Politecnico di Torino

Master's Degree in Engineering and Management

Academic Year 2024/2025

Graduation Session October 2025

Digitalisation and Productivity in Italian Enterprises

Patterns of ICT Adoption and Firm Performance

Supervisor:

Carlo Cambini

Candidate:

Anna Tardella

Co-Supervisor:

Sander Smit

TU Delft

Abstract

As a driver of competitiveness, digitalisation is central to Europe's economic agenda, yet adoption patterns remain heterogeneous, particularly between large firms and Small and Medium-sized Enterprises (SMEs). This thesis explores the micro-level associations between Information and Communication Technology (ICT) adoption with productivity in Italian enterprises, measured through value added per employee, thereby contributing to address a gap in the literature that is often limited to macro indicators or single-technology studies.
The analysis integrates three official Istat 2022-2023 datasets — ICT Survey, Structural Business Statistics (SBS) Frame, and the Statistical Register of Enterprise Groups — into a unified cross-sectional framework. Methodologically, Multiple Correspondence Analysis (MCA) and Exploratory Factor Analysis (EFA) have been applied. The results are used respectively for two Ordinary Least Square (OLS) regression models to examine associations with productivity.
Findings reveal complementarities in ICT adoption, such as basic cloud services combined with ICT security, and Customer Relationship Management (CRM) with Business Intelligence (BI), data analytics and sharing. Italy's digitalisation emerges as highly polarised: large, multinational, high-tech firms adopt more advanced tools and achieve higher productivity, while SMEs and traditional sectors lag behind. The first OLS results confirm that digital sophistication and decision-support technologies are positively associated with performance, though infrastructural tools alone may contribute less if adopted in isolation. The second OLS further highlights that integrated use of Enterprise Resource Planning (ERP) and BI software, websites for professionals and market reach, infrastructural cloud solutions, and customer data exploitation are positively linked to productivity. Artificial Intelligence (AI) adoption, often hindered by skill shortages and high costs, shows contrasting effects. However, self-learning AI in R&D shows a positive relationship.
Overall, this thesis contributes to understanding how digital technologies cluster and influence firm performance, offering policy and managerial implications focused on technological integration, and SME digital diffusion. Future research should adopt a panel approach to assess causality and explore whether the timing of AI-driven productivity gains varies across industries and application domains.

Acknowledgements

Thanks to my supervisor and co-supervisor, who guided me through this journey by providing suggestions and motivation. Every meeting and discussion we shared has been a fundamental pillar in building this thesis.
To my family, my mum, dad, and brother, who will forever be my first supporters in life. To my mother in particular, I owe a special acknowledgment for always being present while also enabling me to fly far from home when I needed to.
To all my friends spread across the world: during these years of moving to different places, I like to think there are several places I can truly call "casa".
A special thanks to my lifelong friends, Costanza, Giorgia, Ludovica, Gabriele and Edoardo, the guardians of my personal growth and my oldest memories.
To my university friends in Rome and Turin, and in particular to Francesca, Cecilia, and Agnese, with whom I have shared so many adventures and trips over the last five years. Without a doubt, you have made this university path less difficult and every exam more enjoyable.
An acknowledgment is also due to all the people I met in the Netherlands. To Alissa, Ines, and Sander, who taught me that it is possible to find soulmates even at this age. From you, I learned that there are bonds that transcend language and cultural barriers.
To Indy: throughout this crazy and chaotic dance between Belgium and Italy, I have discovered the beauty of being independent while having you walking alongside me on this journey.
Grazie di tutto.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Digital transformation has become a cornerstone of the European Union's policy agenda, formalized through the *Digital Decade* initiative with strategic targets set for 2030. Despite this ambition, the Eurostat Report (2025) highlights significant structural challenges hindering progress. The most critical barrier is the persistent digital divide between large enterprises and small-to-medium enterprises (SMEs), with adoption rates differing by 20–30 percentage points across key technologies. This is particularly concerning, as SMEs represent 99% of EU businesses yet only 73% have reached a basic level of digital intensity, far from the 90% target for 2030 [1]. Geographic disparities further complicate progress: Nordic countries lead digital adoption, while Southern and Eastern Europe lag significantly behind. The skills gap adds another major obstacle: only 22% of EU firms provide ICT training, falling to 21% among SMEs compared to 73% of large enterprises. Europe currently counts around 10 million ICT specialists, only halfway toward its 2030 target of 20 million. Technology adoption also varies widely: while 95% of enterprises use internet connectivity, only 13% implement Artificial Intelligence (AI) solutions. In today's volatile global environment, digitalisation is both a response to disruption and a of change, enabling firms to adapt quickly, optimize operations, and foster innovation through data-driven decision-making and new business models [2]. For companies aiming to achieve resilience, competitiveness, and international expansion, embracing digital technologies is no longer optional but essential [3].

The earliest known use of the term dates back to 1876 in the *Medical Times & Gazette* [4]. Business adoption, however, began in the 1950s and accelerated in the 1990s with the rise of the Internet, profoundly reshaping both economic and social activities. How can this multidisciplinary phenomenon be defined? The *Oxford English Dictionary (OED)* defines it as the process of converting something into a digital form, or the use of digital technologies to optimize processes, improve customer experience, and drive innovation. According to Verhoef et al. (2021), digital change unfolds in three stages: digitization, digitalisation, and digital transformation [5]. Digitization refers to the conversion of analog information into digital format, enabling computers to store, process, and transmit data. Digitalisation involves leveraging digital technologies to transform specific business operations, creating new opportunities by reshaping processes such as communication, management, and distribution. For example, firms may launch online platforms that fundamentally alter customer relationships. Digital transformation, by contrast, goes beyond isolated processes and represents a company-wide phenomenon that reshapes organizational strategy, builds new capabilities, and enables business model innovation such as Product-as-a-Service (PaaS), data-driven models, or digital platforms. For incumbents, however, this process is particularly challenging due to legacy structures and barriers to innovation.

Beyond firm-level considerations, digitalisation is also shaped by exogenous drivers. The COVID-19 pandemic accelerated technology adoption, forcing enterprises to adopt digital tools to ensure business continuity and remote working. In an increasingly globalized context, digitalisation plays a crucial role in developing flexible and resilient supply chains that can better withstand disruptions in international trade and logistics [6]. Furthermore, digital tools support the transition toward sustainable business practices, from optimising resource use to enabling energy monitoring and circular economy models [7, 8, 9]. These dynamics underline that digital transformation is not only an economic challenge but also a response to societal and environmental pressures.

In parallel, debates on digital sovereignty have gained prominence, stressing the need for Europe to reduce dependence on non-European providers in critical areas such as cloud infrastructures, data governance, and Artificial Intelligence. Initiatives such as GAIA-X and the European Data Strategy aim to foster a secure, interoperable, and sovereign digital ecosystem, while governative measures across Member States increasingly align with the EU's broader vision of technological resilience and autonomy [10, 11, 12]. Embedding the firm-level analysis of ICT adoption within this broader context allows for a better understanding of how micro-level choices connect to macro-level objectives such as productivity growth, competitiveness, sustainability, and strategic independence.

Against this backdrop, this thesis asks a central question: How do different patterns of ICT adoption among Italian enterprises relate to firm-level productivity? This objective guides the overall structure of the work. Following this introductory discussion on digitalisation, with particular attention to the Italian context, its main challenges, and its strategic relevance, a comprehensive overview of the state of the art on this topic is provided in Chapter 2. In the latter, technological, productivity-related, governative, and structural aspects are analyzed.

The data design and methodological approaches are detailed in Chapter 3. Leveraging the unique opportunity to access detailed 2022–2023 microdata collected

by Istat, several descriptive analyses were conducted. From a methodological perspective, Multiple Correspondence Analysis (MCA) is applied to synthesise two latent dimensions of digital behaviour, while Exploratory Factor Analysis (EFA) is employed for clustering variables around underlying thematic constructs, related to connectivity, data usage, sharing and analysis, cloud computing and Artificial Intelligence (AI). The two techniques are complementary: the MCA provides a broad overview of digital adoption patterns, whereas the EFA allows for a more detailed examination of specific technological clusters.
The outputs of these two techniques are then used to construct two weighted Ordinary Least Squares (OLS) regression models, with value added per employee as the dependent variable (y). Chapter 4 presents and discusses the results obtained. Finally, the main findings, limitations, and avenues for further improvement, along with policy and managerial implications, are summarised in Chapter 5.
This thesis contributes to the broader debate on digital transformation by identifying how digital dimensions cluster together and which of them co-vary most significantly with productivity, offering insights relevant to policy design and managerial practice. The implications focus particularly on skills development, technological integration, and the diffusion of digital tools among small and medium-sized enterprises (SMEs), while maintaining appropriate interpretative caution given the data configuration.

1.1 Digitalisation in Italian Enterprise

Within this European framework, Italy presents a particularly relevant case. In Italy, approximately four million SMEs (99.9%) represent the backbone of the productive system: they employ around 13 million individuals and generate more than 65% of the country's total value added, thereby acting as the primary engine of economic development [13].
Italian enterprises have been investing in a diverse range of digital technologies, including Internet of Things (IoT), 3D printing, robotics, cloud computing, e-sales platforms, social media, data analytics, virtual and augmented reality, ICT security, artificial intelligence, and ICT training. However, a digital investment spectrum can be identified. In fact, the investment patterns reveal significant disparities across industries, regions, and company sizes. These challenges risk slowing down progress toward the Digital Decade objectives and reducing potential productivity gains.
Presenting Italy's digital investment landscape is the main objective of this section, particularly for the fact that the data object of this study are related to the Italian enterprise context. All graphs and descriptive analyses presented in this work are based on macrodata derived from the CNEL tables [14], which are themselves elaborations of Istat microdata collected through the ICT-2024 survey. All histograms

were produced by the author using the RStudio development environment. Moreover, the selection of analytical topics and graphical representations was guided by the 2024 ICT Statistical Report. Crucial insights are derived on how Italian enterprises have approached digitalisation investments between 2021 and 2024, while also revealing their future investment intentions for 2025-2026 [15]. Data are related to companies with more than 9 employees across four key sectors: energy and other utilities, services (financial excluded), manufacturing, and construction, analyzing their investments across eleven critical digital technologies. Services and manufacturing are then divided into several and more specific economical activities. For example, the service sector has information related specifically to enterprises related to information technologies and telecommunication. It is important to clarify that this analysis does not focus on the amount invested (e.g., euros spent on AI tools) by each enterprise, whether already or planned in the future, but solely on whether or not the enterprise has invested (or will go to invest) in that specific technological field. To make this more clear, an enterprise answering to the survey is answering to the question:

- Have you invested on technology X during the years 2021-2024? (Yes/No)
- Will you invest on technology X in the next years (2025-2026)? (Yes/No)

Specifying this is important in order to not misinterpret the outcomes derived from the data analysis.

1.1.1 Industry-Specific Investment Patterns

Figure 1.1 provides a sector-based overview of the different percentages of enterprises technological investing. Particularly, sectors for a matter of visualization have been renamed with capital letters: construction (C), manufacturing (M), services (S), energy and other utilities (U). The construction sector emerges as the most digitally conservative industry, with a striking 40% of businesses having made no-investments in any of the examined technologies between 2021 and 2024. Moreover, it shows the lowest overall investment rate at just 42.4%, highlighting a significant opportunity for digital transformation. Contrarily, the services sector demonstrates more progressive digitalisation efforts, particularly in specialized areas. For example, information technology industry lead with an impressive 36.7% of enterprises adopting AI technologies, followed by film, video, and television production (28.3%), and telecommunications services (27.6%). This kind of industries represent the vanguard of Italy's digital transformation. Figure 1.2 shows how, in comparison to the overall percentages related to all the enterprises that answered to the survey, the percentage of IT and TLC enterprises that have invested in the considered technological fields is higher.

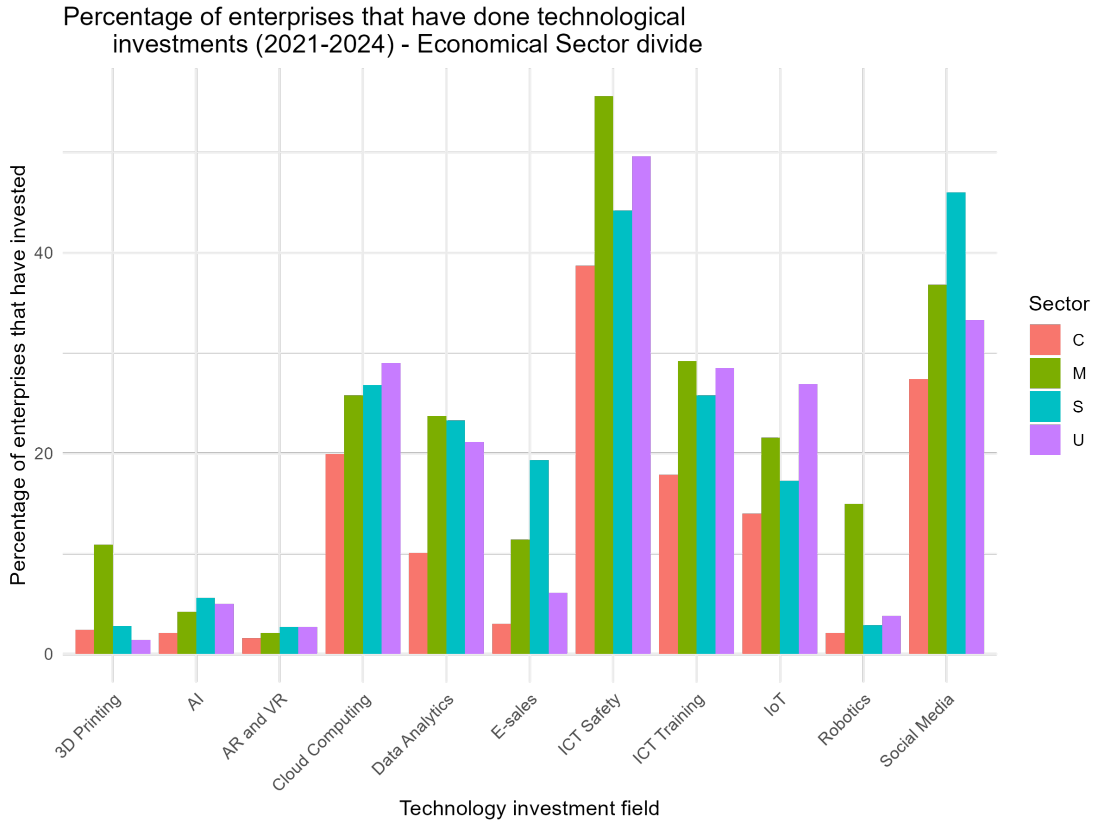

Figura 1.1: Percentage of enterprises that have done technological investments (2021-2024) - Sectoral Divide [14].

Looking at the manufacturing enterprises, varied adoption patterns are shown, with some specific economical activities embracing digital technologies while others lag behind. Traditional industries like textile manufacturing (2.1%) and food, beverage, and tobacco manufacturing (3.4%) show particularly low AI adoption rates, contrasting sharply with more technology-forward manufacturing segments, like computers or medical equipments producers (14.7%). Related to the evidences emerged in this section two hypotheses wanted to be tested.

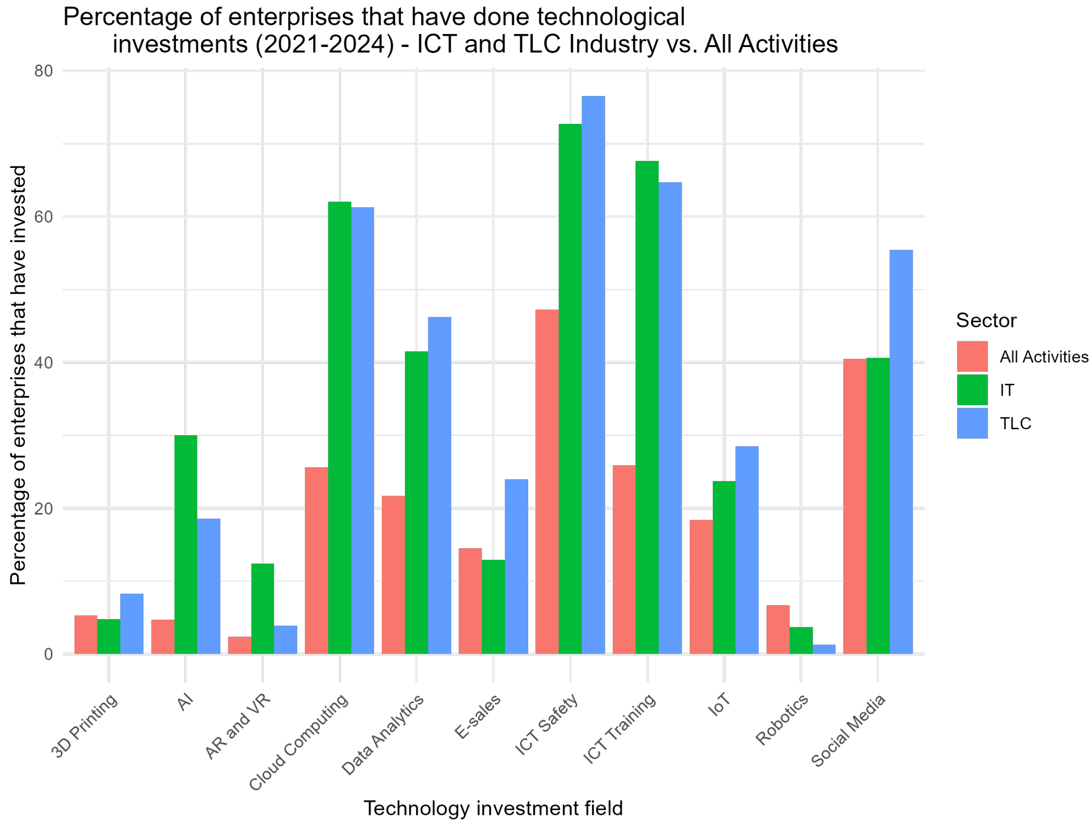

Figura 1.2: Percentage of enterprises that have done technological investments (2021-2024) - ICT and TLC industry vs. all the economical activities [14].

1.1.2 Geographic Digital Divide

Italy's digital investment patterns reveal a pronounced north-south divide that extends beyond simple economic disparities. Both North-West and North-East regions demonstrate superior investment rates across most digital technologies as shown in Figure 1.3. This geographic advantage is particularly pronounced in advanced technologies such as ICT security, ICT training, artificial intelligence, data analytics, and cloud computing, where northern enterprises significantly outpace their southern counterparts. Interestingly, the South and Islands regions show higher investment rates in e-sales platforms compared to northern regions, suggesting a strategic focus on digital commerce to overcome geographic market limitations. 40% of businesses in the South and Islands have invested in none of the examined digital technologies, approximately 10 pp higher of all the other geographical areas considered (i.e., North-West, North-East and Centrum). A visualization of this aspect can be found in Figure 1.4 This represents a massive untapped potential for economic development and competitiveness enhancement. Another aspect to mention, in order to avoid misinterpretation is that likely this geographical divide is due to the fact that in the North there are more technology and service focused enterprises than in the Sounth. Furthermore, geographical information must be managed and interpreted carefully: since an enterprise can belong to a group, its reported origin is that of the group, which may not always coincide with its actual location.

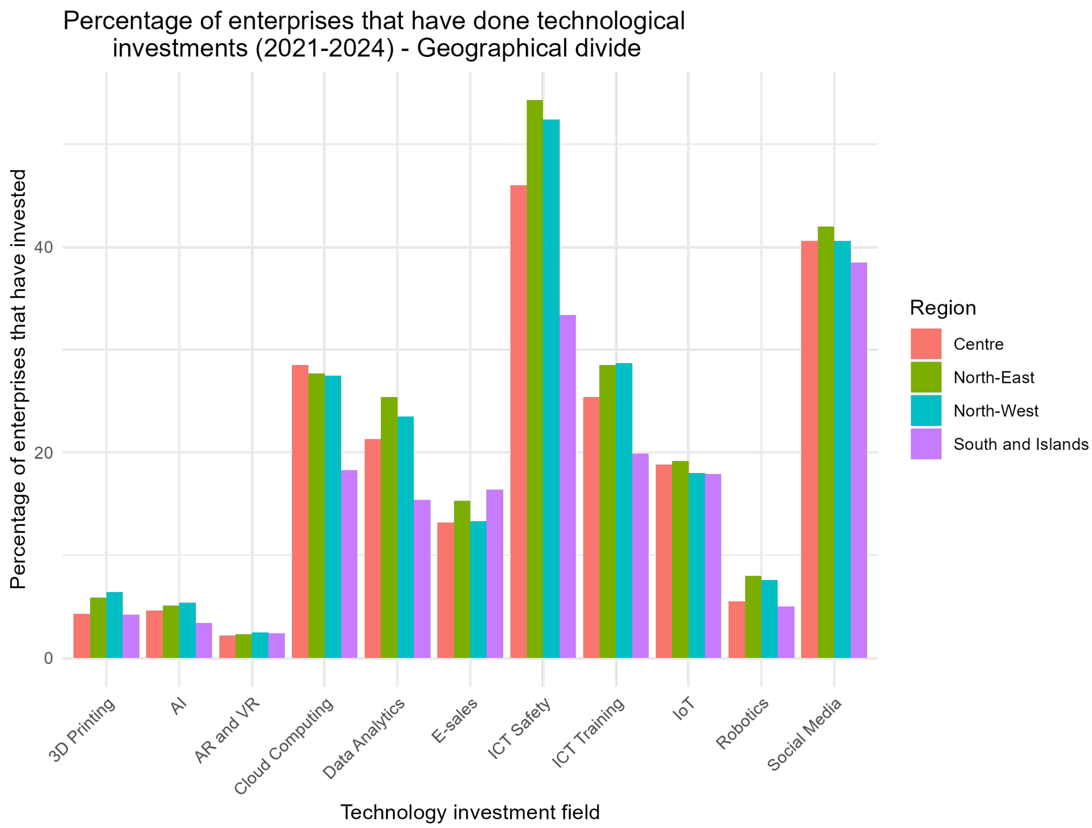

Figura 1.3: Percentage of enterprises that have done technological investments (2021-2024) - Geographical Divide [14].

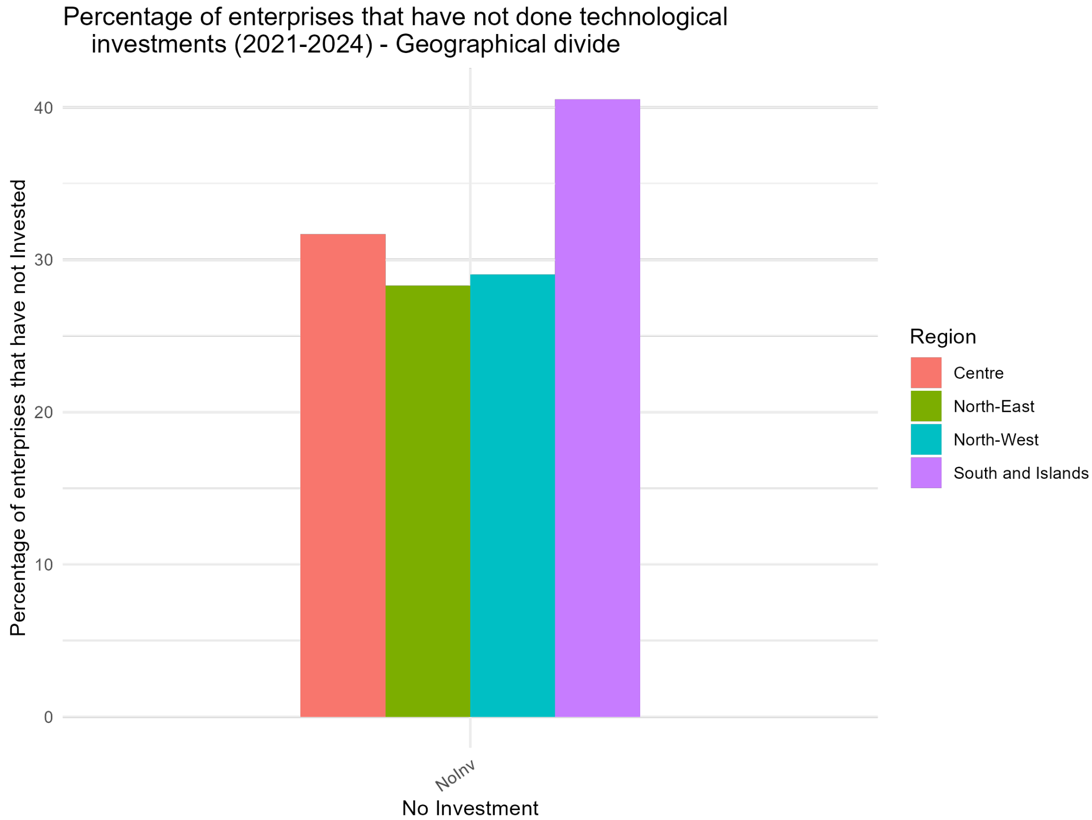

Figura 1.4: Percentage of enterprises that have not done technological investments (2021-2024) - Geographical Divide [14].

1.1.3 Size Matters: The SMEs Investment Challenge

Exactly as what it has been underlined for the European situation, the relationship between company size and digital investment reveals critical insights for Italy's predominantly SME-based economy. The main insights collected looking to the data observation are listed below:

- Companies with 10-49 employees show the highest resistance to digital investment, with 36.3% having made no investments in any examined technologies, as shown in Figure 1.5. After them it is possible to find businesses within the respective range number of employee: 50-99 (16.8%), 100-249 (11.4%). This contrasts dramatically with large enterprises (249+), where only 8% have avoided digital investments entirely.
- The data reveals distinct investment intensity patterns by company size. Particularly, considering the enterprises with 10+ employees, 52.6% have

invested in 1-4 digital areas (2021-2024), with 38% planning similar investments for 2025-2026. But if only the businesses with 250+ are analyzed, 51.9% have invested in 4-7 areas (2021-2024), with an ambitious 50.3% planning investments in 6-9 areas for 2025-2026. This underlines that in general, bigger enterprises are more oriented to digital investments. In Figure 1.6 is shown where different size enterprises have decided to invest, remarking that smaller is the dimension, smaller is the percentage of enterprises that have invested in a specific technological field.

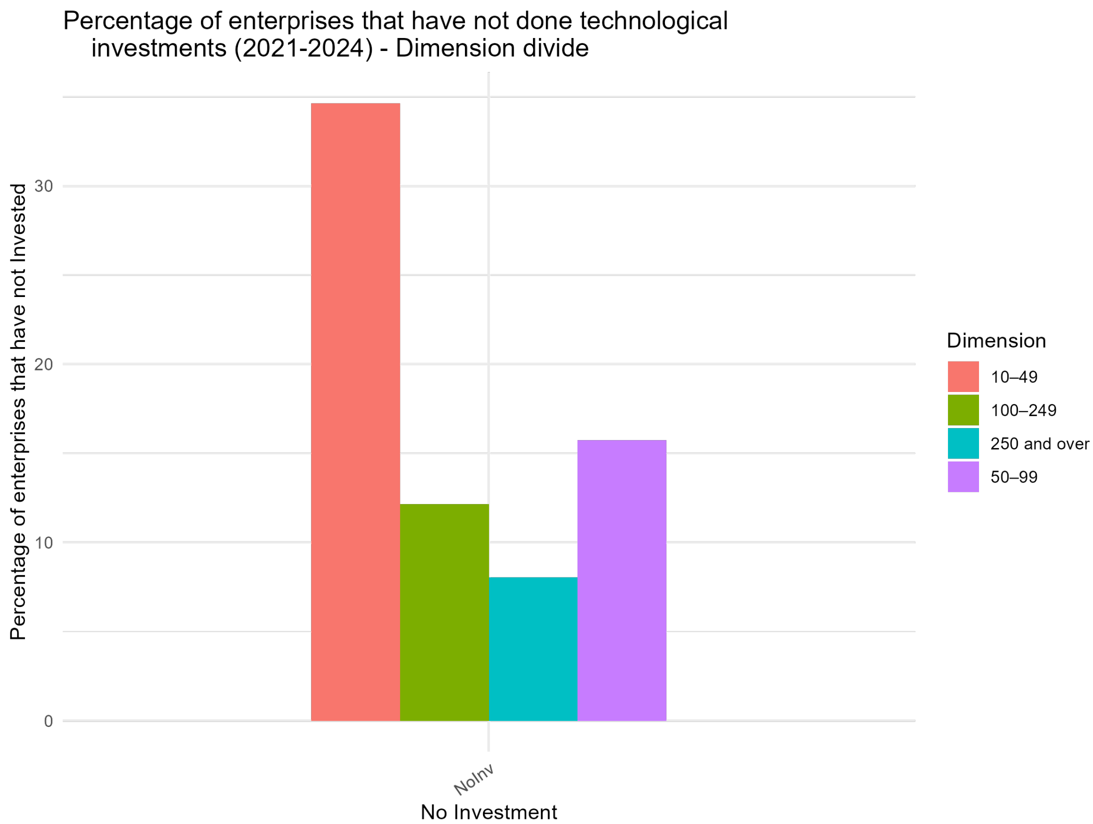

Figura 1.5: Percentage of enterprises that have not done technological investments (2021-2024) - Dimension Divide [14].

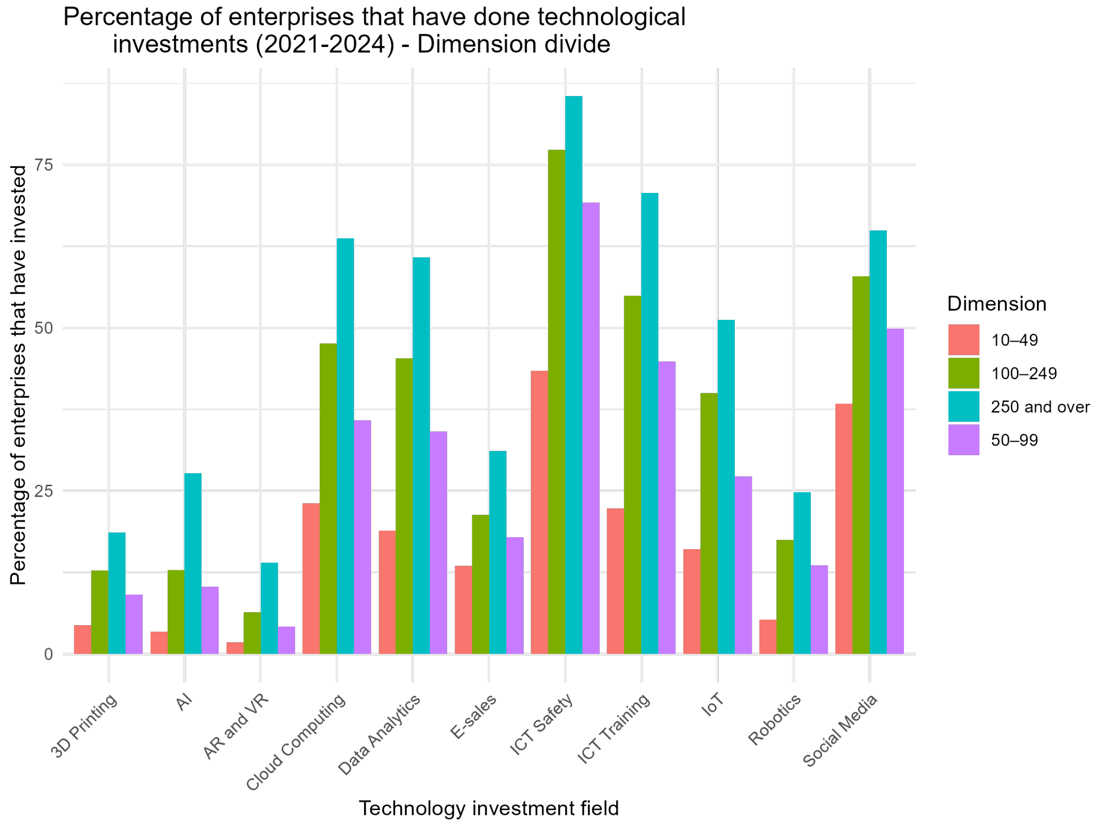

Figura 1.6: Percentage of enterprises that have done technological investments (2021-2024) - Dimension Divide [14].

