Theoretical and Applied Economics
Volume XXXI (2024), No. 4(641), Winter, pp. 5-16

Digitalization and innovation: a regional perspective on their interaction

Monica-Mihaela MAER MATEI
National Scientific Research Institute for Labour and Social Protection Bucharest, Romania
Bucharest University of Economic Studies, Romania
monica.maer@incsmps.ro
Andreea-Monica MUNTEANU
National Scientific Research Institute for Labour and Social Protection Bucharest, Romania
Bucharest University of Economic Studies, Romania
andreea.munteanu@incsmps.ro

Abstract. *Digitalization influences innovation in both positive and negative ways. It can create new opportunities for companies to develop solutions, but it also poses challenges, such as security concerns and the digital divide. This study analyses the correlation using regional data from the Regional Innovation Scoreboard, the most comprehensive innovation measurement tool. Our findings are based on results from canonical correlation analysis and correspondence analysis. We have identified a direct correlation between the two canonical dimensions, which may be weakened by significant regional diversity. The findings show that lower digitalization levels are linked to reduced innovation performance, while the most innovative regions tend to be highly digitalized.*

Keywords: SMEs, digitalization, innovation, NUTS2.

JEL Classification: O30, R58, C38.

Introduction

The impact of digitalization on innovation is complex. On one hand, digitalization can represent a trigger for innovation by providing new opportunities for companies to create and apply new solutions. However, on the other hand, digitalization may also imply challenges and barriers to innovation, such as issues related to security, and the digital divide.

In the last few decades, companies' activities have been affected by external or internal events. According to Eurofound, the COVID-19 pandemic influences firms' normal activity. To overcome these effects, companies adapt and intensify the use of information technologies in their activities. More than 70% of SMEs are increasing the utilization of digital technologies due to the COVID-19 pandemic (OECD, 2021).

Small and medium enterprises need to be present in the market and have competitive advantages. First, digitalization is crucial. Integrating information technologies in firms will lead to cost savings, simplifying tasks, increasing productivity, and developing customer satisfaction. Second, companies should be innovative. Innovation can be quantified by the innovative products or services added to the company's portfolio or by integrating innovative processes that contribute to increasing efficiency.

In this study, we examine the correlation between these two dimensions using data at the regional level, specifically employing the Regional Innovation Scoreboard (RIS), the most comprehensive tool for measuring innovation.

The purpose of this investigation is to utilize the indicators in RIS to gather data on digitalization and innovation. This will help us comprehend the connection between these two aspects. Additionally, we aim to examine how the level of digitalization in each region relates to the innovation performance as defined by RIS.

Literature

The association between digitalization and innovation in small and medium enterprises is currently investigated at country level: Germany (Radicic and Petkovic, 2023; Hassan et al., 2024), Austria (Eller et al., 2020), Finland (Nasiri et al., 2023) or at aggregate level: (i) regional level (Rexhepi Mahmutaj and Jusufi, 2023) or European level (Skare et al., 2023; Rojas et al., 2024).

The level of innovation is greatly influenced by human resources, which can be measured through training activities (Rexhepi Mahmutaj and Jusufi, 2023) and employee skills (Eller et al., 2020; Nasiri et al., 2023), especially digital skills (Rexhepi Mahmutaj and Jusufi, 2023; Rojas et al., 2024), as well as the presence of ICT employees (Rojas et al., 2024).

Rojas et al. (2024) identify a relationship between digitalization and innovation for EU member states using DESI (Digital Economy and Society Index) and EIS (European Innovation Scoreboard) indicators. According to their performance in innovation and digitalization, EU member states are classified as advanced, moderate and less developed countries.

The importance of digital skills in generating innovation is evaluated on 998 Western Balkans SMEs by Rexhepi Mahmutaj and Jusufi (2023). Developing innovative opportunities among Western Balkans SMEs is associated with intensifying training activities among employees and the degree of digital skills of employees. Based on their results, manufacturing SMEs are more likely to introduce innovation in their activities reported to the reference category.

