangible actions so there is a participant, the actor, who does the deed, which may be confined to the actor itself (e.g. "I went away") or may be extended to another entity, the goal (e.g. "I made a cake"). So we may have two participants: the actor, the doer of the process, and the goal, the participant affected by the process. We can also have material clauses that represent abstract events such as in the sentence "The scholar developed a new approach".

Mental clauses encode processes of cognition and perception expressed by cognitive and perceptive verbs (e.g. "I think you're right", "I feel exhausted"), and affection given by desiderative and emotive verbs (e.g.. "Ann liked the film", "I hate spiders"). These processes are concerned with events that take place in the world of our consciousness and are characterized by two participants: the senser, who is always a being endowed with consciousness, and a phenomenon, which is the entity being sensed, i.e. thought, felt, seen, wanted or perceived.

The third main type of process is the relational one. Relational clauses build the relationships of being and having between two participants. They construct our experience as "being" rather than as "doing" (material clauses) and "sensing" (mental clauses). The concept of "being" is expressed through two distinct modes – attributive and identifying – with different participant roles. In the attributive processes we find a carrier (a noun or noun phrase) who is ascribed or attributed to an attribute (a quality or classification), for example "Ann is kind". In the identifying mode, one entity is used to identify another, so there is an Identified/Token (the element that is being identified) and an identifier/value (the element that defines), for example "Today's meeting is the last opportunity for a deal".

Apart from these three major types of process in the English clause, we can identify other three process types: behavioural (between material and mental processes), verbal (between mental and relational processes) and existential (between relational and material processes).

As stated above, research has widely demonstrated that empathy can be considered a relevant communicative goal in doctor-patient interaction. An empirical study on the topic has been conducted by Pounds et al. (2017) through a discourse-pragmatic approach. Two written tests were designed and trialled with a sample of 58 student volunteers at the University of East Anglia to develop a new empathy-specific admission test for applicants to medical schools. Humanities students rather than medical students were invited to take part in the experiment to prevent from any bias induced by previous empathy training of medical students. Consequently, based on this research, we have decided to use the same scenarios in a class of political science students in order to verify their level of empathy on some sensitive real-life issues.

3. A SURVEY TO INVESTIGATE EMPATHETIC PERFORMANCE

The following section is devoted to the description of the questionnaire and of its two scenarios (Section 3.1). On this basis, the linguistic analysis of the statements and responses based on Halliday's transitivity model is presented in Section 3.2.

3.1 THE EMPATHY QUESTIONNAIRE

In the context of the course on English language for communication, freshmen in political sciences were involved in a survey study to investigate their empathetic performance. Each participant was asked to rank four reaction options to a series of triggers, from the most preferred to the least one, for two different imaginative scenarios.

In both scenarios the student has to imagine to be a doctor who is meeting a patient in a medical consultation to discuss about two problems: the failing of his/her relationship (Scenario 1) and the death of his/her dog (Scenario 2). The patient makes the statements listed in Table 1 playing the role of empathetic triggers. The students filled in the test for both scenarios:

- Scenario 1: the patient reports the failing of his/her relationship
- Scenario 2: the patient reports the death of his/her pet dog

	Scenario 1	Scenario 2
T1	My girlfriend/boyfriend and I have just broken up and s/he asked me to take my things and leave.	I was walking my dog the other day and he ran away and was hit by a car. We took him to the vet but he died...
T2	I am really devastated that it has ended like this.	I am really upset. I grew up with Charlie. He was such a good dog.
T3	I really love her/him. I don't know how I'll cope.	I really loved him. I don't know what I'll do without him.
T4	I'm such an idiot. You did tell me to be careful.	I feel so guilty. I should have paid more attention. He was a bit blind...

Table 1: Empathetic triggers - scenario 1 and scenario 2

Possible Student/Doctor's reactions for each trigger are reported in Tables 2-5.

	Scenario 1	Scenario 2
A	I'm really sorry to hear this. Let's talk about this later. How did you get on with your diary?	Oh, what a shame... I am really sorry to hear this. These things happen.
B	I'm sorry but this is not surprising to me… I did not think your relationship was very healthy.	Well, I think it is best not to keep dogs in a city. This is bound to happen sooner or later.
C	I'm so sorry! Would you like to talk about it?	Oh dear! How terrible! How long have you had Charlie?
D	I'm so sorry! Do you have anywhere to go? Can anyone help you move out?	Oh no!.. Perhaps you could make enquiries in case anyone has a new puppy they may want to sell.

Table 2: Reactions to Trigger 1 (R-T1) - scenario 1 and scenario 2

	Scenario 1	Scenario 2
A	Sure, but there is no point being upset now. You have to accept the situation.	I understand but it is not like... a relative.
B	You can't be that upset ... After all that s/he has put you through.	I thought you said you didn't like taking him out for walks so early in the morning.
C	Yes, of course, you were not expecting this.	Yes, of course. It's like losing a best friend!
D	Ok, but try not to be upset. You will feel better soon.	Ok, but try not to think about it. Put away all his things so you will not be reminded.

Table 3: Reactions to Trigger 2 (R-T2) - scenario 1 and scenario 2

	Scenario 1	Scenario 2
A	You may have felt like you loved him/her but, actually, you probably just needed him/her!	Are you sure? You can get another dog.
B	I know how you felt about him/her and this must be very hard.	I know he was like a friend to you and this is really hard.
C	Ok, but you need to move on now and there's no point feeling down about it.	I know it's very sad but you didn't see him much now that you are at Uni.
D	I understand but it will be easier once you have moved out of her/his flat.	I know it's sad but you have a lot going for you at the moment. You are at university, making new friends.

Table 4: Reactions to Trigger 3 (R-T3) - scenario 1 and scenario 2

	Scenario 1	Scenario 2
A	You cannot see things straight when you are in love.	It's difficult when the dog is as big as Charlie and pulls in all directions.
B	I know. I did try to warn you.	Well, perhaps you should have... but it is not worth thinking about it now.
C	It's normal for you to feel like this now. You wanted the relationship to work.	Anyone would feel guilty now but, actually, he was lucky to have you as his owner.
D	Maybe…..but I think s/he just took advantage of your good nature.	Perhaps… but you took such good care of Charlie.

Table 5: Reactions to Trigger 4 (R-T4) - scenario 1 and scenario 2

3.2 TRANSITIVITY LINGUISTIC ANALYSIS OF THE QUESTIONNAIRE

In both scenarios, the first trigger is a material process expressed by action verbs (to break up, to take, to leave, to walk, to run away, to die), while the other 3 triggers are mainly mental and relational (to love, to know, to be devastated, to be upset, to feel guilty) reinforced by the standard emphasiser "really".

By comparing all the response options in both scenarios, it can be noted that most clauses encode processes of cognition and perception expressed by cognitive and emotive verbs (to think, to be sorry).

Finally, by merging the main processes identified and the scenario structure elaborated by Pound et al. (2017: 169), Table 6 shows that only in the trigger "patient's reporting a loss" it is possible to identify material processes. Mental and relational processes are dominant in all other triggers and sought response options.

	Trigger	Main verbal processes	Sought response	Main verbal processes
1	Patient's reporting of loss	Material	Eliciting feelings	Mental/relational
2	Patient's explicit expression of feelings	Mental/relational	Acknowledging feelings	Mental/relational
3	Patient's explicit expression of feelings	Mental/relational	Acknowledging feelings	Mental/relational
4	Patient's expression of self-blame and self-deprecation	Mental/relational	Expressing positive regard and neutral support	Mental/relational

Table 6: Scenario structure and Halliday's processes

4. EMPIRICAL EVIDENCE FROM SURVEY DATA

For each of the four triggers listed in Table 1 (each corresponding to a specific dimension of empathy, and thus to the expected reactions schematized in Tables 2-5), four possible statements (coded with letters from A-D) were to be ranked according to their suitability to convey respondents' reaction to the trigger. For both scenarios and for each trigger, the most empathic ranking is $R^*=(C,D,A,B)$, meaning that reaction C is ranked first, reaction D as second, reaction A as third one and B as the last one. Equivalently, we will consider positional order, so that $R^*=(3,4,1,2)$.

For the pilot study discussed in the present contribution, n=131 bachelor students in political science participated in the survey concerning Scenario 1 (n=116 after omitting partial rankings, namely incomplete rankings of at least one of the set of 4 reactions), whereas n=122 students participated in the survey concerning Scenario 2 (n=112 after omitting partial rankings). No ties were allowed neither for Scenario 1 nor for Scenario 2, resulting in a data matrix of linear orderings of m=4 reactions for each possible trigger. Thus, for each of the four triggers t=1,2,3,4, let $R_i^{(t)}=(R_{iA}^{(t)}, R_{iB}^{(t)}, R_{iC}^{(t)}, R_{iD}^{(t)})$ be the ranking provided by the i-th respondent to the 4 possible reactions A-D of the t-th trigger (so that $R_{iX}^{(t)} = 1$ if X is the most suitable reaction, $R_{iX}^{(t)} = 2$ if X is the second best reaction. $R_{iX}^{(t)} = 3$ if X is the second to last suitable reaction and $R_{iX}^{(t)} = 4$ if X is the least suitable reaction).

The goal of the statistical analysis hereafter pursued was first to run a concise introductory exploratory data analysis to illustrate the data at hand: for this purpose, marginal rankings were considered. Then, the following research questions were tackled:

- RQ1: is there any statistically significant difference between the empathetic reactions raised by the two scenarios?
- RQ2: For each scenario, is there any statistically significant difference between responses provided by political science Italian students and those provided by humanities English students?
- RQ3: How far are the observed reactions, to both scenarios, to the most empathetic one $\mathbf{R}^*$?
- RQ4: With reference to the Italian study, which is the most representative ranking for each scenario and each trigger? For each trigger, how far are the corresponding representative rankings in the two scenarios?

Suitable methods to address RQ1 and RQ2 have to be adopted given the ordered nature of the response variable (marginal rankings): in particular, we resorted to Quantile ANOVA (Mair and Wilcox 2020) to determine if there is any significant difference in each marginal ranking distribution at low, medium and high level of the scale, in terms of low, medium and higher order quantiles. Indeed, quantiles are location measures that can always be defined for ordered variables, and they do not depend on the numerical scores given to categories. We used the function ‘Qanova’ implemented within the R package WRS2, resorting to the Harrel and Davis estimator of quantiles, which is suitable to work also with tied data, as in the case of ordinal outcomes.

Kemeny’s distance between rankings was used to tackle RQ3 and RQ4 since it allows to uniquely solve the so-called consensus ranking problem (see Kemeny and Snell 1962; Emond and Mason 2002): this distance equals 0 if and only if two rankings $\mathbf{R}_i^{(t)}$and $\mathbf{R}_j^{(t)}$ perfectly agree.

In order for the presentation to be self-comprehensive, a concise summary of these statistical tools is provided in a devoted Appendix at the end of the paper (see Section 6).

4.1 A COMPARATIVE ANALYSIS BETWEEN THE TWO SCENARIOS AND WITH RESPECT TO THE ENGLISH PILOT STUDY

Table 7 reports the p-values (adjusted for multiple testing) for the Quantile ANOVA run on the marginal rankings corresponding to each combination of triggers (t=1,2,3,4) and reaction (A,B,C,D). Chosen quantiles are first and ninth decile (D_1 and D_9), first, second and third quartile (Q_1, Q_2, Q_3). It follows that significant differences (at level α=0.05) are found at least in some location of the response for all combination of triggers and reactions, except for T1-C, T3-A, T3-D. For T1-B, T2-A, T2-D, T3-C, the differences are found at the lowest quantiles, indicating that the main differences correspond to the top position of the rankings. Conversely, for T2-C T3-B, T4-C, significant differences between scenarios emerges only at the bottom positions of the rankings (since the test is significant only at the selected higher order quantiles). For other marginal rankings, instead, there is evidence for differences along the entire ranking scale.

In order to perform a comparative analysis with the English pilot study, for each observed ranking – given a trigger – we count the number of positional agreements with the benchmark ranking $\mathbf{R}^*$. More precisely, for each scenario and each trigger t=1,2,3,4, we define $C_i^{(t)}$ as the number of perfect matchings of $\mathbf{R}_i^{(t)}$ with $\mathbf{R}^*$ for the i-th respondent (so that $C_i^{(t)} = 0$ if no matching is found, and $C_i^{(t)} = 4$ in case of a perfect matching: notice that 3 exact matchings imply necessarily $C_i^{(t)} = 4$). Then, we define an individual specific score of empathetic performance as $C_i = C_i^{(1)} + C_i^{(2)} + C_i^{(3)} + C_i^{(4)}$, ranging in set of the positive integers from 0 to 16. For each scenario, boxplots displayed in Figure 2 show the distributions of the empathetic score for the Italian and the English pilot studies in a comparative perspective.

	D_1	Q_1	Q_2	Q_3	D_9
T1 – A	0.000	0.002	0.000	0.000	0.000
T1 – B	0.000	0.000	0.000	0.432	0.432
T1 – C	-	-	0.915	0.190	0.915
T1 – D	0.000	0.238	0.000	0.005	0.040
T2 – A	0.000	0.000	0.000	0.455	0.455
T2 – B	0.843	0.843	0.843	0.000	0.8433
T2 – C	0.455	0.455	0.000	0.000	0.000
T2 – D	0.000	0.000	0.687	0.063	0.040
T3 – A	0.053	0.175	0.465	0.465	-
T3 – B	-	0..4933	0.007	0.000	0.000
T3 – C	0.000	0.000	0.448	0.245	0.448
T3 – D	0.857	0.857	0.857	0.857	0.857
T4 – A	0.000	0.000	0.000	0.000	0.000
T4 – B	0.712	0.170	0.712	-	-
T4 – C	0.477	0.477	0.000	0.000	0.000
T4 – D	0.100	0.397	0.000	0.397	0.145

Table 7: p-values for Quantile ANOVA to compare the results with Pounds et al.'s findings

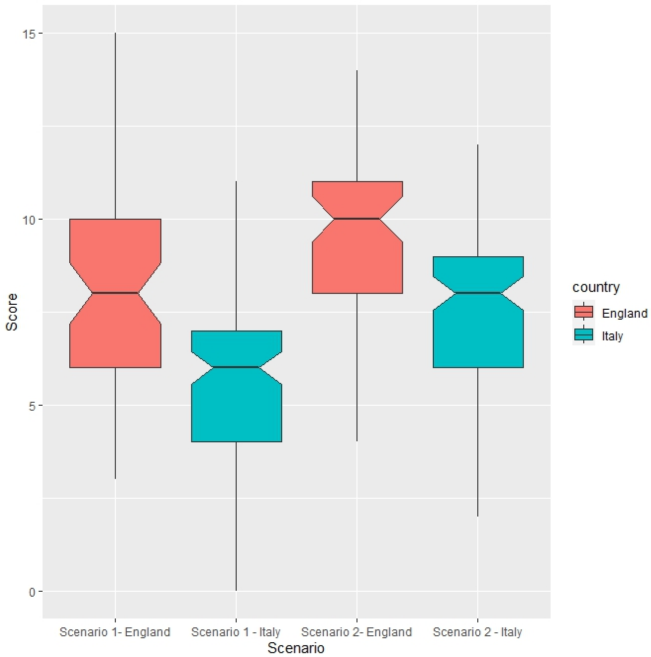

Figure 2: Graphical comparison of the distribution of the empathetic score for the Italian and English pilot study

From this exploratory analysis, it follows that empathetic reactions are poorer for Italian than for English students for both scenarios (to a greater extent for the first scenario). Then, we further investigate this evidence by performing a Quantile ANOVA on the merged score samples with an auxiliary dummy variable D_i flagging English scores (D_i=1) against Italian scores (D_i = 0). Results indicate highly significant differences at each of the selected quantiles for both scenarios (except for the first decile of the first scenario): for the sake of completeness, Table 8 reports the observed quantiles. With reference to the Italian study, it is worth noticing also that observed differences in the total empathetic score between the first and second scenario are statistically significant as well.

	D_1	Q_1	Q_2	Q_3	D_9
S1 – ITA	3	4	6	7	8
S1 – ENG	4	6	8	10	13
S2 - ITA	4	6	8	9	10
S2 - ENG	6	8	10	11	12

Table 8: Selected quantiles of the total empathetic score, for each scenario: comparison between the Italian and English pilot study

4.2 KEMENY'S DISTANCE WITH FULL EMPATHETIC RESPONSE

Figure 3 shows the boxplots of the distribution of Kemeny's distance between the observed rankings and the benchmark ranking **R*** corresponding to a perfectly empathetic reaction, for each trigger. It follows that there are no particularly manifest differences between the two scenarios, except for a more empathetic reaction raised by the first and last triggers (elicitation of feeling and expression of positive regard - mental support) under Scenario 2 than under Scenario 1.

Similarly, it can be observed that distances to the fully empathetic reaction tend to be larger under Scenario 1 than under Scenario 2 for the primary trigger of acknowledgment of feeling (T2), whereas the converse is true for the secondary trigger of this dimension (T3).