1.1.4 Priority Investment Areas

Across all company sizes and regions, certain digital investment areas have emerged as priorities both for the past and, even more, for the future. Figure 1.8 shows a clear visualization of which are the intentions for the future of the Italian enterprises, making a distinction based on dimension. The key findings that are possible to observe are:

- ICT security tops the investment priority list, with 47.2% of companies having invested in the 2021-2024 period and 53.8% planning investments for 2025-2026. This reflects growing awareness of cybersecurity threats and the need for robust digital protection (Figure 1.7).

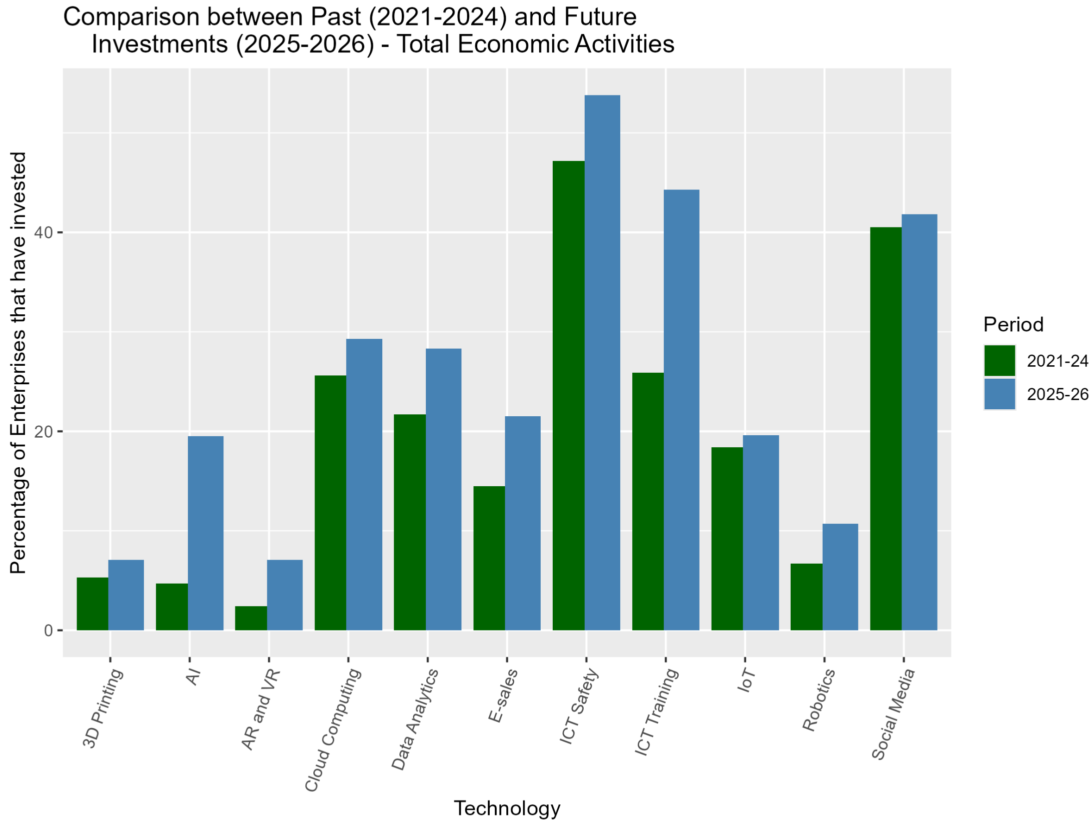

Figura 1.7: Changes in the technological future investments (2025-2026) of the Italian enterprises, considering all the economical activities [14].

- Looking to the percentages of businesses that will not invest in none of these technologies, it is noteworthy that the reduction is just of 3.5 pp (from 35.2% to 31.7%).

- Social media investments remain consistent priority (40.5% past, 41.8% future), while cloud computing demonstrates growing recognition (25.6% past, 29.3% future). ICT training shows the most dramatic growth trajectory, jumping from 25.9% to 44.3% planned investment, an 18 percentage point increase. It is noteworthy that even small companies (between 10-49 employees) declares ICT training investment plans increased from 22.3% to 40.5% (Figure 1.8). This underlines a bigger understanding, even for these small business realities, of the importance nowadays of acquiring and developing a digital know-how.

- In general, looking to Figure 1.8, it is shown that, even if in smaller percentages, the investments' priorities are shared independently from the enterprise size (ICT safety and training, cloud computing, data analytics and social media).

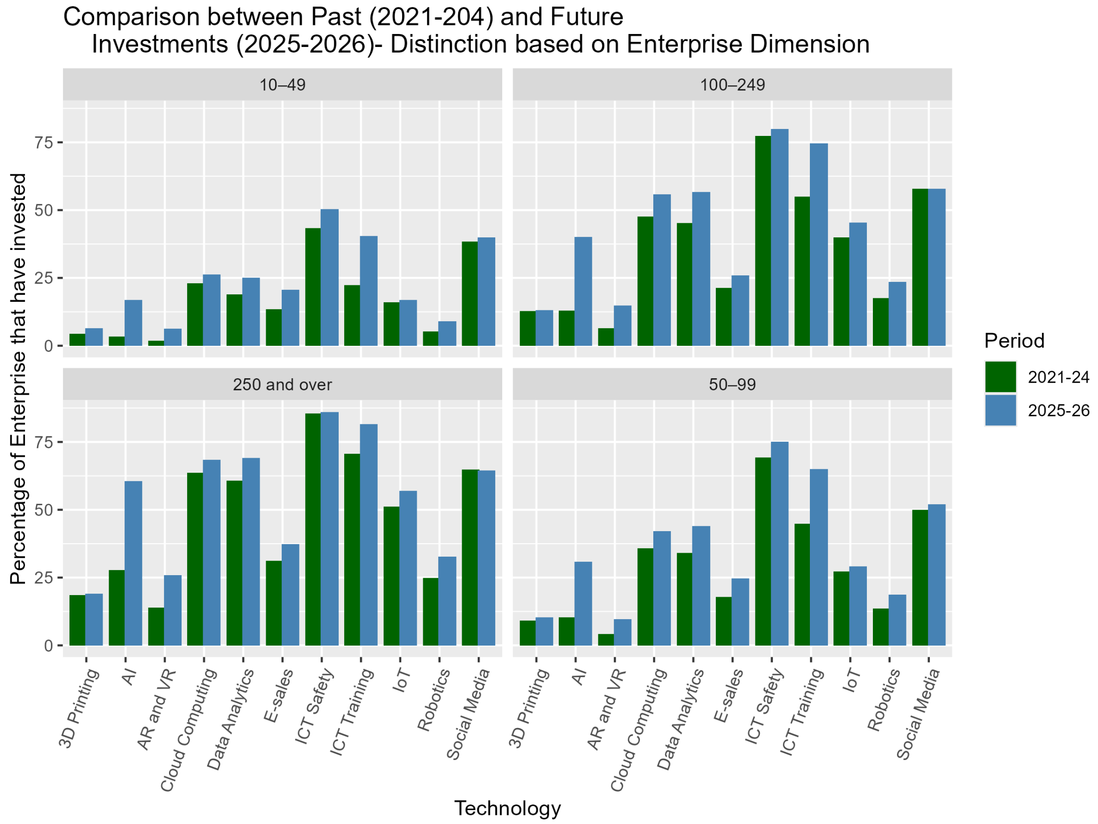

Figura 1.8: Changes in the technological future investments (2025-2026) of the Italian enterprises, with a size-based distinction [14].

1.1.5 Current AI Adoption Landscape

Italy's AI adoption story reveals both promise and challenge. In 2024, 8.2% of Italian enterprises with at least 10 employees utilized at least one AI technology, representing significant improvement from 5.0% in 2023 [16]. This growth trajectory suggests accelerating AI awareness and implementation. Nevertheless, is also important to specify that the country still lag in comparison with the EU average adoption of this type of technology (13.5%). Even if it is possible to notice a size-based AI adoption pattern, both medium and large enterprise are substantially increasing the adoption of AI tools. Particularly, medium enterprises (50-99 employees) have improved from 5.6% (2023) to 14.0% (2024) while, large enterprises (more than 99 employees) from 24.1% (2023) to 32.5% (2024). More enterprises, independently from their dimensions, declared their intention of investing more in AI tools, the percentage increased from 8.2% to 20.6%. Particularly for the businesses with more than 250 employees (that were already using more this technology) the percentage of enterprises that have declared a willingness to invest in this technology changes from 27.7% to 60.5% (Figure 1.8). Considering only the enterprises adopting AI and the seven key main usage areas inserted in the 2024 ICT survey, it is possible to notice that there are different adoption rates (Figure 1.9):

- text mining, extracting knowledge and information from a text document through (54.5%);
- natural language generation, generating written or spoken language (45.3%);
- speech recognition, converting spoken language into a format readable by a computer system (39.9%);
- analyze data through machine learning, deep learning, neural networks and data analysis (31.3%);
- automate workflows or support decision-making processes using Robotic Process Automation (RPA) or software robots that use AI technology to automate human activities (28.1%);
- image processing and object recognition, identifying objects or people based on images (25.4%);
- enable physical movement of machines through autonomous decisions based on the observation of the surrounding environment (autonomous robots or drones, self-driving vehicles) (10.4 %).

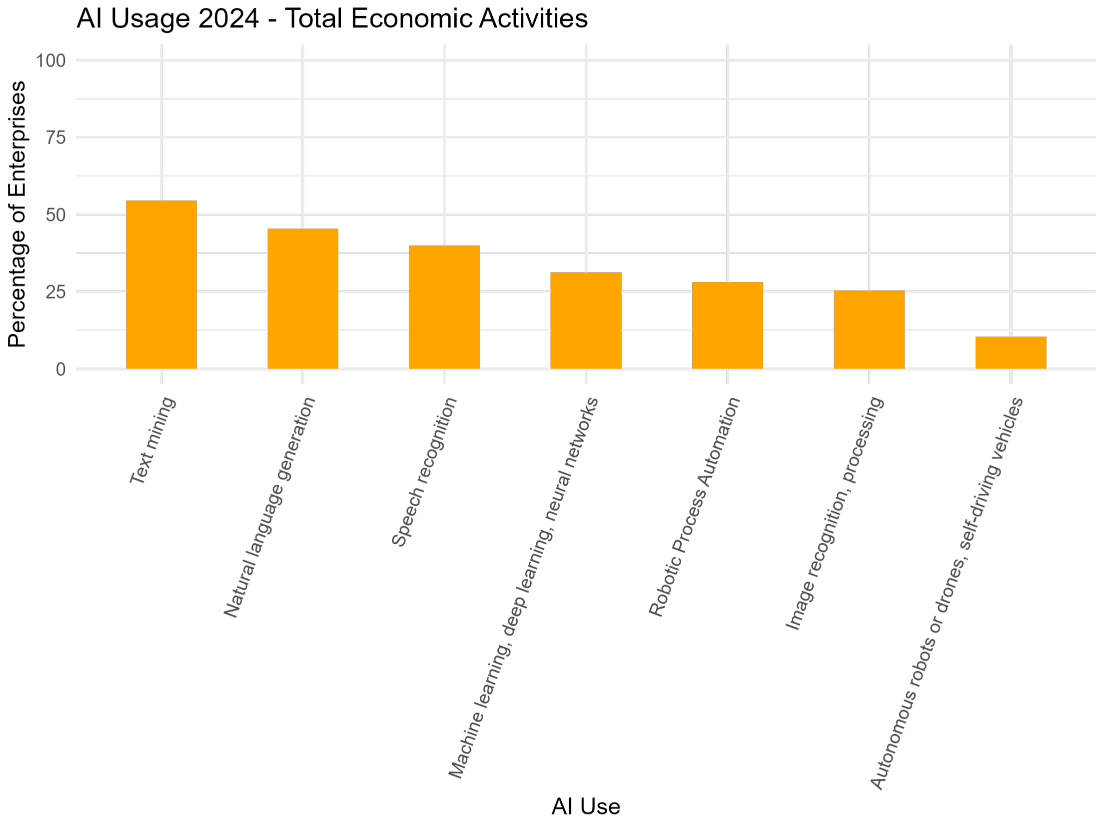

Figura 1.9: Changes in the technological future investments (2025-2026) of the Italian enterprises, considering all the economical activities [14].

Industry-specific AI leadership are visible from the data collected. Indeed, the IT industry (that is part of the service sector) leads AI adoption at 36.7% (up from 23.6% in 2023), followed by telecommunications at 27.6% and media production at 28.3%. Both of this industries, declared to significantly using AI for text mining and deep learning, machine learning and neural networks activities. Conversely, traditional sectors like restoration services (0.9%) and textile manufacturing (4.6%) show minimal AI engagement. Between the current AI users, 70.3% plan additional AI investments in 2025-2026. On the other hand, looking at the non-AI users, only 15.0% plan to begin AI investments.

Chapter 2

Literature Review

This section presents a systematic literature review of recent research on digitalisation within European enterprises. The analysis focuses on contemporary publications addressing digitalisation's relationship with productivity, supply chains, artificial intelligence adoption, sustainability initiatives, and governmental support mechanisms. The review complements the previous descriptive analysis based on macrodata from the 2024 ICT Survey conducted by Istat. This dual approach, combining theoretical insights from recent literature with empirical evidence derived from official statistical sources, provides a comprehensive, data-driven overview of the digital landscape across European, focusing on Italian enterprises. In this way, an essential foundation is established for the subsequent 2023 data analysis and interpretation carried out in this study.

2.1 Economics Benefit from Digitalisation

A wide body of evidence highlights how digitalisation can positively impact productivity at both the firm and macroeconomic level. Particularly, the Productivity Report made by Consiglio Nazionale dell'Economia e del Lavoro (CNEL), Istituto Nazionale di Statistica (Istat) and Banca d'Italia (2025) provides new descriptive and econometric evidence on Italy's productivity dynamics, emphasizing how firm heterogeneity by size, governance, innovation profile, and export orientation maps into sizeable productivity gaps [17]. A central finding is the strong positive association between digital intensity (e.g., Digital Intensity Index tiers) and labour productivity, observable both across and within size classes and sectors; similar premia emerge for the adoption of general-purpose technologies such as AI [17]. On the policy side, it highlights mutually reinforcing territorial and structural frictions (skills shortages, small scale, risk aversion), especially in the Mezzogiorno, and argues for multi-pillar interventions, such as infrastructure and human capital investment, improved public services, and better firm support policies, to unlock the returns to digital complements [17]. Overall, digital adoption (including AI) appears as a key correlate of productivity conditional on complementary assets (skills, organizational capabilities, scale), motivating coordinated policy designs rather than single-instrument remedies [17]. For instance, Nucci et al. (2023) analyse the effect of digital technology adoption (e.g., broadband connectivity, IoT, robotics, automation, 3D printing) on Italian firms' total factor productivity (TFP) using Istat data (2016–2018) [18]. Across different criteria for defining digital adopters, the authors consistently find a positive and statistically significant effect. When adoption is defined as having implemented at least one ICT technology from the Istat survey, the effect is around one percentage point (0.97). Unsurprisingly, the largest effect occurs when firms have invested in AI-related technologies, confirming the high potential of AI for productivity gains.

Evidence also shows that intangible assets strongly complement digitalisation. An Organization for Economic Co-Operation and Development (OECD) study (2021) on Dutch firms provides robust evidence that digital skills and ICT-specialised workforce enhance productivity outcomes [19]. Increasing the share of software experts and ICT professionals by one standard deviation raises labour productivity by about 10.3% and 1.3% annually. Digitalisation and intangible investments appear particularly beneficial for younger and laggard firms, as well as for service industries, suggesting that reskilling and upskilling are crucial to unlock the productivity potential of digital technologies.

From a macroeconomic perspective, the European Central Bank (2024) observes heterogeneous productivity responses across euro area countries [20]. A lack of strong institutions and governance structures has been identified as a possible reason in obtaining so different outcomes. Nevertheless, digitalisation is one of the main forces driving structural and organisational changes in the euro area and the global economy. The productivity-enhancing effects of digitalisation have generated increased interest in the promotion of digital technologies. However, euro area countries still lag behind the United States (US) in terms of digital innovation [20]. This could explain why changes brought by digitalisation are not much beneficial for the EU firms in comparison to the US ones [12]. Clearly, this aspect has an effect on the amount of gain that can be extracted by adopting more digital solutions on a firm level. However, looking to both US and EU's productivity, a shared phenomenon has been pointed out by the literature and it is the so-called "productivity puzzle". It refers to the unexplained slowdown in productivity growth experienced by many developed economies, particularly after the 2008 financial crisis. This open the door to a digital productivity paradox, since, despite technological advancements and increased investment in some areas, overall output per worker has not increased as much as anticipated or as much as it did in previous periods. Moreover, the investigation reveals that labor productivity (LP) and total factor productivity (TFP) increased five years post-digital integration when contrasted with enterprises having minimal digital technology reliance [20]. It is noteworthy that in the adoption year itself, the effect is zero for LP and negative for TFP, presumably due to organizational adjustments. To conclude, the authors underline how complementary assets like skill enhancement investments or related to other intangible assets are fundamental for having a significant positive impact on productivity growth.

2.1.1 Broadband and Ultra-Fast Broadband

Within the digitalisation debate, several studies have examined specifically the role of broadband infrastructures. Koutroumpis (2019) shows, through a logit model with GDP per capita as the dependent variable, that both fixed and mobile broadband adoption by households and firms positively affect growth [21]. A 1% increase in fixed broadband uptake raises GDP per capita by 0.026–0.034%, while a 1% increase in mobile broadband generates between 0.092% and 0.102% growth. Mobile broadband therefore produces stronger immediate effects, whereas fixed broadband yields more robust long-term impacts and larger cumulative benefits, particularly during later rollout phases. Policy recommendations from this study stress that public funding should not focus only on supply-side expansion, but also on stimulating demand through subsidies, vouchers, or tax reliefs, accompanied by investments in ICT skills.

At the firm level, Cambini et al. (2023) examine Italian data (2013–2019) and show that ultra-fast broadband (UFB) access raises both TFP and labour productivity [22]. The positive effects are most evident for service-sector firms and those located in the North-West and South of Italy. Moreover, distinguishing between full-fiber and mixed copper–fiber connections, they find that full-fiber has a significantly stronger effect on productivity. Using Labour Force Survey data, the study also suggests that part of the productivity gains from UFB may stem from structural changes in the workforce, such as the adoption of younger and more digitally skilled employees.

Taking a broader OECD member states perspective, Briglauer et al. (2025) disentangle the effects of broadband deployment from adoption [23]. They find that mere network rollout has only minor effects on GDP, whereas the adoption of broadband services,both fixed and mobile, exerts substantial and significant impacts. Mobile broadband adoption generates contemporaneous GDP growth effects nearly three times larger than fixed broadband, although fixed broadband shows stronger cumulative and long-term contributions. This confirms that adoption, rather than infrastructure deployment alone, is the true driver of productivity and growth, and that the economic benefits of digital technologies are realised primarily when they are effectively integrated and utilised by firms and households.

2.1.2 Data Analysis and Business Software

Recently, organizations have embraced the idea that data is a core asset, fostering a data-driven culture that enables more effective business processes [24]. According to the Eurostat glossary, data analytics refers to the use of technologies, techniques, or software tools to extract patterns, trends, and insights from data in order to support conclusions, predictions, and decision-making, ultimately improving performance through productivity gains or cost reductions [25]. Four main types of analytics are distinguished: descriptive, diagnostic, predictive, and prescriptive, each suited to addressing different managerial questions.

The literature provides robust evidence of the impact of data-driven practices. Braganza et al. (2022) show that analytics capabilities extend beyond efficiency improvements, acting as strategic enablers of resilience and long-term value creation. Similarly, the OECD report on SMEs emphasizes that smaller firms, despite resource limitations, can achieve substantial productivity gains through analytics adoption, provided appropriate policy support and awareness mechanisms are in place [26]. These contributions converge on a central idea: while infrastructures are necessary, it is the integration of analytics into organizational routines, culture, and capabilities that determines their true performance impact.

Business Intelligence (BI) tools represent a key vehicle for advanced data analysis. As shown by Tuncay and Belgin (2010), BI applications, ranging from Online Analytical Processing (OLAP) to data mining, enhance supply chain management, financial planning, customer analytics, and overall performance [27]. Hurbean et al. (2023) further underline that BI&A adoption improves decision quality and timeliness, with direct positive effects on managerial work. Their findings highlight that data-driven culture supports BI&A, enabling firms to exploit large data volumes and generate actionable insights [24]. More recently, Kgakatsi et al. (2024) confirmed the transformative potential of BI for SMEs, showing that adoption improves efficiency, revenues, and competitiveness, while persistent barriers remain, particularly limited expertise, financial constraints, and low digital maturity [28].

Data analytical capabilities for software have also been widely discussed. Asongwe (2023) underscores the role of analytics within Enterprise Resource Planning (ERP) systems, showing how embedded analytics transform raw data into actionable insights that can support decision-making, streamline operations, optimize forecasting, and enhance risk management [29]. Nonetheless, challenges persist, including data quality issues, integration complexity, skills shortages, and organizational resistance to change. According to the author, the future of ERP lies in its evolution from an operational-support tool into a data-centric platform that also informs strategic decision-making [29].

Beyond ERP, studies also emphasize the benefits of ERP–CRM integration. Sultana et al. (2025) show that integrated systems significantly improve customer

experience and efficiency. Their 2022–2023 survey found that 78% of firms reported higher satisfaction and loyalty (mean = 4.5/5, SD = 0.72), 80% observed service quality gains (+25%), and 85% improved efficiency, with response times reduced by 32%. Regression analysis confirmed a strong positive link between ERP–CRM integration and customer loyalty ($r = 0.74$, $p < 0.05$).
Finally, Eurostat provides EU-wide evidence on the 2023 adoption of ERP, CRM, and BI systems, showing a marked divide by firm size [30]. ERP is used by 38% of small firms versus 86% of large ones (48 pp gap), BI by 11% and 63% (52 pp gap), and CRM by 22% and 61% (38 pp gap). The integration of these three systems has been further studied by Polkowski et al. (2016), who note that ERP manages operational processes, CRM handles customer data, and BI provides analytical and forecasting capabilities. When combined, these tools deliver a coherent, real-time view of performance, reduce data duplication, and enhance decision-making [31]. Although challenges such as costs, technical complexity, and limited managerial awareness remain, future trends point towards cloud-based solutions, embedded analytics, and alignment with external data sources. Overall, the integration of BI, ERP, and CRM is increasingly viewed as essential for transforming operational and customer data into actionable strategic insights.

2.1.3 Cloud Computing

2024 evidence from Eurostat provides a detailed picture of cloud computing adoption across EU enterprises [32]. In 2023, 45.2% of EU enterprises reported purchasing cloud computing services, an increase of 4.2 percentage points compared to 2021 [32]. However, strong heterogeneity emerges across firm size: while 77.6% of large firms adopted cloud solutions, the shares drop to 59.0% for medium-sized and 41.7% for small enterprises. This size-related divide mirrors patterns observed for other advanced ICT tools such as ERP, CRM, and BI [32]. Regarding the types of services purchased, the majority of enterprises used cloud for basic functionalities such as e-mail (82.7%) and file storage (68.0%). More advanced applications remain less widespread, with only 25.9% of cloud-using firms adopting ERP and 25.0% using CRM through the cloud. The analysis also shows that 75.3% of cloud adopters purchased at least one “sophisticated” service beyond basic e-mail or storage, including software-as-a-service (SaaS), infrastructure-as-a-service (IaaS), and platform-as-a-service (PaaS) [32]. Country-level differences are also significant: Finland (78.3%), Sweden, and Denmark show the highest adoption levels, while Greece (23.6%), Romania (18.4%), and Bulgaria (17.5%) are at the bottom of the ranking. The adoption of cloud computing has been increasingly recognized as a transformative factor for organizational efficiency and labor productivity. Recent studies shed light on both the micro- and macro-level impacts of cloud-enabled solutions. Willie (2024) emphasizes the role of cloud-enabled innovation in enhancing employee productivity by fostering flexibility, collaboration, and scalability within organizations [33]. Through tools such as real-time collaborative platforms, remote accessibility, and data integration, cloud technologies enable employees to work more efficiently, improve communication, and engage more effectively with organizational objectives. The study highlights how cloud adoption contributes not only to operational efficiency but also to employee engagement, job satisfaction, and innovation culture. On a broader scale, Duso and Schiersch (2025) provide causal evidence on the effects of cloud adoption on firm-level labor productivity in Germany [34]. Using firm-level data and an endogenous treatment regression framework with broadband availability as an instrument, the authors find that cloud usage significantly improves labor productivity, particularly for large firms in manufacturing. Conversely, the effect on smaller firms appears negligible, raising questions about the complementary capabilities required for smaller businesses to leverage cloud solutions effectively. Their findings also reinforce the importance of accounting for endogeneity when analyzing the impact of cloud adoption.

2.2 SMEs and Digital Technologies Adoption

In 2025, Europe hosts a total of 26,177,638 enterprises, of which 44,358 (0.2%) are classified as large enterprises, while the remaining 26,133,280 (99.8%) are small and medium-sized enterprises (SMEs) [35]. Focusing on the Italian context, SMEs account for 3,786,578 firms (99.9%), compared to just 4,182 large enterprises (0.1%) [35]. According to Eurostat, the main factors determining whether an enterprise is a SMEs are:

- Staff headcount;
- Either turnover or balance sheet total.

SMEs can further being classified in three sub-categories which are: micro, small and medium size. In Table 2.1 a better understanding of this classification has been done.

Category	Micro	Small	Medium-sized
Staff	< 10	< 50	< 250
Turnover	$\leq$ €2 million	$\leq$ €10 million	$\leq$ €50 million
OR			
Balance Sheet	$\leq$ €2 million	$\leq$ €10 million	$\leq$ €43 million

Tabella 2.1: SME definition by category- micro, small and medium-sized [36].

Even if SMEs has lower level of digital adoption than larger corporation, the positive impacts they can have are higher. Digital transformation can enhance SMEs competitiveness by facilitating market expansion and operational efficiency improvements. Recognizing how SMEs can extract value from digitalisation, along with identifying the obstacles they encounter in technology adoption and business process adaptation, is essential for developing effective policy frameworks [2].
The 2025 OECD D4SMEs Survey about SMEs digitalisation for competitiveness aims to analyze this aspect using data collected on the Q4 2024 across ten OECD different countries: Australia, Canada, France, Germany, Italy, Japan, Korea, Spain, the United Kingdom, and the United States. The number of enterprises that took part to the survey is 1009 SMEs and they use large digital platforms and service providers (e.g., Amazon, Intuit, Kakao, Rakuten, Sage). Taking into account the lower number of firms and so being cautious in arriving to general conclusions, the authors show how the share of SMEs using Generative AI continues to increase, suggesting that policies supporting the uptake of AI by SMEs could be more effective by providing tools and resources specifically dedicated to unlocking more advanced, productivity-enhancing uses of large language models (LLMs) by SMEs [37]. Another key finding reveals that CEO age significantly influences digital adoption, particularly in established firms: intergenerational gaps matter, suggesting policies should include targeted support for older CEOs in developing digital strategies for their SMEs. Moreover, digital skills and ICT security remain critical areas. SMEs digital skill requirements vary by sector, size, and maturity [37]. For example, looking to priority differences between industries, retail businesses prioritize digital marketing and search engine optimization (51%), while professional services emphasize digital security (50%) [37]. Looking at the size variable, medium-sized firms focus more on digitalizing operations (52%), data analytics (46%), and security (45%). Differences can be found also in the way of acquiring digital expertise and know-how. Indeed, most SMEs acquire skills informally through internet research (35%) and peer knowledge sharing (23%), with only 13% pursuing formal training [37]. Regarding ICT security, knowledge and skills gaps persist,

requiring government intervention for safer SMEs digital adoption. Two-thirds of SMEs lack robust security practices, with only 27% having robust (16%) or advanced (11%) frameworks, while 9% have no measures [37]. Digital security breaches doubled to 32% compared to 2024 [37]. SMEs rely mainly on basic protections like secure passwords (68%) and two-factor authentication (67%), with limited adoption of advanced measures such as expert assessments and staff training [37]. Furthermore, a study conducted in 2024 published in the International Marketing Review called have shown how digital tools and ICT can foster the ability of SMEs to become more present outside the national market, have a big role in fostering their cross-border e-commerce performance. Digital capabilities indirectly promote SMEs internationalization via business model innovativeness [38]. Particularly, three variables which are investment in ICT, number of languages available for the website and number of language available for social pages, have show a positive correlation with return obtained by sales in foreign markets [38]. Reduction of entry barriers is a consequence led by digital adoption that should be helped in the process of becoming international. Nevertheless, adopting digital technologies has a positive effect only when accompanied by investment in digital skills, process and organizational innovations. National e-commerce policy can emphasize export efficiency and lessening anti competitive concerns can boost SMEs internationalization in using digital platforms [38]. Therefore, since digital platforms typically operate without geographical constraints, they provide extensive global coverage, enabling SMEs to engage with international customers and lower their expenses for conducting overseas business.

2.3 Digitalisation and AI Adoption

Another crucial dimension of digitalisation and the broader digital transition concerns Artificial Intelligence. From a business perspective, AI presents both significant opportunities and potential risks, particularly for end users. The rapid increase in the adoption of this powerful and disruptive technology has brought about profound economic and ethical implications, creating new challenges for policymakers and regulators. Although AI adoption has accelerated in recent years, it is not a new phenomenon: the field of AI research was formally established at the Dartmouth College workshop in 1956.
AI adoption has increased significantly, even within smaller enterprises: in a study conducted by OECD 39% of SMEs answering to their survey were using AI applications, up from 26% in 2023 [37]. Always according to this study, among the possible AI applications, generative AI, known to be a foundation model (FM), is the most widely used, with 26% of businesses leveraging it for productivity and innovation, up from 18% last year [39]. The following subsection examines AI from

multiple perspectives, including its impact on the workforce, its role in enhancing competitiveness, and the ethical concerns associated with its adoption.