Similarly, Eller et al. (2020) conclude that information technology, employee skills and digital strategy have a positive impact on digitalization. Furthermore, mediation analysis is performed to identify the potential influence of information technology, employee skills and digital strategy between digitalization and financial performance. Only digitalization positively mediates the impact of information technology on financial performance. Otherwise, the employee skills, along with digital strategy, do not positively mediate the connection between digitalization and financial achievement.

In the same manner, the digital transformation of European SMEs according to their several effects (SMEs activity, SMEs financing opportunities, enhancing SMEs competitiveness, etc) are depicted by Skare et al. (2023). Digital transformation improves the business activities of European SMEs, and it is influencing the overall business activity and competitiveness, lowering input costs, widening SMEs' access to financial resources, and even improving the adaptability and flexibility to market shocks.

The impact of digitalization on introducing new or improved products, services or internal processes is evaluated differently depending on the dimension of the German companies by Radicic and Petkovic (2023). Based on their results, digital production and logistics are drivers for process innovations in micro and medium enterprises. On one hand, big data represents another stimulus for product innovation in small and medium companies. On the other hand, big data does not have any significant effect on product and process innovation in micro firms. Similarly, Hassan et al. (2024) discussed the association between digitalization and innovation among German SMEs and concluded that digital diffusion is an important driver of innovation.

Nasiri et al. (2023) evaluated the impact of digital capabilities on digital innovation. They classified these capabilities into human, collaboration, technical, and innovation capabilities. According to their study, technical skills, human skills, and innovation skills have a positive impact on digital market offerings. Additionally, human skills, collaborative skills, and technical skills influence the digitalization of the business process.

This study expands the existing literature by providing evidence of these correlations at the regional level.

Data

Regional Innovation Index is built on 21 indicators of the 32 indicators utilized for the European Innovation Scoreboard. Indicators characterizing both scoreboards are grouped in mainly four dimensions defined as Framework Conditions, Investments, Innovation activities and Impact.

The RIS comprises 239 regions located in 22 EU countries and Norway, Serbia, Switzerland and UK.

For the purpose of this paper, we have selected the following indicators:

- Individuals who have above basic overall digital skills. This is included in the category Framework conditions in RIS methodology.
- Employed ICT specialists, an indicator belonging to Investments dimension in RIS, showing the use of information technologies.
- Innovation activities are quantified by (i) SMEs introducing product innovations as a percentage of SMEs and (ii) SMEs introducing business process innovations as a percentage of SMEs. These indicators are included in Innovations activities dimension in RIS methodology.
- The outputs of the innovation process we consider are: (i) Employment in innovative enterprises and (ii) Sales of new-to-market and new-to-enterprise innovations as a percentage of total turnover, which are classified under the Impact dimension.

Table 1 presents the definitions of these indicators from the Regional Innovation Scoreboard 2023—Methodology Report, along with their formulas in Table 2.

Table 1. *Variables description*

Name	Codification	Description
Individuals who have above basic overall digital skills	DigitalSkills	Number of persons aged 16-74 who have above basic overall digital skills quantified by their activities performed on the Internet in the last 3 months classified in four categories (communication, information, problem solving and content provision)
ICT specialists (as a percentage of total employment)	ITs	The percentage of working ICT specialists in the total employment
SMEs introducing product innovations as percentage of SMEs	ProductInnnov	The percentage of SMEs who added at least one product innovation in their enterprises
SMEs introducing business process innovations as percentage of SMEs	BusinessInov	The percentage of SMEs who establish at least one business process innovation in their company who is considered state-of-the-art at the company level or at the market level
Employment in innovative SMEs	EmploymentSME	Number of individuals working in innovative SMEs reported to total employment
Sales of new-to-market and new-to-firm innovations in SMEs as percentage of turnover	SalesNew	Turnover of state-of-the-art products (newly products to market) and propagation of existing innovative products (new to companies products)