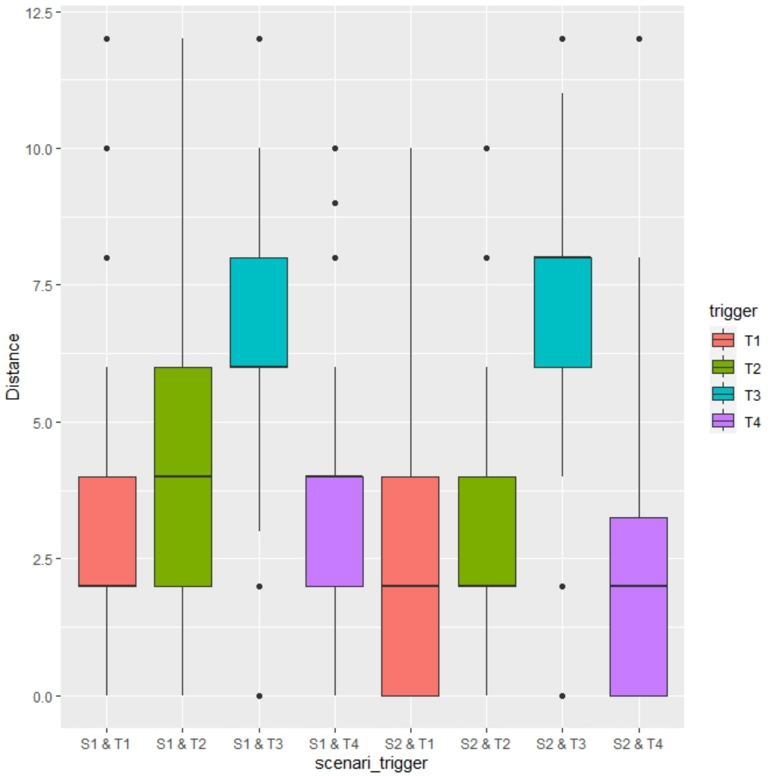

Figure 3: Boxplot of Kemeny's distance between observed ranking and full empathetic response, for each trigger and scenario

4.3 DETERMINING TRIGGERS' CONSENSUS RANKING AND THEIR DISTANCE

Table 9 reports the consensus ranking for each combination of trigger and scenario, so that Tx & Sy corresponds to the combination of the x-th trigger, for x=1,2,3,4, with the y-th scenario, y=1,2. Each row in the Table refers to one of the listed reactions. The consensus ranking[3] – namely the ranking that best represents the observed ones – has been obtained by means of the algorithm implemented in the R package 'ConsRank' (D'Ambrosio, 2021).

[3] See Amodio et al. 2016 for an overview of the consensus ranking problem and of some recent solutions.

	T1&S1	T1&S2	T2&S1	T2&S2	T3&S1	T3&S2	T4&S1	T4&S2
A	4	2	3	4	4	4	1	3
B	3	4	4	3	1	1	4	4
C	1	1	1	1	3	3	2	1
D	2	3	2	2	2	2	3	2

Table 9: Consensus ranking for each trigger and for each scenario

It follows that the consensus ranking fully matches with **R*** only for T2&S1 (acknowledging feeling for Scenario 1) and for T4&S2 (Expressing positive regard and neutral support for Scenario 2). Finally, Kemeny's distance has been computed between the consensus rankings of the two scenarios, for each trigger (see Table 10). It follows that the consensus reaction elicited by the third trigger (secondary acknowledgement of feeling) is the same between the two scenarios, whereas the two scenarios differ mostly, starting from the major consensus, for the elicitation of feeling (T1) and the expression of positive regard and neutral support (T4).

T1	T2	T3	T4
4	2	0	4

Table 10: Kemeny's distance between the consensus rankings of the two scenarios, for each trigger

Notice that it has not been possible to run a comparative analysis between the two studies in terms of consensus ranking, having only the distribution of empathetic scores in Pound et al. (2017), and not the individual ranking responses.

5. CONCLUSIONS

The goal of the present paper was to propose a combined linguistic and statistical analysis of survey data to assess empathetic performances among university students. Halliday's transitivity model allowed to identify the main verbal processes present in the questionnaire. The linguistic analysis revealed that, apart from the trigger "patient's reporting a loss" which contains material processes, all other triggers and sought responses consist of mental and relational clauses, encoding processes of cognition and perception. The statistical data analysis demonstrated that the two scenarios elicit fairy different empathetic reactions only with respect to elicitation of feeling and expression of positive regard – mental support. In particular, the first scenario seems to entail poorer empathetic performances: in general, empathetic reactions are poorer for Italian than for English students for both scenarios.

As empathy is more and more acknowledged as a desired professional skill in all domains, from medicine to politics, this research may be the starting point of further interdisciplinary and empirical studies to measure university students' empathic communicative performance and suggest appropriate interventions that can improve students' ability to communicate empathetically.

6. APPENDIX

This section is meant to provide a concise explanation of the main statistical methods that have been used for the present analysis. The reader is referred to the bibliography items quoted within the text for details.

6.1 KEMENY'S DISTANCE AND CONSENSUS RANKING

Given m objects $o_1,o_2,\dots,o_m$, there are m! rankings, each of which correspond to a permutation of the objects (in our case, objects are the m=4 possible reactions to a given trigger). To determine Kemeny's distance between a pair of rankings, first, for every ranking R, one defines an m-dimensional score matrix S_R so that, for each pair of positions i,j=1,…m, $S_R(i,j)=1$ if o_i is preferred to o_j, $S_R(i,j)=-1$ if o_j is preferred to o_i, and $S_R(i,j)=0$ in case of ties. Then Kemeny's distance between two rankings R and T, is given by:

$$d(R,T) = \sum_{i=1}^{m}\sum_{j=1}^{m} |S_R(i,j) - S_T(i,j)|$$

assuming Kemeny's axiomatic approach to distance between two rankings, and that all positions are equally weighted (namely, differences are treated equally if they occur at the top, at the center or at the end of the scale). In this framework, given n rankings of the 4 possible reactions to a given trigger, the search for the consensus ranking aims at determining the ordering of the reactions that best represents the consensus opinion. Specifically, given Kemeny's distance, a ranking P is defined the consensus ranking if it corresponds to the minimum sum of distances d(R,P) over all possible rankings R. Several algorithms have been proposed in the literature to pursue this task. For our analysis and for illustration purposes, we have resorted to the methods implemented in the R package 'ConsRank' (D'Ambrosio 2021). For a recent discussion on a re-characterization of Kemeny's distance and its properties for general weak orderings, see Can and Storcken (2018).

6.2 QUANTILE ANOVA

Quantile ANOVA is a non-parametric statistical test of hypothesis to assess if two groups of responses of a given (numeric or ordered) variable differ at a given set of locations determined by quantiles (Mair and Wilcox 2020). Significant results imply that the distributions of the two groups at the tested quantiles are genuinely different: for instance, if the test is performed to check the difference in three quartiles but only the difference at the first quartile is significant, then one concludes that there is evidence of the distributions of the two groups to differ only at the lower tail. Being a non-parametric method, its application does not require any assumption on the probability distribution of the response. To give some computational details, the function 'Qanova' relies on a test for the equality of linear contrasts of selected location measures among J independent groups of observations. In our analysis, we considered the simple differences of quantiles between J=2 groups. For each quantile, the test generates bootstrap replicates of the sample to obtain replications of the quantile differences in the two groups; then, it computes a given distance (for instance, Mahalanobis) from the observed quantile difference to each bootstrap difference and to the benchmark zero vector (corresponding to the null hypothesis of no difference). Then, the bootstrap p-value is determined as the number of times the distances based on bootstrapped differences are lower than the benchmark distance between the zero vector and the observed quantile difference in the two groups. For more than one quantiles, these p-values are then corrected for multiple testing, using Hochberg procedure, for instance. No assumption is required on the distribution of the response, nor on the differences in quantiles.

A further approach would be to assess the significance of quantile differences with bootstrap confidence intervals, by means of the function 'qcomhd' implemented in the same R package. Both methods allow to use Harrel-Davis estimates of quantiles and are suitable when tied values may occur. For our analysis, these two approaches provided equivalent conclusions, yet 'qcomhd' is slightly more demanding than 'Qanova' in terms of computational times. A further approach would be to consider the R package 'Qtools' (Geraci, 2016; Geraci and Farcomeni, 2023) which implements the mid-conditional quantile regression suitable to deal with discrete variables.

Disclosure statement: *The authors report there are no competing interests to declare.*

Data availability statement: *Data are available upon request from authors.*

REFERENCES

Amodio, S., D'Ambrosio, A. and Siciliano, R. (2016). Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the Kemeny axiomatic approach. *European Journal of Operational Research*, 249(2): 667-676.

Baker-Eck, B., Bull, R. and Walsh, D. (2020). Investigative empathy: A strength scale of empathy based on European police perspectives. *Psychiatry, Psychology and Law*, 27:3: 412-427.

Barnhill Bayne, H. (2011). Training medical students in empathic communication. *The Journal for Specialists in Group Work*, 36(4): 316-329.

Can, B., Storcken, T. (2018). A re-characterization of the Kemeny distance. *Journal of Mathematical Economics*, 79:112-116.

D'Ambrosio, A. (2021). ConsRank: Compute the median ranking(s) according to the Kemeny axiomatic approach. *R package version 2.1.2.* https://CRAN.R-project.org/package=ConsRank

Edlins, M. (2021). Developing a model of empathy for public administration. *Administrative Theory & Praxis*, 43(1): 22-41.

Emond, E.J. and Mason, D.W. (2002). A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis,* 11:17-28.

Geraci, M. (2016). Qtools: A collection of models and tools for quantile inference. *The R Journal* 8(2): 117-138.

Geraci, M., Farcomeni, A. (2022). Mid-quantile regression for discrete responses. *Statistical Methods in Medical Research*, 31(5):821-838.

Greeno, E. J., Ting, L., Pecukonis, E., Hodorowicz, M. and Wade, K. (2017). The role of empathy in training social work students in motivational interviewing. *Social Work Education*, 36(7): 794-808.

Hall, J. A. and Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology*, 159(3): 225-243

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Arnold.

Halliday, M. A. K. and Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar*, 3rd edn. London: Arnold.

Kemeny, J. G. and Snell, L. J. (1962). Preference ranking: an axiomatic approach.

Mathematical Models in the Social Sciences, 9-23.

Kim, KJ. (2020). Project-based learning approach to increase medical student empathy. *Medical Education Online*, 25(1): 1-6.

Mair, P. and Wilcox, R. R. (2020). Robust statistical methods in R Using the WRS2 package. *Behavior Research Methods*, 52: 464-488

Morgan, B. (2012). Critical empathy: Vernon Lee's aesthetics and the origins of close reading. *Victorian Studies*, 55(1): 31-56.

Pounds, G., Salter, C., Platt, M. J. and Bryant, P. (2017). Developing a new empathy-specific admissions test for applicants to medical school: A discourse pragmatic approach. *Communication and Medicine*, 14(2): 165-180.

Rasoal, C., Danielsson, H. and Jungert, T. (2012). Empathy among students in engineering programmes. *European Journal of Engineering Education*, 5: 427-435.

doi.org/10.26398/IJAS.327

THE IMPACT OF COVID-19 ON ITALIAN FIRMS'PROFITABILITY: A PANEL EVENT STUDY

Desiré De Luca, Giorgio Tassinari
Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Italy

Abstract. *Using data from the AIDA dataset carried out by Bureau van Dijk, we empirically analyze the effect of COVID-19 pandemic on the financial performance of enterprises, specifically focusing on their operational efficiency and capacity to generate profits. The panel structure of the dataset enables us to execute a panel event study, thereby furnishing empirical insights into the significant repercussions of the COVID-19 outbreak on the profitability of Italian companies. The findings demonstrate a noteworthy and enduring influence of the pandemic on businesses, albeit with varying degrees of severity contingent upon the industrial sector and geographic location of the firms. The heterogeneous results are indicative of the diverse lockdown measures on economic activities and the substantial regional economic disparities prevalent within our country.*

Keywords: *Profitability, ROI, ROE, COVID-19 pandemic, panel event study*

1. INTRODUCTION

In 2020, Italy was the first European country to be affected by COVID-19, with very high rates of contagion and death, and the Italian government was the first of Western governments to implement measures progressively stricter in terms of duration and severity, such as lockdowns, curfews, and limitations on face-to-face interactions, to reduce infection and hospitalization rates. Therefore, beyond the impact on public health, the COVID-19 pandemic has had profound economic impacts on people's livelihoods, but also on enterprises, which faced several economic hardships because of diminished demand, disruptions in supply chains, and production slowdowns associated with unsafe work environments. The health regulations, forcing the social distancing between people and the complete or partial closure of many activities in presence and direct contact with customers, have hampered company sales, generating dramatic problems of liquidity and profitability.

Several studies have attempted to investigate the immediate impact of the COVID-19 outbreak on financial outcomes and/or stock markets (Liu et al., 2020; Nicola et al., 2020; Zhang et al., 2020). However, the main limitation of this existing literature is the restricted dataset and the fact that the COVID-19 pandemic was still ongoing at the time of the research, thus not considering subsequent effects. Therefore, given the impossibility of assessing and drawing definitive conclusions about the impact of the health crisis on corporate financial performance, the results of these studies can be considered preliminary.

The main goal of this paper is therefore twofold: first, to assess the intensity of effects of the pandemic outbreak on business profitability. We wish to contribute to the existing literature by offering a more comprehensive assessment of the effect of COVID-19 on the economic activities in Italy, providing an analysis that is based on panel data spanning from 2013 to 2021 encompassing about 150,000 Italian's firms. Second, we show that these effects tend to be uneven across sectors and regions. Since the outbreak tends to be more impactful in sectors where companies classified as non-essential, which had to be shut down when remote work was not possible, for which stronger effects are plausibly expected. In addition, the evident territorial differences that the virus presented in its expansion and spread and the notable duality of the Italian economic landscape, explains the heterogeneous results across regions.

Moreover, our study adds to the literature that use the event study methodology to evaluate the impact of a non-corporate event such as the outbreak of the disease on financial metrics and/or stock markets (Chen et al., 2007, 2018; Heyden and Heyden, 2021; Liu et al., 2020; Pendell and Cho, 2013). Indeed, we empirically test the different and scarring effect of the outbreak of the pandemic on the profitability of the Italians' firms, implementing a quasi-experimental method, i.e. the panel event study. This design is an effective empirical tool to identify the impact of pandemic on financial outcomes, while taking into account pre-event trends and confounding factors that may affect this relationship. We identify pre- and post- COVID-19 periods, and we use two different profitability outcomes the Return on Investment (henceforth ROI) and the Return on Equity (henceforth ROE), adopting the specifications of dynamic fixed effects models. Estimates are carried out separately for sectors and geographical location of the operational headquarters, due to differences in the stringency of the anticontagion policies adopted by the government depending on the severity of the virus spread.

The rest of the paper is organized as follows. Section 2 contains a review of the existing literature on the impact of the COVID-19 pandemic on the profitability of firms; Section 3 describes the data used; the empirical strategy, with a brief illustration of the panel event study methodology are described in Section 4; while Section 5 summarizes the results, and Section 6 concludes.

2. LITERATURE REVIEW

The pandemic's rapid spread had a profound impact on economies and financial markets on a global scale, affecting practically every business sector and industry.

Several studies have explored the influence of COVID-19 at the macro level, particularly on national stock markets' performance and found an adverse, strong relationship. Baker et al. (2020) suggest that no previous infectious disease outbreak, including the Spanish flu, has affected the U.S. stock market as forcefully as the COVID-19 pandemic. However, the global pandemic of COVID-19 has generated negative shocks on the equity markets across the globe. Harjoto et al. (2021), using an event study method, reveal that the adverse impact of COVID-19 on the equity markets is greater for emerging countries than developed countries.

Meanwhile, a growing number of studies take a closer, micro-level examination of the variations in the profitability of firms under the COVID-19 crisis context, revealing that the severity of the shock due to the pandemic suffered by firms is closely correlated by their size and the sector in which they operate. According to Baldwin and Weder di Mauro (2020), Yan (2020), small firms experienced greater negative shocks from COVID-19 relative to large firms, due to their lower competitive power, worse access to capital, experience and operational efficiency. On the other hand, firms operating in different sectors showed different responses to the COVID-19 shock, as documented by Bartik et al. (2020), Shen et al. (2020), Fahlenbrach et al. (2021) and Golubeva (2021). Particularly, companies facing financial constraints within the manufacturing, retail industries and services domains have experienced more severe repercussions from the outbreak and face an elevated risk of operational closure. Nayak et al. (2022) present a detailed explanation of the impact of COVID-19 on six different industries, and they conclude that not only the major areas such as global supply chains, trade, agriculture industry, transportation and tourism industry, and so on, have been severely disrupted because of the outbreak of the pandemic, but also the economy of various other sectors such as aviation industry, entertainment industry, sports industry, have been severely hampered all over the world due to lockdown.

However, these papers offer initial findings as they have carried out an examination during the initial stages of the pandemic, not considering the more intricate effects the COVID-19 outbreak had on companies in later periods. Therefore, we contribute to this literature providing a more detailed investigation on the impact of the outbreak on the profitability of firms, using an effective empirical tool that considers any pre-event trends and confounding factors that might influence this relationship, namely the panel event study method.