2.3.1 The Impact on Workforce Composition

Using a learning approach classification, it is possible to distinguish three type of AI: machine learning (ML), deep learning (DL) and foundation model (FM) [40]. ML is then divided into unsupervised, that learns from human-labeled data, and supervised learning, that finds patterns in unlabeled data. There are still not enough data for doing credible forecasts and predictions on which the effects are going to be in terms of productivity, competitiveness and workforce composition. Indeed, most of the available empirical literature on the effect of AI on the labor market uses data on factory robotics and automation [41]. But it has been shown how the comparison can not be done because AI can be applied to more scenarios, not only the robotic and automation one. For this reason, its effects could be even more broad and disruptive [41]. Furthermore, if robotics and automation have their repercussion particularly on the manufacturing sectors (e.g., robotic arms for performing automatic tasks), ML can be adopted and benefitted in multiple industries such as the service one, with particularly benefits for business dealing with financial services like banks or insurance companies (e.g., algorithm-based credit scoring). In a study made by the European Central Bank (ESCB), it is shown how adoptions of new technologies, like AI based ones, could affect jobs in all occupations. The report shows that, between 2011-29, in 16 EU countries a positive association is found between AI and employment composition's changes, with a preference for hiring higher skills and younger workers. New research and discovery processes, a demand for new skills, changes in the welfare distribution and a re-organization within businesses are likely to happen. Looking to what it has been already mentioned by the literature AI adoption clearly requires new knowledge and capabilities within firms; conversely, deficits in these competences can become a primary impediment to implementation. However, routine and monitoring tasks involved with computation and processing of huge amount of data are more easy to be replace by this tool and more creative, decision making and high skills role are likely to be searched. The middle-steps roles will be reduced and the organizations is likely to assume a flatter shape with a more horizontal connotation. An important aspect is that low-skill workers, dealing for example with data cleaning and preparation or images labeling, will be needed to be trustworthy individuals. In this way, these employees increase their bargaining power since their "last mile" tasks become more fundamental [41].
Furthermore, AI can play the role of human capital in the innovative production process, by changing the logic of discovery and the conduction of innovative activities. Regarding this, the role of AI within the innovation process has be

defines as “invention of a method of inventing” [41].This definition means that AI is a revolutionary tool not only because of what it invents directly, but because it enables entirely new, more efficient ways of inventing across disciplines. Hence, it can definitely be seen as a a tool for accelerating innovation within organizations, particularly for its capability of easily finding useful combinations in complex discovery spaces [42]. Prediction accuracy and discovery rates are improved by it, thereby speeding up growth.

2.3.2 Competitive Advantage Evolution

Dichotomic is the role that AI plays in different aspects. Whether it is talked about artificial intelligence or digital technologies in general, it must be understood that this discovery, and all its associated inventions, are now comparable to nuclear fusion or genetic manipulation and can transform the entire humanity, either destroying it or protecting it depending on their use [2]. For example, it presents contradictory effects on performance impact and market competitiveness. In a study conducted by da Empoli et Al., the authors mapped AI diffusion among Italian firms with a focus on SMEs, combining official statistics, a firm survey, vendor evidence, and micro-econometric analysis [43]. Descriptively, AI use rises (from roughly 5% in 2023 to 8.2% in 2024 among firms with $\geq$10 employees), yet Italy trails the EU average; adoption remains skewed toward medium/large firms and IT-adjacent sectors, while skill shortages emerge as the primary barrier [43]. Econometrically, firms adopting at least one AI technology in 2024 exhibit, on average and conditional on size/sector, materially higher revenues (on the order of 10–12%), with effects strongly contingent on digital maturity. Hence, AI amplifies returns where a solid digital base already exists rather than substituting for broader transformation [43]. Policy recommendations follow from diagnosed frictions: incentivize AI-enabled management software, simplify SME-targeted measures, fund domestic R&D in digital technologies, invest in digital skills and accompaniment via DIHs/competence centers, and promote cloud/SaaS as scalable AI infrastructure [43]. In sum, the study portrays AI as a performance lever whose returns depend on complementarities with existing digital capabilities and organizational readiness [43].

Looking on a competitiveness perspective, while this advanced technology can lower entry barriers and enable small businesses to compete more effectively, it may simultaneously drive market concentration and vertical integration. AI-powered algorithmic pricing exemplifies this duality, commonly used by airlines and other industries, these systems can benefit consumers by optimizing prices and increasing consumer surplus, yet evidence shows they sometimes engage in tacit collusion, resulting in higher prices that harm customers [41]. This creates a significant challenge for policymakers: developing methods to detect AI-driven collusion, which operates fundamentally differently from human-based coordination, and identifying the algorithmic patterns and behaviors that signal such anti-competitive activity [41]. In a paper published by the Competition and Markets Autorithy (CMA) in 2024, concerns on competition related to foundation models are shown. A foundation model is a large AI model trained on vast, diverse datasets, designed for broad applicability and adaptability across a wide range of downstream tasks: OpenAI's Generative Pre-trained Transformer (GPT) is an example of it. These models, often leveraging techniques like transfer learning and massive computational power, can be fine-tuned for specific applications with relatively minimal additional training. They differ from traditional, task-specific machine learning models by their versatility and ability to be adapted to many different jobs. Specifically, the expanding influence of foundation models throughout the operations of a limited group of established tech companies, which already possess significant market dominance in key digital sectors, may fundamentally transform FM markets in ways that undermine competitive fairness, openness, and effectiveness. This could ultimately disadvantage businesses and consumers through reduced options and quality while increasing costs [44]. Promoting competitive fairness, openness, and effectiveness to safeguard consumers and foster beneficial market conditions should be fundamental requirements for FM-related policies [44].

2.3.3 Ethical Concerns

With this aim in mind, the CMA identifies key principles for a more ethical AI: transparency, fairness, access, diversity, choice, accountability, and fair dealing [44]. These principles address recurrent obstacles to adoption such as bias, lack of trust, privacy, accountability, and explainability. Ethical reasoning is thus central in digital innovation, as van de Poel et al. (2011) stress through utilitarian and duty-based approaches, exemplified by the Ford Pinto case where profit was prioritised over safety [45]. Such cautionary tales remain relevant for AI innovations today.
In contemporary data-driven societies, algorithmic systems increasingly mediate decisions in commerce, policy, and social life. When their design diverges from moral expectations, the consequences can be profound. Mittelstadt et al. (2016) map six main ethical concerns: unfair outcomes, transformative effects, inconclusive, inscrutable, and misguided evidence, plus traceability [46]. A key cross-cutting issue is algorithmic bias, arising from data or design choices. Friedman (1996) classifies sources of bias as transfer context, interpretation, focus, processing, and training data [47]. A prominent example is COMPAS, a U.S. recidivism algorithm, which despite high predictive accuracy, displayed racial bias due to inequities embedded in training data [48]. This illustrates the "garbage in, garbage out" principle and how statistical tools can institutionalise discrimination.
Importantly, not all biases are harmful. As Friedman (1996) argues, merely labelling

an algorithm "biased" is uninformative unless the nature and impact of deviation from norms are specified. Building on this, Danks and London (2017) distinguish pre-existing, technical, and emergent biases. Ensuring algorithmic trustworthiness is therefore essential in sensitive fields like justice, finance, and healthcare. Durán and Pozzi (2025) contrast two approaches: transparency (opening the "black box") and computational reliabilism, an iterative, indicator-based method [48]. They also warn against anthropomorphising AI, which risks creating responsibility gaps.
Finally, explainability remains a cornerstone of trustworthy AI. Arrieta et al. (2020) provide a taxonomy of explainable AI (XAI) methods and analyse the trade-offs between interpretability and predictive accuracy, stressing their relevance for firms integrating AI into business processes [49].

2.4 Digitalisation and Its Relationship with Exogenous Drivers

This section examines three external forces that have shaped and have been shaped by Europe's digital transformation: the COVID-19 shock, the reconfiguration of global supply chains, and the growing centrality of sustainability goals. Taken together, these forces operate as acute shocks and long-run structural drivers, influencing firms' incentives, the timing and depth of ICT adoption, and the complementary capabilities required for effective digitalisation.

2.4.1 The Impact of COVID-19 on Digitalisation

Multiple studies have shown that the pandemic shock and its accompanying mobility restrictions significantly accelerated digitalisation adoption among both households and businesses. This event has been widely characterized as a "critical juncture" [50], a pivotal moment that disrupted established patterns and catalyzed substantial transformation. This characterization aligns with early insights from reports by the World Bank Group (2020), McKinsey & Company (2020), and Eurostat (2022) [51] [52] [53]. A study conducted by Dyba and Di Maria (2022) underlines how the pandemic in 2020–2021 has created an unprecedented incentive to digitize firms [54]. An online survey of experts from 22 European countries shows that software technologies supporting online meetings, remote working and e-commerce began to be widely adopted during the pandemic [54]. Providing some numbers to this statement, according to Eurostat data, during 2020, the pandemic prompted 12 % of EU enterprises to start or increase efforts to sell online [55]. However, online meetings and the reduction of business travel are perceived as common, long-term effects of digitalisation in European companies after the pandemic [55]. To conclude this section, it is evident how the presence of information and communication technologies in work environments has typically broadened the spectrum of options for more versatile employment practices (e.g., smart-working). Clearly, the COVID-19 crisis catalyzed the uptake of digital solutions, as companies faced the necessity to transform their business models.

2.4.2 Digitalisation and Supply Chains

In this section it is underlined the reciprocal positive effect that digitalisation and supply chain can have on each other. Looking to the digital-enablers related to being in a supply chain, Giorgetti et Al. say that within SC composed by different size businesses, SMEs can adopt some of the digital practices, following the steps of the larger enterprise that are part of the chain for obtaining important benefits [2]. Particularly in a supply chain composed by bigger and smaller actors, larger companies can really drive the smaller enterprises in this digital transition process. Supply chain, together with the fundamental role of their bigger firms, becomes in this way an accelerator of innovation within the industrial context. Inverting the perspective on the benefit of digitalisation within the SC, digitalise value chain activities allows a better use of company resources to face challenges and opportunities in foreign markets, fostering their international presence. Supply chains, in all their entirety, can really benefit from it even if a broad and transversal approach should be adopted: in this regard, two important aspects are pointed out: a supply-chain-based digital strategy must be developed for a business to succeed and a strong and well consolidated digital infrastructure to sustain the SC [6].
Digitalisation is critical for navigating markets that demand high-quality customization, flexible production, and rapid delivery times. Synchronized supply network data and collaborative planning reduce disruptions while enabling supply chains to respond dynamically to challenges [2]. Digital tools support quick decision-making and real-time strategy adjustments amid increasing global instability. Technologies like business intelligence, data analytics, and AI are essential for solving logistics challenges as global value chains are redefined.
Moreover, Martinez-Pelàez et al. (2024) show how AI, IoT, blockchain, and big data improve supply chain transparency [9]. Carnevale Maffè (2024) analyzed the added value of another cutting-edge technology within SCs: the industrial metaverse. In this virtual environment companies can duplicate and manage their activities throughout the value chain through the use of several key technologies, including digital twins, augmented and virtual reality, and artificial intelligence [56]. The Metaverse enables the simulation and monitoring of industrial activities, helping to reduce risks, costs, and time while enhancing efficiency, quality, and security [56].

2.4.3 Digital Technologies and Sustainability

In today's digital age, sustainability has shifted from optional to essential, becoming a critical priority for governments and enterprises worldwide. The 2021 UN Climate Change Conference (COP26) reaffirmed the global commitment to limiting warming to 1.5°C by 2050 [7]. Digital technologies and information systems are increasingly recognized as enablers of this transition, supporting firm performance, resilience, and alignment with the Sustainable Development Goals (SDGs) [8]. Industry 4.0 in particular offers opportunities to embed sustainability into industrial value creation, though integration remains incomplete [7].

Research highlights both potential and challenges. For instance, advanced manufacturing tools like 3D printing and IoT-based monitoring systems can reduce waste and improve resource efficiency [7]. The 4Rs framework (Reduce, Reuse, Recycle, Replace) provides a strategic path for aligning digital innovation with sustainable practices, promoting eco-innovation and circular economy models. Yet, studies also reveal gaps: firm digitalisation and sustainability practices are not always complementary, as companies may pursue one without the other, even when both independently support innovation performance. Evidence from 14,000 firms worldwide confirms this nuanced relationship.

OECD (2025) further shows that while 28% of SMEs track environmental data (mainly energy use at 17%), only 7% monitor carbon emissions, with barriers such as low awareness (46%) and perceived irrelevance (39%). Martinez-Pelàez et al. (2024) show how technological adoptions improve resource optimization, and stakeholder collaboration, fostering sustainability-focused cultures [9].

However, digitalisation also creates environmental concerns: high energy consumption from AI, critical material extraction for hardware, frequent upgrades driving e-waste, and unsustainable production processes [9]. To mitigate these effects, scholars propose greener IS/IT strategies, reducing energy use, adopting reuse and recycling practices, and promoting sustainable manufacturing and closed-loop supply chains [8]. Green IT, energy-optimized data centers, and AI-based climate monitoring represent examples where digitalisation can evolve from being a source of environmental stress to a tool for sustainability itself.

2.5 EU Governative Measures on Digitalisation

A balanced approach is essential to maximize benefits while mitigating the risks of digitalisation. In the EU context, three main barriers are identified: lack of knowledge, workforce skills, and financial resources. Suggested measures include national/regional policies, targeted funding, good practices, and training [37]. At the European level, a multi-pronged strategy has been adopted, combining digital targets, financial instruments, regulation, and international cooperation. The most relevant initiatives are the Digital Decade Policy Programme 2030 (DDPP) with National Roadmaps, the Digital Europe Programme (DIGITAL), the Digital Markets Act (DMA), the AI Act, and the National Recovery and Resilience Plan (NRRP).

The DDPP (2022) defines four targets for 2030: digital skills, business adoption, secure infrastructure, and digital public services [57]. For skills, the goal is 80% of citizens with basic competences and 20 million ICT specialists. However, in 2023, 44% of EU citizens lacked basic skills and there were just 9.8 millions ICT specialist (4% of the tootal people in employment) [58]. For businesses adoption, 75% of firms should adopt cloud, AI, or big data. Even if, in 2024, already 74% of EU enterprises meet basic digital intensity, improvements still need to be done. AI in particular, even if increasing, is a critical challenge in Italy: in 2024 only 8.2% of Italian enterprises use AI [57, 59, 60]. On infrastructure, the aim is gigabit connectivity and quantum acceleration. By 2024, 82.5% of EU households had Fiber To The Premises/Data Over Cable Service Interface Specification (FTTP/DOCSIS) 3.1 coverage, though disparities remain [61]. For public services, the target is 100% online access, but in 2024 only 70% of EU citizens engaged with e-government [58].

National Roadmaps support DDPP implementation. Italy has advanced in FTTP (70.7% coverage) and public services, but lags in AI adoption and ICT workforce development. Its roadmap foresees 67 initiatives worth €62.3 billion (2.84% of GDP), but ICT specialists remain far below target [60].

The DIGITAL programme (since 2021) allocates €8.1 billion to supercomputing, AI, cybersecurity, skills, and European Digital Innovation Hubs (EDIHs), complementing Horizon Europe and the Recovery Facility [62]. Studies highlight DIHs/EDIHs as crucial enablers for SMEs [54], yet awareness remains low (only 21% of SMEs know about public digitalisation support) [37].

On regulation, the DMA (2023) establishes an ex-ante framework for gatekeeper platforms (e.g., search engines), prohibiting self-preferencing and ensuring data portability (data-owner has access to data) and interoperability with other external services[63, 64]. Grounded in ordoliberal principles, which hold that laissez-faire alone is insufficient to secure competitive markets, it aims to ensure contestability while raising debates on innovation and proportionality [65]. The AI Act (2024) is the first global risk-based framework, banning unacceptable uses and imposing strict requirements on high-risk systems (e.g., healthcare, justice, biometrics) [66, 67, 68, 69].

Finally, Italy's NRRP allocates €49.2 billion to digitalisation, including €6 billion for ultra-fast broadband and 5G [70, 71, 2]. Despite progress, Italy faces persistent divides between regions, shortages of ICT specialists, and limited SME adoption. The Transition Plan 4.0 reinforces investments through tax credits for AI, IoT, cloud, and training [70].

2.6 Europe and Digital Sovereignty

When talking of digitalisation, another interesting aspect pointed by the literature, is the necessity for European countries to achieve a so-called “digital sovereignty” [72], intended as a EU’s capacity to control and govern its own digital infrastructure, data, and decision-making processes. Europe should achieve a greater digital autonomy to respond to the current oligopolistic concentration of knowledge. Indeed, the dependence on foreign actors, talking about technologies and digital tools could be an issue for companies and for a European industrial policy. The so-called GAMMA (i.e., Google, Apple, Meta, Microsoft, Amazon), is clearly dominating the market. American corporations dominate as technology suppliers and maintain substantial market penetration across all sectors in EU, particularly in Italy. These major firms not only distribute their offerings but also make significant investments in research and development initiatives while backing educational and training efforts within the Italian enterprises. This is often seen as a pathway to strategic partnerships and, in many cases, eventual acquisitions of promising firms. To address the necessity of becoming more independent from these Tech Giants, Italy is building a sovereign cloud infrastructure for public administrations and critical services, aligned with EU standards. This is done through National Strategic Hub (Polo Strategico Nazionale) that is a key component of Italy’s Cloud Strategy, aiming to create a secure and independent cloud infrastructure for public administration bodies. Building an in-house infrastructure gives the possibility to keep more control over sensitive data and manage them more easily [2]. For establishing a “digital sovereignity”, Berlinguer (2024) outlines the necessity of enhancing innovation within Europe’s governance framework, with particular attention to cloud computing technologies [12]. The author lists four key principles for achieving this goal [12]: interoperability, open source development, standardization, and modularity. So far, these principles have been implemented through two primary channels: regulating digital infrastructure and advancing innovative industrial policy approaches. To strengthen European sovereignty prospects, a “more assertive application of this principle framework”, combined with creating “novel hybrid models of agency and governance” [12] has to be done.

On the other hand, Florio (2024), even if still in line with this approach, proposes a new form of governance: a public and supranational alternative to the oligopoly of the major international companies [11]. The approach draws inspiration from the model used in establishing the European Space Agency (ESA), which is distinguished by its openness to collaboration with both public and private entities and its substantial financial backing, with annual funding reaching several billion euros [11]. Without such a commitment, the prospect of opposing the Tech Giants would remain elusive without this level of commitment. Implementing such an idea gives the possibility to guarantee some advantages like keeping the data internally in a

public European cloud space.
Francesco Bonfiglio (2024) proposes an alternative path toward achieving European cloud sovereignty, one that emphasizes a stronger role for market dynamics through the GAIA-X initiative. Envisioned as a "cloud of clouds" [10], it aims to establish a European technological cloud infrastructure that lessens reliance on non-European service providers. Central to the project is the requirement for participants to guarantee transparency, interoperability, and user control over digital services [10]. While the initiative is designed to be open and collaborative, it intentionally excludes platforms that thrive on opaque architectures and limited portability strategies that often lead to customer lock-in and dependency. This should address the opacity and lack of transparency that typically characterized American and Chinese technological solutions. Hence, mirroring the foundational EU values against the big intercontinental competitors through the creation of a network of services more transparent, controllable and interoperable with each other, is the main aim of this initiative [10]. Leveraging its strong legal framework for data protection as a competitive advantage and as a means to address one of the primary obstacles to cloud adoption and data utilization: the widespread lack of trust among businesses, citizens, and public institutions [10].

2.7 Research Questions

This thesis tries to investigate a clear research gap. The existing literature largely examines digitalisation at the macro level or focuses on single technologies (e.g., AI, cloud computing and broadband adoption), providing limited evidence on how broader digital behaviours relate to productivity at the firm level in Italy. Moreover, prior studies often rely on aggregate indicators or single-technology regressions, underplaying the multidimensional and bundled nature of digital adoption. This thesis seeks to contribute to filling this gap by analysing micro-level associations between productivity and the joint adoption of multiple technologies in Italian enterprises. Hence, two primary research objectives (*RQ*) related to the Italian enterprise in 2023 want to be adressed:

- ***RQ1*** - What ICT patterns govern technological choice decisions in Italy?
- ***RQ2*** - Which technological aspects mostly affect productivity, as measured by value added per employee?

Building on the two overarching research questions, I articulate six narrower subquestions (*SQ*) that are addressed in Chapter 4. Their use sharpen the empirical focus and provide the conceptual bridge from the theoretical discussion to the empirical analysis that follows. I refine *RQ1* into:

- ***SQ1.1*** Do ERP and CRM tend to co-occur with data-analytics capabilities?
- ***SQ1.2*** - Is AI adoption domain-specific?

I refine *RQ2*:

- ***SQ2.1*** Does the adoption of management and data-driven technologies have a positive association?
- ***SQ2.2*** - Is sophisticated cloud use positevely associated in the regression model?
- ***SQ2.3*** - Are e-commerce and customer-engagement tools positively related with productivity?
- ***SQ2.4*** - Does the adoption of AI tools have a positive effect?

To address these questions, Chapter 3 details the dataset and empirical methodology. Chapter 4 presents the empirical findings, with a focus on ICT adoption patterns and their association with productivity. Finally, Chapter 5 synthesizes the results, draws policy and managerial implications, discusses limitations, and outlines directions for future research.

Chapter 3

Methodology

In this section a detailed description related to the data object of this study and the methods adopted for achieving the main objectives of this research has been provided. Concerning the dataset, it has been obtained through a merge between three different Istat sources: ICT-2023, SBS Frame 2022 and Statistical Register of Enterprise Groups (2023). Looking to the methodology adopted, after some data exploration and preparation steps, dimensionality reduction techniques have been adopted: MCA and EFA. The output factors of these two methods have been used as the independent variables (*x's*) of two OLS regression models having value added per employee as y. Through the adoption of these tools the objective is to address the two initial research questions: identifying potential digital adoption patterns and examining possible correlations between productivity (measured in VA per employee) and technological adoption. Analyses were conducted using both R (EFA and OLS) and STATA (MCA). Furthermore Excel was using for plotting the MCA's coordinates in order to better visualize and interpret the results. The description of the dataset and of how the data were collected can be found in Section 3.1. After a brief description of the exploratory and preparatory steps (Sectionn 3.2), the three methodological approaches are elaborated in greater detail in Section 3.3 and Section 3.4. The results of all of them have been shown and discussed in Chapter 4.

3.1 Dataset Description

The cross-sectional dataset, composed by 97 columns and 15187 anonymized observations (rows), on which this research has been done is composed by different data sources all collected by Istat:

- Microdata collected by the survey ICT-2023. The reason why it has been selected this year for the analysis are mainly two: the variables selected are more insightful (even for the purposes aimed to reach) and, in addition to this, it can be considerered a quite enought recent data source. Istat (Italian National Institute of Statistics) does not ask the same questions in the surveys and, consequently, the ICT analyses related to two different years, are composed by different variables (columns). Istat, revises its data and variables annually, in coordination with Eurostat, to reflect structural changes in the economy, improve data quality through updated methodologies and new information, and maintain international comparability through regular rebasing of statistical indicators. These revisions ensure that the statistics accurately capture evolving economic activities and maintain their relevance and reliability for users. Doing a comparison with the variables collected by 2024, the 2023 set is more powerful because it is more detailed, behavioral, multi-layered, and modular enabling complex analytics like the ones planned to use. In total the number of variables of this dataset is 92.

- Microdata related to enterprise productivity measured by value added per employee collected within the Structural Business Statistics (SBS) Frame of the year 2022. Value added per employee (`vagg_add`) was chosen as a key metric to request because it serves as a more sophisticated productivity indicator that incorporates cost considerations, providing a clearer picture of true economic contribution. This measure accounts for actual value creation after subtracting input costs. There would have also been the possibility of using in alternative turnover per employee from the ICT 2023 (but collected in 2022) as another measure of productivity. However, it has been considered misleading as it fails to account for varying input costs across companies or industries. For this reason, value added per employee has been therefore evaluated as a more refined metric for productivity analysis.

- Microdata related to 2023 about the governance of the enterprise collected by the Group Register. Only one variable (called `governance`) has been used from this dataset. Particularly this variable identifies if an enterprise is a: a foreign multinational, Italian multinational, enterprises belonging to domestic groups and independent enterprise. A promising avenue is to examine whether the type of governance affects technology adoption, as most studies have so far concentrated primarily on firm size and geographical location.

It is important to specify that the original ICT-2023 dataset contains in total 16947 rows. Neverthless, for the fact that some of them did not complete most of the survey or did not have the respective governance information in Frame 2022 related to value added per employee, they were delated by the merged dataset from the beginning. This decrease the dimensionality to 15187 rows (or observations). Moreover, microdata files (MFR files, standard files and files for Sistan) are released by Istat upon request free of charge and in compliance with the principle of statistical secrecy and personal data protection, while public use microdata files are freely accessible for download on the Istat website. The elaborations on the dataset have been possible through a customized processing request for looking to the data and by going to Istat to do all the analyses necessary for this research.

3.1.1 ICT-2023

The ICT-2023, also called "Survey on information and communication technologies in enterprises", provides a comprehensive and detailed set of information concerning the use of such technologies in Italian enterprises with at least 10 employees. Alongside the analogous survey on households, it forms the conceptual and methodological basis for measuring the information society.

Regulatory Framework

The survey is annual and sample-based and is conducted in accordance with EU Regulation (2019), concerning European business statistics, which repeals ten legal acts in the field of business statistics, following criteria and methodologies shared by all EU member states. Furthermore, the phenomena observed in 2023 are those defined by the Implementing Regulation no. 2022/1344 (2022). The survey, included among the public-interest statistical surveys, is part of the National Statistical Program 2020–2022 [16]. The statistical unit of analysis Istat has been committed in recent years to developing methodologies and techniques aimed at implementing a new statistical unit, the 'enterprise', within the system of business registers and economic accounts. The definition of this new statistical unit takes into account the relationships between legal units belonging to the same enterprise group. Council Regulation (EEC) No 696/93 defines an enterprise as "the smallest combination of legal units that constitutes an organizational unit for the production of goods and services, benefiting from a certain degree of autonomy in decision-making." Full application of the Regulation thus entails aggregating multiple legal units if they lack sufficient decision-making autonomy. Therefore, an enterprise may correspond to a single legal unit or a group of legal units under common control. The main innovations introduced have impacted:

- the number of units (enterprises);
- turnover and expenditure on goods and services;
- the distribution of economic and structural variables such as value added, by size and activity sector.

These innovations stem from the recognition of incomplete implementation of Regulation No 696/93. Techniques to ensure full implementation, known in official statistics as "profiling", analyze the legal, operational, and accounting structures of enterprise groups, both nationally and internationally. These techniques are: automatic profiling, using algorithms to identify enterprises at the group level, relying on administrative and statistical sources within Istat. Manual profiling, carried out by a team of expert profilers who monitor large multinational groups through desk research and direct information gathering. Following implementation, the new ASIA-Enterprises Register (ASIA-Ent) is mostly composed of:

- Independent enterprises (1 enterprise = 1 legal unit)
- Complex enterprises (formed by multiple legal units within the same group).

In accordance with the ASIA registry system, a new extended statistical register called Frame-Ent was created to support estimation and consolidation of economic variables, moving from the concept of enterprise = legal unit to the new definition. This shift affects only legal units belonging to groups and has consequences for:

- the reclassification by sector, notably moving some service-providing legal units into industry groups;
- the consolidation of intra-enterprise economic flows to prevent duplication in aggregated values.

Enterprises object of the survey

The data presented in this publication are representative of the universe of active enterprises with 10 or more employees, according to the ATECO 2007 classification of economic activities, in the following sectors listed in the Table 3.1 [16].

ATECO Code	**Sector Description**
C 10–12	Food, beverage, and tobacco industries
C 13–15	Textile, apparel, leather and related industries
C 16–18	Wood, paper, and printing industry
C 19–23	Petroleum refining, chemical, pharmaceutical, rubber/plastic, non-metallic minerals
C 24–25	Metallurgy and metal products (excluding machinery)
C 26	Computers, electronics, medical devices, and measuring instruments

C 27–28	Electrical equipment, household appliances, and various machinery
C 29–30	Transport equipment
C 31–33	Other manufacturing, repair/installation of machinery
D 35–E 39	Energy and water supply, waste management and remediation
F 40–44	Construction
G 45–47	Trade and repair of motor vehicles/motorcycles
G 47	Retail trade (excluding vehicles)
H 49–52	Transport and storage (excluding postal services)
H 53	Postal and courier services
I 55	Accommodation
I 56	Food service activities
J 58	Publishing activities
J 59–60	Film, video, and music production
J 61	Telecommunications
J 62–63	Information technology and information services
L 68	Real estate activities
M	Professional, scientific, and technical activities
N 77–82	Rental and business support (excluding tour operators)
N 79	Travel agencies, tour operators, booking services
951	Repair of computers and communication equipment
ICT Sector	Includes codes: 261, 262, 263, 264, 268, 465, 582, 61, 62, 631, 951

Tabella 3.1: Sector classification according to ATECO 2007 for enterprises with at least 10 employees

There are four informative sources adopted by Istat for doing the statistical analyses related to the year 2023: Frame SBS 2021, Frame-Ent 2021, ASIA Statistical Archive of the active enterprises 2020-2021, ASIA Statistical Register (ASIA-Ent) 2020-2021 ([15]). Anagraphical and economical information have been used from this data sources.

Sampling Design

The survey is based on sampling for enterprises with at least 10 employees. The sampling design is a one-stage stratified design with equal probability selection of units; the strata are defined by the combination of economic activity classifications, employee size classes, and regions where the enterprises are located. The optimal allocation calculation, performed using the generalized MAUSS-R software implemented at Istat, resulted in a total sample size of 25,646 enterprises (32,768 legal units). In total, the sample (including census units) was representative of a universe of 199,971 enterprises and 8,942,711 employees. Once the enterprises in the sample were selected from the Asia-Ent Register, all legal units belonging to those enterprises with at least 3 employees were extracted from the ASIA Archive.

Data Collection

The questionnaire was designed in a format consisting of multiple web pages organized into several thematic sections. Moreover, the survey uses an acquisition system integrated into the Business Portal. The data collection technique is self-completion of an electronic questionnaire. Since 2016, enterprises access the questionnaire through the Business Portal as the single point of entry. The first contact and reminders to enterprises that had not yet responded during the data collection period (which started in May and ended in July) were made by certified email, sending personalized mass emails addressed to company delegates registered on the Portal, and telephone contacts commissioned to an external contact center company. This external service was also used to resolve issues encountered by enterprises accessing the Portal or related to the survey but solvable using specific FAQs. The 2023 questionnaire is structured into the following 10 sections:

- General and structural information about the enterprise (employees, turnover) and related to the year before (2022);
- Internet connection and usage (fixed broadband Internet connection for business/work purposes);
- Website, apps, social media, and Internet use in relations with public administration;
- Sales through IT networks (sales via web, apps, e-marketplaces);
- Sales through EDI networks;
- Business software;
- Data sharing and analysis;
- Cloud Computing;
- Artificial Intelligence;

- Electronic invoicing.