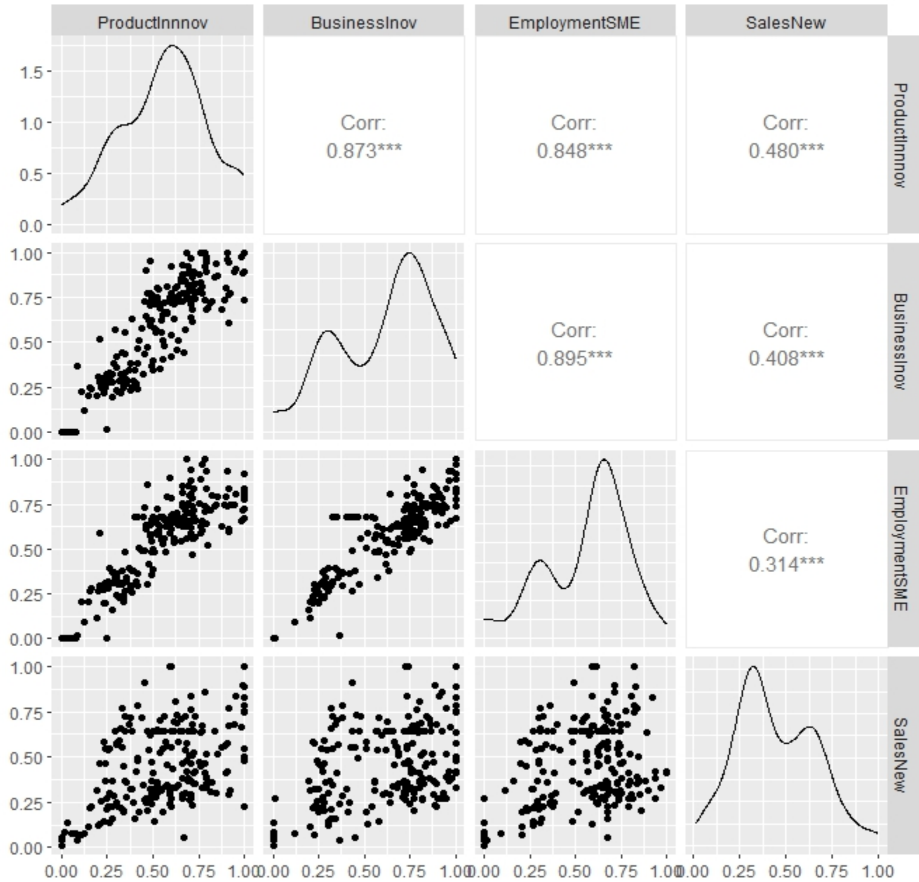
Table 2. *Variables computation formula*

Name	Formula
DigitalSkills	$\frac{Number\ of\ individuals\ with\ above\ basic\ overall\ digital\ skills}{Total\ number\ of\ individuals\ aged\ 16\ to\ 74}$
ITs	$\frac{Number\ of\ employed\ ICT\ specialists}{Total\ employment}$
ProductInnnov	$\frac{Number\ of\ SMEs\ who\ introduced\ at\ least\ one\ product\ innovation}{Total\ number\ of\ SMEs}$
BusinessInov	$\frac{Number\ of\ SMEs\ who\ introduced\ at\ least\ one\ business\ process\ innovation\ either\ new\ to\ the\ enterprise\ or\ to\ their\ market}{Total\ number\ of\ SMEs}$
EmploymentSME	$\frac{Number\ of\ employed\ persons\ in\ innovative\ SMEs}{Total\ employment}$
SalesNew	$\frac{Sum\ of\ total\ turnover\ of\ new\ or\ significantly\ improved\ products\ for\ SMEs}{Total\ turnover\ of\ SMEs}$

All the indicators in RIS are normalized using the min-max procedure, resulting in a range of [0,1] for all variables.

All the variables measuring the innovation dimension are positively correlated. Employment in Innovative SMEs is strongly positively correlated with Business Process Innovators and Product Process Innovators. Therefore, higher employment in innovative SMEs is associated with higher innovators at the regional level.