3. DATA

In order to conduct our study, we use data from the AIDA dataset carried out by Bureau van Dijk, which contain balance sheet/income statement information on Italian firms, providing objective information given that balance sheet and income statement are compulsory, and they are compiled according to transparent and standardized criteria by all firms (except banking, insurance sector and the public sector entities).

The dataset spans the period 2013-2021 of a balanced sample at the firm level, i.e., we only use firms with data for all sample periods.

In addition, we restrict our focus to small-medium firms (less than 100 employees), as large firms are able to mitigate the negative shock of the pandemic outbreak due to their greater competitive power than small firms due to larger market share, better access to capital, experience, and operational efficiency (Ichev and Marinc, 2018).

Tables 1 and 2 in the Appendix provide some descriptive statistics.

4. METHODOLOGY: PANEL EVENT STUDY

In this study, the empirical work is based on the panel event study methodology, as we seek to reveal how Italian firms, particularly their profitability, behave during and after the outbreak of the coronavirus. The design estimates the impact of some events that occur by considering the variation in outcomes around the adoption of the event compared with a baseline reference period, one can estimate both event leads and lags, which allows us to have a clear visual representation of the event's causal impact (Clarke et al., 2021).

The key assumption underlying consistent estimation in panel event study model is that the occurrence of the event is not systematically related to the changes in levels that would have occurred in the future in the absence of the event. In particular, this methodology has been borne out of older difference-in-differences (DD) designs, or two-way fixed-effects models, to overcome their limits, such as the parallel trend assumption. Therefore, it can be used, also in cases where events occur at the same time in each unit.

We identify pre- and post- COVID-19 periods, which coincide with the year 2020. Subsequently, we estimate a dynamic model of the form:

$$y_{it} = \alpha + \sum_{j=2}^{J} \beta_j Lag_{it}^{j} + \sum_{k=1}^{K} \gamma_k Lead_{it}^{k} + \lambda_i + \mu_t + \varepsilon_{it} \quad . \quad (1)$$

Where the y_{it} is the profitability outcome of firm *i* at the time *t*. We use two different profitability outcomes: i) the Return on Invested (ROI), which is a metric that may be used to assess a company's profitability as well as reveal the origins of its competitive advantages; ii) the Return on Equity (ROE), which is a financial metric that reveals the ability of a company to convert its equity financing into profits. When contrasted to ROI, it represents the total return on all capital invested in an asset, whereas ROE solely evaluates the equity component (Damodaran, 2007).

Lag_{it}^{j} and $Lead_{it}^{k}$ are the *j-lag* and *k-lead* set of dummies denoting the time distance away from the event – the outbreak of COVID-19 - λ_i the firms fixed effect, μ_t the time fixed effect.

The terms β_j and γ_k are parameters to be estimated denoting how financial measures vary in periods before and after the COVID-19 outbreak (compared to the year prior to the event, in this case the year of 2019).

In general, when policies are assigned by a group, such as a state, and outcomes are followed over time within these groups, a standard inference problem arises, related to the potential serial correlation of the outcome variable over time. However, the standard solution is to allow for within-cluster autocorrelation by using a cluster-robust variance-covariance estimator (CRVE) to estimate standard errors and CIs on regression parameters (Wooldridge, 2010). Therefore, to overcome this problem, we adopt a dynamic fixed effects model specifications with standard errors clustered at the firm level.

5. RESULTS

To analyze the impact of the COVID-19 pandemic on the profitability of Italian companies, we estimate Eq. (1) separately for ROI and ROE. In addition, considering that some companies were closed for several months due to government decisions (e.g., the March 22, 2020 DPCM), while others were able to continue operating during the pandemic, albeit not as usual, but with rather limited restrictions related to mitigating the spread of the virus, we make separate estimates by industry sectors. Furthermore, we estimate the impact of the pandemic by also distinguishing companies by geographic area (north, central, and south), so as to consider the different economic starting conditions that companies face in relation to the location of the company, but also to the different level of spread of the COVID-19 virus regionally. [1]

Considering that, the leads and as the fixed effects included in the model allow us to control for any linear and non-linear trends of unobservable that may affect profitability in a given year. It is important to note that in the estimates reported below, in which all possible leads are included, some significant differences are observed in the pre-COVID 19 period, but these are sufficiently far removed from the pandemic period, so these significant differences are probably due to changes in the composition of variables and not to temporal trends.

[1] All event-study results were conducted using the Stata program *eventdd* by Clarke and Schythe, (2023).

In Tables 3a and 3b and Tables 4a and 4b, we report estimates on the impact of the pandemic on the accommodation and catering services sector and the rental, travel agencies, business support services, respectively, that are nested in the hospitality industry. Our findings suggest that, in the context of the hospitality industry, activities of accommodation and catering services sector were the most affected by the outbreak of pandemic. In fact, the ROI of Italians firms operating in this sector (column 1 of Table 3a), compared to the baseline year -i.e., 2019 the period immediately prior to the occurrence of the pandemic- declines by about 6.8 pp at the timing of the pandemic outbreak, and deteriorates further in the following year by about 1.8 pp. Table 3b, column 1, shows the impact of COVID-19 on profitability measured by the ROE; the results suggest a decrease of about 11 pp in 2020, but an increase in the following year, which could be due to an increase in the level of debt due to emergency loans, resulting in higher borrowing costs and consequently higher ROE.

Table 3a: The impact of COVID-19 on the ROI of firms operating in the accommodation and catering services sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-2.318	***	-2.433	***	-1.889	***	-2.629	***
	(0.184)		(0.276)		(0.355)		(0.323)	
t-6	-1.361	*	-1.060	*	-1.244	*	-1.968	**
	(0.177)		(0.266)		(0.342)		(0.312)	
t-5	-0.493	*	0.144		-0.962	**	-1.007	*
	(0.171)		(0.259)		(0.326)		(0.306)	
t-4	-0.138	*	0.275		-0.554	*	-0.341	
	(0.162)		(0.248)		(0.308)		(0.283)	
t-3	-0.0200		0.255		-0.100		-0.415	
	(0.153)		(0.236)		(0.283)		(0.271)	
t-2	0.158		0.293		0.0157		0.0193	
	(0.139)		(0.214)		(0.256)		(0.251)	
Event	-6.769	***	-7.200	***	-7.047	***	-5.792	***
	(0.166)		(0.251)		(0.315)		(0.296)	
t+1	-1.803	***	-1.861	***	-2.360	***	-1.270	***
	(0.156)		(0.240)		(0.300)		(0.271)	
Intercept	6.424	***	5.855	***	6.452	***	7.227	***
	(0.100)		(0.154)		(0.188)		(0.178)	
Obs	69,530		29,756		20,544		20,800	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 3b: The impact of COVID-19 on the ROE of firms operating in the accommodation and catering services sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-1.833	***	-2.837	***	-1.317		-0.726	
	(0.458)		(0.721)		(0.821)		(0.809)	
t-6	1.068	**	0.755		0.722		1.637	**
	(0.427)		(0.674)		(0.769)		(0.752)	
t-5	4.121	*	5.230	**	3.161	**	3.4846	*
	(0.408)		(0.649)		(0.725)		(0.727)	
t-4	3.604	**	4.305	**	2.799	*	3.340	**
	(0.392)		(0.621)		(0.703)		(0.695)	
t-3	2.647		3.691		1.238		2.653	
	(0.362)		(0.566)		(0.650)		(0.657)	
t-2	1.829		1.785		1.212		2.386	
	(0.324)		(0.507)		(0.584)		(0.587)	
Event	-11.48	***	-11.60	***	-11.67	***	-10.91	***
	(0.394)		(0.605)		(0.745)		(0.694)	
t+1	1.069	***	1.844	***	0.117		0.861	
	(0.373)		(0.583)		(0.694)		(0.652)	
Intercept	9.188	***	7.536	***	9.496	***	11.29	***
	(0.224)		(0.355)		(0.398)		(0.398)	
Obs	137,085		57,183		42,055		40,971	

Note. The Table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Similarly, in Tables 4a and 4b we note that the effect of the spread of the virus has the same negative impact on the profitability indicators for firms operating in the rental, travel agency and business support services sectors in Italy.

Table 4a: The impact of COVID-19 on the ROI of firms operating in the rental, travel agencies, business support services sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-0.553	**	-0.836	**	0.0880		-0.820	
	(0.257)		(0.342)		(0.500)		(0.574)	
t-6	-0.156		-0.214		0.279		-0.634	
	(0.254)		(0.335)		(0.503)		(0.565)	
t-5	0.290		0.170		0.436		0.216	
	(0.238)		(0.317)		(0.468)		(0.522)	
t-4	0.421	*	0.303		0.630		0.232	
	(0.221)		(0.309)		(0.415)		(0.460)	
t-3	0.273		0.319		0.249		0.0420	
	(0.212)		(0.287)		(0.401)		(0.460)	
t-2	0.449		0.503		0.309		0.306	
	(0.193)		(0.264)		(0.376)		(0.408)	
Event	-3.053	***	-3.325	***	-3.253	***	-2.224	***
	(0.220)		(0.305)		(0.415)		(0.462)	
t+1	-0.972	***	-1.076	***	-0.893	**	-0.820	*
	(0.213)		(0.297)		(0.403)		(0.449)	
Intercept	8.581	***	8.619	***	8.296	***	8.922	***
	(0.142)		(0.196)		(0.271)		(0.302)	
Obs	36,388		18,516		9,864		8,789	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. Significance: *** p<.01, ** p<.05, *p<.10

The pandemic has also severely affected the arts, cultural and intellectual activities sector, which were subjected to a harsh and lengthy lockout: in fact, as of March 2020, all cultural facilities were closed, and on-site activities were suspended. Although demand for cultural and creative content intensified during the lockdown period, and digital access became more critical than ever before (Radermecker, 2020), this sector has been one of the hardest hits and probably one of the slowest to recover.

The estimates shown in Tables 5a and 5b suggest a significant negative impact of COVID-19, which has indiscriminately affected businesses operating in these sectors throughout the peninsula. The indicator of the ROI declined by about 6 pp during the first year of the pandemic and a further decline of just under

2 pp in 2021. As regards ROE, this declined by about 7 pp in 2020, with no statistically significant recovery in the second year of the pandemic.

Table 4b: The impact of COVID-19 on the ROE of firms operating in the rental, travel agencies, business support services sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-0.0151		-0.418		-0.640		1.410	
	(0.532)		(0.716)		(1.025)		(1.178)	
t-6	1.611	***	1.406	**	0.482		3.070	***
	(0.505)		(0.683)		(0.972)		(1.096)	
t-5	3.708	***	3.535	***	3.460	***	4.070	***
	(0.472)		(0.647)		(0.905)		(1.010)	
t-4	2.793	*	3.367	**	2.058	**	2.205	**
	(0.450)		(0.617)		(0.846)		(0.975)	
t-3	2.382		2.757		1.824	*	2.199	
	(0.428)		(0.568)		(0.835)		(0.933)	
t-2	1.390		1.960		0.630		0.765	
	(0.395)		(0.522)		(0.757)		(0.881)	
Event	-3.670	***	-3.938	***	-4.033	***	-2.524	***
	(0.425)		(0.590)		(0.808)		(0.883)	
t+1	0.211	***	0.293	***	-0.0751		0.619	
	(0.429)		(0.586)		(0.824)		(0.914)	
Intercept	12.17	***	11.80	***	11.84	***	13.25	***
	(0.273)		(0.372)		(0.508)		(0.597)	
Obs	77,985		38,078		21,696		19,769	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 5a: The impact of COVID-19 on the ROI of firms operating in the artistic, sports, and entertainment activities sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-0.731		-0.734		-1.183		-0.0685	
	(0.470)		(0.690)		(0.866)		(0.915)	
t-6	-0.806	*	-0.928		-1.384		0.136	
	(0.474)		(0.702)		(0.842)		(0.924)	
t-5	-0.280		0.616		-1.236		-0.878	
	(0.454)		(0.699)		(0.802)		(0.832)	
t-4	0.274		0.721		0.452		-0.523	
	(0.411)		(0.610)		(0.729)		(0.794)	
t-3	0.739*		0.672		0.625		0.978	
	(0.388)		(0.571)		(0.738)		(0.732)	
t-2	0.689		1.016		0.592		0.612	
	(0.376)		(0.536)		(0.726)		(0.737)	
Event	-5.874	***	-7.019	***	-5.821	***	-4.060	***
	(0.441)		(0.659)		(0.835)		(0.795)	
t+1	-2.138	***	-2.847	***	-1.706	**	-1.258	
	(0.433)		(0.643)		(0.845)		(0.773)	
Intercept	5.759	***	5.473	***	5.634	***	6.060	***
	(0.270)		(0.397)		(0.503)		(0.512)	
Obs	11,496		5,226		3,122		3,433	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 5b: The impact of COVID-19 on the ROE of firms operating in the artistic, sports, entertainment, and entertainment activities sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-0.663		-0.459		-3.447	*	0.726	
	(1.095)		(1.674)		(1.993)		(1.971)	
t-6	2.617	**	2.263		-1.959		6.594	***
	(1.017)		(1.533)		(1.959)		(1.803)	
t-5	4.873	***	5.715	***	2.383		5.161	***
	(0.946)		(1.363)		(1.815)		(1.808)	
t-4	3.794	***	4.349	**	3.424	**	2.791	
	(0.900)		(1.347)		(1.636)		(1.711)	
t-3	5.022		4.685	*	3.906		5.867	
	(0.816)		(1.200)		(1.522)		(1.543)	
t-2	2.671		3.300		0.364		3.138	
	(0.770)		(1.120)		(1.495)		(1.430)	
Event	-9.371	***	-12.38	***	-7.527	***	-7.327	***
	(0.904)		(1.455)		(1.626)		(1.541)	
t+1	-0.363		0.278		-0.726		-1.198	
	(0.868)		(1.365)		(1.575)		(1.523)	
Intercept	7.563	***	6.179	***	7.050	***	10.25	***
	(0.527)		(0.798)		(0.989)		(0.957)	
Obs	24,650		10,775		6,767		7,756	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Similar but smaller results are obtained, if we look at the impact of the pandemic on the firms operating in the professional, scientific, and technical activities sector, which are reported in Tables 6a and 6b.

Table 6a: The impact of COVID-19 on the ROI of firms operating in the professional, scientific, and technical activities sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	0.315		0.314		0.857	*	-0.238	
	(0.197)		(0.242)		(0.439)		(0.511)	
t-6	0.390	**	0.349		0.780	*	0.304	
	(0.194)		(0.238)		(0.435)		(0.498)	
t-5	0.593	**	0.685	**	0.602		0.356	
	(0.184)		(0.227)		(0.404)		(0.475)	
t-4	0.639	**	1.015	**	0.0801		-0.107	
	(0.178)		(0.224)		(0.387)		(0.427)	
t-3	0.718		1.008		0.256		0.253	
	(0.163)		(0.206)		(0.347)		(0.394)	
t-2	0.517		0.612		0.775		-0.126	
	(0.154)		(0.194)		(0.321)		(0.376)	
Event	-2.103	***	-2.155	***	-2.243	***	-1.560	***
	(0.163)		(0.208)		(0.362)		(0.372)	
t+1	-0.686	***	-0.578	**	-1.062	***	-0.477	
	(0.164)		(0.210)		(0.351)		(0.378)	
Intercept	7.915	***	7.675	***	8.091	***	8.448	***
	(0.111)		(0.140)		(0.245)		(0.261)	
Obs	51,764		32,152		11,452		8,997	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 6b: The impact of COVID-19 on the ROE of firms operating in the professional, scientific, and technical activities sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-0.181		-0.854	*	0.130		2.569	***
	(0.393)		(0.489)		(0.853)		(0.992)	
t-6	0.827	**	0.387		1.108		2.695	***
	(0.378)		(0.476)		(0.787)		(0.957)	
t-5	2.832	***	2.312	***	2.494		4.941	***
	(0.354)		(0.443)		(0.759)		(0.897)	
t-4	1.856	***	1.891	***	1.170		2.722	***
	(0.337)		(0.422)		(0.734)		(0.826)	
t-3	2.362	**	2.578	**	1.575	*	2.182	**
	(0.317)		(0.399)		(0.685)		(0.769)	
t-2	2.140		2.388		1.650		1.948	*
	(0.289)		(0.364)		(0.610)		(0.722)	
Event	-2.253	***	-2.635	***	-2.435	***	-0.754	
	(0.298)		(0.381)		(0.627)		(0.716)	
t+1	0.449		0.292		-0.241		2.274	***
	(0.309)		(0.389)		(0.663)		(0.758)	
Intercept	12.38	***	12.67	***	12.49	***	10.98	***
	(0.203)		(0.256)		(0.433)		(0.493)	
Obs	108,201		66,087		24,935		18,825	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

In addition, the decision to limit international, regional, and local travel with the objective of carrying out health controls, have also jeopardized the transportation industry, which heavily depends on the mobility of people (Yang et al., 2020). Overall, both passenger transport and freight have suffered severe setbacks from the pandemic. Our findings reported in Tables 7a and 7b suggest that the impact of the COVID-19 outbreak was negative on the profitability indicators. However, a slight recovery in 2021 in the ability to generate profits is estimated for firms operating in the north; in fact, the coefficient is not statistically different from zero when looking at the impact on ROI, while it is positive and statistically significant when looking at ROE.