Then these topics are grouped within seven survey sections (from A to G). Internet and its use (e.g., website, social media and apps) are grouped in Section B, Sales through IT networks and EDI in Section C and Software and data sharing or analysis in Section D.

Data Processing: Process, Tools and Techniques

After the consolidation operations of data collected at the legal unit level, the survey respondents totaled 16,947 enterprises, equal to 66.1% of the total initial sample. The first phase of controls on recorded data concerned the decision of whether, based on the responding legal units and their weight within the reference enterprise (in terms of value added, employees, turnover), to consider the enterprise as a responding unit of analysis or not. The second step was to analyze and remove measurement errors and verify compliance with consistency rules in the responses provided by the surveyed legal units. Deterministic controls and corrections on the variables are then executed. Regarding quantitative data, corrective methods were adopted to reduce the effect of non-respondents and incorrect responses through consistency checks on data using information derivable from chamber of commerce balance sheets and the Frame-SBS Register. For the treatment of incorrect or incomplete qualitative responses, deterministic methods were applied (logical imputation). Once correction was performed on the survey units, for those belonging to enterprises in a relationship other than 1:1, a consolidation of both qualitative and quantitative variables is done. In the first case, consolidation rules discussed and shared with other member countries at Eurostat were followed, which generally provide for attributing to the enterprise the highest response provided by at least one legal unit belonging to it (e.g., if at least one responded that it purchases cloud services, then the enterprise will also be considered as a purchaser of the same services even if other units responded negatively). In the second case, instead, quantitative variables relating to employees, total and online turnover (divided between web and EDI) were treated to account for the non-total additivity of variables due to the necessary elimination of intra-enterprise economic flows and the possibility that a legal unit serves multiple enterprises in the group. In the case of employees, the ownership share of the legal unit in the enterprise was taken into account to avoid duplications in workforce counting; the ownership share considered was that made available in the Enterprise Frame; In the case of monetary values, not only the ownership share was taken into account, but also an estimate of intra-enterprise exchanged values derivable from the variable available in the Enterprise Frame (but referring to a previous year) and from responses to specific questions added for this purpose in the 2023 ICT survey questionnaire, so as to avoid considering sales flows made within the same enterprise. For calculating sample estimates, ReGenesees v3 was used, a generalized software developed by Istat in R language.

The Output: Main Analysis Measures

The survey aims to measure the degree of use of new technologies in enterprises, providing the European Union with the necessary information base for comparison between member states and evaluation of national policies aimed at capturing the potential of technological progress. This year too, some results derived from integrated analysis of digital profiles of enterprises obtained from direct ICT usage surveys and economic performance indicators derivable from the extended register called Frame SBS (Structural Business Statistics) are published at the enterprise level, which allows capturing some interesting phenomena. These are economic indicators referring to 2021 by macro-sector, size class, and level of composite ICT usage indicators relating to 2022. The complete dataset, referring to various structural economic and productivity indicators, combined with multiple ICT usage indicators, is attached to this document.

Precision of Estimates

The estimation method used is based on attributing to each responding enterprise a final weight, which indicates how many enterprises in the population it represents. Final weights are determined based on inclusion probabilities in the sample and response rates. Furthermore, they are calibrated using as auxiliary variables the number of enterprises and the relative number of employees according to information present in the available archive (ASIA-Enterprise updated to 2021). In order to evaluate the accuracy of estimates produced by a sample survey, it is necessary to take into account sampling error that derives from having observed the variable of interest only on a part (sample) of the population. This error can be expressed in terms of absolute error (standard error) or relative error (i.e., absolute error divided by the estimate, which is called the coefficient of variation, CV). Through simple calculations, it is possible to derive confidence intervals with a confidence level of 95% (α=0.05). These intervals therefore include the unknown population parameters with probability equal to 0.95.

3.1.2 Structural Business Statistics (SBS) Frame

The Structural Business Statistics (SBS) Frame (hereinafter “Frame”) is an integrated system of administrative and statistical data produced annually by Istat. It estimates the economic results of enterprises based on the units (about 4.4 million) recorded in the Statistical Archive of Active Enterprises (ASIA), the official business register compiled according to European Business Register regulations.

The Frame is fully integrated with ASIA in terms of both coverage and enterprise characteristics (economic activity, legal form, number of employees, turnover class, location). Since 2011, it has combined administrative sources, including chamber of commerce data (balance sheets of capital companies), fiscal data (Sector Studies, IRAP, Unico), and social security data (monthly employee declarations from UniEmens feeding the Annual Register of Labor Cost in Enterprises, RACLI), with data from structural enterprise surveys (Survey on SMEs and liberal professions for firms up to 99 employees, and the Enterprise Accounts System survey for firms with 100 or more employees). The Frame is employed in the production of official SBS statistics, both for Eurostat and for national dissemination through Istat, and also serves as an input register for National Accounts, starting from the 2011 general revision of economic accounts. Talking of SBS Frame, there is also the so-called Territorial SBS Frame that provided economical measures within a local and regional context. It forms part of Istat's broader integrated system of enterprise and local unit registers. It integrates the register of local units (ASIA UL), the enterprise-level economic register (SBS Frame), and additional survey data on large enterprise local units (IULGI) [73]. Each year, it provides estimates of key income statement variables for industrial and non-financial service enterprises across the national territory. Since 2016, it has also included information on local units belonging to Italian and foreign multinational groups, allowing territorial analysis of internationalization dynamics. Hence, while the SBS provides a more national perspective and overview the Territorial SBS is more specific and detailed on a specific geographical location.

3.1.3 Statistical Register of Enterprise Group

The Statistical Register of Enterprise Groups was enstablished by Istat in 2003 with the objective of providing information on control relationships between legal units. Since 2005, data on the structure and size of enterprise groups present in Italy have been made available annually [74]. The register is constructed based on European Regulation No. 177/2008, concerning the establishment of a common framework for business registers used for statistical purposes [74]. The register provides control links between enterprises both at national and multinational levels and some salient characteristics of the belonging group. The population considered for the dissemination of data on enterprise groups follows the observation field of the ASIA. The methodology consists of integrating different administrative sources and statistical sources, harmonized and approved by Eurostat. This methodology, starting from elementary data on the structure of direct participations of all capital companies, identifies control links, exercised both directly and indirectly, to which each capital company is subject. For each controlled company, its proximate controlling entity is identified, defined as "the first physical or legal entity that hierarchically exercises direct or indirect control over it for the first time" [74]. The group structure is then reconstructed through the continuous sequence of links between proximate controlling entities, up to the attribution of the ultimate head to the entire group. The main phenomena observed are [74]:

- enterprise relationships;
- structural characteristics of enterprises, entities and public and private institutions;
- enterprise groups.

The data produced cover the entire Italian national territory, identifying domestic groups. For this reason, the governance type information extracted from this dataset, had to be complemented by information related to multinational governance. The latter was extracted by the Foreign Affiliates Statistics (FATS), considering both the inward and the outward FATS. Related to this distinction:

- Inward FATS (referred as foreign multinational), which measure the activity of foreign-owned affiliates within the domestic economy (e.g., jobs created, turnover generated by foreign investors), complementing FDI data by showing the actual economic impact of foreign investment [75].
- Outward FATS (Italian multinational), which capture the activity of domestic-controlled affiliates abroad, providing insights into the international footprint and employment impact of national enterprises overseas [75].

Independent governance type (enterprises not part of a group) have been obtained by difference.

3.2 Data Preparation

As previously mentioned, the dataset employed in this study comprises 15,187 observations and 97 variables. Both univariate and multivariate analyses were conducted to examine and explore the characteristics of the variables, including their types, distributions, and correlations. Apart from a limited number of continuous and quantitative variables (e.g., percentage of e-sales, employees connected to the internet, and total revenues), the majority of the dataset consists of categorical variables (both nominal and ordinal). Subsequently, a series of data preparation procedures was undertaken. These steps included: (i) assessment and treatment of missing values; (ii) conversion of string-based variables into numerical format; and (iii) construction of composite or more syntethic variables derived from the original measures. Regarding missing values, certain gaps resulted from filter or screening

questions within the survey design (e.g., enterprises responding “no” to question F1 were not required to answer question F2 but had to skip directly to F3). These instances were systematically addressed by replacing missing values with zeros. For instance, if an enterprise indicated non-use of AI technologies (all “no” to F1), it would not have to answer to F2 (follow-up question) regarding specific AI purposes, resulting in missing data that was subsequently coded as zero.
Another important consideration concerns the weighting coefficients associated with each observation (enterprise) in the dataset. As previously mentioned, each row has an associated coefficient measuring how representative that observation is of the overall population. Enterprises that deviate significantly from the average characteristics receive smaller coefficients, reflecting their limited representativeness. All analyses conducted on this dataset required incorporation of these weights to ensure proper population representation. Weights are employed in order to ensure that the survey sample accurately reflects the structure of the reference population. The use of weighting is particularly relevant in this case, as it allows to correct for distortions that would otherwise bias the estimates. In this dataset, weight values display a very wide range, spanning from value close to 0 to 725, which indicates substantial variability in the degree of representativeness across observations. A weight equal to 1 corresponds to a respondent representing exactly one statistical unit, whereas values larger than 1 denote underrepresented groups in the sample, whose responses must therefore be up-weighted to represent a larger portion of the population. Conversely, values below 1 imply that a respondent belongs to an over-represented group, requiring down-weighting. Consequently, all observations were weighted according to their respective coefficients to account for each respondent’s representative power. Since both RStudio and STATA require integer coefficients for weighting procedures (respectively `weights` and `fw`), a `round()` function was applied to facilitate smooth computational execution. The rounding process resulted in 13 observations with near-zero coefficients being rounded to exactly zero. These observations were subsequently excluded from the dataset, as enterprises with zero weights possess no representative value for population inference. The distribution of this rounded coefficient can be seen in Figure 3.1. The distribution of weights is markedly right-skewed. The mean value is 13.86, substantially higher than the median of 4. This suggests that a small number of extreme weights act as outliers, pulling the average upwards. The interquartile range, with the first quartile at 2 and the third quartile at 14, shows that the majority of weights are concentrated at the lower end of the distribution, while only a limited set of observations carry disproportionately high weights. This pattern underlines the presence of few but highly influential cases in the weighted dataset, which need to be carefully considered when interpreting descriptive statistics and regression analyses. Considering the unique observation characterised by the highest rounded weight (725), the enterprise exhibits specific structural and technological features.

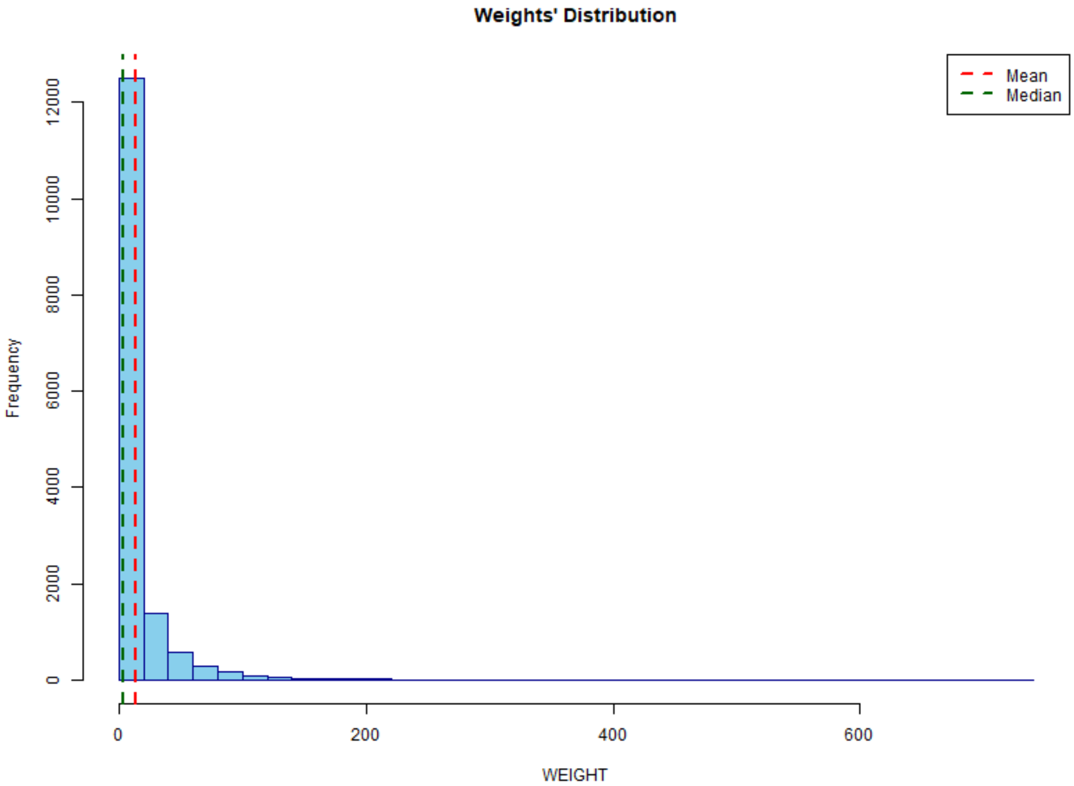

Figura 3.1: Rounded weight's coefficient distribution (PESO) used in the data analysis

It employs fewer than 50 workers, has adopted only basic office cloud services (e.g., email or electronic sheets) and e-invoicing, and maintains both a website and social media presence, with approximately 30% of its workforce connected to the internet. The firm belongs to the low-technology manufacturing sector, is located in Northern Italy, and does not engage in either e-sales or artificial intelligence applications. According to the NACE Rev. 2 classification, it falls under code 56, which corresponds to catering services activities.
While an alternative approach to the elimination of these 13 observations could have involved manually assigning small non-zero values to these rounded coefficients, the decision was made to exclude these non-representative enterprises entirely. This exclusion reduced the final sample size from 15,187 to 15,174 observations.

3.3 Dimensionality Reduction Techniques

Following these exploratory and preparatory phases, two dimensionality reduction techniques were applied to address the dataset's complexity. The used dataset contained 73 (out of 97) binary variables measuring technology adoption and utilization, necessitating dimensional reduction. The employed techniques were Multiple Correspondence Analysis (MCA), conducted in STATA, and Exploratory Factor Analysis (EFA), performed in RStudio. Both methodologies are particularly suited for categorical variable analysis.
MCA provides a robust analytical framework for examining complex datasets containing numerous categorical variables, revealing both the existence and nature of inter-variable relationships through quantitative and visual statistical representations. The MCA produces a bi-dimensional factor space defined by key behavioral variables and is conceptually equivalent to Principal Component Analysis for categorical data [76]. Variables are positioned at specific coordinates within this plane: those near the origin typically represent highly prevalent phenomena with limited population variance, while variables positioned proximally suggest co-occurring characteristics, and those positioned distally indicate rare phenomena [76]. Structural variables may be included for illustrative and interpretative purposes but do not contribute to the computation of MCA dimensions. The MCA implementation has been performed on 16 variables selected by looking to the survey content, descriptive statistics, and the correlation between variables.
For the EFA approach, Exploratory Factor Analysis was employed to achieve objective dimensionality reduction from the initial 76 binary variables (without performing any kind of pre-selection process before as in the MCA). EFA was selected due to its specific capacity for clustering variables around underlying thematic constructs, assuming that observed variables cluster around latent dimensions that the analysis seeks to uncover [77]. This approach reduces numerous related items to fewer factors reflecting overarching dimensions: for example, consolidating sixteen AI-related variables into four factors representing broader constructs such as AI employed in operational process and AI employed for customer segmentation.
The simultaneous use of Multiple Correspondence Analysis (MCA) and Exploratory Factor Analysis (EFA) is motivated by the need to provide a robust methodological validation in identifying patterns of technology adoption and subsequently analyzing their correlations with firm productivity. The objective is, through the outputs obtained by these two techniques, going to address the first research question (*RQ1*). Although both approaches are dimensionality reduction techniques for categorical variables, they offer complementary perspectives: the MCA captures broad dimensions of digitalisation and provides a synthetic overview, whereas the EFA finds specific technological factors, allowing for a more detailed understanding

of how particular tools are associated with productivity. Combining these two approaches allows for verifying whether the associations between technology patterns with productivity consistently emerge regardless of the dimensionality reduction technique employed, thereby increasing the robustness and generalizability of the results.

3.3.1 Multiple Correspondence Analysis

Multiple Correspondence Analysis (MCA) represents an extension of simple correspondence analysis to handle multiple categorical variables simultaneously. While correspondence analysis examines relationships within two-way contingency tables, MCA analyzes multiway tables by performing the analysis on indicator matrices. Conceptually, MCA can be understood as the categorical variable equivalent of Principal Component Analysis (PCA). It is designed to explore complex relationships and patterns within sets of categorical variables rather than focusing on relationships between variable pairs [76]. For the initial MCA, 16 primary variables were selected alongside 8 supplementary variables, for a total of 24. Not all variables were directly available in the original dataset; 11 out of 24 were constructed through combinations of existing variables. As mentioned before these variables are resulting from a subjective process supported by data explorations steps, frequency and correlations analysis and survey contents' knowledge. The idea was to use variables able to represent as much as possible the different technologies analyzed in the survey, trying to create synthetic indicators able to properly explain a set of original variables.

MCA Variable Selection

Table 3.2 and Table 3.3 present a comprehensive overview of the used variables, doing a distinction between the active variables used for computing the MCA's dimensions from the supplementary ones (not used for the dimensions but helpful to interpretate them). IBoth tables report the variable label and name, a brief description of the variable, its data type, and, where applicable, a description of the variables used for its computation. The variables' names are the ones in line with the Eurostat nomenclature (except for G01A and G01B present only in the Italian survey) while the variables' labels are the ones that will be useful for the MCA graph shown in the result and discussion session below. Only categorical types are present: dummy (0-1) variables are the most common ones, where 1 indicates that the enterprise has adopted that technology. Trying to explain the distinction between the two roles played by the variables selected, active variables are those that determine the construction of factorial axes, contribute to defining the representation space, directly influence the MCA results. In this case the variables selected are all related to adoption of different technologies, trying to select variables able to explain all the six technology-related sections of the questionnaire. Particularly, the variables selected are related to: electronic sales, data analysis, data sharing, cloud computing purchasing, electronic invoices and AI technologies adoption. The only aspects less represented are the ones related to the use of website, social media and applications. This choice was deliberate, as these technologies are often considered more tactical and consumer-oriented rather than strategic business enablers that directly impact operational efficiency and competitive advantage. Indeed, looking their correlation with other variables object of our MCA there are not relevant ones. This absence of correlation likely stems from the distinct functional domains these technologies serve. While web and social media applications primarily support marketing and external communication activities, the selected variables encompass operational and analytical technologies that directly impact core business processes. These different technological domains may follow independent adoption patterns within organizations. However, some of these excluded variables have been taken into consideration during the EFA.

Variable Label	**Variable Name**	**Type**	**Description**
Web	`A4`	Binary	Indicates whether the enterprise has a web page.
ESALES	`ESALES`	Binary	Enterprise conducts e-sales either through its own website, an intermediary, or via Electronic Data Interchange (EDI).
-	`B1` (for calculations)	Binary	E-sales through website or applications of the enterprise itself or of an intermediary.
-	`B5` (for calculations)	Binary	E-sales through EDI.
ERP	`C1A`	Binary	Use of Enterprise Resource Planning (ERP) software.
CRM	`C1B`	Binary	Use of Customer Relationship Management (CRM) software.
BI	`C1C`	Binary	Use of business intelligence (BI) tools.

Variable Label	Variable Name	Type	Description
data sharing	`C2`	Binary	Use of sharing data electronically with suppliers or customers within the supply chain (e.g., via websites or apps, EDI systems, real-time sensors, or monitoring).
cc_basic	`cc_basic`	Binary	Use of cloud services for office, email, or accounting software.
-	`D2A` (for calculations)	Binary	Use of cloud computing for e-mail services, certified email (PEC).
-	`D2B` (for calculations)	Binary	Use of cloud computing for office software (e.g., word processing programs, spreadsheets) .
-	`D2C` (for calculations)	Binary	Use of cloud computing for finance and accounting software applications.
cc_management	`cc_management`	Binary	Use of cloud services for CRM or ERP software.
-	`D2D` (for calculations)	Binary	Use of ERP.
-	`D2E` (for calculations)	Binary	Use of CRM.
cc_security	`D2F`	Binary	Use of cloud for security applications (e.g., antivirus, network access control).
cc_DBhosting	`D2G`	Binary	Use of cloud services for hosting company databases.
cc_security	`D2H`	Binary	Use of cloud services for file storage.
cc_computation	`D2I`	Binary	Use of cloud services for computing capacity to run company software.
cc_platform	`D2J`	Binary	Use of cloud platform for application development and testing (e.g., APIs, reusable modules).

Variable Label	Variable Name	Type	Description
data analytics	`ANALYTICS`	Binary	Use of data analysis, either internal (C3) or outsourced (C5).
-	`C3` (for calculations)	Binary	Use of data analysis performed internally.
-	`C5` (for calculations)	Binary	Use of data analysis performed by third parties.
AI	`AIANY`	Binary	Use of at least one AI technology (between E1A-E1G).
-	`E1A` (for calculations)	Binary	Use of AI technologies for analyzing text documents (e.g., text mining).
-	`E1G` (for calculations)	Binary	Use of AI technologies for converting spoken language into a format readable by a computer (speech recognition).
-	`E1C` (for calculations)	Binary	AI technologies for generating written or spoken language (natural language generation, speech synthesis).
-	`E1D` (for calculations)	Binary	Use of AI technologies for identifying objects or people based on images or video (recognition, image processing).
-	`E1E` (for calculations)	Binary	Use of AI technologies for data analysis through machine learning (e.g., machine learning, deep learning, neural networks).
-	`E1F` (for calculations)	Binary	Use of AI technologies for automating workflows or supporting decision-making processes (e.g., process automation, robotic software using AI technologies to automate human activities).

Variable Label	Variable Name	Type	Description
-	E1G (for calculations)	Binary	Use of AI technologies enabling physical movement of machines through autonomous decisions based on observation of the surrounding environment (autonomous robots or drones, self-driving vehicles).
e-invoice	F1A	Binary	Use of electronic invoice to other industries (G01A) or to public administration (PA, G02B).
-	G01A	Binary	Use of electronic invoice to other industries or privates.
-	G01B	Binary	Use of electronic invoice to PA.

Tabella 3.2: List of active or principal variables, labels, and descriptions used in the MCA

Supplementary variables are variables that do not actively participate in constructing the factorial axes and representation space, but are passively projected onto it for interpretative purposes. These variables do not influence the determination of the analysis' main dimensions, but are used to enrich the interpretation of results: they are only passively projected to facilitate the interpretation of results. Their position in the factorial space is calculated ex-post, allowing observation of how they relate to active variables and patterns identified by the MCA, thus providing additional elements to understand underlying relationships in the data without altering the structure of the main analysis (Table 3.3). As it is possible to see from the table these are mostly structural variables: size, level of internet connection, productivity measured by three-clusters of value added per employee, intensity in the AI technologies' adoption, technological intensity related to the industry and governance type. A potential structural variable for inclusion could have been the geographic region of enterprise location. However, this variable was excluded due to the complexity arising from multi-location organizational structures, where legal units may belong to an enterprise with headquarters situated in one region while maintaining production facilities in another location. However, for giving information about this dimension, the geographic distribution demonstrated regional concentration in northwest Italy (32%), followed by northeast regions (25.3%), southern Italy and islands (20.31%), and central Italy (20.3%).

Variable Label	Variable Name	Type	Description
SIZE	`CL4`	Categorical	Firm size: 10–49 (1); 50–99 (2); 100–249 (3); 250+ employees (4).
cl_IUSE	`cl_IUSE2`	Categorical (1–3)	Cluster based on share of personnel using connected computers (obtained by dividing the number of employees connected to internet with the total number of employees): low (3), medium (2), high (1).
-	`A1`	Numerical	The number of employees using at least one device among computers, laptops, PDAs, tablets, iPads, smartphones, or other portable devices connected to the Internet (via fixed or mobile connection) to perform their work
-	`X2`	Numerical	Number of Employees
cl_vagg_add	`cl_vagg_add`	Categorical (1–3)	Cluster of enterprises based on value added per employee, derived from Frame 2022 data. Identification of high performance, low performance and normal performance enterprise.
-	`vagg_ad`	Numerical	Value added per employee.
intensity_AI	`intensity_AI`	Categorical	Intensity of AI adoption: no AI technologies (none), 1–2 (low), 3–4 (medium), 5+ (high).

Variable Label	Variable Name	Type	Description
-	`E1A` (for calculations)	Binary	Use of AI technologies for analyzing text documents (e.g., text mining).
-	`E1G` (for calculations)	Binary	Use of AI technologies for converting spoken language into a format readable by a computer (speech recognition).
-	`E1C` (for calculations)	Binary	AI technologies for generating written or spoken language (natural language generation, speech synthesis).
-	`E1D` (for calculations)	Binary	Use of AI technologies for identifying objects or people based on images or video (recognition, image processing).
-	`E1E` (for calculations)	Binary	Use of AI technologies for data analysis through machine learning (e.g., machine learning, deep learning, neural networks).
-	`E1F` (for calculations)	Binary	Use of AI technologies for automating workflows or supporting decision-making processes (e.g., process automation, robotic software using AI technologies to automate human activities).

Variable Label	Variable Name	Type	Description
-	`E1G` (for calculations)	Binary	Use of AI technologies enabling physical movement of machines through autonomous decisions based on observation of the surrounding environment (autonomous robots or drones, self-driving vehicles).
ERP_h	`C1A,D2D`	Binary	Indicates use of ERP software only locally (not cloud-based). Obtained by the difference between `C1A` and `D2D`.
-	`C1A`	Binary	Use of ERP software
-	`D2D`	Binary	Purchasing of cloud computing for ERP
CRM_h	`C1B,D2E`	Binary	Indicates use of CRM software only locally (not cloud-based). Obtained by the difference of `C1B` with `D2E`.
-	`C1B`	Binary	Use of CRM software
-	`D2E`	Binary	Purchasing of cloud computing for CRM

Variable Label	Variable Name	Type	Description
tech_intensity	`tech_intensity`	Categorical	Two-digit industries grouped by techno-economic classification (Eurostat), identifying: low-tech manufacturing (1), mid-tech manufacturing (2), high-tech manufacturing (3), other industries including energy, utilities, and construction (10), low-tech services (11),knowledge intensive service (12), high-tech services (13) [78].

Variable Label	Variable Name	Type	Description
-	`X1`	Categorical	5 digits NACE (ATECO) code . The first two ones were used following the Eurostat classification for grouping enterprise based on their technological intensity. Particularly, looking to the first two numbers, this is the logic of the classification: LTM (ISIC 10 to 18, 31, 32), MTM (ISIC 19, 20, 22 to 25, 27 to 30, 33), HTM (ISI 21, 26), other_ind (ISIC 35 to 39, 41 to 43), LTS (45 to 47, 49, 52, 53, 55, 56, 68, 77, 79, 81, 82, 95), KIS (ISIC 50, 51, 58, 69 to 71, 73 to 75, 78, 80), HTS (ISIC 59 to 63; 72). Financial services are not considered. [78] [79].
governance	`governance`	Categorical (1–4)	Governance type: Foreign multinational (1), Italian multinational (2), Companies belonging to domestic groups (3), Independent (4).

Tabella 3.3: List of supplementary variables, labels, and descriptions used in the analysis

Analysis of Variables Frequencies and Correlations

Before jumping into running the MCA, the frequency of each of these variables (both active and supplementary) and their correlations have been analyzed. Looking to the weighted frequency on STATA (`tab var. [fw=PESO]`) of these variables it is possible to do several observations. Examining the weighted frequency distributions, the 2023 data reveals that a substantial proportion of enterprises (39.38%) maintained low internet connectivity rates among their workforce. Electronic commerce penetration remained limited at 19%. Enterprise resource planning and customer relationship management systems showed moderate adoption rates, with approximately 50% of enterprises implementing CRM or ERP solutions, contrasting with lower business intelligence adoption at 17%. Notably, cloud-based deployment dominated among CRM and ERP in-house adopters, indicating a preference for remote hosting solutions over local infrastructure. Data-driven practices showed varied adoption patterns: data sharing initiatives remained nascent at 14%, while data analyses were more prevalent (26%). Artificial intelligence integration was minimal, with only 5.3% of enterprises utilizing AI technologies (Figure 3.2), and a mere 0.2% (of the total respondents) representing high-end AI users employing more than five AI technologies.

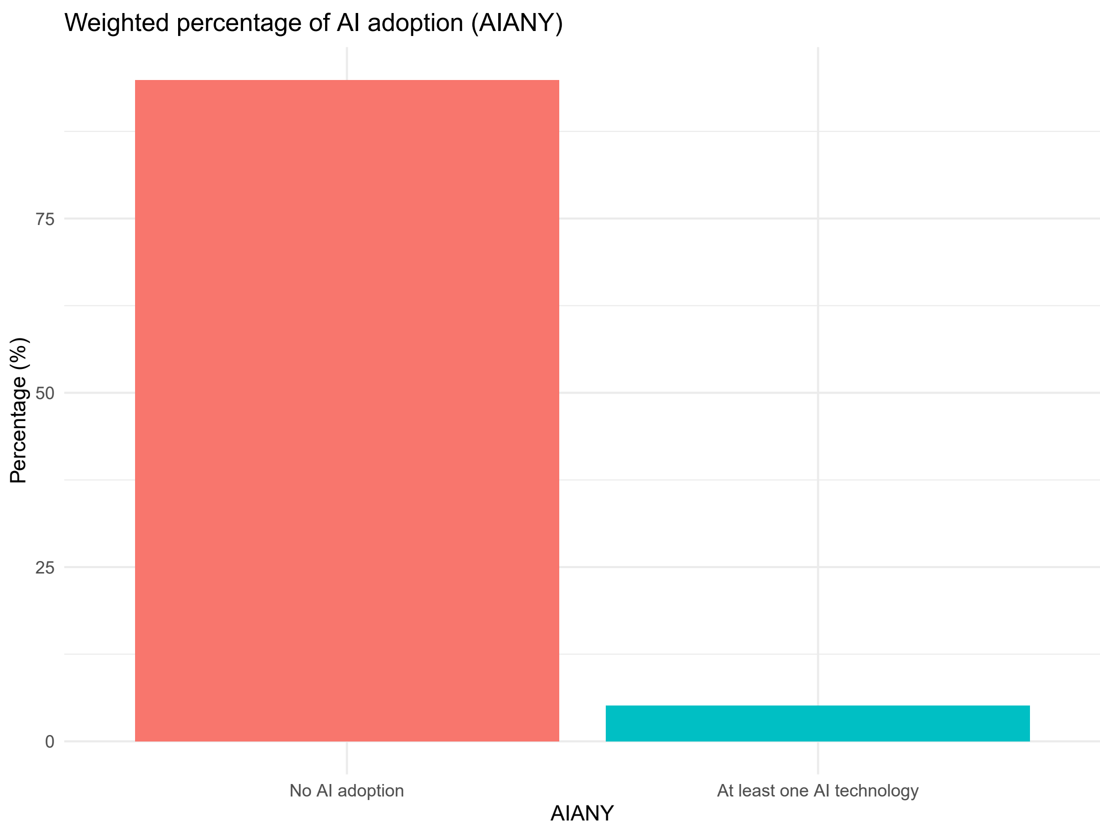

Figura 3.2: Percentage of AI and no-AI adopters (2023).