Figure 1. *Correlations for Y variables*

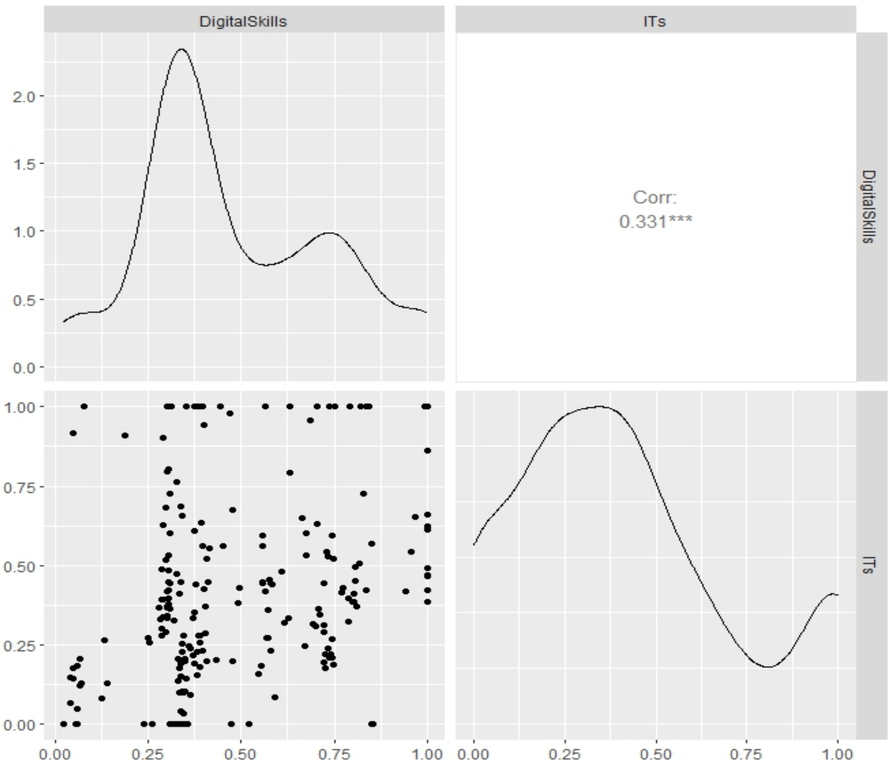


Source: Author's computation in RStudio.

The distributions of these variables are relatively smooth. The highest variability is encountered for Sales of new-to-market and new-to-enterprise innovations. Obviously, the SMEs introducing modern products in their companies and distributing existing innovations at the market level are concentrated within specific regions.

The two variables selected to depict the digitalization dimension are positively correlated. The relatively weak association between these two variables could also be explained by the high share of regions encountering the maximum value for the IT specialists variable.

Figure 2. *Correlations for X variables*



Source: Author's computation in RStudio.

Method

Canonical correlations analysis is a technique used to analyze the relationships between two sets of numerical variables, usually labelled as X and Y, observed for a sample of n units. This method shares similarities with dimensionality reduction techniques, which involve defining principal components for a single set of correlated variables, as well as regression analysis, which emphasizes the relationship between a set of variables X and a variable Y. In this case, both X and Y are matrices that have at least two columns: X_{np}, Y_{nq}.