Table 7a: The impact of COVID-19 on the ROI of firms operating in the transport and warehousing sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	0.232		0.220		-0.0435		0.546	
	(0.251)		(0.335)		(0.575)		(0.477)	
t-6	0.802	***	0.939	*	0.527		0.858	*
	(0.249)		(0.339)		(0.574)		(0.453)	
t-5	1.661	***	1.895	***	0.851		1.902	***
	(0.233)		(0.315)		(0.540)		(0.422)	
t-4	1.896	***	2.130	***	1.378	***	1.928	**
	(0.218)		(0.310)		(0.473)		(0.386)	
t-3	1.389	*	1.781	*	1.232	**	1.039	*
	(0.204)		(0.288)		(0.457)		(0.356)	
t-2	0.230		0.382		0.129		0.123	
	(0.190)		(0.269)		(0.425)		(0.330)	
Event	-1.964	***	-1.746	***	-2.848	***	-1.565	***
	(0.207)		(0.298)		(0.438)		(0.364)	
t+1	-0.310		0.364		-0.866	**	-0.845	**
	(0.204)		(0.291)		(0.441)		(0.359)	
Intercept	7.409	***	6.974	***	7.618	***	7.863	***
	(0.137)		(0.195)		(0.305)		(0.236)	
Obs	35,651		16,680		7,853		11,820	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 7b: The impact of COVID-19 on the ROE of firms operating in the transport and warehousing sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	2.673	***	0.245		4.146		5.304***	**
	(0.554)		(0.814)		(1.147)		(0.967)	
t-6	5.200	***	3.032	***	5.707		8.365	***
	(0.518)		(0.742)		(1.114)		(0.914)	
t-5	7.495	***	6.320	***	9.390	***	7.905	***
	(0.490)		(0.713)		(1.028)		(0.860)	
t-4	6.253	***	5.240	***	7.294	***	7.003	***
	(0.467)		(0.697)		(0.941)		(0.813)	
t-3	4.388	**	4.028	*	5.593	*	4.298	*
	(0.437)		(0.651)		(0.900)		(0.749)	
t-2	1.294		1.030		2.453		1.042	
	(0.403)		(0.597)		(0.846)		(0.689)	
Event	-3.683	***	-3.779	***	-4.733	***	-2.510	***
	(0.448)		(0.663)		(0.980)		(0.742)	
t+1	0.286		1.552	**	0.216		-1.240	*
	(0.445)		(0.652)		(0.993)		(0.735)	
Intercept	10.16	***	9.610	***	8.986	***	11.39	***
	(0.288)		(0.433)		(0.605)		(0.483)	
Obs	65,395		28,908		14,955		22,842	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

The outbreak of pandemic affected the manufacturing sector, as expected. The extent of disruption was largely twofold: an endogenous disruption of production processes and systems and extreme shifts in supply and demand caused by an exogenous supply chain disruption. Due to supply chain disruption and the unavailability of raw material, some industries, such as electronics, have put new product development on hold and have also reduced production quantities (Ivanov and Dolgui, 2020). Therefore, the pandemic has paralyzed the manufacturing sector, negatively impacting firms' profitability. Tables 8a show that ROI decreased not only during the first year of the health crisis, but also during the following year, while the results reported in Table 8b suggest, that profitability as measured by the ROE indicator would appear to have declined by about 5 pp; while a more pronounced reduction is estimated for firms operating in central Italy, with a 6 pp decrease from the year 2019.

Table 8a: The impact of COVID-19 on the ROI of firms operating in the manufacturing sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-0.278	***	-0.290	***	0.144		-0.602	***
	(0.0895)		(0.112)		(0.207)		(0.197)	
t-6	0.318	***	0.450	***	0.421	**	-0.250	
	(0.0861)		(0.108)		(0.198)		(0.194)	
t-5	0.697	***	0.789	***	0.832	***	0.261	
	(0.0830)		(0.104)		(0.191)		(0.185)	
t-4	1.034	***	1.294	***	0.983	***	0.165	
	(0.0792)		(0.100)		(0.181)		(0.176)	
t-3	1.069	**	1.421	*	0.973	**	-0.0892	
	(0.0747)		(0.0948)		(0.172)		(0.163)	
t-2	0.834		1.126	*	0.682	*	-0.0230	
	(0.0683)		(0.0862)		(0.155)		(0.153)	
Event	-2.958	***	-2.937	***	-3.381	***	-2.517	***
	(0.0756)		(0.0946)		(0.176)		(0.170)	
t+1	-0.290	***	-0.138		-0.580	***	-0.496	***
	(0.0755)		(0.0952)		(0.174)		(0.168)	
Intercept	7.927	***	7.884	***	7.879	***	8.021	***
	(0.0509)		(0.0648)		(0.116)		(0.109)	
Obs	190,287		120,482		37,796		47,875	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 8b: The impact of COVID-19 on the ROE of firms operating in the manufacturing sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-1.075	***	-1.680	***	-0.264		0.631	
	(0.205)		(0.259)		(0.458)		(0.470)	
t-6	0.974	**	0.400		1.678	***	2.311	***
	(0.197)		(0.248)		(0.447)		(0.452)	
t-5	3.212	***	2.786	***	3.552	***	4.331	***
	(0.188)		(0.236)		(0.426)		(0.431)	
t-4	3.203	***	3.195	***	3.637	***	2.883	***
	(0.178)		(0.223)		(0.393)		(0.415)	
t-3	3.879	*	4.380	**	3.805	*	2.402	**
	(0.171)		(0.214)		(0.379)		(0.403)	
t-2	2.481		3.036		2.229		1.105	
	(0.154)		(0.194)		(0.339)		(0.358)	
Event	-5.358	***	-5.426	***	-5.951	***	-4.282	***
	(0.174)		(0.218)		(0.391)		(0.401)	
t+1	-0.0803		0.346	***	-0.561		-0.757	*
	(0.170)		(0.213)		(0.380)		(0.400)	
Intercept	10.38	***	10.04	***	10.20	***	11.42	***
	(0.112)		(0.142)		(0.248)		(0.254)	
Obs	314,518		190,805		68,774		61,094	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

The pandemic and subsequent lockdown resulted in social distancing and isolation that had a detrimental effect on the wholesale and retail sector, despite sales of small and medium-sized supermarket chains increased in a short period of time, as severe pandemic measures led to the closure of high population densities malls and supermarkets. In addition, the unprecedented systematic uncertainty resulting from the combination of uncertainty about the duration of the crisis, the expectations for income and employment caused people to reduce current consumption and increase savings, making the situation worse. Results reported in Tables 9a and 9b, suggest that the pandemic outbreak has had a large impact on the Italian firms of the wholesale and retail sector, with ROI falling by

2.5 pp in 2020, and a reduction of ROE of about 3 pp. In addition to all the problems caused by the pandemic, it has accelerated the online presence of retail enterprises. However, many traditional retail enterprises were unable to develop online platforms, and carry on the business, which resulted in cash flow constraints that have brought many businesses to the brink of bankruptcy. As a result, the government has taken a series of measures for economic support especially for the wholesale and retail sector, which has been particularly hard hit by the COVID-19 shock, and this could explain the estimated improvement in both profitability indicators, in 2021.

Table 9a: The impact of COVID-19 on the ROI of firms operating in the wholesale and retail, repair of motor vehicles and motorcycles sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-0.578	***	-0.876	***	-0.394	**	-0.176	
	(0.0926)		(0.131)		(0.196)		(0.172)	
t-6	-0.0828		-0.140		-0.0665		0.0230	
	(0.0891)		(0.127)		(0.186)		(0.163)	
t-5	0.512	***	0.484	***	0.343	*	0.695	***
	(0.0838)		(0.119)		(0.179)		(0.152)	
t-4	0.779	***	0.922	***	0.502	***	0.753	***
	(0.0792)		(0.113)		(0.169)		(0.143)	
t-3	0.440	**	0.826	**	0.118		0.0257	
	(0.0744)		(0.109)		(0.154)		(0.132)	
t-2	0.369		0.538		0.229	*	0.200	*
	(0.0666)		(0.0968)		(0.139)		(0.117)	
Event	-2.470	***	-2.348	***	-3.095	***	-2.177	***
	(0.0764)		(0.109)		(0.165)		(0.134)	
t+1	0.0320		0.502	***	-0.336	**	-0.397	
	(0.0746)		(0.110)		(0.157)		(0.129)	
Intercept	7.986	***	7.690	***	7.866	***	8.551	***
	(0.0495)		(0.0726)		(0.104)		(0.0860)	
Obs	188,434		89,106		45,352		57,480	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 9b: The impact of COVID-19 on the ROE of firms operating in the wholesale and retail, repair of motor vehicles and motorcycles sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-0.761	***	-1.835	***	-0.618		1.158	***
	(0.214)		(0.311)		(0.439)		(0.386)	
t-6	1.063	***	0.0293		0.867	**	3.073	***
	(0.201)		(0.293)		(0.411)		(0.359)	
t-5	3.298	***	2.767	**	3.087	**	4.454	***
	(0.192)		(0.280)		(0.398)		(0.336)	
t-4	2.728	**	2.720	***	2.540	***	2.958	**
	(0.185)		(0.269)		(0.381)		(0.326)	
t-3	2.264		2.830	*	1.896	*	1.755	*
	(0.175)		(0.259)		(0.361)		(0.306)	
t-2	1.303		1.604		1.125		1.045	
	(0.159)		(0.234)		(0.329)		(0.275)	
Event	-3.244	***	-2.453	***	-4.132	***	-3.758	***
	(0.178)		(0.260)		(0.369)		(0.310)	
t+1	0.884	***	2.232	***	0.205		-0.408	
	(0.173)		(0.254)		(0.359)		(0.299)	
Intercept	10.48	***	9.292	***	10.08	***	12.44	***
	(0.113)		(0.168)		(0.234)		(0.192)	
Obs	326,309		146,391		83,463		102,839	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

In addition, COVID-19 has completely disrupted any previous daily practice also in the construction industry, for example architects left the office and began completing the design stage remotely. Therefore, the pandemic has had a severe impact on the ability of contractors to work on-site and to meet deadlines. Some sites were suspended, there have been delays in payments and in the delivery of materials, all this has led to a lack of cash, manpower, and resources in general, creating a chain of delays, loss of productivity and profitability, as shown in Tables 10a and 10b.

However, since July 2020, with the desire to restart the construction sector strongly penalized by the pandemic, it has been implemented a public grant policy, so called "*Superbonus 110%*", (Italian Law. L. 17 July, 2020) that by cutting costs for property owners strongly encouraged energy efficiency works. Indeed, as we can see from the Tables 10a and 10b, in 2021 than 2019, the profitability of the construction industries increases.

Table 10a: The impact of COVID-19 on the ROI of firms operating in the construction sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-0.0679		-0.396	**	0.572	**	-0.146	
	(0.128)		(0.195)		(0.254)		(0.217)	
t-6	-0.242	*	-0.502	***	-0.0231		-0.0497	
	(0.126)		(0.192)		(0.253)		(0.212)	
t-5	0.411	***	0.311	*	0.251		0.646	***
	(0.121)		(0.179)		(0.243)		(0.211)	
t-4	0.462	***	0.476	***	0.788	***	0.221	
	(0.116)		(0.176)		(0.236)		(0.192)	
t-3	0.285	*	0.551	**	0.352		-0.103	
	(0.111)		(0.167)		(0.225)		(0.186)	
t-2	0.422		0.545		0.361		0.267	
	(0.101)		(0.152)		(0.209)		(0.165)	
Event	-1.845	***	-1.927	***	-2.120	***	-1.557	***
	(0.108)		(0.165)		(0.219)		(0.180)	
t+1	0.910	***	0.830	***	0.697	***	1.104	***
	(0.110)		(0.167)		(0.224)		(0.185)	
Intercept	8.865	***	8.835	***	8.687	***	9.053	***
	(0.0747)		(0.114)		(0.149)		(0.124)	
Obs	103,855		45,819		25,593		35,414	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 10b: The impact of COVID-19 on the ROE of firms operating in the construction sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-0.0185		-1.360	***	0.257	***	1.544***	
	(0.274)		(0.422)		(0.533)		(0.469)	
t-6	0.439	*	-1.217	***	0.118		2.869	**
	(0.260)		(0.399)		(0.516)		(0.434)	
t-5	4.104	***	2.922	***	3.574	***	5.942	***
	(0.251)		(0.383)		(0.507)		(0.427)	
t-4	2.691	***	3.183	***	2.593	***	2.150	***
	(0.239)		(0.367)		(0.463)		(0.411)	
t-3	2.238		2.734		2.246		1.769	
	(0.226)		(0.349)		(0.441)		(0.382)	
t-2	1.469		2.042		1.691		0.758	
	(0.206)		(0.306)		(0.413)		(0.356)	
Event	-3.278	***	-3.462	***	-4.409	***	-2.136	***
	(0.212)		(0.322)		(0.426)		(0.360)	
t+1	4.012	***	3.250	***	4.254	***	4.812	***
	(0.216)		(0.336)		(0.426)		(0.363)	
Intercept	14.24	***	14.21	***	14.27	***	14.18	***
	(0.144)		(0.222)		(0.278)		(0.244)	
Obs	212,817		90,963		55,183		72,426	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

The analysis proceeds, distinguishing these branches of the economy from other sectors such as agriculture and mining, which were able to continue operating during the COVID-19 pandemic, albeit not as usual, but with rather limited restrictions. Despite government restrictions on agricultural labor mobility, the establishment of safety protocols to prevent virus transmission, the trade and provision of essential items has been ensured and normalized. Indeed, results reported in Tables 11a and 11b, suggest that the pandemic slightly reduced the profitability of the firms operating in the agriculture, forestry and fishing sector, but with no statistically significant impact on ROE for enterprises operating in northern and southern Italy. In contrast, for firms operating in central

Italy, for which the pandemic caused a reduction in profitability, this effect may have been driven by the complete closure of only those firms operating in the forestry and logging subsector, which are more prevalent in central Italy.

Table 11a: The impact of COVID-19 on the ROI of firms operating in the agriculture, forestry, and fishing sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	0.201		0.121		-0.641	*	0.816	*
	(0.238)		(0.383)		(0.389)		(0.422)	
t-6	-0.174		0.216		-0.489		-0.292	
	(0.238)		(0.388)		(0.372)		(0.423)	
t-5	-0.168		-0.0502		-0.500		-0.0858	
	(0.232)		(0.372)		(0.390)		(0.409)	
t-4	0.409	*	0.452		0.401		0.359	
	(0.217)		(0.351)		(0.341)		(0.384)	
t-3	-0.0494		0.146		-0.441		0.00742	
	(0.212)		(0.342)		(0.316)		(0.381)	
t-2	0.0273		0.322		0.204		-0.302	
	(0.189)		(0.298)		(0.277)		(0.341)	
Event	-0.703	***	-0.516	*	-0.559	*	-0.901	**
	(0.200)		(0.286)		(0.339)		(0.360)	
t+1	0.0864		0.177		0.304		-0.210	
	(0.204)		(0.320)		(0.341)		(0.357)	
Intercept	2.518	***	2.297	***	0.879	***	3.73	***
	(0.140)		(0.221)		(0.221)		(0.251)	
Obs	21,146		6,787		5,803		8,975	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** $p<.01$, ** $p<.05$, *$p<.10$

Table 11b: The impact of COVID-19 on the ROE of firms operating in the agriculture, forestry, and fishing sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	2.234	***	1.438		0.698		3.584	***
	(0.720)		(1.135)		(1.356)		(1.176)	
t-6	0.920		-0.155		-0.383		2.303	**
	(0.710)		(1.196)		(1.344)		(1.119)	
t-5	1.791	***	0.501		1.889		2.643	**
	(0.659)		(1.078)		(1.235)		(1.050)	
t-4	2.738	***	2.843	***	1.729		3.249	***
	(0.663)		(1.023)		(1.257)		(1.076)	
t-3	1.831	*	2.490	*	0.160		2.135	*
	(0.611)		(1.002)		(1.132)		(0.975)	
t-2	0.878		1.687		1.961		-0.258	
	(0.564)		(0.885)		(1.041)		(0.914)	
Event	-1.026	*	-0.264		-1.810	*	-0.941	
	(0.578)		(0.938)		(1.074)		(0.925)	
t+1	0.570		1.130		0.122		0.495	
	(0.579)		(0.942)		(1.150)		(0.900)	
Intercept	3.094	***	2.210	***	-2.01	***	6.562	***
	(0.396)		(0.625)		(0.753)		(0.633)	
Obs	32,613		9,873		8,479		14,873	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The p-value of the F-test is less than the significance level, the model is significant. The baseline year is 2019. Significance: *** p<.01, ** p<.05, *p<.10

In Tables 12a and 12b we can see estimates for the mining industry, which are negatively affected by the COVID-19 crisis, although many mines have remained operational and productive during the pandemic, despite having less people on site. However, business continuity has come at a cost due to the added expenses of new processes, procedures, health protocols, and so on, (Gałaś, et al., 2021; Jowitt, 2021).