Figure 3.3 reports the distribution of AI adopters by number of technologies implemented, ranging from one to nine.
Moreover, among the 95% of enterprises not adopting AI, only 4.2% declared in 2023 an intention to invest in these technologies but did not proceed. This evidence suggests that the vast majority of non-adopters are currently not interested in AI. Such reluctance may reflect the need to achieve a preliminary level of digitalisation, given that AI is unlikely to represent a priority if enabling tools such as cloud services are not yet in place, or it may indicate a persistent skepticism regarding the potential benefits of AI.
Looking to the enterprises that did not invest at all even if they wanted to, an histogram with the reasons of this choice is shown in Figure 3.4. Approximately 50% of respondents indicated the lack of in-house expertise as the main obstacle, highlighting the central importance of digital training and skilled workforce recruitment for Italian firms. Around 45% identified high-costs as a barrier, suggesting that targeted economic incentives could play a crucial role in fostering adoption. Lastly, 23% reported ethical concerns as a limiting factor for AI diffusion, underlining that this is not the biggest investment barrier perceived. Figure 3.5 emphasises that obstacles to adoption are often cumulative: in 80% of the cases, enterprises report more than one reason simultaneously. From an interpretive perspective, studying the correlations between the variables that analyze the barriers to AI adoption (all below 0.02), it emerges that there is no systematic combination of obstacles, but rather each barrier acts fairly autonomously.

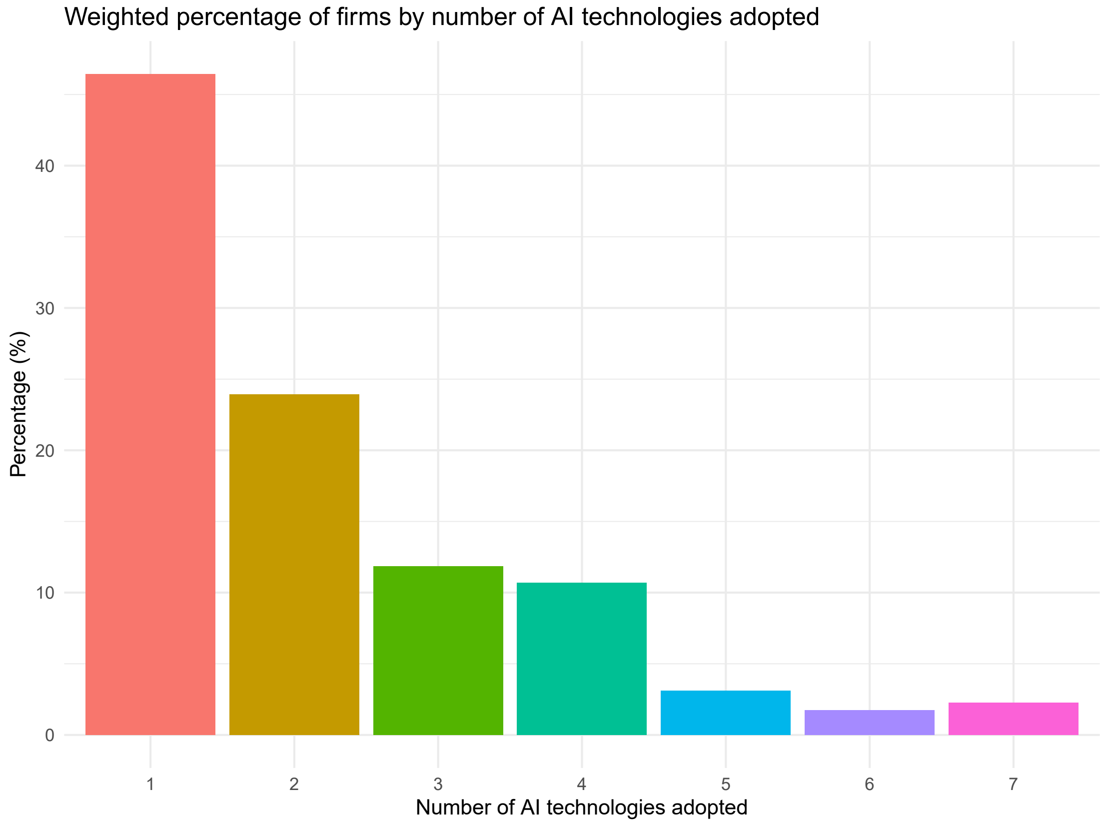

Figura 3.3: Percentage of AI users adopting 1, 2, 3, 4, 5 or 6 technologies (2023).

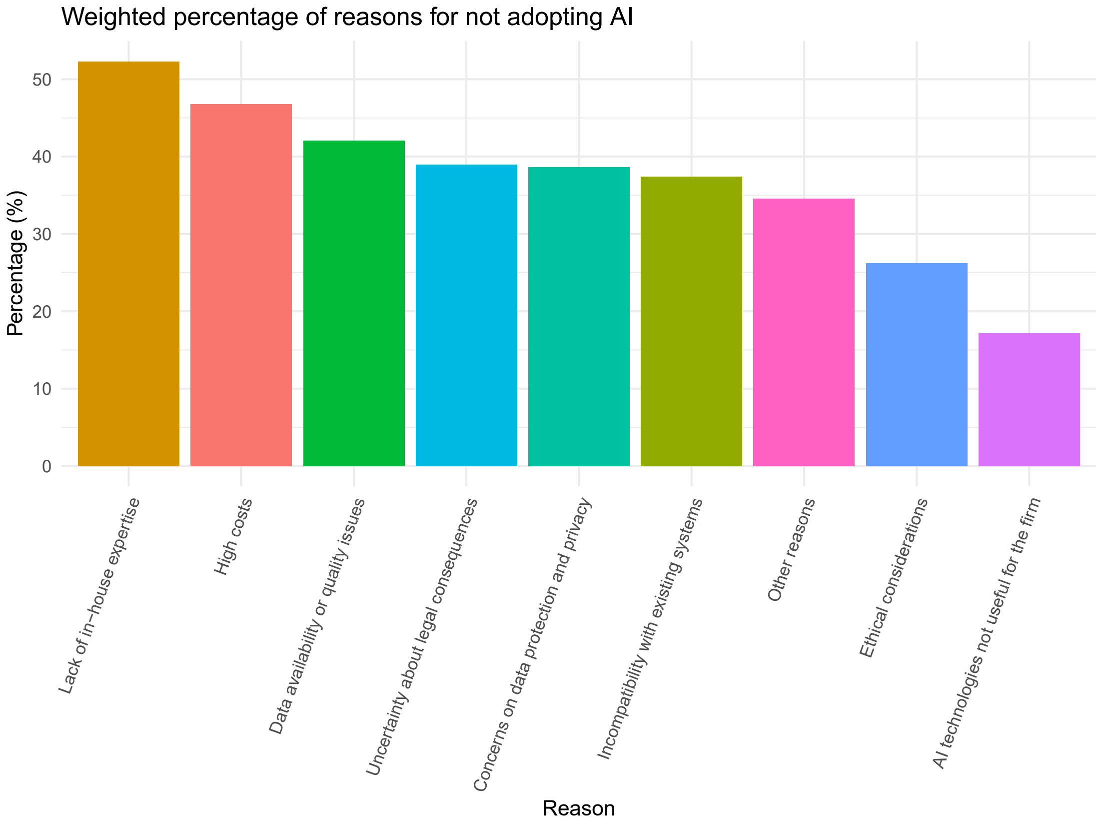

Figura 3.4: Reasons why enterprises did not adopt AI (2023).

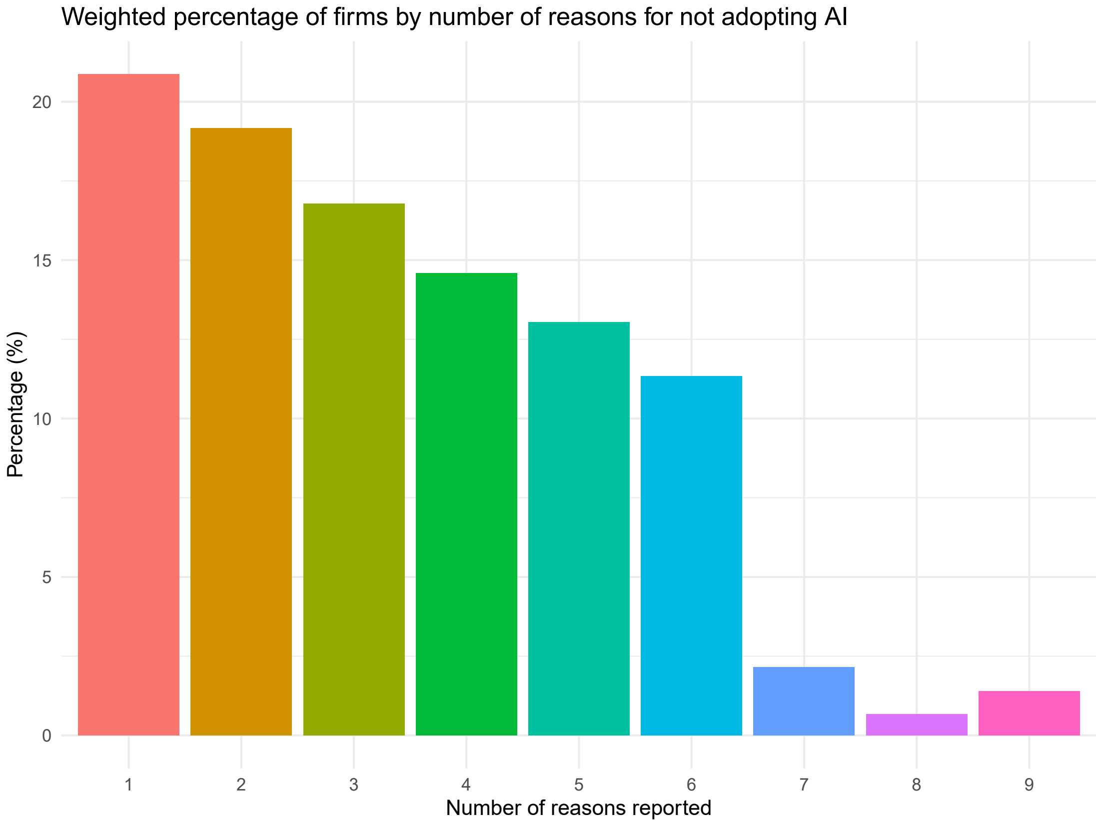

Figura 3.5: Distribution of enterprises by number of reported reasons against AI adoption (2023).

Moreover, structural characteristics revealed a concentrated size distribution, with 87.58% of enterprises employing between 10 and 49 personnel, while large enterprises exceeding 250 employees constituted less than 2% . Organizational governance structures were predominantly characterized by independent companies (73.85%), followed by subsidiaries of domestic corporate groups (18.28%). Sectoral analysis indicated significant representation of low-technology services (45%) and medium-technology manufacturing (19%). In particular, when considering enterprises with fewer than 50 employees, independent governance, and belonging to low-tech manufacturing or service sectors, the percentage is approximately 60% of the total population. This indicates that the majority of Italian enterprises are small in size and characterised by a limited level of technological advancement. After having observed the frequencies and having explored deeper the variable selected, the correlation between the active variables is studied, in order to gain some insights of possible technological adoption patterns the MCA could show afterwards. The correlation table of the variables selected has been run on STATA using the following code:

```
corr A4 ESALES C1A C1B C1C C2 cc_basic cc_management D2F
D2G D2H D2I D2J ANALYTICS AIANY F1A [fweight=PESO]
```

The correlations have been collected in Table 3.4. The outcomes reveal that the strongest relationships are concentrated among variables related to management software and analytical tools usage. BI is correlated with CRM, indicating that a company using Customer Relationship Management software is likely to adopt business intelligence. Furthermore, BI software's adoption shows a correlation with data analysis usage. Looking to the cloud related variables: the purchase of cloud applications for data management shows significant correlations (0.429) with the one of cloud for basic applications like electronic sheets or e-mail (cc_basic). Moreover, the highest correlation (around 0.76) is observed between some cloud ICT security applications and basic services, demonstrating that also with non-advanced applications, the enterprises consider important guarantee safety. Additionally, purchase of cloud file storage revealed a strong correlation with different clouds: basic cloud apps (0.598), database hosting (0.523) or ICT security cloud (0.599). These correlation patterns in cloud computing may also reflect strategic bundling practices by technology vendors, who often facilitate integrated adoption through convenience-based package offerings and complementary pricing structures, thereby influencing enterprise technology adoption decisions beyond pure operational requirements. Overall, the correlation pattern indicates clusters of technologies adopted together, while others, particularly the AI usage, remain isolated. This information is useful for interpreting data structure and for subsequent MCA.

	Web	ESALES	ERP	CRM	BI	data sharing	cc_basic	cc_mngt	cc_security	cc_DB hosting	cc_file space	cc_computation	cc_platform	ANALYTICS	AIANY	e-invoice
Web	1	0.211	0.248	0.188	0.176	0.093	0.167	0.176	0.177	0.202	0.154	0.114	0.124	0.215	0.091	0.108
ESALES		1	0.125	0.198	0.214	0.223	0.046	0.135	0.064	0.124	0.076	0.105	0.113	0.215	0.117	0.049
ERP			1	0.356	0.341	0.200	0.188	0.369	0.207	0.207	0.198	0.163	0.160	0.351	0.130	0.075
CRM				1	0.438*	0.200	0.152	0.429*	0.174	0.243	0.212	0.224	0.212	0.330	0.144	0.053
BI					1	0.250	0.158	0.288	0.170	0.212	0.183	0.216	0.232	0.418	0.185	0.003
data sharing						1	0.139	0.166	0.156	0.179	0.186	0.177	0.185	0.280	0.144	0.034
cc_basic							1	0.327	0.761*	0.435	0.598*	0.263	0.229	0.205	0.112	0.102
cc_mngt								1	0.341	0.390	0.330	0.301	0.294	0.310	0.147	0.043
cc_security									1	0.460*	0.599*	0.312	0.219	0.215	0.128	0.080
cc_DB hosting										1	0.523*	0.387	0.319	0.236	0.160	0.047
cc_file space											1	0.383	0.269	0.210	0.148	0.076
cc_computation												1	0.429*	0.208	0.191	0.028
cc_platform													1	0.219	0.263	0.026
ANALYTICS														1	0.169	0.065
AIANY															1	0.026
e-invoice																1

Tabella 3.4: Correlation matrix of MCA active variables (* indicates $r > 0.4$).

Looking to the supplementary variables, the only significative correlation outputs are:

- A negative correlation of -0.3753 between `governance` and `SIZE`. This reflects the fact that higher values of `governance` correspond to more independent firms (as opposed to multinational enterprises), while higher values of `SIZE` correspond to larger firms in terms of employees. The negative coefficient therefore indicates that independent firms are more likely to be smaller in size, whereas multinational firms are more likely to be larger.

- A positive correlation of +0.3158 between in-house adoption of ERP and CRM systems (`ERP_h` and `CRM_h`) and the absence of cloud computing services for these functionalities. This suggests that, for example, enterprises which are using ERP internally (no cloud use) are less likely to outsource the CRM one via cloud solutions: the in-house decision is transversally adopted.

How to run an MCA on STATA

The code for running a Multiple Correspondence Analysis on STATA is:

```
ssc install mca
mca A4 ESALES C1A C1B C1C C2
cc_basic cc_management D2F D2G
D2H D2I D2J ANALYTICS
AIANY F1A [fw=PESO], method(joint) supp(cl_IUSE cl_vagg_add
tech_intensity intensity_AI
ERP_h CRM_h governance)
```

This code installs and runs a MCA in STATA. It first installs the mca package since not already available, then analyzes a set of specified variables, treating some as active and others as supplementary (`supp()`). The `method(joint)` option indicates that categories are analyzed together, while `[fw=PESO]` applies the related weight (rounded) coefficient to each of the observations [76]. Supplementary variables are included for interpretation but, as mentioned before, do not influence the principal dimensions. In this case, the analysis has been computed on active variables related to the use of different technologies and then for interpret it some supplementary variables have been projected: the adoption of internet, the value added per employee, the technological intensity of the sector, the intensity of AI adoption, the use of local ERP and CRM are adopted for helping us in reading the results. For saving the factors resulting by the MCA (coordinates of each observations), the command on STATA is:

```
predict fac1 fac2
replace fac1 = -fac1
replace fac2 = - fac2
```

The inversion requirement stems from the need to enhance interpretability of the results and to align the factors signs with the plot done with the two dimensions resulting from the MCA. This coordinate transformation ensures that the factor loadings align with theoretically expected directions, facilitating meaningful interpretation. Saving the factors in this manner is essential for their subsequent use as independent variables in the linear regression models.

3.3.2 Exploratory Factor Analysis

The EFA implementation followed the methodological framework established by Field's (2012) [77], which provides comprehensive procedures for various analytical techniques. In statistics and psychometrics, factor analysis constitutes a statistical technique designed to uncover underlying latent constructs (in psychometrics) or factors and dimensions (in statistics) that cannot be directly measured, within a set of directly observable variables (sometimes referred to as indicator variables or instrumental variables) that are theoretically related to these latent constructs. These underlying dimensions are characterized by internal theoretical coherence, meaning that factors must not only represent statistical associations among variables but must also be substantively interpretable from a scientific and rational perspective. Factor analysis can be conducted for both exploratory purposes (EFA) and confirmatory purposes (Confirmatory Factor Analysis - CFA). Particularly, in the exploratory approach, factors are derived empirically from the data patterns. The EFA performed has been conducted for each of the 6 digital-adoption-related sections in the survey:

1. Web, internet connection, and social media — hereinafter referred to as **Section 1** (corresponding to Section B of the survey)

2. E-sales — hereinafter referred to as **Section 2** (corresponding to Section C of the survey)

3. Data use, analysis, and sharing — hereinafter referred to as **Section 3** (corresponding to Section D of the survey)

4. Cloud computing — hereinafter referred to as **Section 4** (corresponding to Section E of the survey)

5. Artificial intelligence — hereinafter referred to as **Section 5** (corresponding to Section E of the survey)

6. E-invoices — hereinafter referred to as **Section 6** (corresponding to Section G of the survey)

All the variables considered are binary 0-1 (dummies).

Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure

Before performing the EFA on RStudio for each of these sections, two tests have been done to verify whether the data selected for each section are suitable for factor analysis: Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure [77].
Bartlett's Test of Sphericity evaluates whether the correlation matrix significantly differs from an identity matrix (where correlations between variables equal zero) [77]. The null hypothesis states that variables are uncorrelated in the population. A significant result ($p < 0.05$) indicates sufficient correlations exist among variables to justify factor analysis, while a non-significant result ($p > 0.05$) suggests factor analysis is inappropriate.
On the other side, the Kaiser-Meyer-Olkin (KMO) measure assesses sampling adequacy by comparing the magnitude of observed correlation coefficients to partial correlation coefficients [77]. To better explain what this menas, partial correlations measure the relationship between two variables while controlling for the influence of all other variables in the dataset. If the observed correlations are substantially larger than the partial correlations, it indicates that the variables share common underlying factors. This condition supports the suitability of the data for Exploratory Factor Analysis (EFA). KMO values range from 0 to 1, where higher values indicate greater suitability for factor analysis. Generally the used thresholds are: KMO ≥ 0.90 (marvelous), 0.80–0.89 (meritorious), 0.70–0.79 (middling), 0.60–0.69 (acceptable), and < 0.60 (unacceptable) [77]. Hence, values below 0.50 suggest factor analysis should not be conducted [77] and this is the logic applied in this study.
Both tests serve as prerequisite diagnostics: Bartlett's test confirms that correlations exist among variables, while KMO ensures these correlations are strong enough. Together, they provide statistical justification for proceeding with exploratory factor analysis.
With respect to the dataset analyzed, Bartlett's test consistently rejected the null hypothesis (p-value < 0.05), confirming the presence of sufficient correlations among the variables. By contrast, the results of the KMO test led to two main consequences: (i) the elimination of specific variables within certain sections (i.e., the first and third sections), and (ii) the removal of entire sections (i.e., the second and sixth), where even after selecting subsets of variables, the adequacy measure remained unsatisfactory. In the next paragraphs, the initial variables selected for each section (using the Eurostat nomenclature, beside for G01A and G01B) and the ones preserved after the Barlett and KMO's test have been shown and explained. Furthermore, a summary of all these process, can be found in Table 3.5 : from 59 variables selected before the KMO, 43 were then used for performing the different EFAs.
Below it is shown the RStudio code for doing these two tests already explained,

considering also the weighting process for taking into account the representativeness of each observation.

```r
# Packages to install
install.packages("wCorr")
install.packages("psych")
install.packages("dplyr")

# Packages loading
library(wCorr)
library(psych)
library(dplyr)

# Variable selection (Here the example for Section 4)
dati4 <- dati2023 %>% select(D2A, D2B, D2C, D2E, D2D, D2F, D2G, D2H, D2I, PESO)
vars <- dati4 %>% select(-PESO)  # excluding PESO
pesi <- dati4$PESO               # vector with weights

# Weighted matrix creation
n_vars <- ncol(vars)
weighted_cor_matrix <- matrix(NA, n_vars, n_vars)
colnames(weighted_cor_matrix) <- colnames(vars)
rownames(weighted_cor_matrix) <- colnames(vars)
for (i in 1:n_vars) {
  for (j in i:n_vars) {
    weighted_cor_matrix[i, j] <- weightedCorr(vars[[i]], vars[[j]], weights = pesi, method = "Pearson")
    weighted_cor_matrix[j, i] <- weighted_cor_matrix[i, j] # simmetria
  }
}

# Visualization of the weighted matrix
weighted_cor_matrix

# Bartlett Test
cortest.bartlett(weighted_cor_matrix, n = nrow(dati4))

# KMO Measure
KMO(weighted_cor_matrix)
```

Listing 3.1: Variable selection through Barlett test and KMO measure

KMO Variable Selection – Section 1: Connection and Internet Use

The 13 variables related to this section are:

- A2 - At least one Internet Fix Connection (0-1)

- A5A - Description of goods and services offered, information about prices
- A5B - Possibility to place orders or reservations online (e.g., online shopping cart)
- A5C - Online order tracking
- A5D - Ability to customize website content for returning visitors
- A5E - Ability to personalize or design goods and services for website visitors
- A5F - A chat service for customer support (provided by a chatbot, virtual agent, or human responding to customers)
- A5G - Posting of job vacancies or possibility to submit job applications online
- A5H - Website content available in at least two languages (consider a multilingual website, e.g., Italian and English, either within a single domain like ".com" or multiple domains in different languages, e.g., ".it" and ".uk")
- A6 - Mobile App for clients (loyalty programs, e-commerce, customer services)
- A7A - Social networks (e.g., Facebook, LinkedIn, MySpace, Google+, Xing, Viadeo, Yammer)
- A7B - Company blog or microblog (e.g., Twitter, Tumblr)
- A7C - Websites or apps for sharing multimedia content (e.g., YouTube, Instagram, Spotify, Pinterest, Flickr, SlideShare, Snapchat, TikTok)

It can be noticeds how the A5 variables are related to website usages whie A7 ones are on the social media adoptions. After the two tests, all of them have been preserved except for A7A, A7B, A7C all related to social media usage. Hence, the total amount of variables kept is 10. This choice is related to the fact that, removing them, the KMO produced a result of 0.74 against the 0.5 obtained with them.

KMO Variable Selection – Section 2: Sales through Computer Networks

Here sales through different digital tools and the ones performed through EDI are investigated. Particularly, the initial 4 variables selected have been:

- B1 - E-sales trough its applications or website or through intermediaries
- B1A - E-sales through its applications or website
- B1B - E-sales through intermediaries

- B5 - Sales through EDI

After having selected the proper variables for doing an EFA a correlation matrix is computed. For the non-satisfactory results of the KMO, even with subset of these variables, all of them have been removed and an EFA has not been performed for this section.

KMO Variable Selection – Section 3: Data Use, Share and Analysis

In this section of the survey, questions related to usages of software, data integration, analysis and sharing are made. The 13 associated variables are the following:

- C1A - ERP use
- C1B - CRM use
- C1C - BI use
- D03 - data used by software(s) are memorized in just one relational database
- C2 - data sharing along the SC (website, EDI, sensors etc..)
- C4A - Analysis of data from transactions such as detailed information on sales and payments (e.g., from the Enterprise Resource Planning – ERP system, or from the company's online store)
- C4B - Analysis of customer data such as purchase information, location, preferences, reviews, searches (e.g., from the Customer Relationship Management – CRM system, or from the company's website)
- C4C - Analysis of data from social media, including the company's own social media profiles (e.g., personal information, comments, videos, audio, images)
- C4D - Analysis of web data (e.g., search engine trends, data from web scraping, i.e., software programs for extracting data from websites)
- C4E - Analysis of location data from the use of mobile devices or vehicles (e.g., mobile devices using cellular networks, wireless connections, or GPS)
- C4F - Analysis of data from smart devices or sensors (e.g., sensors installed in machinery, production sensors, smart meters, RFID tags)
- C4G - Analysis of open data from government authorities (e.g., public registries, weather conditions, topographic conditions, transport data, housing or building data)

- C4H - Analysis of satellite data (e.g., satellite images, navigation or positioning signals)

Three main recurring topic-blocks are evident in all these variables: all the C1 are related to the use of software (i.e., CRM, ERP, BI), C2 to the data sharing and C4 to data analysis' usages. All of the 13 variables have been kept since the KMO produced 0.83 as score.

KMO Variable Selection – Section 4: Cloud Computing

There the 11 initial variables are:

- D1 - purchase of cloud computing services
- D2A - Email services, certified email (PEC)
- D2B - Office software (e.g., word processing programs, spreadsheets)
- D2C - Finance and accounting software applications
- D2D - ERP software applications
- D2E - CRM software applications
- D2F - Security software applications (e.g., antivirus programs, network access control)
- D2G - Hosting of company databases
- D2H - File storage
- D2I - Computing capacity to run the company's software
- D2J - IT platform that provides an environment for developing, testing, and deploying applications (e.g., reusable software modules, application programming interfaces – APIs)

For obtaining a good KMO's score (0.89) D2J has been excluded from our analysis since not showing a good correlation with all the other 10 variables.

KMO Variable Selection – Section 5: Artificial Intelligence

14 original variables, related to AI technologies and their contexts of adoption, have been selected and then all kept in the EFA (KMO=0.88):

- E1A – AI technologies for analyzing text documents (e.g., text mining)

- E1B – AI technologies for converting spoken language into a format readable by a computer (speech recognition)
- E1C – AI technologies for generating written or spoken language (natural language generation, speech synthesis)
- E1D – AI technologies for identifying objects or people based on images or videos (image recognition and processing)
- E1E – AI technologies for data analysis through machine learning (e.g., machine learning, deep learning, neural networks)
- E1F – AI technologies for automating workflows or supporting decision-making processes (e.g., process automation, software robots using AI to automate human tasks)
- E1G – AI technologies enabling physical movement of machines through autonomous decisions based on observation of the surrounding environment (autonomous robots or drones, self-driving vehicles)
- E2A - Use of AI in marketing or sales, e.g., chatbots based on natural language processing for customer support or profiling; price optimization, personalized marketing offers, market analysis using machine learning; autonomous robots for order processing
- E2B - Use of AI in production processes of goods or services, e.g., predictive maintenance or process optimization based on machine learning; tools to classify products or detect defects using computer vision; autonomous drones for monitoring, security, and production inspection; assembly tasks performed by autonomous robots
- E2C - Use of AI for organizing or managing business administration processes, e.g., using virtual assistants based on machine learning or natural language processing for document drafting; analyzing data and making strategic decisions using machine learning (e.g., risk assessment); planning or forecasting with machine learning; human resource management using machine learning or natural language processing (e.g., candidate pre-screening, employee profiling, or performance analysis)
- E2D - Use of AI in logistics, e.g., solutions for picking items from shelves and packing parcels using autonomous robots; tracking, distribution, or sorting; route optimization based on machine learning
- E2E - Use of AI for ICT security, e.g., facial recognition using computer vision for user authentication; detection and prevention of cyberattacks using machine learning

- E2F - Use of AI for accounting, control, or financial management, e.g., machine learning to analyze data supporting financial decisions; invoice processing using machine learning; machine learning or natural language processing for accounting documents

- E2G - Use of AI for Research & Development (R&D) or innovation (excluding AI research), e.g., machine learning to analyze data for research, or to develop a new or significantly improved product/service

KMO Variable Selection – Section 6: Business Invoicing

Originally, 4 variables have been selected:

- G01A : Electronic invoices sent to other businesses or private individuals in a standard format suitable for automatic processing (e.g., XML, EDI, UBL)

- G01B : Electronic invoices sent to the Public Administration in a standard format suitable for automatic processing (XML, FatturaPA)

- F1B : Invoices sent in an electronic format not suitable for automatic processing (e.g., email or email attachments in PDF, TIF, JPEG, or other formats)

- F1C: Paper invoices

All of them produce a negative result (below 0.5) in the KMO: for the same reason of section 2, section 5 has also been delated by the EFA. Summing up below, Table 3.5 showns, for each section, which variables have been kept for performing the EFA after having run a Barlett and KMO test.

Determination of the Factors' Optimal Number

After having selected the subset of 43 variables on which performing the EFA, the number of factor has to be determined in order to insert it in the EFA's code chunk. How can the factors' number be determined in a systematic way, rather than leaving the decision to chance? Usually before running an EFA, it is asked to R of determining the optimal number of factors. This can be accomplished through several criteria, such as: the Kaiser criterion, the Cattell criterion and the Parallel analysis. The latter one (`fa.parallel`) was adopted in all the EFAs performed. This method compares a scree plot of eigenvalues derived from the actual data with those obtained from randomly generated data. The number of factors to retain is identified at the point where the eigenvalues from the real data remain larger than those from the random data. It is important to note that eigenvalues correspond to the proportion of variability "explained" by each factor and assume descending values from the first factor to the last. `fa.parallel` recommendation

Section	Variables Selected for KMO	KMO Score	Variables Kept for EFA	Final KMO Score
1	A5A:A5H, A2, A6, A7A:A7C	<0.6	A5A:A5H, A2, A6	0.74
2	B1A, B1B, B5	<0.6	-	-
3	C4A:C4H, D03, C1A:C1C	0.83	C4A:C4H, D03, C1A:C1C	0.83
4	D2A:D2J	<0.6	D2A:D2I	0.89
5	E2A:E2G, E1A:E1G	0.88	E2A:E2G, E1A:E1G	0.88
6	F1B, F1C, G01A, G01B	<0.6	-	-

Tabella 3.5: Summary of variables selected for the EFA and KMO scores

serves as guidance rather than prescriptive rules, with the final decision contingent upon the interpretability of the resulting factors. In fact, EFA constitutes an inherently iterative process wherein the primary objective is achieving exploratory and interpretable factorial structures that provide substantive insights into the underlying data patterns.