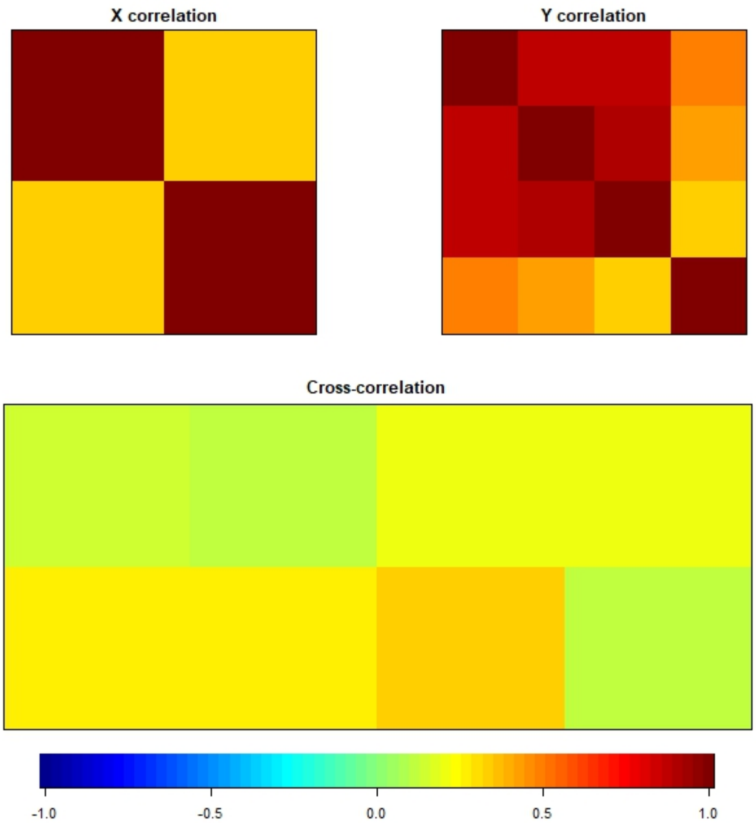
The set X contains a number of variables that is less than or equal to the number in set Y: $p \leq q$. In canonical correlation analysis, the problem is generally formulated as the estimation of two vectors of coefficients. The first vector defines a factor that extracts information from set X, while the second vector constructs a factor that synthesizes information from set Y. These coefficients are estimated by solving an optimum problem that maximizes the correlation between the two extracted factors, with the constraint that the variances of the two factors are 1. In total, p canonical pairs can be extracted; the variables in a canonical pair are uncorrelated with the other canonical variables in the same category (Hardle, Simar, 2007).

Correspondence analysis is the second multivariate analysis technique we used in this article to understand the associations between the two dimensions represented as categorical variables. This technique involves the decomposition of information in a contingency table. It represents the rows and columns of the contingency table in a new reduced dimensional space. The new coordinates are extracted based on a singular-value decomposition of the transformed contingency table. The transformation highlights the contribution of each cell in the matrix to the total variation in the data, denoted by inertia. The main findings will be derived from a two-dimensional graph representing all the contingency table's rows and columns (Hardle, Simar, 2007).

Results

In our investigation, the set X includes the variables that measure the level of digitalization, while Y is the matrix of the variables that quantify regional innovation. The representation in Figure 3 highlights the strength of the correlations within each set of indicators as well as the correlations between them. The cross-correlation heatmap shows moderate to weak positive correlations between the variables within the two sets.

Figure 3. *Correlation matrices within and between two data matrices*



Source: Author’s computation in RStudio.

After examining the heatmap in Figure 3, we expect to find weak canonical correlations as revealed by the canonical correlations analysis: 0.38 and 0.21. These coefficients indicate that the highest correlation between a factor derived from the digitalization observed variables and another factor derived from the innovation-specific variables is 0.38.

The standardized estimated coefficients used to compute these factors are presented in Table 3 for the X variables and in Table 4 for the Y variables. These values show how each observed variable influences the canonical variables. Given that we are reporting the standardized canonical coefficients, we can interpret that a one-standard-deviation increase in Digital Skills leads to a 0.66-standard-decrease in the score on the first canonical variate for set X when the other variables in the model are held constant.

Table 3. *The estimated canonical coefficients for the X variables*

Variable	First canonical pair-Xscores1	Second canonical pair-Xscore2
DigitalSkills	-0.66	0.83
ITs	-0.56	-0.90

Source: Author's computation in RStudio.

The employment in innovative SMEs variable has a stronger influence on the first canonical dimension, while the share of SMEs introducing business process innovations (BusinessInov) contributes a higher amount to the second dimension.

Table 4. *The estimated canonical coefficients for the Y variables*

Variable	First canonical pair-Yscores1	Second canonical pair-Yscores2
ProductInnnov	0.09	-0.67
BusinessInov	1.15	-1.07
EmploymentSME	-1.78	0.70
SalesNew	-0.47	0.86

Source: Author's computation in RStudio.

The correlations between canonical variables and the observed variables show the associations between each new dimension and the observed X and Y sets. The cross coefficients are rather low intensity, as anticipated by the canonical correlation coefficients.