Table 12a: The impact of COVID-19 on the ROI of firms operating in the extraction of minerals from quarries and mines

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-0.804		-1.868	*	-0.323		-0.0881	
	(0.742)		(1.108)		(1.310)		(1.392)	
t-6	-0.108		-0.795		0.0349		0.301	
	(0.752)		(1.096)		(1.398)		(1.449)	
t-5	-0.605		-1.390		-0.744		0.887	
	(0.713)		(1.056)		(1.451)		(1.202)	
t-4	-0.131		-1.866	*	0.689		1.284	
	(0.698)		(1.072)		(1.240)		(1.242)	
t-3	-0.689		-1.290		-0.143		-0.407	
	(0.628)		(0.910)		(0.910)		(1.343)	
t-2	-0.0848		-0.428		0.834		-0.320	
	(0.537)		(0.737)		(1.034)		(1.046)	
Event	-1.52	**	-1.804	**	-2.855	**	-0.219	
	(0.601)		(0.811)		(1.250)		(1.081)	
t+1	-0.169		-0.381		-1.869		1.262	
	(0.662)		(0.901)		(1.246)		(1.304)	
Intercept	4.374	***	4.772***	***	4.734	***	3.885	***
	(0.446)		(0.661)		(0.750)		(0.858)	
Obs	2,350		1,108		598		723	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 12b: The impact of COVID-19 on the ROE of the firms operating in the extraction of minerals from quarries and mines

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-0.793		-0.0644		-0.876		-1.324	
	(1.817)		(2.169)		(3.434)		(4.033)	
t-6	2.778	*	1.556		2.908		5.636	
	(1.677)		(2.370)		(2.982)		(3.507)	
t-5	3.234	*	-0.120		4.091		8.358	***
	(1.663)		(2.518)		(3.348)		(2.946)	
t-4	-0.858		-3.777	*	0.252		4.006	
	(1.700)		(2.274)		(3.411)		(3.420)	
t-3	0.352		1.391		0.336		-0.311	
	(1.582)		(1.888)		(3.245)		(3.472)	
t-2	2.209		0.859		6.125	*	2.138	
	(1.416)		(1.941)		(2.594)		(2.941)	
Event	-2.625		-2.890		-2.280		-1.684	
	(1.613)		(2.163)		(2.902)		(3.335)	
t+1	1.790		1.842		-1.424		4.979	
	(1.694)		(2.467)		(3.050)		(3.278)	
Intercept	2.360	**	2.780	**	1.857		1.129	
	(1.043)		(1.336)		(2.064)		(2.208)	
Obs	3,459		1,550		917		1,103	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

As for the power sector, the pandemic had a relatively small impact on these economic activities (Tables 13a and 13b); and similarly, for the water supply sector (Tables 14a and 14b). Indeed, while the lockdown measures depressed consumption in the commercial and industrial sectors, they increased demand in the residential sector (Renukappa et al., 2021).

Table 13a: The impact of COVID-19 on the ROI of firms operating in the supply of electricity, gas, steam and air conditioning sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	1.334	*	1.809	*	1.060		1.036	
	(0.707)		(0.958)		(1.552)		(1.319)	
t-6	1.276	**	1.722	**	0.281		1.199	
	(0.618)		(0.840)		(1.153)		(1.236)	
t-5	0.181		0.304		-0.542		0.931	
	(0.620)		(0.804)		(1.469)		(1.312)	
t-4	0.343		0.0572		0.976		0.740	
	(0.564)		(0.766)		(1.165)		(1.093)	
t-3	0.767		0.463		1.304		1.613	
	(0.525)		(0.720)		(1.054)		(1.048)	
t-2	0.340		-0.349		1.878	**	0.732	
	(0.485)		(0.666)		(0.908)		(0.970)	
Event	-1.214	***	-1.065	*	-1.556		-1.732	**
	(0.422)		(0.547)		(1.009)		(0.869)	
t+1	-1.385	**	-0.456	***	-3.064	**	-2.881	***
	(0.551)		(0.761)		(1.187)		(1.022)	
Intercept	6.684	***	6.636***	***	6.788	***	6.835	***
	(0.356)		(0.478)		(0.731)		(0.735)	
Obs	3,371		1,982		647		849	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 13b: The impact of COVID-19 on the ROE of firms operating in the supply of electricity, gas, steam and air conditioning sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	3.697	**	2.529		6.125		2.924	
	(1.624)		(1.975)		(4.545)		(3.307)	
t-6	1.331		-0.749		6.079		2.451	
	(1.613)		(2.071)		(4.003)		(2.989)	
t-5	0.198		-1.653		4.474		1.798	
	(1.410)		(1.745)		(3.857)		(2.574)	
t-4	0.619		-0.811		5.772		0.904	
	(1.458)		(1.794)		(3.666)		(2.977)	
t-3	0.476		-0.610		4.271		0.680	
	(1.282)		(1.626)		(3.396)		(2.295)	
t-2	0.781		-1.471		9.987	**	-0.599	
	(1.235)		(1.544)		(3.437)		(2.018)	
Event	-2.979	**	-3.830	**	1.822		-5.175	**
	(1.264)		(1.637)		(3.074)		(2.390)	
t+1	-1.387		-3.475	*	5.122		-2.181	
	(1.445)		(1.905)		(3.574)		(2.441)	
Intercept	12.19	***	13.26	***	8.807*	***	12.70	***
	(0.881)		(1.110)		(2.358)		(1.501)	
Obs	5,006		2,935		944		1,267	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 14a: The impact of COVID-19 on the ROI of firms operating in the water supply; sewage networks, waste treatment and remediation activities sector

Time relative to COVID-19 event	ROI		ROI North		ROI Centre		ROI South	
t-7	-0.821	*	-1.439	**	0.218	**	-0.175	
	(0.450)		(0.621)		(0.964)		(0.809)	
t-6	-0.946	**	-1.557	**	0.787		-0.865	
	(0.445)		(0.619)		(1.064)		(0.744)	
t-5	-1.270	**	-1.424	*	0.0175		-1.575	*
	(0.439)		(0.556)		(1.051)		(0.842)	
t-4	-0.629		-1.003		0.646		-0.850	
	(0.444)		(0.586)		(1.010)		(0.823)	
t-3	0.712		0.177		2.490		0.628	
	(0.415)		(0.543)		(1.067)		(0.735)	
t-2	0.399		-0.0256		2.176	*	0.299	
	(0.391)		(0.537)		(0.857)		(0.694)	
Event	-0.632	*	-1.074	**	-0.613		-0.0762	
	(0.371)		(0.496)		(0.902)		(0.645)	
t+1	1.982	***	1.595	***	3.355	***	1.779	***
	(0.403)		(0.573)		(0.952)		(0.648)	
Intercept	8.184	***	8.757	***	6.741***	***	7.882***	***
	(0.269)		(0.377)		(0.649)		(0.439)	
Obs	7,400		3,731		1,467		2,465	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROI. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Table 14b: The impact of COVID-19 on the ROE of firms operating in the water supply; sewage networks, waste treatment and remediation activities sector

Time relative to COVID-19 event	ROE		ROE North		ROE Centre		ROE South	
t-7	-1.961	*	-3.674	**	0.562		-0.291	
	(1.001)		(1.449)		(2.214)		(1.656)	
t-6	-1.129		-2.463		-0.420		0.864	
	(1.023)		(1.479)		(2.342)		(1.638)	
t-5	-1.599		-3.410	***	2.010		-0.728	
	(0.974)		(1.309)		(2.280)		(1.713)	
t-4	-2.826	**	-3.601	***	-0.892		-2.636	
	(0.951)		(1.314)		(2.040)		(1.722)	
t-3	0.619		-0.354		1.448		1.345	
	(0.926)		(1.366)		(2.076)		(1.566)	
t-2	1.453	*	0.581		2.101		2.248	
	(0.830)		(1.178)		(1.667)		(1.470)	
Event	-1.187		-0.546		-3.450	*	-1.312	
	(0.848)		(1.120)		(2.066)		(1.536)	
t+1	4.238	***	5.876	***	4.134	**	2.269	
	(0.850)		(1.240)		(1.863)		(1.427)	
Intercept	11.59	***	10.95	***	10.06	***	12.89	***
	(0.566)		(0.836)		(1.224)		(0.932)	
Obs	12,241		5,759		2,546		4,343	

Note. The table reports the panel event study estimates using a dynamic fixed-effect model specification for ROE. In parentheses, standard errors clustered at the firm level. The baseline year is 2019. The p-value of the F-test is less than the significance level, the model is significant. Significance: *** p<.01, ** p<.05, *p<.10

Overall, all companies have been severely affected by the pandemic outbreak, which is reflected mainly in declining stock prices, revenues, and profits. At the same time, however, government interventions to support the economic and financial balance of enterprises and the maintenance of employment levels, and the relaxation of lockdown measures mitigated the negative effects of the pandemic outbreak. The public subsidies and tax relief measures somewhat mitigated firm losses and had significant effect, but relatively mild compared to the size of the economic shock. Therefore, results also highlight the importance of public support measures in helping firms cope with the pandemic. In the absence of such programs, pronounced sales losses would have threatened the liquidity and the survivability of firms and hence also the job security of employees (Janzen and Radulescu, 2022).

6. CONCLUSIONS

The COVID-19 pandemic had a considerable influence on business activities in almost all economic activities on the national level. However, the disruptions of the pandemic did not affect all enterprises equally. Therefore, it is necessary to explore how severely the industry sectors have been affected by the health crisis. To this end, we focus on investigating and analyzing the impacts of the pandemic at the sectoral and territorial levels on the quality of profitability, as measured by key financial metrics, such as ROI and ROE, of Italian firms. We use a robust empirical methodology in our analysis, the panel event study, which allows us to discern the influence of COVID-19 disruptions on the financial performance of Italian companies, by considering previous trends and potential variables that could distort this correlation.

We estimate that in 2020, the profitability indicators of all the considered sectors declined, but those that suffered most from the pandemic were the accommodation and food services, transportation, manufacturing, and cultural sectors, with a reduction in profitability ranging from 4 to 11 pp. In addition, the results highlight that the COVID-19 crisis was undoubtedly a regional crisis, with spatially uneven impacts, and with heavy negative effects especially for firms in the above sectors operating in the Northern and particularly in central Italy. When moving to 2021, we see a moderate recovery for Italian industries; the economic blow has been cushioned by the various government interventions to support the economy, with particular attention to the hardest hit sectors by the pandemic.

Overall, the results can be largely explained by the strictness of the anti-contagious policies, which caused disruptions in supply chains, prevented some purchases, and highlighted the inability of many industries in several sectors to transition to remote work, and online sales. In fact, these measures had a significant impact on several sectors, such as accommodation and food services, rental, travel agencies, business support services, and the arts and culture sector, which were forced to close, and on those businesses that could not benefit from remote work.

REFERENCES

Abadie, A., Athey, S., Imbens, G. W. and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1), 1-35.

Alsamhi, M. H., Al-Ofairi, F. A., Farhan, N. H., Al-Ahdal, W. M. and Siddiqui, A. (2022). Impact of COVID-19 on firms' performance: Empirical evidence from India. *Cogent Business & Management*, 9(1), 2044593.

Baldwin, R. and Di Mauro, B. W. (2020). Economics in the time of COVID-19: A new eBook. *Vox CEPR Policy Portal*, 2(3).

Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M. and Viratyosin, T. (2020). The unprecedented stock market reaction to COVID-19. *The Review of Asset Pricing Studies*, 10(4), 742-758.

Bartik, A. W., Bertrand, M., Cullen, Z. B., Glaeser, E. L., Luca, M. and Stanton, C. T. (2020). How are small businesses adjusting to COVID-19? Early evidence from a survey (No. w26989). *National Bureau of Economic Research*.

Chen, M. H., Jang, S. S. and Kim, W. G. (2007). The impact of the SARS outbreak on Taiwanese hotel stock performance: An event-study approach. *International Journal of Hospitality Management*, 26(1), 200-212.

Chen, M. P., Lee, C. C., Lin, Y. H. and Chen, W. Y. (2018). Did the SARS epidemic weaken the integration of Asian stock markets? Evidence from smooth time-varying cointegration analysis. *Economic research-Ekonomska istraživanja*, 31(1), 908-926.

Clarke, D. and Tapia-Schythe, K. (2021). Implementing the panel event study. *The Stata Journal*, 21(4), 853-884.

Clarke, D. and Tapia Schythe, K. (2023). EVENTDD: Stata module to panel event study models and generate event study plots

Cohen, M. J. (2020). Does the COVID-19 outbreak mark the onset of a sustainable consumption transition?. *Sustainability: Science, Practice and Policy*, 16(1), 1-3.

Ding, W., Levine, R., Lin, C. and Xie, W. (2021). Corporate immunity to the COVID-19 pandemic. *Journal of Financial Economics*, 141(2), 802-830.

Damodaran, A. (2007). Return on capital (ROC), return on invested capital (ROIC) and return on equity (ROE): Measurement and implications. Return on Invested Capital (ROIC) and Return on Equity (ROE): *Measurement and Implications* (July 2007).

Gałaś, A., Kot-Niewiadomska, A., Czerw, H., Simić, V., Tost, M., Wårell, L. and Gałaś, S. (2021). Impact of COVID-19 on the mining sector and raw materials security in selected European countries. *Resources*, 10(5), 39.

Golubeva, O. (2021). Firms' performance during the COVID-19 outbreak: International evidence from 13 countries. *Corporate Governance: The International Journal of Business in Society*, 21(6), 1011-1027.

Harjoto, M. A., Rossi, F. and Paglia, J. K. (2021). COVID-19: Stock market reactions to the shock and the stimulus. *Applied Economics Letters*, 28(10), 795-801.

Heyden, K. J. and Heyden, T. (2021). Market reactions to the arrival and containment of COVID-19: An event study. *Finance research letters*, 38, 101745.

Ichev, R. and Marinč, M. (2018). Stock prices and geographic proximity of information: Evidence from the Ebola outbreak. *International Review of Financial Analysis*, 56, 153-166.

Ivanov, D. and Dolgui, A. (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International journal of production research*, 58(10), 2904-2915.

Janzen, B. and Radulescu, D. (2022). Effects of COVID-19 related government response stringency and support policies: Evidence from European firms. *Economic Analysis and Policy*, 76, 129-145.

Jowitt, S. M. (2020). COVID-19 and the global mining industry. *SEG Newsletter*, (122), 33-41.

Labadze, L. and Sraieb, M. M. (2023). Impact of anti-pandemic policy stringency on firms' profitability during COVID-19. *Sustainability*, 15(3), 1940.

Levy, D. L. (2021). COVID□19 and global governance. *Journal of Management Studies*, 58(2), 562.

Liu, H., Manzoor, A., Wang, C., Zhang, L. and Manzoor, Z. (2020). The COVID-19 outbreak and affected countries stock markets response. *International Journal of Environmental Research and Public Health*, 17(8), 2800.

Lu, L., Peng, J., Wu, J. and Lu, Y. (2021). Perceived impact of the COVID-19 crisis on SMEs in different industry sectors: Evidence from Sichuan, China. *International Journal of Disaster Risk Reduction*, 55, 102085.

Nayak, J., Mishra, M., Naik, B., Swapnarekha, H., Cengiz, K. and Shanmuganathan, V. (2022). An impact study of COVID□19 on six different industries: Automobile, energy and power, agriculture, education, travel and tourism and consumer electronics. *Expert systems*, 39(3), e12677.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., ... and Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International journal of surgery*, 78, 185-193.

Pendell, D. L. and Cho, C. (2013). Stock market reactions to contagious animal disease outbreaks: An event study in Korean foot□and□mouth disease outbreaks. *Agribusiness*, 29(4), 455-468.

Radermecker, A. S. V. (2020). Art and culture in the COVID-19 era for a consumer-oriented approach. *SN Business & Economics*, 1(1), 4.

Renukappa, S., Kamunda, A. and Suresh, S. (2021). Impact of COVID-19 on water sector projects and practices. *Utilities Policy*, 70, 101194.

Van Dijk, B. (2017). Aida database.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Yan, C. (2020). COVID-19 outbreak and stock prices: Evidence from China. *Available* at SSRN 3574374.

Yang, Y., Zhang, H. and Chen, X. (2020). Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak. *Annals of Tourism Research*, 83, 102913.

Zhang, D., Hu, M. and Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance Research Letters*, 36, 101528.