Techniques Adopted in the EFA on RStudio

Once the number of factors for running the EFA has been defined, the two R packages necessary are installed (`library(psych)` and `library(dplyr)`) and a correlation matrix is computed. This matrix can use various computational techniques. For our data the tetrachoric correlation matrix has been selected since it provides a more realistic estimate of associations between binary variables compared to simple Pearson correlation, ensuring that the extracted factors are valid and interpretable. On RStudio this is obtained with:

```
{r}
tetra\_corr1 <- tetrachoric(dati\_vars1)\$rho
```

Among the most frequently employed methods for performing an EFA are principal axis factoring, principal component factoring, and maximum likelihood estimation: the selection of the most appropriate method depends on statistical, methodological, and interpretative considerations. In the R code Minimum Residuals (`minres`) has been adopted. Since factorial solutions are infinite and mathematically equivalent, the results can be subjected to rotation using various methods. Orthogonal rotation methods preserve factor independence (with Varimax being the most commonly used), while oblique rotation methods relax the independence constraint to enhance interpretability (notably Promax). In exploratory factor analysis (EFA), the choice of rotation significantly influences the structure of the obtained factors. In this study, two are the rotation approaches adopted: `varimax` (orthogonal, more independent factors) and `oblimin` (oblique, more correlated and interpretable factors). Respectivily, on RStudio the chunks run for obtaining the factors are:

```
efa_result1 <- fa(tetra_corr1, nfactors = 3, rotate = "varimax", fm =
"minres", scores = "regression")
efa_result1 <- fa(tetra_corr1, nfactors = 3, rotate = "oblimin", fm =
"minres", scores = "regression")
```

When scores = "regression" is set (see the code chunk above), the function computes factor scores for each observation using the regression method, which predicts factor scores as a linear combination of the observed variables, weighted by the factor loadings and unique variances.
Note that the number of factors is inserted through the parameter `nfactors` within the above code. It is possible that after having seen the results, the number of factors is changed (even more than once), always for facilitate the interpretability.

An Example of the Procedure Followed

In the paragraph below, a brief explanation of what it has been done for each section, using as example the R code run for the fourth section (Cloud Computing) in the`oblimin` (oblique) rotation case. Binary survey data (`D2A-D2I`) from the survey third section have been selected by expanding it according to the sampling weights (`PESO`), computing a tetrachoric correlation matrix, and running a parallel analysis to determine the optimal number of latent factors to retain in the Exploratory Factor Analysis. The R code has been reported below.

```
{r}
library(psych)
library(dplyr)

# Select variables + PESO
dati_subset4 <- dati2023 %>%
  select(D2A:D2I, PESO)

# Round PESO to integers if needed
dati_subset4$PESO <- round(dati_subset4$PESO)

# Expand dataset according to PESO
dati_expanded4 <- dati_subset4[rep(1:nrow(dati_subset4),
    dati_subset4$PESO), ]

# Remove PESO before factor analysis
dati_vars4 <- dati_expanded4 %>% select(-PESO)

# Compute tetrachoric correlation matrix
tetra_corr4 <- tetrachoric(dati_vars4)$rho

# Parallel analysis plot
png("fa_parallel4.png", width = 1200, height = 800, res = 150)
fa.parallel(tetra_corr4, n.obs = nrow(dati_vars4), fm = "minres",
    fa = "fa")
dev.off()
```

Listing 3.2: Data Preparation and Parallel Analysis for Section 4

The analysis suggested four factors (Figure 3.6). However, for interpretability and practical considerations, it was deemed more appropriate to retain only two factors in this case. In EFA, the statistically "optimal" number of factors does not always align with the most meaningful or interpretable solution, and reducing the number of factors can simplify the model, highlight the most relevant patterns, and facilitate clearer interpretation of the latent constructs. A two-factor EFA is then performed on the weighted data, extracting factor scores (MR7 and MR8), which are subsequently collapsed back to the original firm-level observations. The final outcome is the initial dataset enriched with two new latent variables (MR7 and MR8) that capture patterns of technology adoption. In the following page, the R code for doing that has been shown.

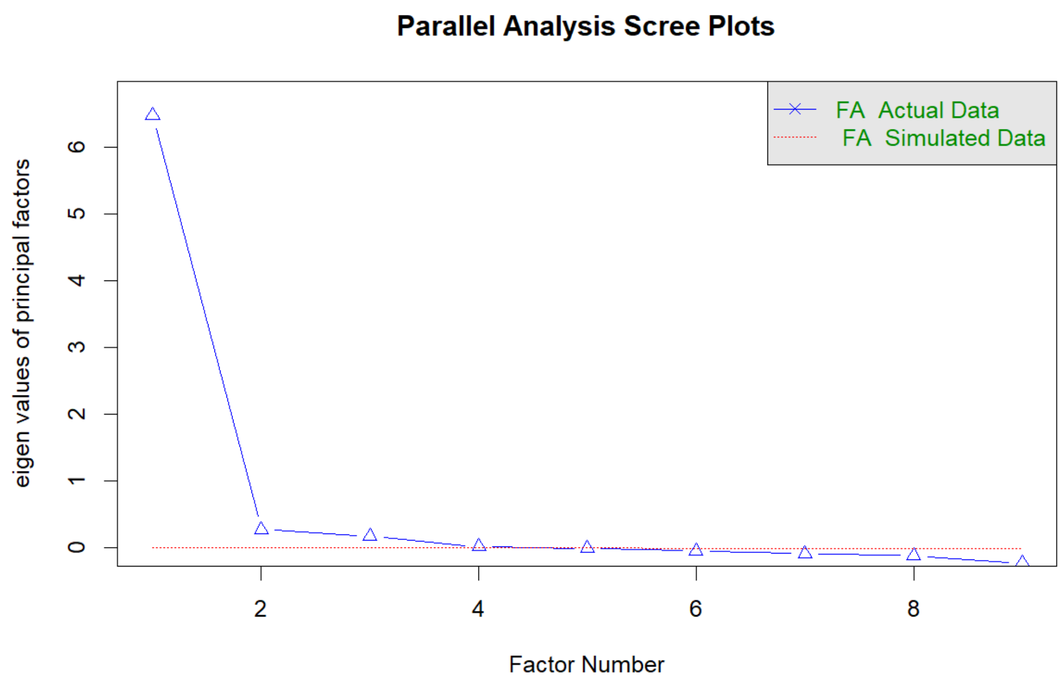

Figura 3.6: Results from the Parallel Analysis performed on the variables selected from Section 4

```
# Run EFA
efa_result4 <- fa(tetra_corr4,
                  nfactors = 2,
                  rotate = "oblimin",
                  fm = "minres",
                  scores = "regression")  # compute factor scores

# Print factor loadings
print(efa_result4$loadings, cutoff = 0.3)

# Save factor diagram
png("fa_plot4.png", width = 1200, height = 800, res = 150)
fa.diagram(efa_result4, cut = 0.3)
dev.off()

# Compute factor scores
factor_scores_expanded <- factor.scores(dati_vars4, efa_result4,
   method = "regression")$scores
factor_scores_expanded_df <- as.data.frame(factor_scores_expanded)
```

```
colnames(factor_scores_expanded_df) <- c("MR7", "MR8")

# Add an ID for original rows
dati_expanded4$orig_id <- rep(1:nrow(dati_subset4), dati_subset4$PESO)

# Combine factor scores with expanded data
dati_expanded4 <- cbind(dati_expanded4, factor_scores_expanded_df)

# Aggregate factor scores by original ID (weighted mean if needed)
factor_scores_original <- dati_expanded4 %>%
  group_by(orig_id) %>%
  summarise(
    MR7 = mean(MR7),
    MR8 = mean(MR8)
  ) %>%
  ungroup()
# Combine with original dataset
dati_with_factors4 <- cbind(dati_subset4 %>% select(-PESO), factor_scores_original[, -1])
dati_with_factors_KMO4 <- cbind(dati_subset4 %>% select(-PESO), factor_scores_original[, -1])
```

Listing 3.3: EFA and Factor Loads computation for Section 4

This weight-based expanding and collapsing procedure addresses a technical limitation of the `fa()` function, which does not properly incorporate sampling weights: by expanding the dataset, weights are effectively translated into frequencies, ensuring that the factors estimated are representative of the target population. Collapsing the data afterward allows to return to the original firm-level structure while preserving the weighted influence of each observation in the factor estimation. The same procedure has been applied for all the other sections.
In summary, the analysis began with 59 variables which, following the application of Bartlett's test and the KMO measure within each of the six survey sections, were reduced to 43. This selection process implied that section 2 (e-sales and EDI) and 6 (business invoicing) have been completely excluded from the analysis (insufficient KMO). For each section, the optimal number of factors was then determined, and Exploratory Factor Analysis (EFA) was conducted using both `varimax` and `oblimin` rotations. In several cases, the number of retained factors was further refined based on interpretability considerations. The results of the parallel analyses, together with the factor loadings and cumulative explained variance from the EFAs, are presented in the Results and Discussion section, where each factor is interpreted in detail.

3.4 OLS Regression Models

The dimensionality reduction performed in two different ways (EFA and MCA) wants to achieve two main objectives: look how technologies are grouped together, identifying behavioral patterns in the technology adoption and trying to use these synthetic factors as co-variates (also called independent variables or x's) for building two weighted Ordinary Least Squares (OLS) regression models. This is a linear regression model estimated through the method of least squares. The dependent variable y is expressed as a linear function of one or more explanatory variables $x_1, x_2, \ldots, x_k$ according to the equation

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i,$$

where ε_i denotes the error term. The OLS method estimates the coefficients β_k by minimizing the sum of squared residuals, that is, the squared differences between the observed values of y and those predicted by the model. In this way, OLS identifies the linear relationship that best fits the data in terms of mean squared error. In this study, the y is the value added per employee (vagg_add) of each enterprise and as x's the following variable:

- MCA factors for the first OLS regression model.
- EFA factors (obtained with the `varimax` rotation) for the second OLS regression model. The use of the output dimensions obtained with `varimax` is for avoiding multi-collinearity issues.

There is a limitation in this approach acknowledged ex-ante that is related to a temporal mismatch between the dependent (y) and independent (x's) variables. Hence, for observation i, the model reads:

$$y_{i,t} = \beta_0 + \beta_1 x_{i1,t+1} + \cdots + \beta_k x_{ik,t+1} + \varepsilon_{i,t}.$$

In fact, the y is collected in 2022 (value added registered by the enterprise on 31/12/2022), while the x's are collected between May and July 2023. Since the covariates are temporally later than the outcome, a causality (cause–effect) relationship cannot be stated. On the other hand, it should also be noted that reports in the literature often do not take into account the time difference between y and x's, implicitly assuming that in the four-month period between the collection of the economical data and the ICT survey administration the adopted technologies remain the same (or alternatevely that the value added per employee remain the same). This non-technological-variation assumption is not completely unsupported: in fact, when talking about technology, it must be considered that it results from collective efforts that are not completed overnight. For example, in a technical report

published by Da Empoli et al. (2025) , the relationship between AI adoption (2024) and revenue (2023) is studied through a OLS regression model, and it is referred to as an “impact” [43]. The digital adoption data used in the aforementioned study are those of ICT-2024 and the economical performance measure is the turnover collected at the end of 2023 (always findable in ICT-2024 dataset). In my study, it has been considered statistically unsupported to use the word “impact” (since cause–effect cannot be inferred due to temporal mismatching), but the cited paper represents an example supporting the decision to use linear regression as a method for cases like the one presented in this thesis.
Moreover another clarification should be done in order to explain the choice of the code chunks presented in the next chapter: the regressions done are weighted OLS regression since the each observations has a different relevance. For doing so, `weights = PESO` in the lm() function has been used. In this way, the model estimates a weighted regression, whereby greater “importance” is assigned to observations with higher weights, as if they were replicated multiple times. This constitutes the appropriate methodology for data sampled with Istat weights. The approach followed in Chapter 4, after showing some descriptive analysis related to value added per employee (y), for each of the two regressions consists of:

- Analyzing x's distribution and range, looking to possible problematic correlations between the x's and the y (that can cause multicollinearity).
- Exploring the factors and value added per employee in relation with structural variables (i.e, size, governance and technology intensity sector)
- Showing and discussing the OLS model outcomes, with a particular focus on the statistically significant associations.

Chapter 4

Results and Discussion

In this section, the main results of our analysis are discussed. In particular, the outcomes of the MCA are presented, together with a graphical representation aimed at better visualizing patterns and associations between technologies. Subsequently, the resulting EFA factors, both the 11 obtained with `oblimin` rotation and the 14 extracted using `varimax`, are introduced and explained. Afterwards, the results of the two OLS models are reported and discussed.

4.1 MCA Results

The MCA objective is to identify relationships and associations between categories of qualitative variables, projecting them into a two-dimensional space that synthesises their variability. Hence, it is possible to effectevely interpret the outcomes plotting the two dimensional coordinates for each variable (active and supplementary) on a bidimensional graph, together with looking at each variable's contribution in building a single dimensions. The results of the Multiple Correspondence Analysis conducted on 15,174 observations, representing a 2023 picture of digitalisation within Italian enterprises, reveal clear adoption patterns and significant structural characteristics.

To better visualize the MCA results, it is important to note that STATA lacks of an automated function for graph involving all the variables object of the analysis. For this reason, all the obtained variables' coordinates for dimensions one (*Dim1*) and two (*Dim2*), were exported to Excel to construct an easy-to-interpret scatter plot. For doing so, the x-axis coordinates were multiplied by -1 to enhance interpretability, positioning technologically advanced enterprises on the right-hand side of the plot. Hence, with this transformation, the values of the first principal component (*Dim1*) were inverted to facilitate more intuitive interpretation. Moreover, a reference point indicating no technology adoption (`no-tech (AVG)`) was incorporated into the visualization, calculated as the average coordinates obtained for the variables equal to 0 (i.e, no adoption of that technology). This baseline serves as a comparative benchmark within the analytical framework and as an interpretation tool. Figure 4.1 represents what it has been obtained afterwards. However, before going on with the description and interpretation of the scatter plot, it is important to look at the information among inertia and variability. As shown in Table 4.1, the particularly low total inertia (0.066338) is indicating limited overall variability and, consequently, a relative homogeneity in the technological behaviour of the enterprises considered. Furthermore, the first dimension (*dim1*) accounts for 88.75% of the total variability, while the second one is contributing only 8.84% to the total variance, completes an interpretative framework covering 97.60% of the overall variability, confirming the adequacy of the two-dimensional representation. Hence, in Figure 4.1, the horizontal axis (*Dim1*) captures the main dimension of variation among enterprises, while the vertical axis (*Dim2*) represents a second orthogonal dimension that explains minor additional differences.

Number of observations:	15,174
Method:	Joint (JCA)
Number of axes:	2
Total inertia:	0.066338

Dimension	Inertia	Percent	Cumulative Percent
Dim 1	0.0588779	88.75	88.75
Dim 2	0.0058662	8.84	97.60
Total	0.066338	100.00	100.00

Tabella 4.1: MCA overall results - Inertia and variability

For distinguishing between the active and supplementary variables in Figure 4.1, the first ones, that were used for computing the MCA dimensions (e.g., ERP, CRM, BI, AI, different cloud services purchasements) are represented with dark-blue dots and labeled under "Technologies". The supplementary variables are associated to other colors. Particularly:

- The light-blue spots represent the cluster of workforce connected to internet (`IUSE`), distinguishing between a low, average and high-level use.
- The orange dots are showing the different `SIZE` groups. The related dashed line wants to underline the relationship between enterprise size and digital choices. It is also relevant to note that in the survey only 1.77% of enterprises have more than 250 employees. This explains why its associated dot is more isolated by the rest.

- Red labels indicate enterprise governance type (`governance`): for example, "ITA multinational" and "non-ITA multinational" suggest that adoption of certain technologies is linked to group affiliation (domestic or foreign). Generally, more advanced technologies are associated with more international enterprises. Furthermore governance is related to size (dots proximity), showing that larger firms are more likely to be part of a group.

- Green labels correspond to activity categories (`tech_intensity` grouped by their technological adoption. Their role is showing how technologies are used by these different industrial clusters.

- In purple there are local software, indicating an in-house usage modality for ERP (`ERP_h`) and CRM (`CRM_h`) are the ones obtaining highest value for the second dimension.

- In lighter-green there are the AI types (`AI_intensity`), all associated with the bigger-sized enterprises but also more isolated (particularly for intermediary and advanced AI), showing that few enterprises are adopting them (very rare phenomenon). This is in line with the AI-related statistic shown before in Chapter 3 where only 5% of the enterprises declared of being adopters of this technology in 2023. Advanced AI is very isolated and having the highest x-value, underlining the particularly sporadic nature of this scenario. This supplementary categorical variable (`AI_intensity`) is doing a distinction between the number of AI technologies adopted (elementary, intermediary and advanced) while the active variable (`AIANY`, referred in the graph as AI) is a binary 0-1 that just looks if at least one technology is adopted. The fact that the latter is close by the elementary-AI category is absolutely coherent with what shown inChapter 3 that is around 75% of the AI adopters use between 1 and 2 technologies.

- Value added per employee (`cl_vagg_add`) are represented through dark-green dots. Star-performance enterprises (with an exceptional value added per employee) can be found close to AI adoption, bigger size and multinational governance, while an average performance is closer to the origin, demonstrating that most of the observations are part of this group.

Focusing on the relationships (spatial proximity) between active and supplementary variables, some adoption patterns can be observed:

- Web-page adoption and electronic invoicing are close by, together with HTM, LTM, LTS, independent governance-type and small (10-49 employees) size. This suggests a strong associative relationship between all these aspects.

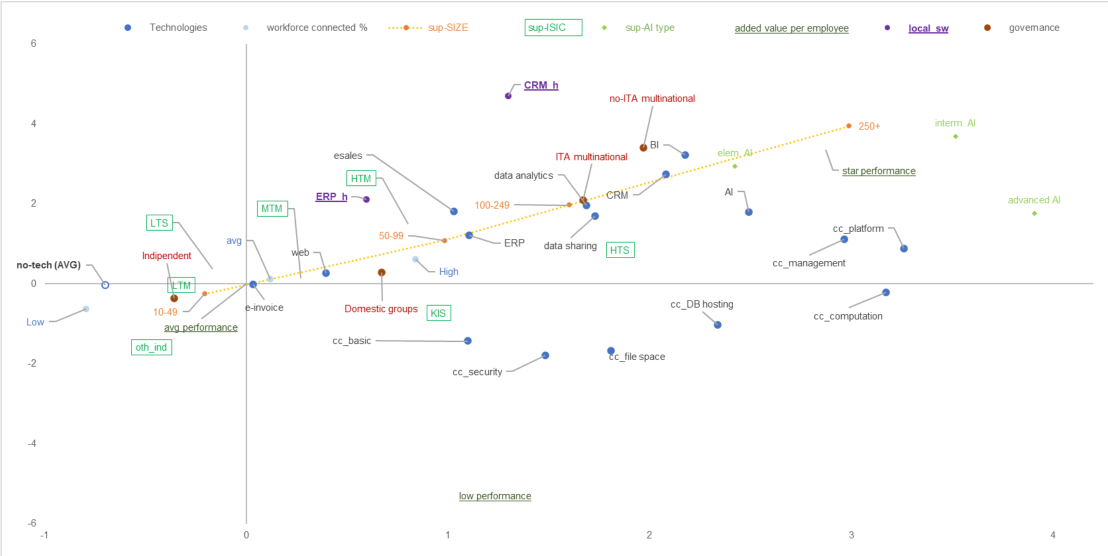

Figura 4.1: MCA Scatterplot

- Data analytics and sharing appear in close proximity on the plot, indicating a strong association. Their position suggests that they are less widely adopted than more basic tools such as websites or e-invoices. Moreover, they are closely linked with Italian multinationals employing between 100 and 249 workers. This shows that small enterprises (majority of the Italian businesses) is still lag in these digital practices.

- CRM and BI are cluster together, indicating a strong link between these practices, particularly adopted by medium-sized foreign multinational. This finding in particular is related to **SQ1.1** that asks if the use of data analytics comes with management software. However, ERP seem too distant for saying that it comes with BI and CRM adoption.

- Some cloud services (cc_basic and cc_security; cc_management and cc_platform) also appear in proximity. Purchasements of cloud for electronic sheets, e-mail or basic activities and for ICT safety measures are bundled. Outsourcing ERP and CRM- related cloud services is often related with the one for platforms development. All these patterns are consistent with the correlation matrix presented in Chapter 3.

- It can also be observed that ERP adoption, e-sales, a high percentage of employees connected to the Internet, belonging to domestic groups and HTM are related, together with a 50-99 employee size range. Interestingly, while internet connection was once associated with highly digitalised enterprises, it now appears more typical and adopted across the majority of enterprises, reflecting the fact that this tool is no longer digitalisation proxy rather a mature technology. Hence, always making reference to **SQ1.1**, ERP software seem more associated with electronic commerce adoption. The fact that ERP and e-sales are closer to the origin rather than other means that they are also more adopted technologies.

- A positive association appears between size, AI adoption and enterprise performance (measured by cluster of value added per employee). AI adoption is close to larger enterprises with higher performance (labelled "star performance" in the plot). Elementary AI is more close to other technologies (i.e, BI and CRM) and to the AI active variable, in line with the fact that around 70% of the 2023 AI adopters were using AI in an elementary way.

In general, a more general distinction can be identified in the plot. Big-sized and multinational enterprises are positioned in the first-right quadrant, adopting BI, CRM, AI, cloud for platforms and management software. High-tech services are associated with this individuated area. Conversely, average productivity performance, smaller and independent-governance enterprises are located in in the second or first-left quadrant are characterised by none or minimal technology adoption that regards using e-invoices and web page with a low percentage of workforce connected. Low technological service or manufacturing services, together with construction and energy industries are part of this group. Furthermore, this second scenario described is where most of the Italian enterprise, looking to the 2023 data, were located. Instead, the first situation described is more unique and rare. Based on examination of Figure 4.1 and each variables' contribution, the following interpretations emerged:

- *Dim1* reflects the degree of digital technology sophistication and intensity. For example, obtaining a low score can be translated into the adoption of minimal or none digital tools while an high one is associated with use of more advanced technologies (e.g, AI, platforms, management softwares with the related cloud computing services)

- *Dim2* captures variation in technological modality patterns, where higher values correspond to decision-support, data-driven and in-house technologies (e.g, ERP or BI), while lower values indicate tendentially infrastructural and outsourced uses (e.g., outsourced cloud services, ICT security and file space).

For the first dimension, an exceptionally high value of variability indicates the existence of a fundamental and clear distinction between groups of enterprises. This concentration suggests a marked dichotomy in technological adoption behaviour, likely reflecting the contrast between enterprises with a high degree of digitalisation

and those with limited or no adoption. The second dimension, since contributes in a minor way to the total variance, indicates that technological modality differences are significantly less relevant than the main distinction.
In general, all the considerations done so far, show somehow that structural dimensions, like size, governance and sector, play a fundamental role when talking of technological adoption and contribute to shape the digital pattern adopted. Lastly, the analysis shows which technologies act as enablers for others. Digitalisation emerges as a gradual process requiring multiple investments: technologies located further to the left often serve as enablers of those positioned to the right.

4.2 EFA Results

For the interpretation of results, two parameters are of primary importance. At the global level, the amount of variability indicating the variance "explained" by the complete set of factors and by each individual factor. Secondly, factor loadings describe the strength of the relationship between a factor and the measured variable; very low loadings (when standardized, with absolute values below 0.30) are typically used to exclude weak associations between variables and factors, thereby simplifying the structure. For this reason, in the R code used for plotting loadings and EFA results (`print(efa_result4$loading, cutoff = 0.3)`), only variables with loadings above 0.3 are displayed. It should be noted that, when discussing this parameter, it is more logically correct to state that "the factor loads on the variable" rather than the reverse. This is because the loading tells how strongly a factor explains or accounts for the variance in a variable.
As mentioned in Chapter 3, four sections of the ICT survey were selected for performing the two EFAs, each of which produced good results in the Bartlett and KMO tests. All the selected variables have been then used in two different EFA rotation methodologies: `oblimin` and `varimax`. The former produced 11 factors, whereas the latter extracted 14. The analytical process and the associated meaning (label) to each of them has been explain in detail along this section. Particularly, the interpretative logic is consistent for both, but the diagrams presented here refer only to the `oblimin` method, for not being repetitive. At the end of this section, however, Table 4.7 summarises the labels of the 14 factors obtained with the varimax rotation and looking at it is fundamental for understanding which variables have been inserted in the second OLS model.
Focusing on the `oblimin` case, the results obtained for each of the four survey sections are now presented. For each section, the plots showing the loadings of each factor on each variable are explained and discussed. These factors are labelled "MR" in the following figures, as they are derived using the Minimum Residual (`minres`) method. In all the EFA diagrams below, a distinction must be made:

- On the left-hand side, numbers represent factor loadings (the ability of the factor of explaining the variance of that variable). Higher is the factor loading, greater is the power of that factor in representing that variable.

- On the right-hand side, numbers represent factor correlations (the relationship between latent factors after oblique rotation).

At the end of this section, Table 4.6 summarises the interpreted meaning of each factor.

4.2.1 Extracted Factors from the First Section of the ICT Survey

For the first section, related to website usage and internet connection, three factors (unobserved constructs) are extracted. As shown in Figure 4.2, MR1 is associated with most of the variables, particularly those related to customer e-sales services (i.e., A5A, A5B), fixed internet connection, mobile applications, and websites used for customer assistance or support. MR2 is related to website use aimed at offering personalised experiences or activities. This factor represents online visibility and communication capabilities, encompassing interactions not only with customers but also with employees. Lastly, MR3 is associated with websites used for international markets and talents (i.e., posting job vacancies) reach. All the factors seem somehow correlated between each other: this depends also on the oblimin rotation method adopted, that aims to underline possible relationships between factors. The most high correlation is between the first and second factor that might show an association between customer-oriented digital activities (i.e., website and mobile app) and customized website experiences.

In the figure, MR factors are ordered from top to bottom according to the total variance explained, in decreasing order: MR1 explains more variance than MR2 and MR3. Taken together, the three factors explain approximately 0.56 of the variance (Table 4.2). This value is acceptable and typical for behavioural data such as those analysed in this study.

	MR1	MR2	MR3
SS Loadings	2.837	1.720	1.058
Proportion Var	0.284	0.172	0.106
Cumulative Var	0.284	0.456	0.562

Tabella 4.2: Factor loadings and explained variance for the three factors extracted from the first ICT survey section.

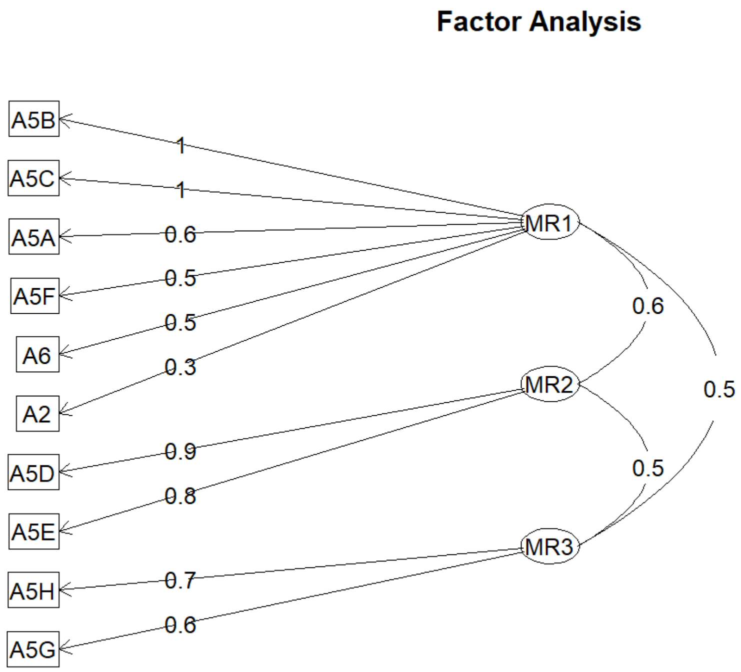

Figura 4.2: Factor loadings for the first section

4.2.2 Extracted Factors from the Third Section of the ICT Survey

Figure 4.3 shows the results of the EFA for the section related to data usage, sharing, and analysis. As in the first section, three factors were extracted: MR4, MR5, and MR6 (referenced in the figure as MR1, MR3, and MR2, respectively - this is due to RStudio settings for returning the output). The first factor (MR4) is associated with the use of CRM and data analysis from customer behavioural data (e.g., purchase history), web data, and social media. The second factor (MR5) is linked to data analysis from satellites, location services (e.g., GPS), sensors, and governmental websites. The third factor (MR6) is associated with three variables indicating the use of BI and ERP software, as well as the presence of a unified database for data collected and processed by these systems. This factor provides an additional information to completely answer to **SQ1.1**: from the MCA is emerged that enterprise tend to use together CRM, BI and data analytics while ERP usage, is less close to the latter tools on the scatterplot. However, from the EFA is showed how it can be found an hidden pattern related to the adoption of BI and ERP within a unified relational database. The use of position and customer data-related

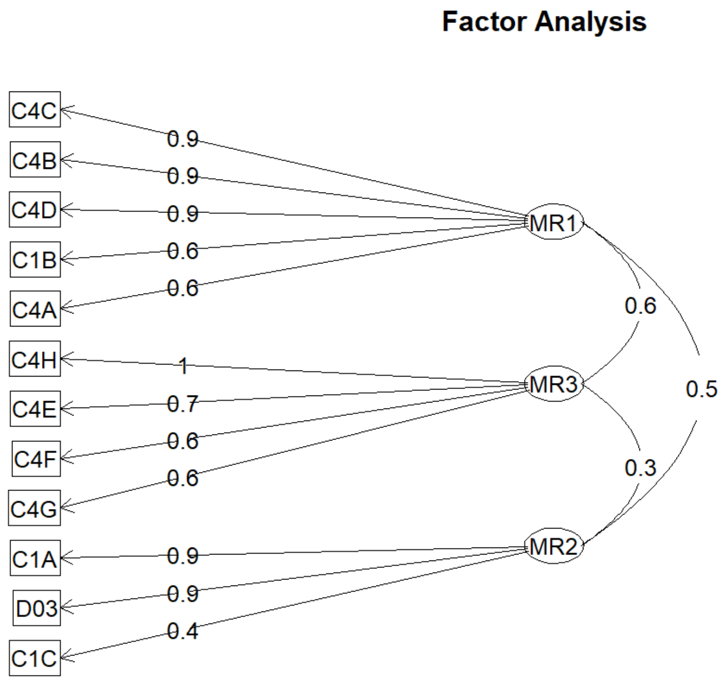

Figura 4.3: Factor loadings for the third section

factors seem to be correlated more than the others.
The cumulative variance explained is approximately 0.68 (Table 4.3).