Asymptotic tests for the statistical significance of canonical correlation coefficients presented in Table 5 show that both coefficients are statistically significant. The first line in Table 5 tests the significance for both coefficients and the second line tests the significance for only the second correlation. The first canonical variate pair captures, to a higher extent, the variance shared by the two sets of variables. The second pair adds additional information but is less important than the first one.

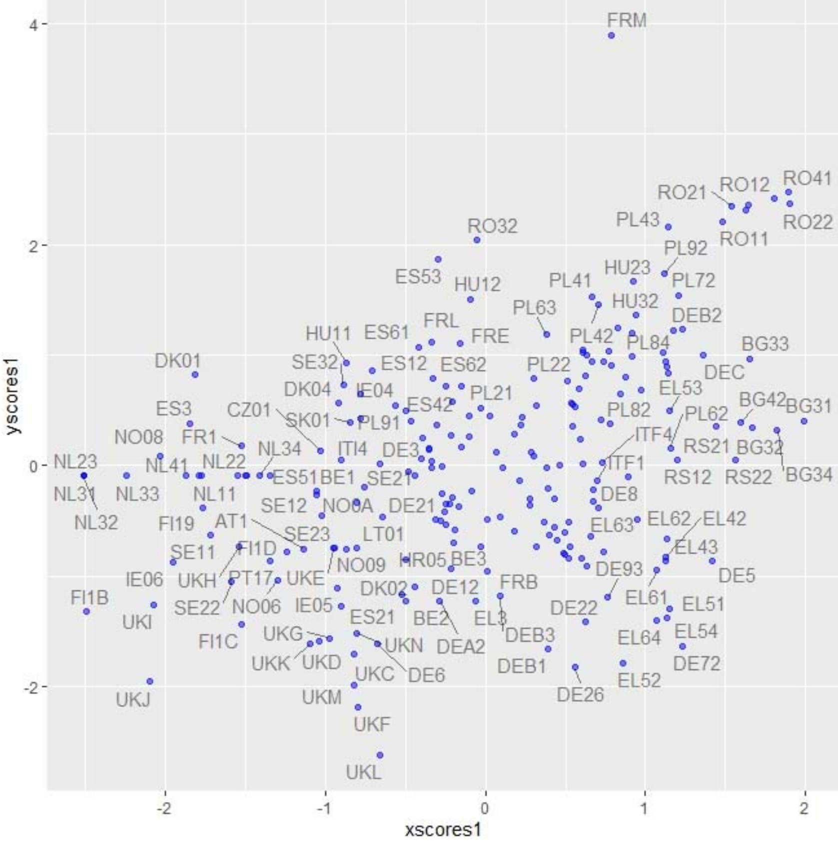
Table 5. *Statistical significance of the correlation coefficients*

	Stat	Approximation	df1	df2	p-value
1 to 2	0.8135448	5.896316	8	434	3.545704e-07
2 to 2	0.9544861	3.465048	3	218	1.711161e-02

Source: Author's computation in RStudio.

The canonical correlations analysis enables us to plot regions in two dimensions (Figure 4), helping us understand their similarities based on digitalization and innovation factors.

Figure 4. *First canonical pair*



Source: Author's computation in RStudio.

The representation of the regions in a scatterplot defined by the first canonical pair reveals the following main findings:

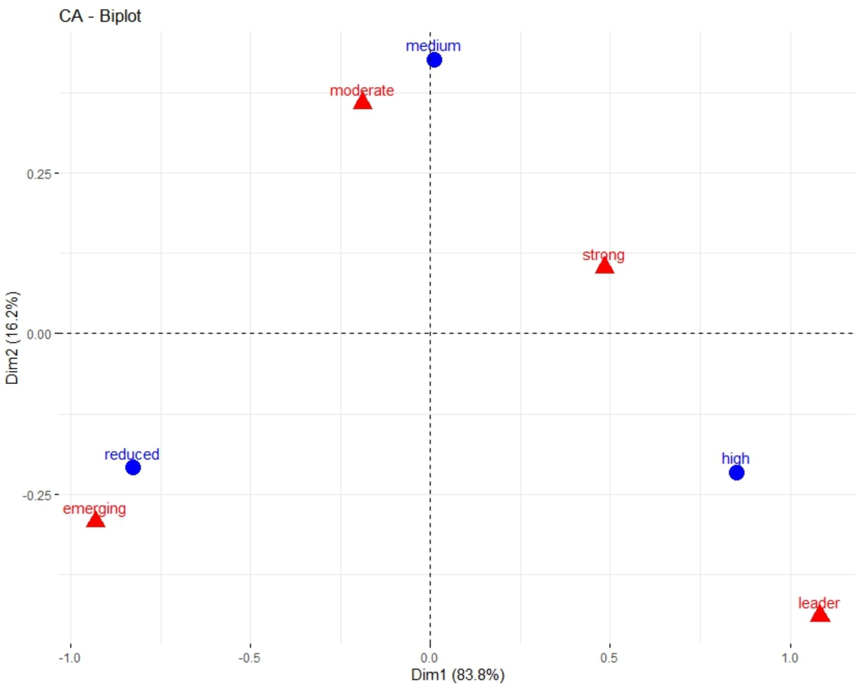
- The Corsica region of France has significantly higher scores compared to other regions due to the values computed from the Y dataset. The high score results from the high values for the share of SMEs introducing business process innovations and low levels of employment in innovative SMEs and sales of new-to-market and new-to-enterprise innovations. On the flip side, the Wales region (UKL) in the United Kingdom has a negative score due to high values of indicators with negative coefficients, such as Employment in innovative SMEs and Sales of new-to-market and new-to-enterprise innovations.
- The homogeneous group of regions outlined on the far right consists of regions from Romania. Their position is determined by the score recorded by the first canonical variable of set X. Considering the negative sign of the canonical coefficients in this set, we understand that these regions have extremely low values for digital skills and ICT specialists. Their scores are similar to those encountered by regions in Poland, Hungary or Bulgaria. This result aligns with previous findings that classify Romania and Bulgaria as less developed in digitalization and innovation (Rajos et al. 2024).

- A direct correlation between the two canonical dimensions exists. It might be weak due to high regional heterogeneity.

Correspondence analysis was used to investigate the association between innovation performance reflected by the RIS score and the level of digitization quantified by the scores estimated by canonical correlation analysis. The first categorical variable has four categories: innovation leaders, strong innovators, moderate innovators and emerging innovators. This categorization is based on the RIS methodology. Regional performance group membership is based on the relative performance compared to the EU average in 2023.

The second categorical variable was constructed based on the scores of the first pair of canonical variables for the digitalization dimension. The tertiles of the scores were used to define the three categories: high, reduced and medium.

Figure 5. *Correspondence Analysis Biplot*



Source: Author's computation in RStudio.

The results indicate a significant association between the two variables, showing that lower digitalization levels are generally linked to reduced innovation performance. Additionally, the leading regions in innovation are highly digitalized. In the same manner, Rojas et al. (2024) identify the connection between digitalization and innovation using the same indicators as a proxy for the digitalization dimension. According to their study, an increase in digitalization is associated with a huge interest in innovation. Similar results are obtained by Hassan et al. (2024) which concluded that digital diffusion is a major driver source of innovation among German SMEs.

Conclusions

In this study, we conducted an analysis at the regional level to investigate the correlation between innovation and digitalization using a quantitative approach based on multivariate analysis techniques. The results were based on data provided by the Regional Innovation Scoreboard (RIS), which is the most comprehensive tool for measuring innovation. In the first step, we applied canonical correlation analysis to reduce the dimensionality in two sets of matrices, measuring the two investigated dimensions and simultaneously highlighting the most important correlation between them. We identified a direct correlation between the two canonical dimensions, which may be affected by significant regional diversity. In the second stage, we applied correspondence analysis to investigate the association between the score obtained for the digitization component and the classification offered by RIS. This approach is beneficial in reducing the limitations of an approach that emphasizes linear correlations. In this way, we were able to highlight a significant association between the two variables, showing that lower digitalization levels are generally linked to reduced innovation performance. Additionally, the leading regions in innovation are highly digitalized. Future work should consider investigating the nonlinear relation between the discussed dimensions using an approach based on the kernel canonical correlation analysis.