APPENDIX

Table 1: Descriptive statistics for ROI

ROI	Mean								
	t-7	t-6	t-5	t-4	t-3	t-2	t-1	Event	t+1
	6.02	6.424	6.943	7.546	7.269	7.727	7.577	5.492	7.692
Sectors:									
Agriculture, forestry, and fishing	1.757	1.509	1.713	2.723	2.304	2.503	2.534	2.223	2.901
Mining of minerals from quarries and mines	2.553	3.540	3.163	3.567	3.423	3.978	4.349	3.328	4.817
Manufacturing	6.942	7.548	7.969	8.518	8.644	8.579	7.989	5.690	8.096
Supply of electricity, gas, steam and air conditioning	5.668	5.405	4.631	5.381	6.273	5.899	6.266	6.140	5.859
Water supply; sewage networks, waste Treatment and remediation activities	7.011	7.205	7.009	7.657	8.252	8.089	7.968	7.274	9.449
Construction	6.745	6.858	7.631	8.179	8.381	8.645	8.624	7.595	9.978
Wholesale and retail, repair of motor vehicles and motorcycles	6.465	7.014	7.696	8.284	8.232	8.398	8.280	6.271	8.765
Transport and Warehousing sector	6.604	7.249	8.201	8.841	8.417	7.686	7.772	6.128	7.822
Activities of accommodation and catering services	3.730	4.585	5.246	5.874	5.989	6.313	6.453	0.325	4.939
Professional, scientific, and technical activities	6.693	6.966	7.460	7.730	7.656	7.792	7.794	6.284	7.584
Rental, travel agencies, business support services	6.998	7.364	7.955	8.378	8.444	8.830	8.555	6.346	8.429
Artistic, sports, entertainment, and entertainment activities	3.936	4.299	4.715	5.346	6.014	5.784	5.795	0.750	4.021

Table 2: Descriptive statistics for ROE

ROE	Mean								
	t-7	t-6	t-5	t-4	t-3	t-2	t-1	Event	t+1
	5.937	7.921	10.972	11.297	12.141	12.286	11.586	8.711	14.186
Sectors:									
Agriculture, forestry, and fishing	1.445	0.728	2.301	3.779	3.968	3.436	3.158	3.159	4.690
Mining of minerals from quarries and mines	-0.484	2.469	3.281	-0.096	1.429	3.007	2.804	1.414	6.328
Manufacturing	6.076	8.503	11.352	12.037	13.302	12.785	11.089	6.872	12.349
Supply of electricity, gas, steam and air conditioning	12.893	8.898	6.329	8.161	7.621	8.581	9.728	8.088	11.253
Water supply; sewage networks, waste Treatment and remediation activities	7.300	8.249	8.477	7.898	10.863	12.188	11.212	11.176	16.262
Construction	7.642	8.976	13.534	13.207	14.237	14.611	14.361	12.491	20.571
Wholesale and retail, repair of motor vehicles and motorcycles	5.976	8.068	11.155	11.458	11.914	12.183	11.779	9.671	14.365
Transport and Warehousing sector	8.070	11.444	14.778	14.279	13.577	11.515	11.630	8.853	13.759
Activities of accommodation and catering services	2.278	5.973	9.699	10.316	10.332	11.037	10.582	0.739	13.108
Professional, scientific, and technical activities	7.973	9.622	12.260	12.285	13.315	14.387	13.243	11.734	15.506
Rental, travel agencies, business support services	7.415	9.691	12.804	12.977	13.154	12.773	13.438	10.755	15.344
Artistic, sports, entertainment, and entertainment activities	1.423	4.944	8.566	8.653	10.734	10.175	9.351	1.672	9.741

doi.org/10.26398/IJAS.335

A COMPARISON OF PERFORMANCE INDICATORS IN FOOTBALL ACROSS THE TOP FIVE EUROPEAN LEAGUES

Ravi Ramcharitar, Brendon Bhagwandeen, Asad Mohammed, Isaac Dialsingh
Department of Mathematics and Statistics, The University of the West Indies, St Augustine, Trinidad and Tobago

Corresponding author: Brendon Bhagwandeen
Email: brendon.bhagwandeen@gmail.com

Abstract. *The appetite for European football has been in continuous increase over the last decade with resulting debates about league supremacy. The purpose of this investigation was to compare the top five European football leagues using aggregated seasonal match data from the Premier League, Bundesliga, La Liga, Serie A and Ligue 1, over nine complete seasons from 2009-2010 to 2017-2018. Multivariate analysis of variance and profile analysis with subsequent univariate tests and post-hoc multiple comparison procedures were carried out on 15 football match performance indicators. It was observed that the Bundesliga had significantly higher averages than Ligue 1 in offensive, defensive and physical profiles. The Premier League averaged more aerial duels than Ligue 1, La Liga, and Serie A but a lower number of tackles and fouls committed. The Premier League had the highest average shots per game and was outperformed only by the Bundesliga in the other offensive metrics. Overall, the study yielded results that may find utility in further comparative football research to understand the differences in attribute profiles across leagues and serve as a basis for insights into the contrasting performances of teams from the respective leagues in European club competitions.*

Keywords: *ANOVA; Aggregated data; European football; Key performance indicators; Profiles.*

1. INTRODUCTION

Association football is the most popular sport in the world, but empirical studies centred around performance analysis in football have been limited to exploration of specific aspects of on-field performance, including influences on match outcomes, factors affecting team performance and physiological estimates of individual player characteristics (Hughes and Franks, 2005; Kubayi, et al., 2017;

Taylor, et al., 2008). However, it has been suggested that emphasis should be placed on the development and the use of key performance indicators (Carling, et al., 2008; M. D. Hughes and Bartlett, 2002), which has been the case in football research of the last decade. Performance indicators have been outlined as the selection and combination of variables that define some aspect of performance and help achieve sporting success (Hughes and Bartlett, 2002; Wright, Carling, et al., 2014). These indicators constitute a profile of ideal performance, and identification of physical and technical parameters that could influence team performances to predict the future behaviour (Fernandez-Navarro, et al., 2016; Jamil, et al., 2021; Jones, et al., 2004; O'Donoghue, 2005; Zhou, et al., 2018).

Although there have been attempts to construct individual performance profiles in team sports such as basketball, baseball, rugby, volleyball, and American football (Boulier and Stekler, 2003; Campos, et al., 2014; Csataljay, et al., 2009; Drikos and Vagenas, 2011; García, et al., 2013; Ibáñez, , 2009; Jones et al., 2004; Ortega,et al., 2009), there has been little research on the construction of these indicators and profiles and their applications in association football (Cefis and Carpita, 2024a, 2024b). Earlier studies have attempted to provide indicators of performance through wins and losses of teams (Hughes and Churchill, 2005; Hughes and Franks, 2005; Jones et al., 2004; Lago-Ballesteros and Lago-Peñas, 2010). However, these provided contradictory findings. Authors previously compared the performance of European and African teams during the 2018 FIFA World Cup matches and found that differences existed, with European teams producing better performance metrics in terms of shots, goals scored, possession, passes and corner kicks (Kubayi and Toriola, 2020). Other authors found that successful teams had longer periods of possession than unsuccessful teams during the UEFA Euro 2016 Tournament (Casal, and al., 2017), and there were homogenous distributions of ball recovery for top European and South American nations at the 2014 FIFA World Cup (Maleki, et al., 2016).

In a comparative study of the divisions of English football, Bradley et al. (2013) discovered several differences between football at the elite level and the lower levels (Bradley et al., 2013). Different styles of play not only existed among football leagues within a country, but across leagues of various countries as well. Di Salvo et al. (2012) found that physical aspects of performance, as well as technical skills varied in players across elite football leagues (Di Salvo, et al., 2013). Other studies, through traditional data analysis, determined that the

Spanish La Liga was characterized by a higher concentration of ball possession and players with high technical skills (Espitia-Escuer and Garcia-CebriAn, 2004; Espitia-Escuer and García-Cebrián, 2006), the English Premier League was characterized by direct, fast play and solid defending (Sarmento et al., 2011), the Italian Serie A became well-known for a highly tactical defence and a well-developed use of the counterattack (Vialli and Marcotti, 2007), while the Bundesliga was known for its high tempo and speed of play (Vialli and Marcotti, 2007). However, due to the limited research conducted to determine the tactical and technical characteristics of each competition, popular speculation and perceived performance characteristics have created an open debate (Wilson, 2013).

The Union of European Football Associations (UEFA) club licensing benchmark report (UEFA, 2020) listed the top five domestic first division leagues as the Premier League (England), Bundesliga (Germany), La Liga (Spain), Serie A (Italy) and Ligue 1 (France). These rankings were based on attendance levels, stability with respect to the UEFA coefficient system, average aggregated revenue, club broadcast revenue, wage growth, and transfers (UEFA, 2020). The variety of issues associated with European football have been the source of recent research in the sport. By virtue of its popularity and sheer magnitude of resources involved in football today, together with the propagation of data mining and data analytics, significant effort into exploration of aspects of club football has taken root.

Anderson and Sally (2013) concluded that the strongest leagues in Europe, i.e., those in England, Germany, Spain, and Italy, are distinctly like each other when it comes to their key traits (Anderson and Sally, 2013). In their comparative analysis of scoring across a decade of playing, they revealed that spectators in the top European football leagues saw an average of between two and three goals per match, regardless of the four countries where the game was played. Their data also showed a similar number of shots, shots on target, corners and penalty kicks per game. They added that the number of free kicks, crosses from open play, and headed goals were also very similar. However, this convergence did not occur in the other leagues, such as the Dutch Eredivisie, Ligue 1 in France, and US Major League Soccer (Anderson and Sally, 2013). García-Aliaga et al. (2022) used machine learning to observe differences across the Premier League, Bundesliga, La Liga, and Serie A in these countries (García-Aliaga et al., 2022). They found that with the evolution of playing styles, these top European leagues appeared to be approaching a level of homogeneity in terms of technical and tactical behaviours. However, distinctions were observed in terms of fewer free kicks and long passes, more errors in ball control but greater success in dribbling in the Premier League (García-Aliaga et al., 2022).

Littlewood et al. (2011) examined trends in player acquisition at clubs belonging to the top five European football leagues between the seasons ending 2005-2009 (Littlewood, et al., 2011). They found that the numbers of home-grown players were decreasing in all leagues. However, four out of the five leagues remained indigenously dominant. The Bundesliga was the only league with most non home-grown players. The overall perception of football in some of Europe's top leagues was explored by Sarmento et al. (2013), where the differences among the English, Spanish and Italian first divisions by coaches' characterizations of the leagues' particularities were investigated (Hugo Sarmento et al., 2013). The style of play was attributed to cultural, strategic, and skill-defined factors with coaches distinguishing the styles as being physically direct, defensively tactical and aesthetically controlled, respectively. To compare quality of players in different positions in different leagues, as well as to ascertain differences in age, stature and body mass, Bloomfield et al. (2005) used non-parametric analyses of variance on data from each of England, Spain, Germany, and Italy for the 2001-2002 season (Bloomfield, et al. , 2005). There were evident differences in playing style, physical demands, as well as in physical conditioning of players from the different leagues. Further comparisons between the English Premier League and Spanish La Liga for 2006-2007 season were made using match performance variables measuring physical activity and technical abilities, including total distance covered with and without possession, heading and ground duels, passing, ball possession and ball touches (Dellal et al., 2011). Analysis was carried out using two-way analysis of variance (ANOVA) with player positions and running speeds for physical activity variables, and one-way ANOVA for technical aspects according to player position. The results indicated significant differences between variables for both leagues, concluding that cultural differences existed between them.

The relationships with elite football players and their playing positions were explored using repeated-measures ANOVA, to examine the muscle strength, anthropometric and cardiovascular profiles in a football club (Herdy et al., 2020). Sapp et al. (2017) used two-way ANOVA with leagues and seasons as independent variables to compare aggressiveness among the top five European leagues (Sapp, et al., 2018). They concluded that England boasted the most aggressive of the five leagues, while there was an overall decreasing trend in aggression over the last decade. Mitrotasios et al. (2019) compared the goal-scoring opportunities in the top four European football leagues. The results reported some differences in the four leagues: Spanish La Liga was good at the combination of offensive methods; English Premier League showed a high degree of direct play; Italian Serie A showed the shortest offensive sequences; and German Bundesliga had the greatest number of counterattacks (Mitrotasios, et al., 2019).

Though univariate ANOVA was a useful mechanism for comparisons, multivariate approaches aimed at distinguishing multifactorial characteristics were also employed in some of the earlier literature. Reilly et al. (2000) applied multivariate analysis of variance (MANOVA) to data, split into groups, and comprised of variables measuring performance on test items designed to assess somatotype, body composition, body size, speed, endurance, technical skill, anticipation, anxiety, and task and ego orientation (Reilly, et al., 2000). The study was geared towards elite football player talent-identification and distinguished between elite and sub-elite groups on which significant multivariate effects were evident. MANOVA was also used to establish whether the footballers from the Premier League exhibited heterogeneity in anthropometric variables according to playing position but revealed that there were no differences between playing positions for overall body mass, stature, fat mass, muscle mass, skeletal mass, residual mass, or lean body mass (Hencken and White, 2006). A similar study was conducted by Chmura et al. (2022) using two seasons of German Bundesliga data (Chmura et al., 2022).

In addition, multivariate analysis of variance had utility in examining data concerning groups of players by position, across time periods and even by injury and fasting status. It allowed for the evaluation of anthropometric and functional characteristics, personality and coping factors, motivation factors as well as injury rates (Carling, et al., 2012; Chamari, et al., 2012; Ivarsson and Johnson, 2010; Mladenovic and Marjano, 2011). Kannekens, et al., (2009) were also able to discern differences between Dutch and Indonesian youth team players in terms of tactical skill variables and competitive metrics facilitated by MANOVA (Kannekens, et al., 2009). Further MANOVA analyses were used to compare game parameters between Italian and Israeli football league matches which revealed that ball movement and attacking efficiency was significantly superior in the Italian Serie A (Elyakim et al., 2020).

Many of the existing studies have not comparatively incorporated match attributes encapsulating overall footballing abilities of European clubs. If differences in the football played among professional divisions were to be identified, then one should consider all metrics related to match play. To our knowledge, there have been no big data studies with a large sample size investigating differences in key match performance indicators across the top European football leagues. There is a need for coaches and analysts to benchmark these elite European leagues to identify performance variables that defined successful team performance across the continent (Winter and Pfeiffer, 2016). Key performance indicators may differ across leagues. Therefore, the aims of this study were to assess the major match attributes by way of seasonal and league performance indicators, and examine the differences, if any, among the English, German, Spanish, Italian and French first division leagues over multiple seasons.

2. METHODOLOGY

2.1 DESIGN AND SAMPLE

A comparative study was conducted to analyze the overall match performance indicators of the top five European football leagues. Our sample included data from nine football seasons across the Premier League, Bundesliga, La Liga, Serie A, and Ligue 1, spanning from 2009/10 to 2017/18. The final metrics for each season across the leagues were recorded, resulting in a sample of 17 variables, with "League" and "Year" used as independent factors. The data was sourced from Whoscored.com (www.whoscored.com), which provides football match data collected by OPTA and made publicly available. OPTA Sports, known for having one of the largest sports databases in European football, has had its reliability tested and verified in previous studies (Liu, et al., 2016).

2.2 PERFORMANCE INDICATORS

Performance indicators were selected based on variables explored in the existing literature (Castellano and Casamichana, 2015; Elyakim et al., 2020; García-Aliaga et al., 2022; Herold,et al., 2021; Kubayi and Toriola, 2020; Velasco and Castán, 2022). These variables were divided into four profiles: offensive (Herold et al., 2021; Velasco and Castán, 2022), defensive (Ruan et al., 2022; Velasco and Castán, 2022), physical (Chmura et al., 2022; Yang, et al., 2018) and control (Casal et al., 2017; Hadji and Benosmane, 2022). Table 1 provided a summary of the variables used in this study.

Table 1: Description of selected performance indicators and independent factors

	Profile	Description
League	-	League from which data were recorded: Premier League, Bundesliga, La Liga, Serie A, Ligue 1
Year	-	Year ending the season from which data were recorded. For example, 2016 corresponds to the 2015/16 season.
Offsides	Offensive	Number of times players were caught in an offside position by match officials per game
Shots		On/off target attempts made to score a goal per match
Shots on Target		On target attempts made to score a goal per match
Goals Scored		Goals scored per game
Dribbles		Number of times players successfully dribbled past an opposition player while retaining ball possession per game
Goals Conceded	Defensive	Goals conceded per game
Shots Conceded		On/off target attempts on goal conceded per game
Shots on Target Conceded		On target attempts on goal conceded per game
Interceptions		Number of times a pass is prevented from reaching a teammate per game
Tackles	Physical	Number of tackles made per game
Fouls Committed		Number of fouls committed per game
Times Fouled		Number of fouls awarded per game
Aerial Duels Won		Number of headers won per game
Possession	Control	Time spent with the ball per game
Pass Completion		Number of passes successfully completed per game

2.3 STATISTICAL ANALYSIS

Multivariate analysis of variance (MANOVA) was used to determine whether there were differences across leagues for performance indicators simultaneously. Since many of the performance indicators are correlated, MANOVA tests for the differences across these variables in a single model to detect differences that might not be apparent if each dependent variable was analyzed separately. First, assumptions of the MANOVA were checked using a quantile-quantile (Q-Q) plot of the squared Mahalanobis distances against the theoretical quantiles of the appropriate chi-square distribution to assess multivariate normality; Box's M test was used to test equal variance-covariance matrices across groups; and multicollinearity was assessed based on a correlation matrix of the performance indicators. Univariate normality was assessed using Q-Q plots, while homogeneity of variances was checked using Levene's test. Multivariate normality was satisfied but departures in the homogeneity of variance assumption were observed. However, MANOVA test statistics are robust to violations of this assumption (Johnson and Field, 1993).