	MR4	MR5	MR6
SS Loadings	3.548	2.390	2.176
Proportion Var	0.296	0.199	0.181
Cumulative Var	0.296	0.495	0.676

Tabella 4.3: Factor loadings and explained variance for the three factors extracted from the third ICT survey section.

4.2.3 Extracted Factors from the Fourth Section of the ICT Survey

Figure 4.4 shows the results of the fourth section, associated with the use and purchase of cloud computing related to several purposes. Enterprises were asked to report the reasons for purchasing cloud computing services. Two factors explain the data: MR7 and MR8 (referred to in the figure and table as MR1 and MR2).

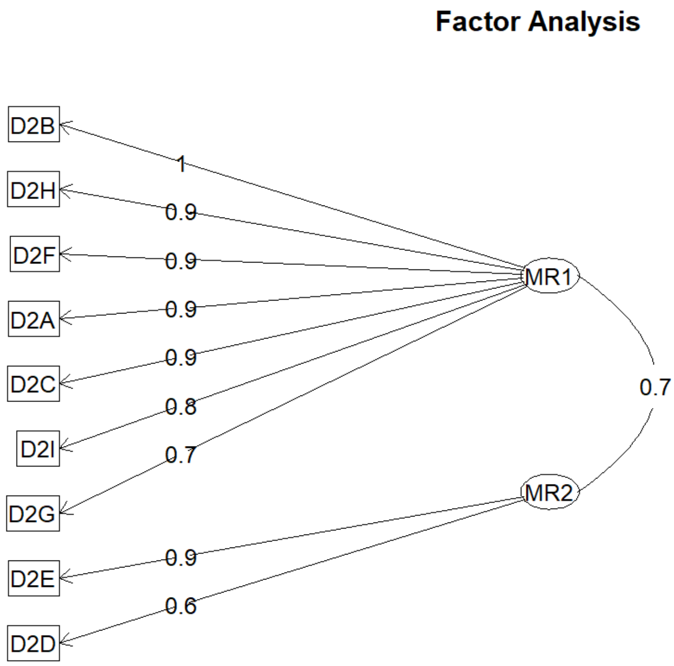

Figura 4.4: Factor loadings for the fourth section

The first factor represents the procurement of cloud services for office and basic functionalities, file storage, database hosting, computational capacity, and security. The second factor encompasses the use of cloud computing for management software applications (i.e., ERP and CRM). It seems that the two factor are somehow correlated between each other: this might be interpret that tendentially a cloud-service buyer, acquires cloud for all (or almost all) purposes. The cumulative variance explained is 0.735, which can be considered a very good result (Table 4.4).

	MR7	MR8
SS Loadings	5.354	1.261
Proportion Var	0.595	0.140
Cumulative Var	0.595	0.735

Tabella 4.4: Factor loadings and explained variance for the two factors extracted from the fourth ICT survey section.

4.2.4 Extracted Factors from the Fifth Section of the ICT Survey

Figure 4.5 shows the loading diagram for the last section analysed with EFA, related to artificial intelligence technologies and their usage. Three factors were extracted: MR9, MR10, and MR11 (referenced in the figure as MR1, MR2, and MR3). The first factor represents the use of cognitive AI for marketing and security applications through technologies for detection, analysis, and conversion. The second factor encompasses the use of AI, particularly automatic tools and image recognition systems, for business processes, including robotics and machine learning. Among the most influential variables, there are E1G and E2D, which correspond to the use of drones or automatic tools and the adoption of AI for logistics activities (e.g., solutions for picking items from shelves and packing parcels using autonomous robots; tracking, distribution, or sorting; route optimization based on machine learning). Other variables capture the use of AI for administrative, accounting, production, and ICT security activities. In general it seems that MR10 represents the use of advanced AI for several business-activities, with a particular relevance on logistics. The third factor reflects the use of self-learning AI (i.e., ML, DL, neural networks) in research and development (R&D) or innovation activities. The cumulative variance explained by the three factors is 0.65 (Table 4.5). These results address **SQ1.2**: independently of the rotation (oblimin or varimax), the factor structure points to domain-specific AI adoption. With oblimin, three coherent patterns emerge: self-learning AI concentrated in R&D/innovation; process-automation and AI software deployed in administrative and financial functions; and cognitive/robotic AI embedded in operations. Under varimax, this last pattern resolves into two distinct uses: cognitive AI supporting primary activities (e.g., manufacturing, safety/security) through recognition or text-generation, and automation-oriented AI in logistics, where robots and drones predominate.

	MR9	**MR10**	**MR11**
SS loadings	4.610	3.306	1.163
Proportion Var	0.329	0.236	0.083
Cumulative Var	0.329	0.565	0.649

Tabella 4.5: Factor loadings and explained variance for the three factors extracted from the fifth ICT survey section.

4.2.5 Final Interpretation

The 11 factors resulting from the first EFA approach (`rotation = oblimin`), together with their respective labels assigned, are summarised in Table 4.6.

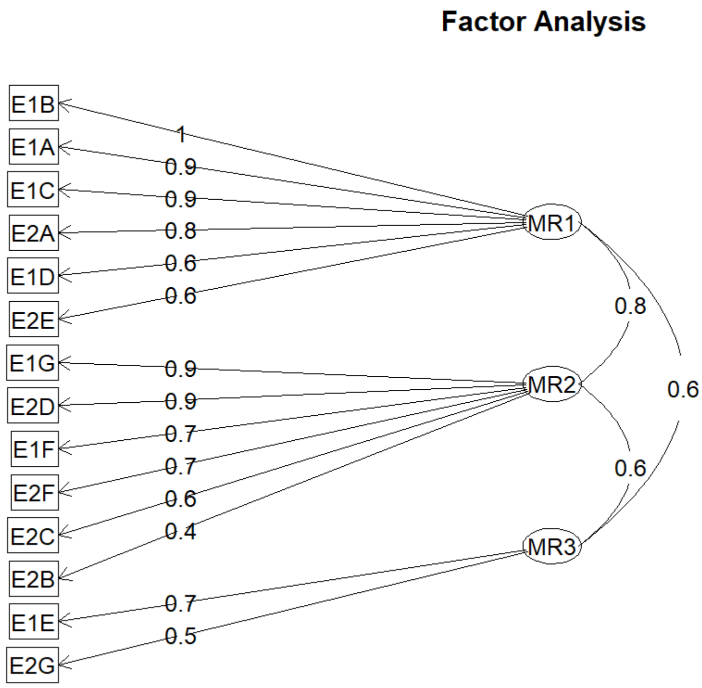

Figura 4.5: Factor loadings for the fifth section

Factor	Variables	Description
MR1	A5B, A5C, A5A, A2, A5F, A6	E-commerce and support website or mobile app usage
MR2	A5D, A5E	Content engagement website usage
MR3	A5G, A5H	Website usage for talent (posting job vacancies) and market reach (international website)
MR4	C4C, C4B, C4D, C1B, C4A	Use of CRM together with customer or market data collection and analytics
MR5	C4H, C4E, C4G, C4F	Geospatial data analytics
MR6	C1A, D03, C1C	Integrated data usage of BI and ERP software
MR7	D2B, D2H, D2F, D2A, D2C, D2I, D2G	Basic and infrastructural cloud services for office, accountability, security or space-storage purposes
MR8	D2E, D2D	Advanced and managerial cloud services related to ERP or CRM
MR9	E1B, E1C, E1A, E2A, E1D, E2E	Cognitive AI for marketing and security (e.g., customer analysis and image recognition)
MR10	E2F, E2C, E2D, E1F, E2B	Automatic or advanced AI for core business functions, especially logistics and robotics
MR11	E1E, E2G	Self-learning AI (ML) for research and innovation

Tabella 4.6: Factors resulting from Exploratory Factor Analysis (EFA) with `oblimin` rotation

The 14 factors resulting from the second approach (`rotation = varimax`), together with their associated labels used in the OLS models section, are reported in Table 4.7 and the explained variance by each factor in Table 4.8.

Factor	**Variables**	**Description**	**Variable Re-name**
MR1	A6, A5F	Customer engagement and supporting tools	`CustEngage_Tools`
MR2	A5B, A5C, A5A, A2	E-commerce presence	`EComm_Presence`
MR3	A5D, A5E	Website for personalization/customization	`Web_Personalization`
MR4	A5G, A5H	Website for talent and market reach	`Web_Talent_MarketReach`
MR5	C4C, C4B, C4D, C1B, C4A	CRM with customer/market data analytics	`CustMarket_Analytics`
MR6	C4H, C4E, C4G, C4F	Geospatial data use	`ExternalData_Use`
MR7	C1A, D03, C1C	Integrated BI and ERP data usage	`IntegratedData_BI_ERP`
MR8	D2H, D2I, D2G, D2F	Infrastructural cloud (storage, ICT security)	`Cloud_Infrastructural`
MR9	D2C, D2B, D2A	Cloud for basic apps (e-mail, spreadsheets)	`Cloud_Basic`
MR10	D2E, D2D	Cloud for enterprise systems (ERP, CRM)	`Cloud_Mgmt`
MR11	E1B, E1C, E1A, E2A, E1D, E2E, E2B	Cognitive AI (text, image, marketing, safety)	`AI_Cogn_Operations`
MR12	E1E, E2G, E1F	Self-learning AI (ML) for R&D	`AI_Adv_Innov`
MR13	E1G, E2D	AI for robotics/automation in logistics	`AI_Smart_Logistic`
MR14	E2F, E2C	AI for admin and management activities	`AI_Mgmt`

Tabella 4.7: Factors resulting from Exploratory Factor Analysis (EFA) with `varimax` rotation

Factor	SS Loadings	Proportion Var	Cumulative Var
EFA for the First Survey Section			
CustEngage_Tools	1.946	0.195	0.195
EComm_Presence	1.926	0.193	0.387
Web_Personalization	1.758	0.176	0.563
Web_Talent_MarketReach	1.234	0.123	0.686
EFA for the Third Survey Section			
CustMarket_Analytics	3.511	0.293	0.293
ExternalData_Use	3.121	0.260	0.553
IntegratedData_BI_ERP	2.656	0.221	0.774
EFA for the Fourth Survey Section			
Cloud_Infrastructural	2.763	0.307	0.307
Cloud_Basic	2.396	0.266	0.573
Cloud_Mgmt	2.118	0.235	0.809
EFA for the Fifth Survey Section			
AI_Cogn_Operations	4.739	0.338	0.338
AI_Adv_Innov	2.781	0.199	0.537
AI_Smart_Logistic	2.416	0.173	0.710
AI_Mgmt	2.266	0.162	0.872

Tabella 4.8: Summary statistics of factors Exploratory Factor Analysis (EFA) with `rotation=varimax`: SS loadings, proportion of variance, and cumulative variance.

The procedure for obtaining factors is the same, with the only differences being:

- the specification of `rotation=varimax` instead of `rotation=oblimin` in the RStudio EFA code chunk;
- the absence of arrows between factors, as these are orthogonal (independent) in the `varimax` approach;
- a factor more extracted for each section beside section 3 that has the same factors of before. This increasement is due interpretability reasons. Furthermore, in this case, it has been noticed, a bigger alignment with `fa.parallel()` optimal factors' number suggestions;
- the AI section, represented by more factors, allows more specific latent behaviours to be identified;
- an higher cumulative variance values obtained (around 10-20% higher for each section).

Employing both rotations (i.e., `oblimin` and `varimax`) is advantageous for assessing the robustness of results: if factor loadings display similar patterns, confidence in the identified structure is strengthened. In this case, the structure and meaning of the factors are highly similar across both approaches. However, when factor scores are used as predictors in linear regression, `varimax` rotation is generally preferable, since the independence of factors reduces multicollinearity's risk and facilitates coefficient interpretation. For this reason, after confirming that the factors obtained with the `oblimin` rotation exhibited way stronger correlations (often above 0.7) and a better interpretability than those from `varimax`, it was decided to use the 14 factors (`rotation = varimax`) for the linear regression models (Table 4.7). Table 4.7 includes an additional column that reports the labels assigned to each factor. These re-named variables are used in the subsequent analysis to improve clarity and facilitate interpretation.

4.3 OLS Regression Results

The following subsections present the main results for both the OLS regressions, complemented before with a description of the independent variables, their distributions, and correlation analyses between the dependent and independent variables. The regression findings are interpreted while carefully considering the temporal mismatch limitation. In fact, as previously mentioned, for covariates measured temporally after the outcome variable (2023 and 2022, respectively), it is possible to discuss only associations or correlations rather than causality. Another aspect to underline ex-ante is that what matters for both the OLS is the significance and direction of the coefficients (positive or negative association), but not their absolute magnitude, because the MCA and EFA factors are standardized and lack a natural unit. Nevertheless, the size of a coefficient can still be compared to the coefficients of the other factors in the same regression model. For example if a co-variate has a much larger coefficient than other ones, it means that its changes are associated with stronger changes in value added than changes in other factors, within the model.
Before going deeper into the specific findings related to the two regression models, a descriptive analysis of the y (`vagg_add`) has been reported since it is present in both the models. Particularly, looking at Table 4.9 below, it is possible to notice that this variable is highly skewed (Skewness= 27.7 and Kurtosis= 1600). There are extreme values (very high or very low outliers) for this variable, which affect the mean and, consequently, both regression models (Figure 4.6). Notice that the number of observations (210,330) reported in the table does not reflect the usual one: this is because the dataset has been expanded by taking into account the rounded weighting coefficients.

Statistic	Value
Observations (Weighted Sum)	210,330
Mean	59,356.84
Standard Deviation	82,952.82
Variance	6.88e+09
Minimum	-2,692,961
1st Percentile	594.43
5th Percentile	15,176.1
10th Percentile	20,926.13
25th Percentile (Q1)	32,166.33
Median (50th Percentile)	47,926.31
75th Percentile (Q3)	70,114.74
90th Percentile	102,178.9
95th Percentile	128,978.9
99th Percentile	259,454.6
Maximum	6,352,824
Skewness	27.70
Kurtosis	1600.99

Tabella 4.9: Weighted summary statistics of `vagg_add`

The value added (VA) per employee is yearly computed with the following formula:

$$\text{(1) Value Added (VA)} = \text{Value of Production of Goods and Services} \\ - \text{Value of Intermediate Goods and Services Consumed}$$

$$\text{(2) Value Added per Employee} = \frac{\text{Value Added (VA)}}{\text{Number of Employees}}$$

Looking to (1), if the first term is smaller than the second one, a negative value added is obtained. The possibility for the y of being below the zero, has also affected our modeling decisions, since logarithmic (or square-root) regression cannot be used. It also limits the possibilities for model tuning and improvement.
As shown in Table 4.9, the minimum and maximum values are respectively: $-2{,}692{,}961$ and 6,352,824. Most observations are concentrated at lower positive values, with a long tail to the right. Indeed, the mean of value added per employee for Italian enterprises (in the energy, construction, no-financial services, and manufacturing sectors) at the end of 2022 is €59,356.84. Comparing this with the same performance indicator published on the Eurostat website, which takes all sectors into account, a value of €79,663.27 is reported against the 2022 EU average of €80,588.25 [80]. This indicates that the excluded sectors like the financial ones are likely bringing the average up and also that in the studied year

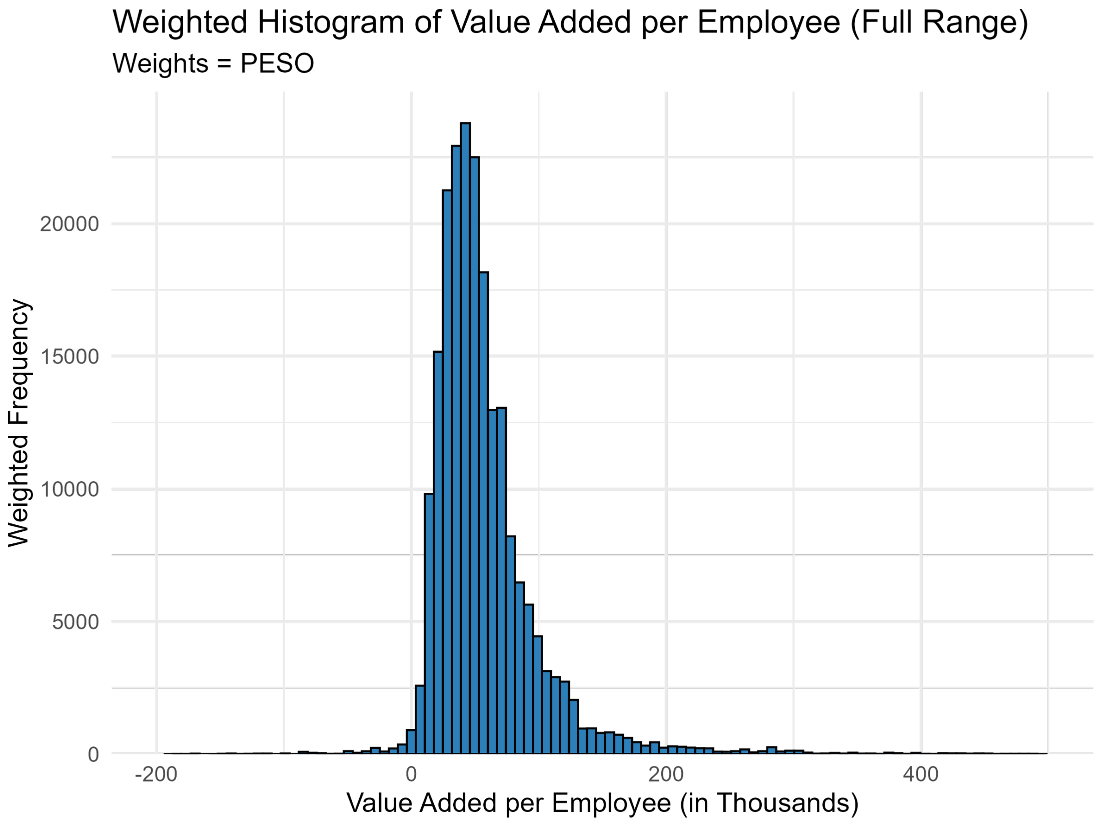

Figura 4.6: Distribution of VA per employee

Italy was underperforming in comparison with EU. In the following paragraphs, it has been analysed how average value added is distributed across different structural variables (sectors grouped by technological intensity, size, and governance). To support the discussion, illustrative histograms are provided to visually highlight the main evidence. Figure 4.7 shows that low-tech services and manufacturing have the lowest `vagg_add`. Conversely, medium-low, medium-high and high-tech manufacturing exhibit the highest performance in terms of value added per employee. The results found are in line with the literature. In fact, from a broader European perspective, in 2022 Italy recorded the second largest share (12. 7%) of the value added generated in the manufacturing sector of the EU [80]. However, these types of enterprises are not the biggest technology adopters. On average, high-tech services exhibit the highest level of technological adoption. This can be related to the fact that digitalisation and technology adoption are not the only factors influencing enterprise productivity.

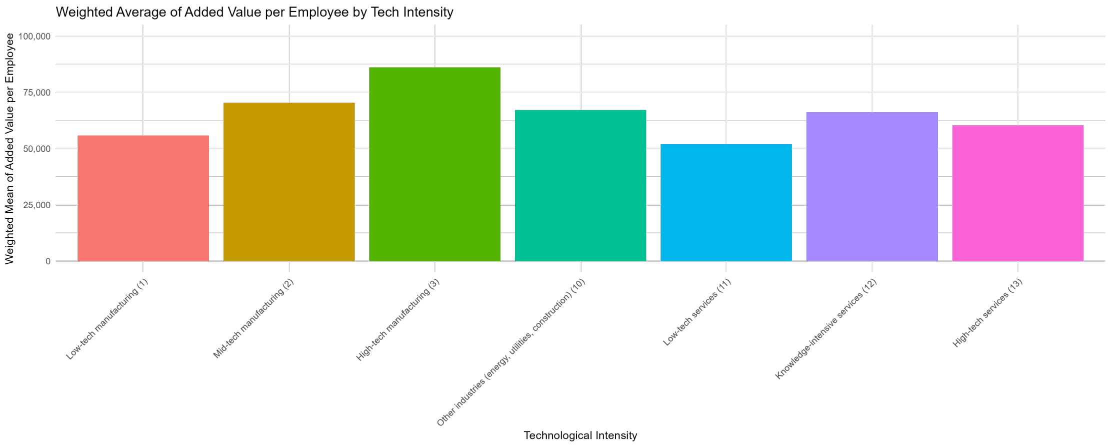

Figura 4.7: Average value added per employee grouped by technology intensity

Figure 4.8 highlights that, in general, larger companies (sizes 3 and 4) exhibit a higher average value added per employee.

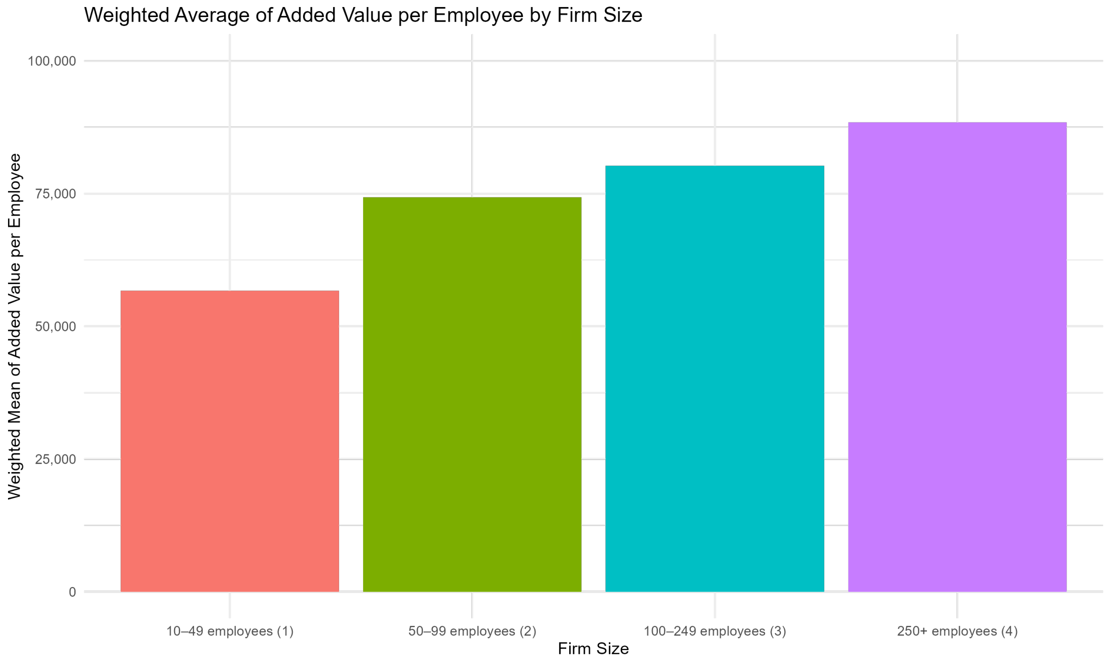

Figura 4.8: Average value added per employee grouped by size

With regard to the relation between governance and average value added in Figure 4.9, the more structured the governance (Italian or foreign multinational), the higher the value added per employee.

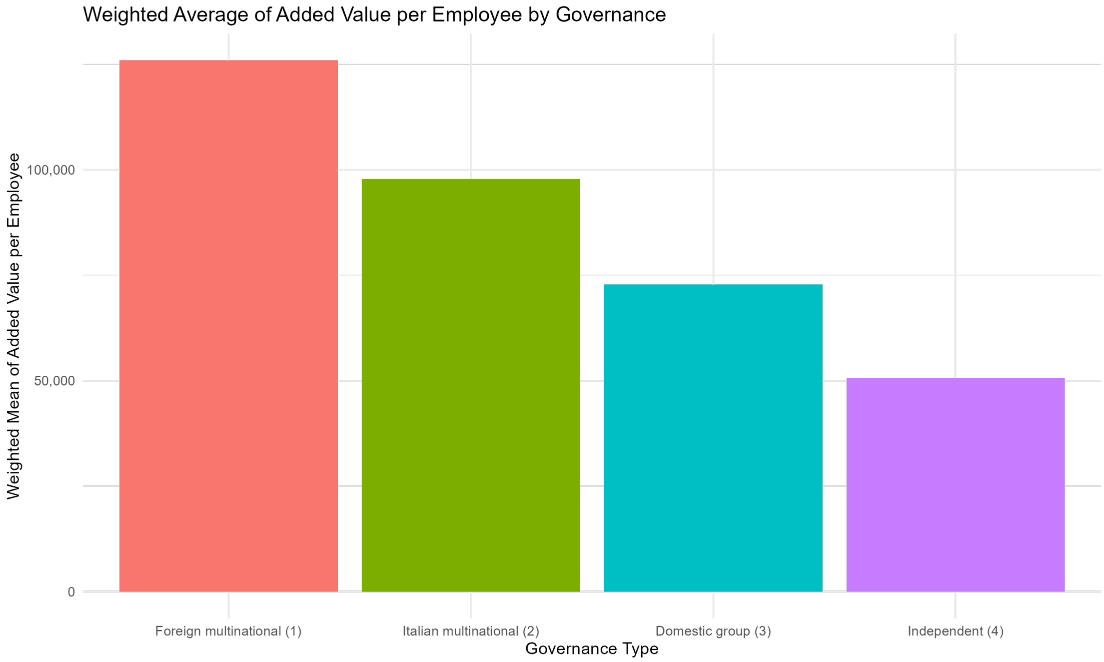

Figura 4.9: Average value added per employee grouped by governance

4.3.1 Findings - First OLS Regression

Looking at the two factors (*fac1* and *fac2*) resulting from the MCA, Figure 4.10 shows that, in this case, the values can be negative, with a range from approximately -2 to $+4$. The first factor represents digital sophistication and intensity, while the second one captures the modality of technology adoption: higher values of *fac2* indicate a preference for data-driven, in-house and decision-support technologies (e.g., BI, AI, CRM), whereas lower values reflect a greater reliance on infrastructural, outsourced and more passive technologies (e.g., cloud computing for file storage). Although dimensions and factor scores should not be confused, the interpretations derived from the two MCA dimensions can be used to assign meaning to the corresponding factorial variables. In fact, while dimensions (axes) are the latent variables interpreted, factor scores are the coordinates of observations on those dimensions.

Table 4.10 reports the values obtained by the correlation study between the single co-variates (*fac1* and *fac2*) with y. The correlation made here is a weighted

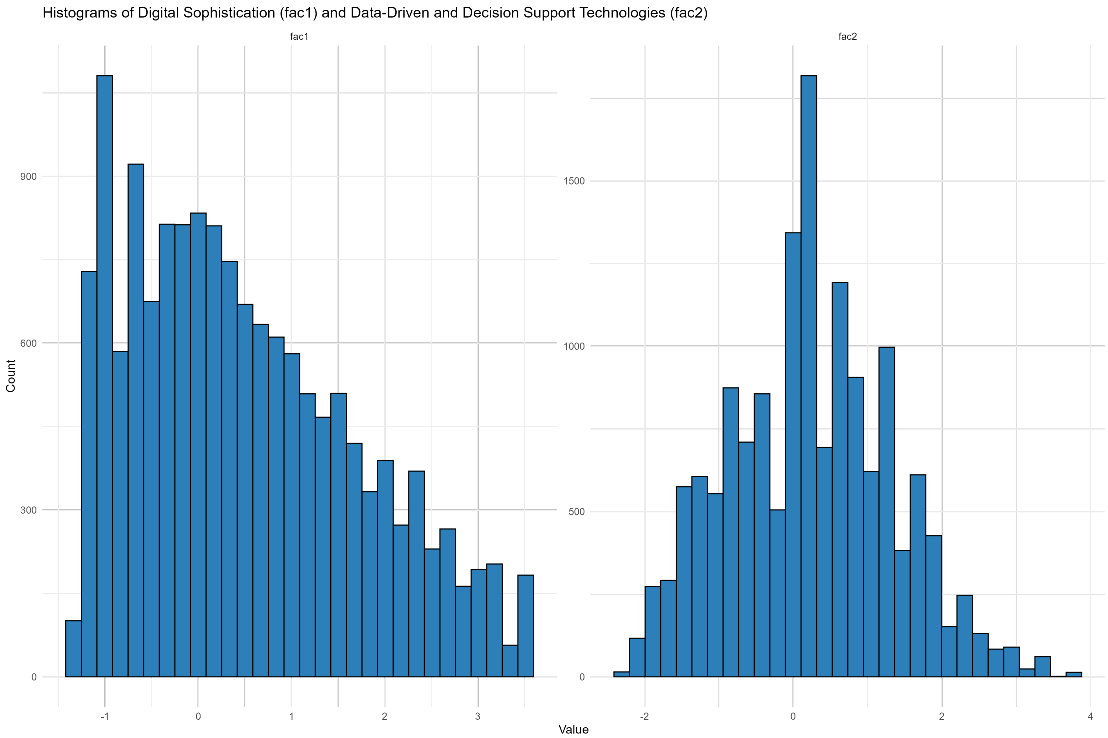

Figura 4.10: Distribution of the two MCA factors

Pearson correlation that takes into account the fact that the co-variates have been already weighted while vagg_add (i.e., y) still is not. As it possible to see, in both the cases, the value obtained is very small. Table 4.11 is showing that

	Correlation	Std. Err.	t-value	p-value
Y - fac1	0.1478	0.0080	18.41	7.0×10^{-75}
Y - fac2	0.0603	0.0081	7.44	1.07×10^{-13}

Tabella 4.10: Correlation results between the co-variates and y

between *fac1* and *fac2* resulting from the MCA the correlation is low (around 0.21). However, the p-value is lower than 0.05, showing a statistically significant correlation. For interpreting this controversial situation, the literature on multicollinearity in regression models suggests focusing on the magnitude of correlations rather than their statistical significance, since with large samples even small correlations often yield p-values below the 0.05 threshold [81, 82]. Correlations below 0.7-0.8 are generally considered not problematic, which supports the inclusion of all the respective factorial covariates in the two models without major concerns about multicollinearity effects. Hence, both variables have been included in the model. After having seen the correlations between the selected variables, *fac1* (i.e.,

Statistic	Value
Correlation Method	Pearson's product-moment correlation
t-statistic	25.856
Degrees of Freedom (df)	15172
p-value	$< 2.2 \times 10^{-16}$
Alternative Hypothesis	true correlation is not equal to 0
95% Confidence Interval	[0.1901, 0.2206]
Sample Estimate	0.2054

Tabella 4.11: Summary of Pearson correlation test between variables $fac1$ and $fac2$

technology adoption and sophistication) and *fac2* (i.e., technology modality) are explored in relation with some structural variables (i.e., technological intensity, size, and governance). The results, presented in Figure 4.11, Figure 4.12, and Figure 4.13, highlight the following main findings:

- Sectors related to utilities and construction have a *fac2* value below zero, indicating that these industries rely more on outsourced and infrastructural technologies. These industries show to have a technological modality extremely different in comparison to all the other ones. Regarding *fac1*, high-tech services achieve the highest scores, followed by high-tech manufacturing. These two sectors seem to be the biggest digital adopters.