Acknowledgement

NUCLEU Project PN 22100103.

References

Eller, R., Alford, P., Kallmünzer, A. and Peters, M., 2020. Antecedents, consequences, and challenges of small and medium-sized enterprise digitalization. *Journal of Business Research*, *112*, pp. 119-127.

Eurofound, n.d. COVID-19 and digitalisation. [Online] Available at: <https://www.eurofound.europa.eu/en/covid-19-and-digitalisation> [Accessed October 2024].

European Comission, 2023. *Regional Innovation Scoreboard 2023. Methodology Report.*

Härdle, W. and Simar, L., 2007. *Applied multivariate statistical analysis*. Vol. II, Springer-Verlag, Heidelberg.

Hassan, S.S., Meisner, K., Krause, K., Bzhalava, L. and Moog, P., 2024. Is digitalization a source of innovation? Exploring the role of digital diffusion in SME innovation performance. *Small Business Economics*, *62*(4), pp. 1469-1491.

Hernández de Rojas, F., Pita, P.R. and Pérez Martínez, J.E., 2024. Assessing the European association between digitalization and innovation.

Nasiri, M., Saunila, M., Ukko, J., Rantala, T. and Rantanen, H., 2023. Shaping digital innovation via digital-related capabilities. *Information Systems Frontiers*, *25*(3), pp. 1063-1080.

Radicic, D. and Petković, S., 2023. Impact of digitalization on technological innovations in small and medium-sized enterprises (SMEs). *Technological Forecasting and Social Change*, 191, p. 122474.

Rexhepi Mahmutaj, L. and Jusufi, N., 2023. The importance of digital skills in firms' innovation: the case of Western Balkans. *Journal of technology management & innovation*, *18*(3), pp.98-102.

Skare, M., de Obesso, M.D.L.M. and Ribeiro-Navarrete, S., 2023. Digital transformation and European small and medium enterprises (SMEs): A comparative study using digital economy and society index data. *International journal of information management*, *68*, p. 102594.

Annex

Table 6. *Correlations*

```
$corr.X.xscores
                   [,1]       [,2]
DigitalSkills -0.8480172  0.5299687
ITs           -0.7804622 -0.6252030

$corr.Y.xscores
                   [,1]        [,2]
ProductInnnov -0.2442191 -0.12679933
BusinessInov  -0.2118716 -0.14340094
EmploymentSME -0.3135992 -0.11760343
SalesNew      -0.1974361  0.06802089

$corr.X.yscores
                   [,1]       [,2]
DigitalSkills -0.3258660  0.1130634
ITs           -0.2999068 -0.1333807

$corr.Y.yscores
                   [,1]       [,2]
ProductInnnov -0.6355434 -0.5943539
BusinessInov  -0.5513640 -0.6721717
EmploymentSME -0.8160946 -0.5512495
SalesNew      -0.5137978  0.3188383
```

Source: Author's computation in RStudio.

Table 7. *Numerical output for correspondence analysis*

```
Principal inertias (eigenvalues):

 dim    value      %   cum%   scree plot
 1      0.471030  83.8  83.8  **********************
 2      0.090727  16.2 100.0  ****
        -------- -----
 Total: 0.561757 100.0


Rows:
    name   mass  qlt  inr    k=1 cor ctr    k=2  cor ctr
1 | high |  327 1000  450 |  852 940 505 | -216   60 168 |
2 | medm |  332 1000  108 |    9   0   0 |  427 1000 668 |
3 | rdcd |  341 1000  442 | -827 940 495 | -209   60 164 |

Columns:
    name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
1 | emrg |  260 1000  442 | -932 910 480 | -294  90 247 |
2 | ledr |  139 1000  338 | 1083 858 346 | -441 142 297 |
3 | mdrt |  296 1000   87 | -188 216  22 |  359 784 420 |
4 | strn |  305 1000  133 |  484 957 152 |  103  43  36 |
```

Source: Author's computation in RStudio.