Following this, MANOVA using Wilks' lambda was performed for a general comparison of the 15 performance indicators across the five leagues while controlling for the season of play as a blocking factor. As our follow-up analysis, we then used two-way analysis of variance (ANOVA) with Bonferroni correction with seasons as the blocking variable to look for differences in each of the performance indicators across the leagues. Tukey's HSD procedure was applied as part of the post-hoc analysis to account for multiple comparisons. Performance indicators were standardised, and profile analysis was also conducted to analyze patterns across the five leagues. Statistical significance was set at $p < 0.05$.

The statistical program R (version 3.5.1) was used to conduct analyses in this study. The core package 'stats' provided several inbuilt functions that generated univariate plots for checking of assumptions, as well as allowing for the implementation of the MANOVA, ANOVA and Tukey's HSD tests. For multivariate normality assessment, the package 'MVN' (Korkmaz, et al., 2014) was used to construct Q-Q plots with Mahalanobis distances. The Box's M test was implemented using the 'biotools' package (da Silva and da Silva, 2017) while Levene's tests were carried out using a combination of user-defined functions and the package 'car' (Fox et al., 2012). Finally, the implementation of the package 'profileR' (Desjardins and Bulut, 2020) enabled the testing of profile hypotheses and the construction of profile plots.

3. RESULTS

3.1 ANALYSIS OF VARIANCE

Table 2 summarised the multivariate analysis of variance results. The MANOVA suggested that there was a significant difference in the performance indicators across the top five European leagues, while controlling for the season of play (Wilks' $\Lambda = 0.095$, $p < 0.001$). The blocking factor (Season) also had a significant effect (Wilks' $\Lambda = 0.076$, $p < 0.001$), indicating that performance indicators varied by season.

Table 3 summarised the analysis of variance results. In identifying which of the performance indicators were subject to these differences, the ANOVA suggested significant differences across the five leagues for each of the attributes ($p < 0.001$) were present, except for the 'Possession' attribute ($p = 0.586$).

Table 2: MANOVA summary

	Degrees of freedom	Wilks' Λ	F-statistic	Numerator degrees of freedom	Denominator degrees of freedom	p-value
League	4	0.095	41.661	60	3023.5	< 0.001
Season	8	0.076	20.120	120	5522.8	< 0.001

Table 3: ANOVA summary

Performance indicator	Source of variation	F-statistic	p-value
Offsides	League	29.160	< 0.001
	Season	20.34	< 0.001
Shots	League	15.793	< 0.001
	Season	3.481	< 0.001
Shots on Target	League	5.500	< 0.001
	Season	1.461	0.168
Goals Scored	League	4.614	0.001
	Season	0.461	0.884
Dribbles	League	32.260	< 0.001
	Season	16.36	< 0.001
Goals Conceded	League	7.372	< 0.001
	Season	0.741	0.655
Shots Conceded	League	11.478	< 0.001
	Season	3.031	0.002
Shots on Target Conceded	League	29.448	< 0.001
	Season	9.396	< 0.001
Interceptions	League	35.200	< 0.001
	Season	76.470	< 0.001
Tackles	League	19.030	< 0.001
	Season	41.210	< 0.001
Fouls Committed	League	188.840	< 0.001
	Season	40.780	< 0.001
Times Fouled	League	197.070	< 0.001
	Season	38.870	< 0.001
Aerial Duels Won	League	41.300	< 0.001
	Season	92.130	< 0.001
Possession	League	0.709	0.586
	Season	0.070	> 0.999
Pass Completion	League	18.459	< 0.001
	Season	6.954	< 0.001

3.2 MULTIPLE COMPARISONS

For performance indicators having significant differences across leagues, post-hoc multiple comparison tests were observed to identify which pairs of leagues presented these differences. Table 4 summarised the results of the Tukey's HSD multiple comparisons of leagues. Tukey's HSD test was chosen since it is a more balanced and conservative approach compared to other post-hoc tests. Tukey's test provides a better balance between controlling Type I errors and maintaining the power to detect significant differences. Pairwise league comparisons for the 15 performance indicators were categorised into 'mild' for up to 5 significant differences, 'moderate' for 6-9 significant differences, and 'severe' for 10 or more significant differences. Table 5 summarised the pairwise comparisons of the leagues based on these categories.

Ligue 1 vs Bundesliga displayed differences for 12 of the 15 performance indicators in this study, highlighting contrasts for offensive, defensive and physical profiles. This was followed by Ligue 1 vs Premier League and La Liga vs Premier League with differences in 10 of the performance indicators. Distinctions in offensive, defensive and physical profiles were prominent between Ligue 1 vs Premier League, while these variations were present in offensive and physical profiles for La Liga vs Premier League. La Liga vs Bundesliga presented the least number of differences with 4 out of 15 performance indicators, primarily exhibiting contrasts in physical profiles. Serie A vs Bundesliga also displayed contrasts in physical profiles, while Ligue 1 vs La Liga had variations in defensive profiles. Both comparisons were significantly different for 5 performance indicators.

3.3 PROFILE ANALYSIS

Analysis revealed that profiles across the five leagues were not parallel (Wilks' $\Lambda = 0.175$, $p < 0.001$). Profile plots have been presented in Figures 1-4. It was observed that the Bundesliga had significantly higher averages than Ligue 1 in offensive, defensive and physical profiles. The Premier League averaged more aerial duels than Ligue 1, La Liga, and Serie A but a lower number of tackles and fouls committed. The Premier League led with average shots per game, with only the Bundesliga having better offensive attributes.

The profile plots and post-hoc pairwise comparisons were suggestive of overall trends among the five leagues. The Bundesliga appeared to be leading in terms of average dribbles completed per game, shots on target, goals scored, and offsides, thus displaying the highest offensive profile. In contrast, Ligue 1 averaged the least shots, shots on target, goals scored; and the second fewest offsides and dribbles completed per game.

For the defensive indicators, the Bundesliga conceded significantly more goals per game on average than the Serie A and League 1, and averaged the second highest number of shots on target conceded per game, with Ligue 1 conceding the fewest. The Premier League conceded the highest number of shots and shots on target on average but ranked with significantly lower interceptions than La Liga, the Bundesliga and Ligue 1.

The Bundesliga profile plot of ‘physical’ indicators was above other leagues with significantly higher averages than three of the remaining four leagues for all such variables. The Premier League averaged lower in every attribute apart from aerial duels won per game. Ligue 1 and La Liga shared similar characteristics in their physical profiles.

There were no significant differences among leagues regarding ball possession. Furthermore, it was the Serie A that boasted the highest average for successful passes per game, followed by the Premier League, Ligue 1, Bundesliga, and La Liga, respectively.

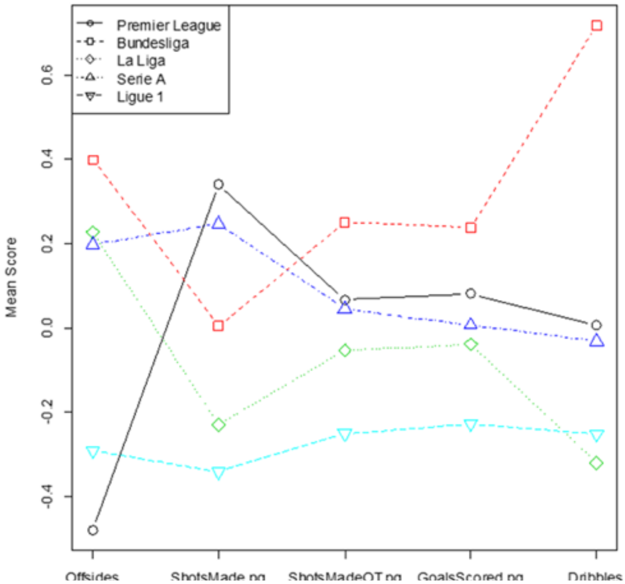

Table 4: Tukey's HSD multiple comparison summary

Profile	Performance indicators	Bundesliga - Premier League	La Liga - Premier League	Serie A - Premier League	Ligue 1 - Premier League	La Liga - Bundesliga	Serie A - Bundesliga	Ligue 1 - Bundesliga	Serie A - La Liga	Ligue 1 - La Liga	Ligue 1 - Serie A
Offensive	Offsides	*	*	*				*		*	*
	Shots	*	*		*			*	*		*
	Shots on Target				*			*			*
	Goals Scored				*			*			
	Dribbles	*	*			*		*	*		
Defensive	Goals Conceded				*		*	*		*	
	Shots Conceded				*			*	*		*
	Shots on Target Conceded	*	*	*	*			*		*	
	Interceptions	*	*		*		*		*	*	
Physical	Tackles	*	*	*			*	*		*	
	Fouls Committed	*	*	*	*	*		*	*		*
	Times Fouled	*	*	*	*	*		*	*		*
	Aerial Duels Won		*	*	*	*	*	*	*		*
Control	Possession		*	*			*		*		*
	Pass Completion										
# Significant pairwise comparisons		**8**	**10**	**7**	**10**	**4**	**5**	**12**	**8**	**5**	**8**

Table 5: Differences in pairwise comparisons of leagues

Mild	Moderate	Severe
La Liga vs Bundesliga Serie A vs Bundesliga Ligue 1 vs La Liga	Bundesliga vs Premier League Serie A vs Premier League Serie A vs La Liga Ligue 1 vs Serie A	La Liga vs Premier League Ligue 1 vs Premier League Ligue 1 vs Bundesliga

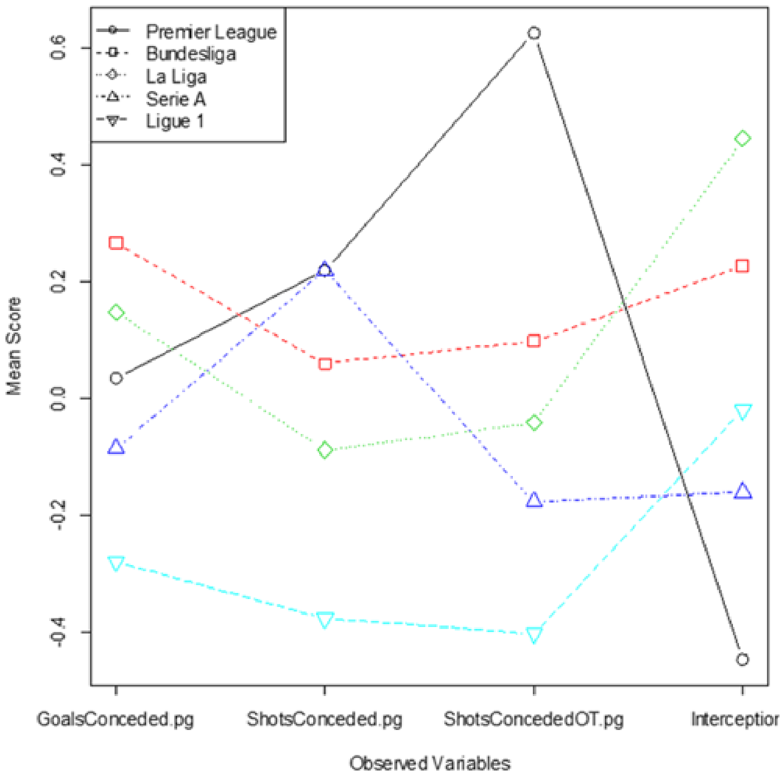

Figure 2: Profile plot of defensive performance indicators

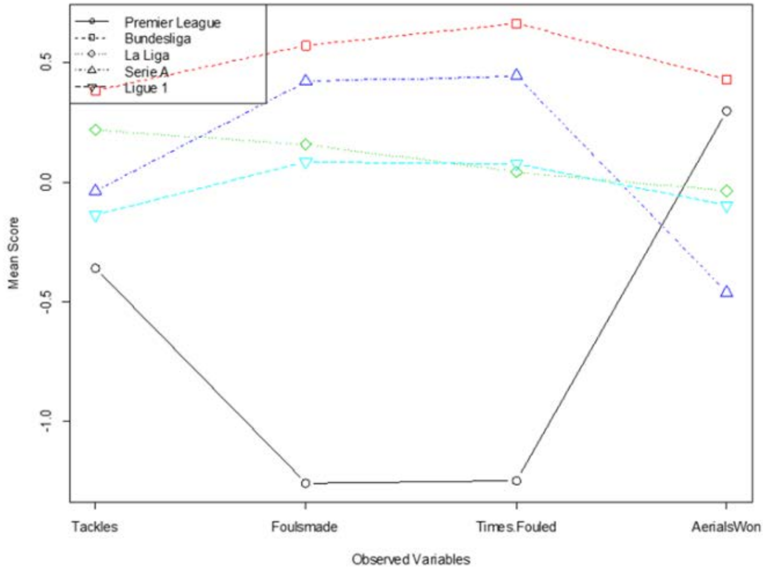

Figure 3: Profile plot of physical performance indicators

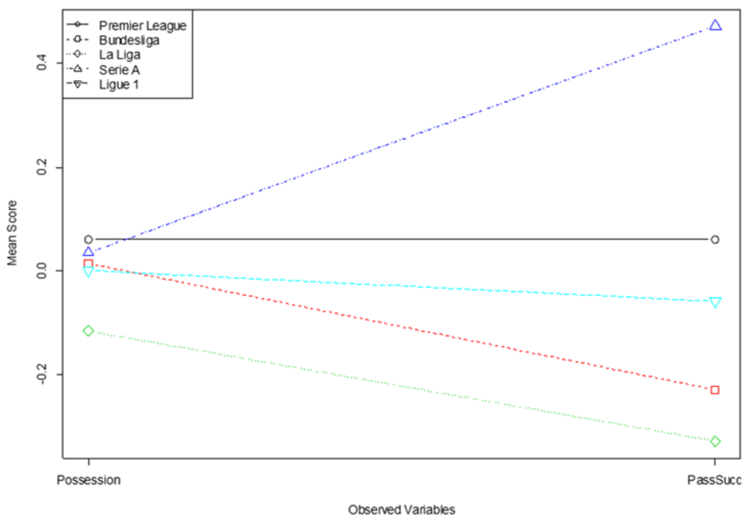

Figure 4: Profile plot of control performance indicators

4. DISCUSSION

In this research, our aim was to perform a comparative study using some of the key performance indicators across the top five European football leagues. The dataset for our analysis included 15 variables accounting for aggregated individual and team performances over nine seasons of league football which took place between 2009 and 2018. In the first step, a general comparison was performed to find attributes that differed across leagues. Much of the previous research explored similar performance indicators in isolation and only investigated leagues in a single country (Castellano and Casamichana, 2015; Chmura et al., 2022; Hadji and Benosmane, 2022; Ruan et al., 2022; Yang et al., 2018) or compared national teams and leagues from different continents (Elyakim et al., 2020; Kubayi and Toriola, 2020; Velasco and Castán, 2022). This study showed similarities but also some contrasts with prior research due to the different methods of analysis and the data sources used. Therefore, this added knowledge relevant to understanding the performance of teams from the top five European leagues.

This analysis revealed that the Bundesliga led the top five European leagues on average in the offensive profile for dribbles completed per game, shots on target, goals scored and offsides per match; and averaged higher attributes in the physical profile over three of the remaining four leagues (except for Serie A). This corroborated with the findings of Chmura et al. (2022) that the German league required high physical match demands; and Vialli, and Marcotti (2006) and Wilson (2013) who alluded to the Bundesliga having a high tempo and face-paced speed of play (Chmura et al., 2022; Vialli and Marcotti, 2007; Wilson, 2013).

Our study contrasted the findings of Vales-Vázquez et al. (2017) who observed La Liga as having the best overall competitive profile (Vales-Vázquez, et al., 2017). Our performance metrics saw La Liga's position fluctuate in each of the four attribute profiles. Other authors reported that La Liga and the Premier League performed comparable match actions, having no considerable differences across the observed performance variables (Velasco and Castán, 2022). Our findings contradicted this notion since La Liga and the Premier league displayed significant differences in 10 of the 15 performance indicators recorded. This represented contrasts in offensive and physical profiles between the two leagues;

with La Liga having higher averages in physical attributes but the Premier League displayed higher offensive metrics. Although both studies used aggregated data for each of the selected performance indicators, we collected data across nine years of league football and conducted analyses which controlled for the variations due to different seasons of play, while Velasco and Castán (2022) used only one season of match data.

Authors have suggested that the differentiation in playing profiles across leagues could be due to a combination of cultural, historical, social, technical, tactical and physical reasons (Galeano, 2003; Mitrotasios et al., 2019; Hugo Sarmento et al., 2013). As such, Spanish football has been identifiable with possession-based play, more control of the ball and less physicality (Gonzalez-Rodenas, et al., 2015). Our findings presented no significant differences in possession across the five leagues which suggested that aspects of playing styles of leagues seem to be equalizing. This has been notably highlighted in the research presented by García-Aliaga et al. (2022).

The Premier League averaged more aerial duels won per game, as well as shots per game and dribbles but reported lower tackles and fouls per game. This supported the previous characterisation of the League as being direct and fast paced but contrasted the notion of constant and hard defending (Liu et al., 2016; Sarmento et al., 2011; Wilson, 2013). Sarmento et al. (2013) associated the Serie A with an emphasis on defensive organisation. Our study provided some support to this claim as we observed Serie A as having the second lowest average for goals conceded per game and shots on target conceded per game, with Ligue 1 having the best defensive profile.

To the best of our knowledge, this is the first study to directly compare match performance indicators across the top five football leagues in Europe. The results of this study can provide important knowledge to footballing institutions in understanding how to approach games against teams from other leagues in continental competitions, as according to Vales-Vázquez et al. (2017), this was key to achieving a competitive advantage.

4.1 LIMITATIONS

Our comparison was restricted to association league football tournaments including only first division teams in the respective domestic leagues. There exists a multi-level hierarchical pyramid of footballing divisions in each of

England, Germany, Spain, Italy and France with up to five professional divisions in some cases. The availability of reliable match data for lower league tournaments was of paramount concern and hence the study was limited to the top tier of these professional football associations only.

In addition, the analyses used data for only nine league seasons while these major European leagues have been in existence for many decades. This may give a limited comparative overall view of football in these countries and, therefore, if taken on their own, findings should only be conservatively generalized to within these leagues for a contained timeframe. It should be noted that there are stark differences in resources between clubs within a first division league and this may be translated into variation among replicates within each treatment-block ('League-Season') combination.

5. CONCLUSION

The analyses conducted in this comparative study of European football leagues found significant differences in football played across England, Germany, Spain, Italy and France based on match attributes. There were significant differences observed in attribute profiles between Ligue 1 and each of the German and English top divisions, as well as between Spain's La Liga and the Premier League. It was also notable that the average time spent in possession per match did not differ across the leagues. Overall, the methods yielded results that may find utility in further comparative football research as well as in peripheral studies which may include viewership and market demands. Understanding the differences in attribute profiles and match performance indicators across leagues may also serve as a basis for insights into the contrasting performances of teams from the different leagues in European club competitions.

5.1 SCOPE FOR FUTURE RESEARCH

Teams in lower league divisions will not necessarily share characteristics of their first division counterparts, having to employ strategic and tactical game plans that are aimed to achieve different objectives for the league season. Their attribute values may tend to reflect this and influence the offensive, defensive and physical image of the leagues in a particular country. To this end, including the comparison of the lower tiers of European association football would serve to give a more holistic view of the differences among nations subject to the availability of analogous datasets.

Furthermore, performing similar analyses on spatial and temporal match aggregates should be considered. It is possible that analyses of more informative match variables may result in the attenuation of stylistic differences among the leagues. The use of designs in the multivariate decomposition of variances to account for different or even more than one blocking factor can also be pursued. The MANOVA model can be made to account for, and reduce variability across leagues due to the disparity in clubs' resources by potentially using net revenue or expenditure by season as another blocking factor. The effect of referee bias on physical-type variables investigated can be considered in subsequent research, and non-parametric multivariate techniques such as permutation multivariate analysis of variance (Anderson, 2001, 2014) can be utilized in the future to ascertain differences among leagues. Cluster analysis and principal component analysis (PCA) can also be considered for detecting variable groups, while also creating and evaluating composite performance indicators (Oliva-Lozano, et al., 2024). This study serves as a first foray into comparatively analyzing match performance indicators in European football and provides a solid methodological framework upon which to build.

REFERENCES

Anderson, C., & Sally, D. (2013). *The Numbers Game: Why Everything you Know about Soccer is Wrong*: Penguin.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology, 26*(1), 32-46.

Anderson, M. J. (2014). Permutational multivariate analysis of variance (PERMANOVA). *Wiley statsref: statistics reference online*, 1-15.

Bloomfield, J., Polman, R., Butterly, R., and O'Donoghue, P. (2005). Analysis of age, stature, body mass, BMI and quality of elite soccer players from 4 European Leagues. *The Journal of Sports Medicine and Phisical Fitness, 45*(1), 58-67.

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting, 19*(2), 257-270.

Bradley, P. S., Carling, C., Diaz, A. G., Hood, P., Barnes, C., Ade, J., . . . Mohr, M. (2013). Match performance and physical capacity of players in the top three competitive standards of English professional soccer. *Human Movement Science, 32*(4), 808-821.

Campos, F. A., Stanganélli, L. C., Pasquarelli, B. N., Campos, L. C., and Gómez, M.-Á. (2014). Performance indicators analysis at Brazilian and Italian women's volleyball leagues according to game location, game outcome, and set number. *Perceptual and Motor Skills, 118*(2), 347-361.

Carling, C., Le Gall, F., and Malina, R. M. (2012). Body size, skeletal maturity, and functional characteristics of elite academy soccer players on entry between 1992 and 2003. *Journal of Sports Sciences, 30*(15), 1683-1693.

Carling, C., Reilly, T., and Williams, A. M. (2008). *Performance Assessment for Field Sports*: Routledge.

Casal, C. A., Maneiro, R., Ardá, T., Marí, F. J., and Losada, J. L. (2017). Possession zone as a performance indicator in football. The game of the best teams. *Frontiers in Psychology, 8*, 1176.

Castellano, J., and Casamichana, D. (2015). What are the differences between first and second divisions of Spanish football teams? *International Journal of Performance Analysis in Sport, 15*(1), 135-146.

Cefis, M. (2022). Football analytics: A bibliometric study about the last decade contributions. *Electronic Journal of Applied Statistical Analysis, 15*(1), 232-248.

Cefis, M., and Carpita, M. (2024a). The higher-order PLS-SEM confirmatory approach for composite indicators of football performance quality. *Computational Statistics, 39*(1), 93-116.

Cefis, M., and Carpita, M. (2024b). A new xG model for football analytics. *Journal of the Operational Research Society*, 1-13.

Chamari, K., Haddad, M., Wong, D. P., Dellal, A., and Chaouachi, A. (2012). Injury rates in professional soccer players during Ramadan. *Journal of Sports Sciences, 30*(sup1), S93-S102.

Chmura, P., Oliva-Lozano, J. M., Muyor, J. M., Andrzejewski, M., Chmura, J., Czarniecki, S., Konefał, M. (2022). Physical performance indicators and team success in the German Soccer League. *Journal of Human Kinetics, 83*(1), 257-265.

Csataljay, G., O'Donoghue, P., Hughes, M., and Dancs, H. (2009). Performance indicators that distinguish winning and losing teams in basketball. *International Journal of Performance Analysis in Sport, 9*(1), 60-66.

da Silva, A. R., and da Silva, M. A. R. (2017). Package 'biotools'. Avaliable online at: https://CRAN. R-project. org/package=~ biotools (accessed December 27, 2020).

Dellal, A., Chamari, K., Wong, D. P., Ahmaidi, S., Keller, D., Barros, R., . . . Carling, C. (2011). Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science, 11*(1), 51-59.

Desjardins, C. D., and Bulut, O. (2020). profileR: An R package for profile analysis. *Journal of Open Source Software, 5*(47), 1941.

Di Salvo, V., Pigozzi, F., Gonzalez-Haro, C., Laughlin, M., and De Witt, J. (2013). Match performance comparison in top English soccer leagues. *International journal of sports medicine, 34*(06), 526-532.

Drikos, S., and Vagenas, G. (2011). Multivariate assessment of selected performance indicators in relation to the type and result of a typical set in men's elite volleyball. *International Journal of Performance Analysis in Sport, 11*(1), 85-95.

Elyakim, E., Morgulev, E., Lidor, R., Meckel, Y., Arnon, M. and Ben-Sira, D. (2020). Comparative analysis of game parameters between Italian league and Israeli league football matches. *International Journal of Performance Analysis in Sport, 20*(2), 165-179.

Espitia-Escuer, M. and Garcia-CebriAn, L. I. (2004). Measuring the efficiency of Spanish first-division soccer teams. *Journal of Sports Economics, 5*(4), 329-346.

Espitia-Escuer, M. and García-Cebrián, L. I. (2006). Performance in sports teams: Results and potential in the professional soccer league in Spain. *Management Decision*.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P. R. and McRobert, A. P. (2016). Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. *Journal of Sports Sciences, 34*(24), 2195-2204.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., . . . Graves, S. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing, 16*.

Galeano, E. (2003). Miracles and anthems The alchemy of soccer. *Harpers, 306*, 63-68.

García-Aliaga, A., Marquina Nieto, M., Coterón, J., Rodríguez-González, A., Gil Ares, J. and Refoyo Román, I. (2022). A longitudinals study on the evolution of the four main football leagues using artificial intelligence: Analysis of the differences in English Premier League teams. *Research Quarterly for Exercise and Sport*, 1-9.

García, J., Ibáñez, S. J., De Santos, R. M., Leite, N. and Sampaio, J. (2013). Identifying basketball performance indicators in regular season and playoff games. *Journal of Human Kinetics, 36*, 161.

Gonzalez-Rodenas, J.,Calabuig , F., Lopez-Bondia, I., Aranda, R., and James, N. (2015). Association between playing tactics and creating scoring opportunities in elite football. A case study in Spanish Football National Team. *Journal of Human Sport and Exercise, 10*(1), 65-80.

Hadji, A. and Benosmane, A. B. (2022). Ball possession and passes as key performance indicators in the Algerian league of soccer (LFP1)*., 19*(1), 106-120.

Hencken, C. and White, C. (2006). Anthropometric assessment of Premiership soccer players in relation to playing position. *European Journal of Sport Science, 6*(4), 205-211.

Herdy, C. V., Figueiredo, T., Costa, G., Galvão, P. V. M., da Souza Vale, R. G. and Simão, R. (2020). Comparison between anthropometry and multi-frequency bioimpedance for body composition evaluation in Brazilian elite U-20 soccer athletes. *Motricidade, 16*(1), 28-38.

Herold, M., Kempe, M., Bauer, P. and Meyer, T. (2021). Attacking key performance indicators in soccer: current practice and perceptions from the elite to youth academy level. *Journal of Sports Science & Medicine, 20*(1), 158.

Hughes, M., and Churchill, S. (2005). Attacking profiles of successful and unsuccessful teams in Copa America 2001. Paper presented at the Science and football V: The proceedings of the fifth world congress on science and football.

Hughes, M., and Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of SportsSciences, 23*(5), 509-514.

Hughes, M. D., and Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences, 20*(10), 739-754.

Ibáñez, S. J., García, J., Feu, S., Lorenzo, A. and Sampaio, J. (2009). Effects of consecutive basketball games on the game-related statistics that discriminate winner and losing teams. *Journal of Sports Science & Medicine, 8*(3), 458.

Ivarsson, A. and Johnson, U. (2010). Psychological factors as predictors of injuries among senior soccer players. A prospective study. *Journal of sports science & Medicine, 9*(2), 347.

Jamil, M., Liu, H., Phatak, A. and Memmert, D. (2021). An investigation identifying which key performance indicators influence the chances of promotion to the elite leagues in professional European football. *International Journal of Performance Analysis in Sport, 21*(4), 641-650.

Johnson, C. R. and Field, C. A. (1993). Using fixed-effects model multivariate analysis of variance in marine biology and ecology. *Oceanography and Marine Biology. An Annual Review, Vol. 31*, 177-221.

Jones, P., James, N. and Mellalieu, S. D. (2004). Possession as a performance indicator in soccer. *International Journal of Performance Analysis in Sport, 4*(1), 98-102.

Kannekens, R., Elferink-Gemser, M. T. and Visscher, C. (2009). Tactical skills of world-class youth soccer teams. *Journal of Sports Sciences, 27*(8), 807-812.

Korkmaz, S., Göksülük, D. and Zararsiz, G. (2014).SMVN: An R package for assessing multivariate normality. *R Journal, 6*(2).

Kubayi, A., Paul, Y., Mahlangu, P. and Toriola, A. (2017). Physical performance and anthropometric characteristics of male South African university soccer players. *Journal of Human kinetics, 60*, 153.

Kubayi, A., & Toriola, A. (2020). Differentiating African teams from European teams: Identifying the key performance indicators in the FIFA World Cup 2018. *Journal of Human Kinetics, 73*, 203.

Lago-Ballesteros, J. and Lago-Peñas, C. (2010). Performance in team sports: Identifying the keys to success in soccer. *Journal of HumanKkinetics, 25*(2010), 85-91.

Littlewood, M., Mullen, C. and Richardson, D. (2011). Football labour migration: An examination of the player recruitment strategies of the 'big five'European football leagues 2004–5 to 2008–9. *Soccer & Society, 12*(6), 788-805.

Liu, H., Gómez, M.-A., Gonçalves, B. and Sampaio, J. (2016). Technical performance and match-to-match variation in elite football teams. *Journal of Sports Sciences, 34*(6), 509-518.

Maleki, M., Dadkhah, K. and Alahvisi, F. (2016). Ball recovery consistency as a performance indicator in elite soccer. *Revista Brasileira de Cineantropometria & Desempenho Humano, 18*, 72-81.

Mitrotasios, M., Gonzalez-Rodenas, J., Armatas, V. and Aranda, R. (2019). The creation of goal scoring opportunities in professional soccer. tactical differences between Spanish la Liga, English Premier League, German Bundesliga and Italian Serie A. *International Journal of Performance Analysis in Sport, 19*(3), 452-465.

Mladenovic, M. and Marjanovic, A. (2011). Some differences in sports motivation of young football players from Russia, Serbia and Montenegro. *SportLogia, 7*(2), 145-153.

O'Donoghue, P. (2005). Normative profiles of sports performance. *International Journal of Performance Analysis in Sport, 5*(1), 104-119.

Oliva-Lozano, J. M., Cefis, M., Fortes, V., Campo, R. L.-D. and Resta, R. (2024). Summarizing physical performance in professional soccer: Development of a new composite index. *Scientific Reports, 14*(1), 14453.

Ortega, E., Villarejo, D. and Palao, J. M. (2009). Differences in game statistics between winning and losing rugby teams in the Six Nations Tournament. *Journal of Sports Science & Medicine, 8*(4), 523.

Reilly, T., Williams, A. M., Nevill, A. and Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences, 18*(9), 695-702.

Ruan, L., Ge, H., Gómez, M.-Á., Shen, Y., Gong, B. and Cui, Y. (2022). Analysis of defensive playing styles in the professional Chinese Football Super League. *Science and Medicine in Football*, 1-9.

Sapp, R. M., Spangenburg, E. E. and Hagberg, J. M. (2018). Trends in aggressive play and refereeing among the top five European soccer leagues. *Journal of Sports Sciences, 36*(12), 1346-1354.

Sarmento, H., Marques, A., Martins, J., Anguera, T., Campaniço, J. and Leitão, J. (2011). Playing tactics in the English Premier League, Spain's La Liga and Italy's Serie A. *British Journal of Sports Medicine, 45*(15), A6-A7.

Sarmento, H., Pereira, A., Matos, N., Campaniço, J., Anguera, T. M. and Leitão, J. (2013). English Premier League, Spaiń's la Liga and Italý's Serie's A–What's different? *International Journal of Performance Analysis in Sport, 13*(3), 773-789.

Taylor, J. B., Mellalieu, S. D., James, N. and Shearer, D. A. (2008). The influence of match location, quality of opposition, and match status on technical performance in professional association football. *Journal of Sports Sciences, 26*(9), 885-895.

UEFA. (2020). Country coefficients for the 2020/2021 season. Retrieved December 14th 2022, from https://es.uefa.com/memberassociations/uefarankings/country/#/yr/2021

Vales-Vázquez, Á., Casal-López, C., Gómez-Rodríguez, P., Blanco-Pita, H. and Serra-Olivares, J. (2017). Competitive profile differences between the best-ranked European football championships. *Human Movement Special Issues, 2017*(5), 97-105.

Velasco, J. M. I. and Castán, J. C. R. (2022). Offensive difference styles and technical situational variables between european and South American elite football leagues. *MHSalud: Movimiento Humano y Salud, 19*(2), 3.

Vialli, G. and Marcotti, G. (2007). *The Italian Job:JA journey to the Heart of Two Great Footballing Cultures*: Random House.

Wilson, J. (2013). *Inverting the Pyramid: the History of Soccer Tactics*: Bold Type Books.

Winter, C. and Pfeiffer, M. (2016). Tactical metrics that discriminate winning, drawing and losing teams in UEFA Euro 2012®. *Journal of Sports Sciences, 34*(6), 486-492.

Wright, C., Carling, C. and Collins, D. (2014). The wider context of performance analysis and it application in the football coaching process. *International Journal of Performance Analysis in Sport, 14*(3), 709-733.

Yang, G., Leicht, A. S., Lago, C. and Gómez, M.-Á. (2018). Key team physical and technical performance indicators indicative of team quality in the soccer Chinese super league. *Research in Sports Medicine, 26*(2), 158-167.

Zhou, C., Zhang, S., Lorenzo Calvo, A. and Cui, Y. (2018). Chinese soccer association super league, 2012–2017: Key performance indicators in balance games. *International Journal of Performance Analysis in Sport, 18*(4), 645-656.