- Larger firms tend to prioritize data-driven and decision-support technologies over infrastructural ones, and the larger the firm, the greater its overall technology adoption and sophistication.

- The governance dimension reflects a pattern similar to that observed for the size of the firm. Moreover, Italian multinationals obtained the highest *fac1* score, resulting in a slightly higher technological intensity in comparison with foreign multinationals.

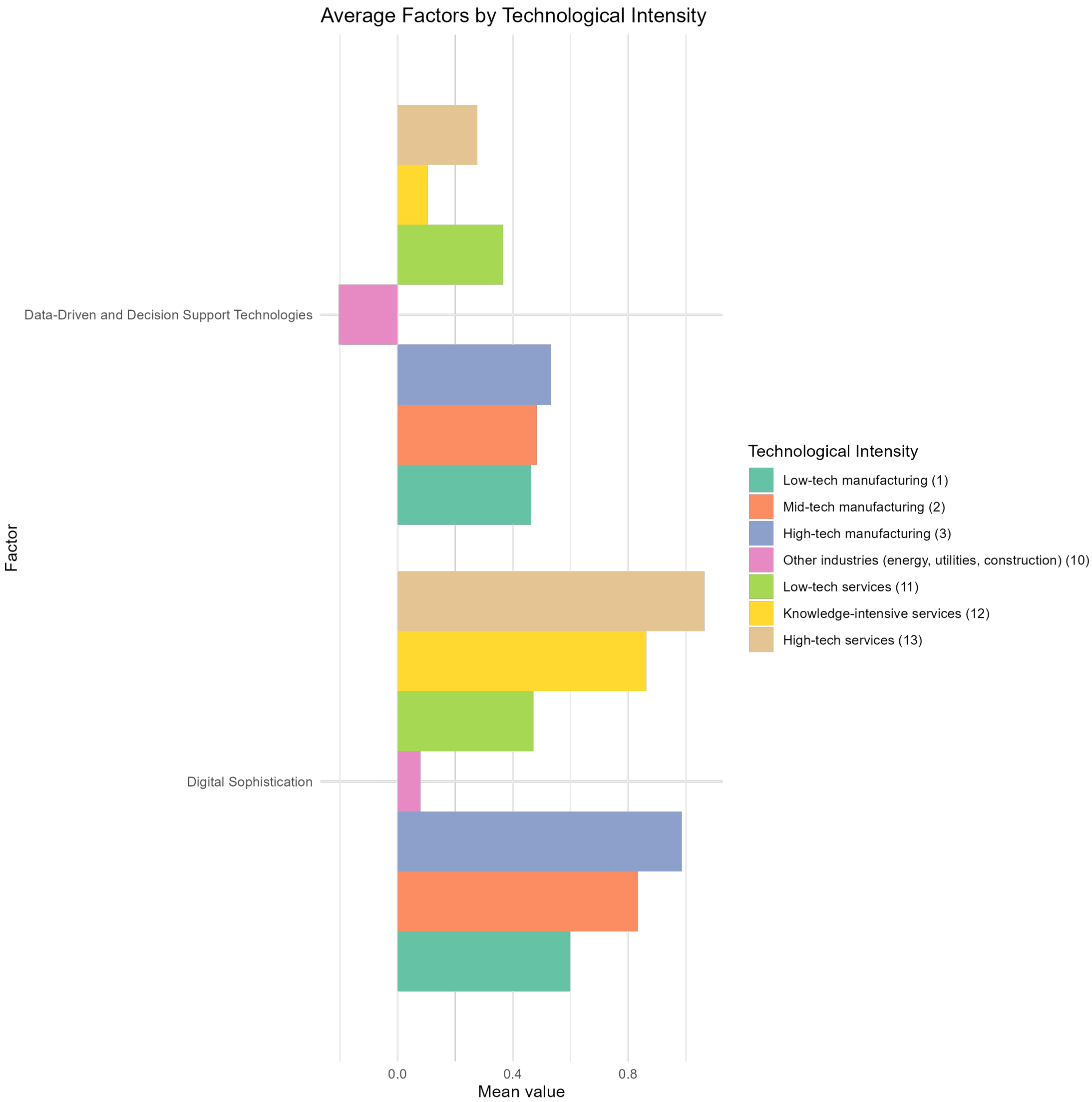

Figura 4.11: Average MCA's factors grouped by technology intensity

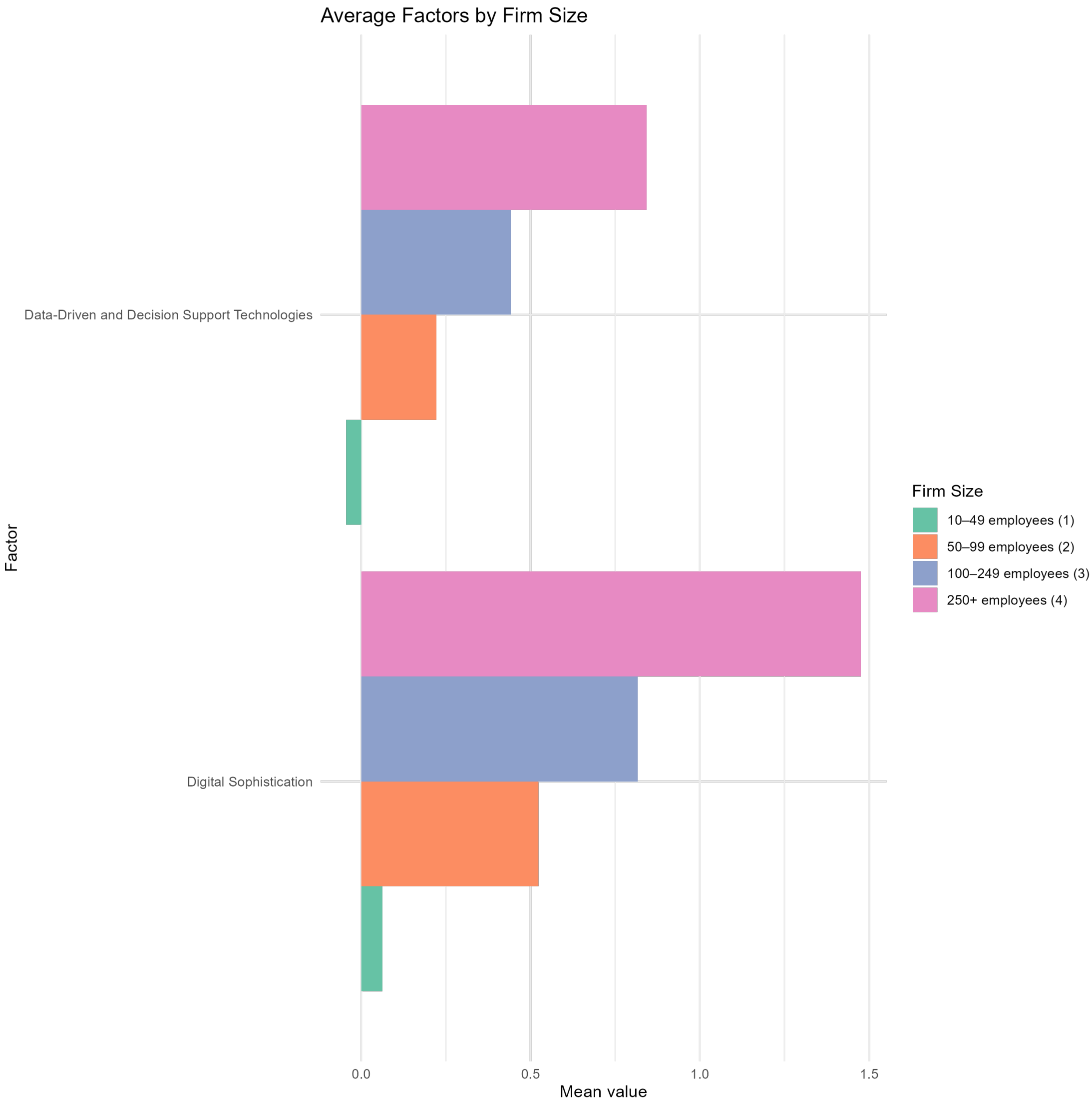

Figura 4.12: Average MCA's factors grouped by size

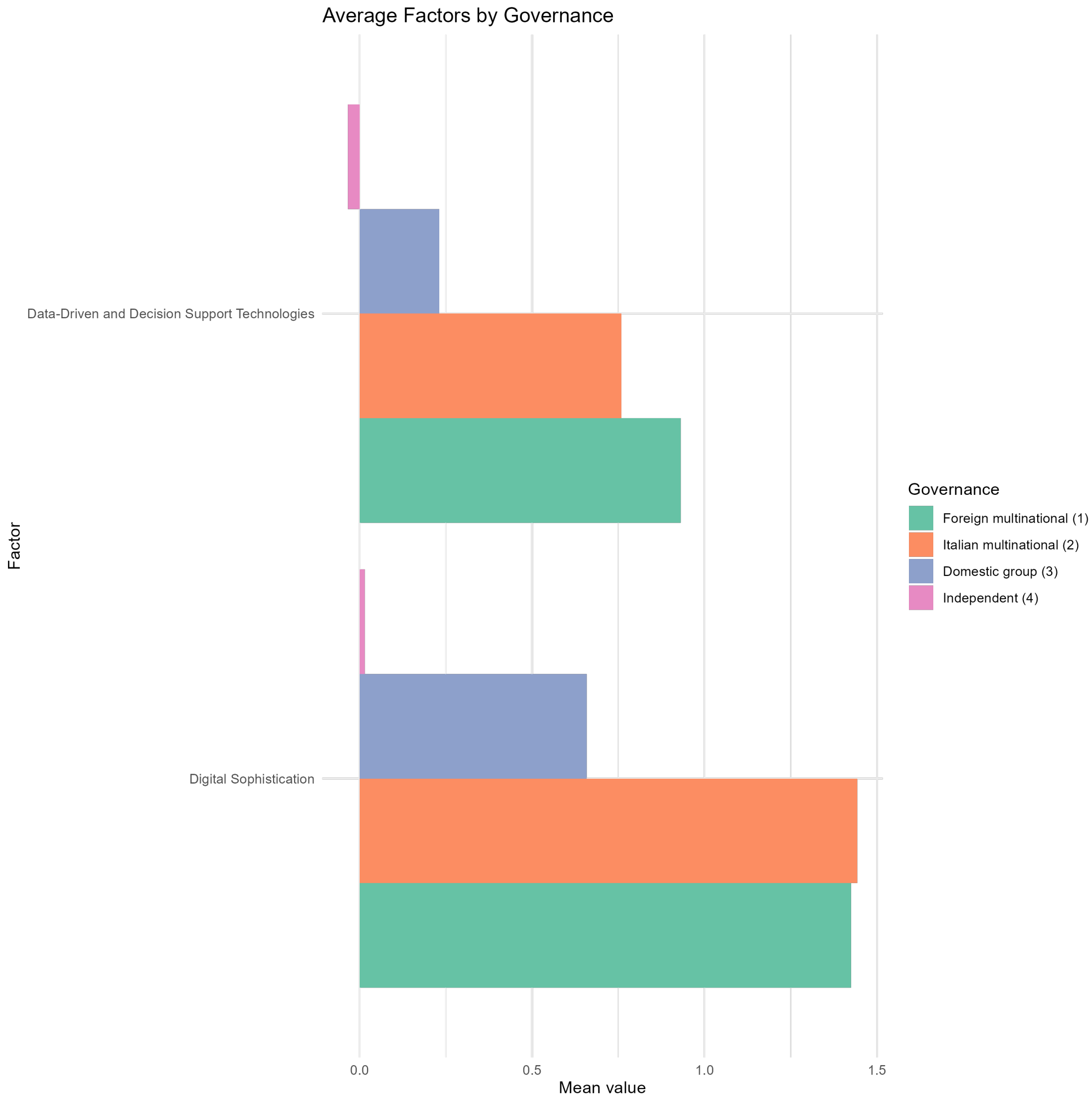

Figura 4.13: Average MCA's factors grouped by governance

First OLS Outcome in RStudio

For performing the first OLS regression the following code chunk is run in RStudio.

```
model_mca_fac <- lm(vagg_add ~ fac1+fac2, data = dati2023,
weights = PESO)
summary(model_mca_fac)
```

It is important to note that in the first regression model the variables have not been standardised, as the MCA procedure inherently produces factors with mean values approximately equal to zero and standard deviations close to one. The outcome of this first OLS model is reported in Table 4.12 below. The regression

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	59,356.8	664.8	89.280	$< 2 \times 10^{-16***}$
Digital Sophistication (fac1)	12,262.7	664.8	18.445	$< 2 \times 10^{-16***}$
Data-driven and decision support tools (fac2)	5,000.8	664.8	7.522	$5.7 \times 10^{-14***}$
Signif. codes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$				

Tabella 4.12: Linear regression results for value added per employee using the two MCA factors.

results highlight the statistically significant contribution of both latent factors (*fac1* and *fac2*) to variations in value added per employee. The coefficient associated with *fac1* ($\hat{\beta}$ = 12,262.7, $p < 0.001$) is positive, indicating that higher levels of technological intensity and sophistication are associated with increases in productivity. This suggests that enterprises adopting more advanced, complex, technological solutions tend to achieve superior efficiency and higher economic performance. Similarly, the coefficient for *fac2* ($\hat{\beta}$ = 5,000.8, $p < 0.001$) is also positive and statistically significant, though of smaller magnitude than *fac1*. This factor captures the adoption of decision-support technologies such as BI, CRM, ERP, and AI. The results indicate that these systems, when effectively integrated, enhance managerial decision-making and resource allocation, thus contributing to improved value creation. By contrast, lower values of *fac2*, which reflect a reliance on infrastructural rather than decision-support technologies (e.g., cloud storage or cloud services for basic office applications), are associated with lower value added per employee. Assuming that these technologies were already in place by the end of 2022 (when value added was measured), thus removing the time-mismatch limitation, the finding may suggest that infrastructural tools, since they are not directly embedded in productive operations, contribute less, or even negatively, to performance when compared to technologies more closely tied to production processes (e.g., AI, ERP, BI, CRM). Sectors that rely more heavily on these infrastructural tools may therefore not experience, at least in the short term, a direct positive effect on productivity. It is important to stress, however, that the literature consistently highlights a time lag between the adoption of new technologies and their measurable impact on economic performance. Overall, the model highlights that the greatest productivity gains arise from the strategic implementation of advanced and sophisticated digital technologies. However, this positive effect may be offset by an extensive reliance on infrastructural digital tools (i.e., cloud for business, storage or database), which tend to have negative effects on performance as they are not directly connected to core business operations. Looking at the adjusted R^2 in Table 4.14, its value is approximately 0.025, indicating that the model has limited explanatory power with respect to the variance in the data.

Statistic	**Value**	**Notes**
Min	-7,008,711	—
1Q	-51,598	First quartile
Median	-11,099	—
3Q	46,381	Third quartile
Max	14,291,569	—

Tabella 4.13: Weighted residuals of the first regression model.

Statistic	**Value**	**Notes**
Residual Standard Error	304,900	df = 15,171
Multiple R^2	0.02549	—
Adjusted R^2	0.02536	—
F-statistic	198.4	on 2 and 15,171 df
p-value	$< 2.2 \times 10^{-16}$	—

Tabella 4.14: Summary statistics of the regression model fit using MCA factors.

Even after applying tuning techniques, such as removing outliers using Cook's distance and adding quadratic (polynomial) terms of the covariates, the value only increases to 0.084, which remains relatively low. Moreover, removing outliers may result in the exclusion of important enterprises which, although not representative of the average population, still provide valuable information. The low R^2 can be explained by the absence of structural or non-digital variables, such as firm size, capital intensity, human capital, and sectoral characteristics, which the literature identifies as key predictors of enterprise performance.

4.3.2 Findings - Second OLS Regression

Below, Figure 4.14 is showing that the EFA's factors' distributions are all between small values (between -5 and +12) and that all the factors can also assume negative values. As it is possible to notice, the distributions are highly heterogeneous. Several factors (e.g., AI-related ones) show extreme right-skewness, with the majority of enterprises scoring near zero and only a small minority displaying very high values, reflecting limited and concentrated adoption. Other factors (particularly cloud-based and customer-engagement or e-commerce or integrated software tools) exhibit more dispersed distributions, indicating more mature and widespread technologies. This asymmetry confirms again a polarised technological landscape, where only a subset of firms act as digital leaders while most remain at a low adoption level.

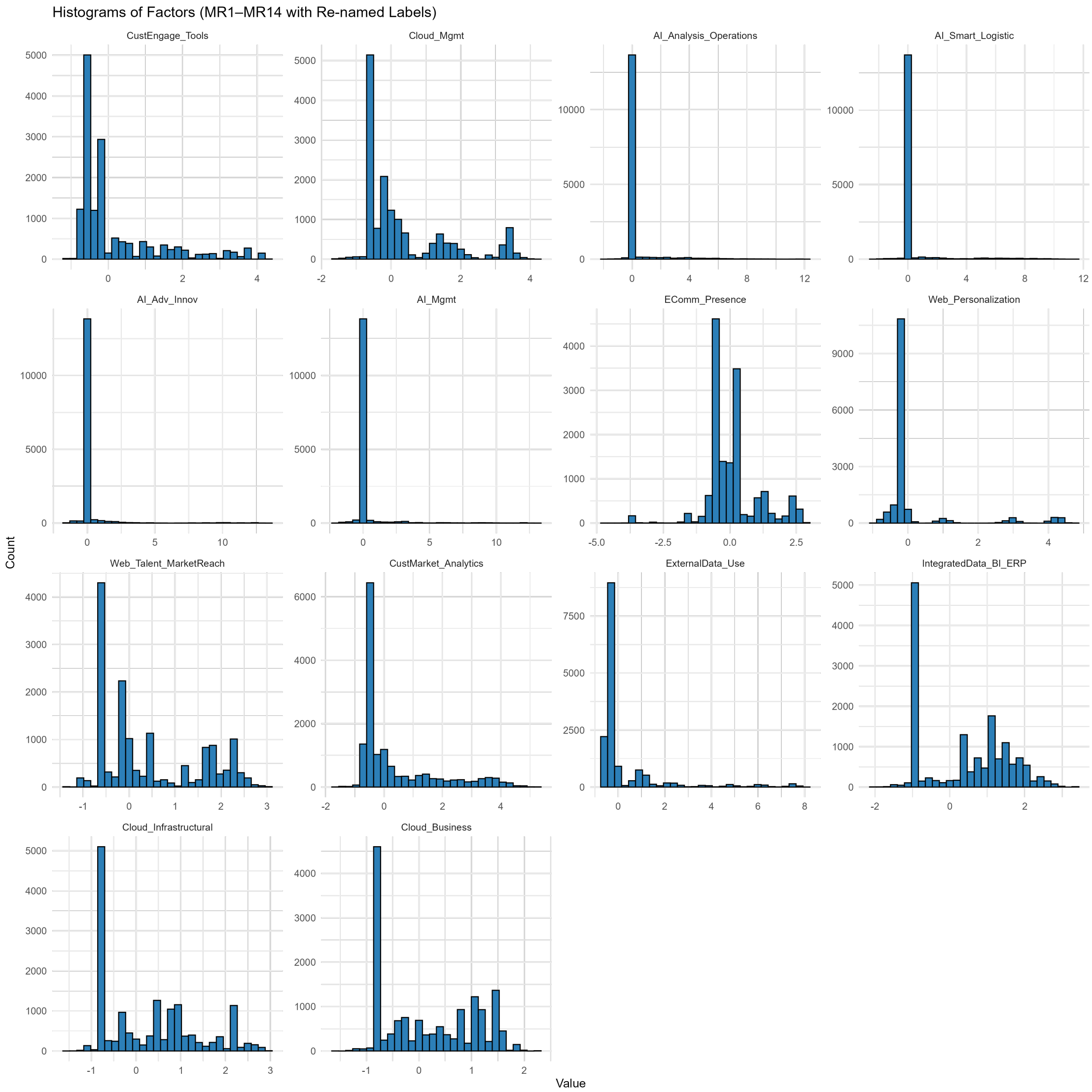

Figura 4.14: Distribution of the 14 EFA Factors

In Table 4.15, the output of a weighted Pearson correlation can be found. This has been done in order to weight with the coefficient (`PESO`) the dependent variable while keeping the independent ones untouched (since they have been computed considering the weight). Hence, in the table, the correlation between y and each co-variate (x) is shown. The values are very small, like in the MCA case, with the highest results obtained by value added with the use of website for talent and market reach (0.13), as well as the use of ERP and BI integrating the data collected in a unique integrated database (0.15). Table 4.16 reports the correlation matrix of

	Correlation	Std. Err.	t-value	p-value
Y - `CustEngage_Tools`	0.00496	0.00812	0.611	0.5412
Y - `EComm_Presence`	0.04117	0.00811	5.075	3.92×10^{-7}
Y - `Web_Personalization`	0.03411	0.00811	4.204	2.63×10^{-5}
Y - `Web_Talent_MarketReach`	0.13206*	0.00805	16.41	5.37×10^{-60}
Y - `CustMarket_Analytics`	0.09426	0.00808	11.66	2.69×10^{-31}
Y - `ExternalData_Use`	0.06049	0.00810	7.465	8.80×10^{-14}
Y - `IntegratedData_BI_ERP`	0.15191*	0.00802	18.93	5.17×10^{-79}
Y - `Cloud_Infrastructural`	0.08742	0.00809	10.81	3.91×10^{-27}
Y - `Cloud_Basic`	0.06028	0.00810	7.438	1.08×10^{-13}
Y - `Cloud_Mgmt`	0.09841	0.00808	12.18	5.63×10^{-34}
Y - `AI_Cogn_Operations`	0.02129	0.00812	2.623	0.00873
Y - `AI_Adv_Innov`	0.08214	0.00809	10.15	3.90×10^{-24}
Y - `AI_Smart_Logistic`	0.01806	0.00812	2.225	0.0261
Y - `AI_Mgmt`	0.03324	0.00811	4.097	4.21×10^{-5}

Tabella 4.15: Weighted Pearson correlations - factors and `vagg_add`. * if r>0.1.

the factors, a threshold of 0.4 was used for marking the correlations with an asterisk in order to facilitate the visualization of the relevant values. Particularly, MR11 and MR12, reaches the highest figure with 0.61 but it is still remaining below the conventional threshold of 0.7 (often adopted by the literature as a rule of thumb). This suggests that the use of AI in operational and analytical activities, particularly for primary business activities like marketing and production (MR11), is more closely associated with its advanced application in R&D activities (MR12) than with any other pair of factors. As in the MCA case, the overall level of correlation is relatively low, but statistical significance is observed due to the large sample size. In line with the previous discussion and the evidence in the literature, this indicates that multicollinearity is not a concern, and all independent variables (x's) can be retained in the model.

	CustEngage_Tools	EComm_Presence	Web_Personalization	Web_Talent_MarketReach	CustMarket_Analytics	ExternalData_Use	IntegratedData_BI_ERP
CustEngage_Tools	1.00	0.42*	0.36	0.26	0.33	0.14	0.27
EComm_Presence	0.42	1.00	0.27	0.18	0.23	0.09	0.23
Web_Personalization	0.36	0.27	1.00	0.28	0.22	0.11	0.19
Web_Talent_MarketReach	0.26	0.18	0.28	1.00	0.35	0.19	0.41*
CustMarket_Analytics	0.33	0.23	0.22	0.35	1.00	0.37	0.36
ExternalData_Use	0.14	0.09	0.11	0.19	0.37	1.00	0.28
IntegratedData_BI_ERP	0.27	0.23	0.19	0.41*	0.36	0.28	1.00
Cloud_Infrastructural	0.19	0.14	0.16	0.32	0.36	0.21	0.33
Cloud_Basic	0.11	0.09	0.09	0.22	0.20	0.15	0.23
Cloud_Mgmt	0.26	0.19	0.19	0.37	0.48*	0.20	0.47*
AI_Cogn_Operations	0.17	0.11	0.15	0.20	0.31	0.21	0.18
AI_Adv_Innov	0.18	0.11	0.16	0.21	0.30	0.22	0.22
AI_Smart_Logistic	0.10	0.06	0.11	0.13	0.15	0.16	0.14
AI_Mgmt	0.11	0.06	0.11	0.15	0.20	0.18	0.15

Correlation Matrix (Columns 1–7)

	Cloud_Infrastructural	Cloud_Basic	Cloud_Mgmt	AI_Cogn_Operations	AI_Adv_Innov	AI_Smart_Logistic	AI_Mgmt
CustEngage_Tools	0.19	0.11	0.26	0.17	0.18	0.10	0.11
EComm_Presence	0.14	0.09	0.19	0.11	0.11	0.06	0.06
Web_Personalization	0.16	0.09	0.19	0.15	0.16	0.11	0.11
Web_Talent_MarketReach	0.32	0.22	0.37	0.20	0.21	0.13	0.15
CustMarket_Analytics	0.36	0.20	0.48	0.31	0.30	0.15	0.20
ExternalData_Use	0.21	0.15	0.20	0.21	0.22	0.16	0.18
IntegratedData_BI_ERP	0.33	0.23	0.47*	0.18	0.22	0.14	0.15
Cloud_Infrastructural	1.00	0.49*	0.50	0.24	0.24	0.13	0.16
Cloud_Basic	0.49	1.00	0.50*	0.15	0.14	0.08	0.10
Cloud_Mgmt	0.50*	0.50*	1.00	0.24	0.25	0.14	0.16
AI_Cogn_Operations	0.24	0.15	0.24	1.00	0.61*	0.43*	0.57*
AI_Adv_Innov	0.24	0.14	0.25	0.61*	1.00	0.43*	0.50*
AI_Smart_Logistic	0.13	0.08	0.14	0.43*	0.43*	1.00	0.46*
AI_Mgmt	0.16	0.10	0.16	0.57*	0.50*	0.46*	1.00

Tabella 4.16: Correlation Matrix of Variables MR1–MR14 (split into two 14×7 blocks), * indicates r > 0.4).

After examining the correlations between the covariates, it was considered noteworthy to explore how the selected variables differ when grouped by categories of structural characteristics (technology intensity, size, and governance). Are certain factors more frequent in larger enterprises? Are specific technologies more commonly adopted within particular technology-intensity sectors? Below, several plots are presented to address these questions. Furthermore, descriptive analyses have been carried out to show how value added (mean value) per employee is distributed across these structural variables. Looking at the results in Figure 4.15, the main observations are the following:

- High-tech manufacturing enterprises (3), like pharmaceutical or electronic companies, primarily adopt an integrated BI and ERP software, advanced AI for R&D, and cloud for ERP or CRM.
- Mid-tech manufacturing (2), like machinery manufacturers, makes substantial use of integrated data collection through ERP and BI (MR7).
- Utilities and construction enterprises (10) have the highest factorial scores in location data use and cloud for basic business applications (MR6 and MR9).
- Knowledge-intensive services (12), such air or water transport, security and investigation activity mostly use infrastructural and ICT security cloud solutions as well as on customer or market data use.

- High-tech services (13) make intensive use of MR12 and MR11. This indicates that sectors such as audiovisual and broadcasting services, telecommunications, ICT-related services (including programming, consultancy, and information activities), and scientific research and development mainly rely on cognitive (e.g, text mining or image recognition) or self-learning AI (e.g., ML) respectively for primary business operations and R&D.

Hence, it is evident that across different sectoral groups the technologies adopted vary considerably. In Figure 4.16, it is clear that enterprises of size 3 primarily have

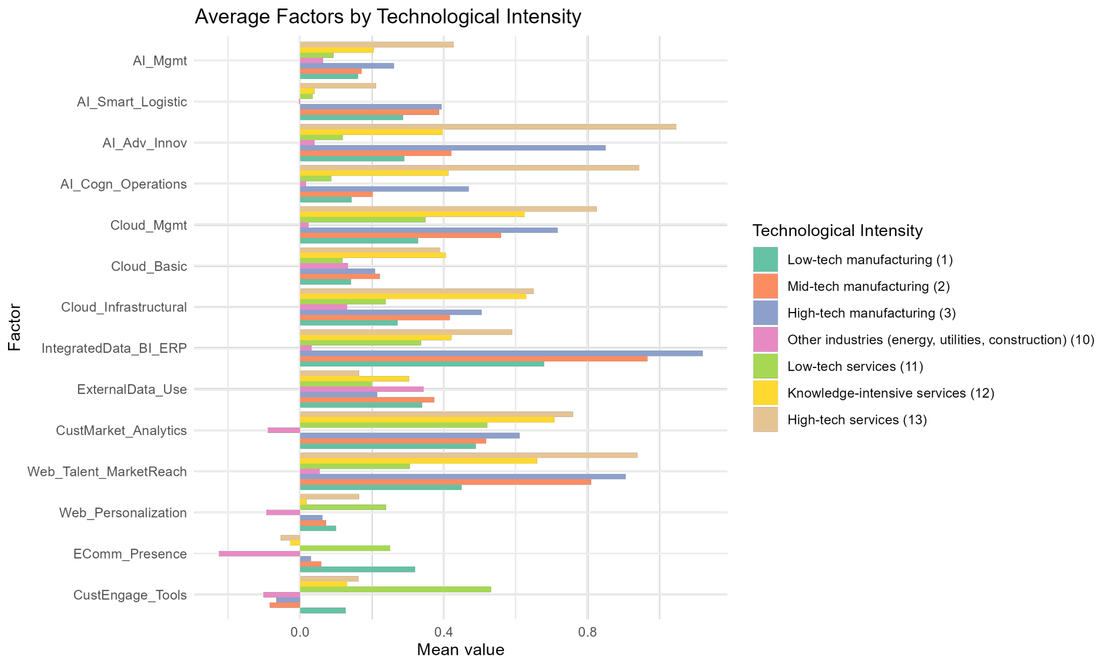

Figura 4.15: Average EFA factors grouped by technology intensity

an ERP and BI integrated database, an intensive customer data use, together with adoption of cloud for ERP and CRM and self-learning AI for R&D (MR7, MR5, MR10 and MR12). The largest enterprises (size 4) shows even greater adoption (higher factors' scores) of the already mentioned technologies. Looking below at Figure 4.17, it can be observed that enterprises with multinational governance (1 and 2) are those adopting the largest numbers for the different digital factors. In particular, both groups show intensive use of ERP and BI with data integration, cloud for ERP and CRM software, and customer or market data use (MR7, MR10, and MR5).

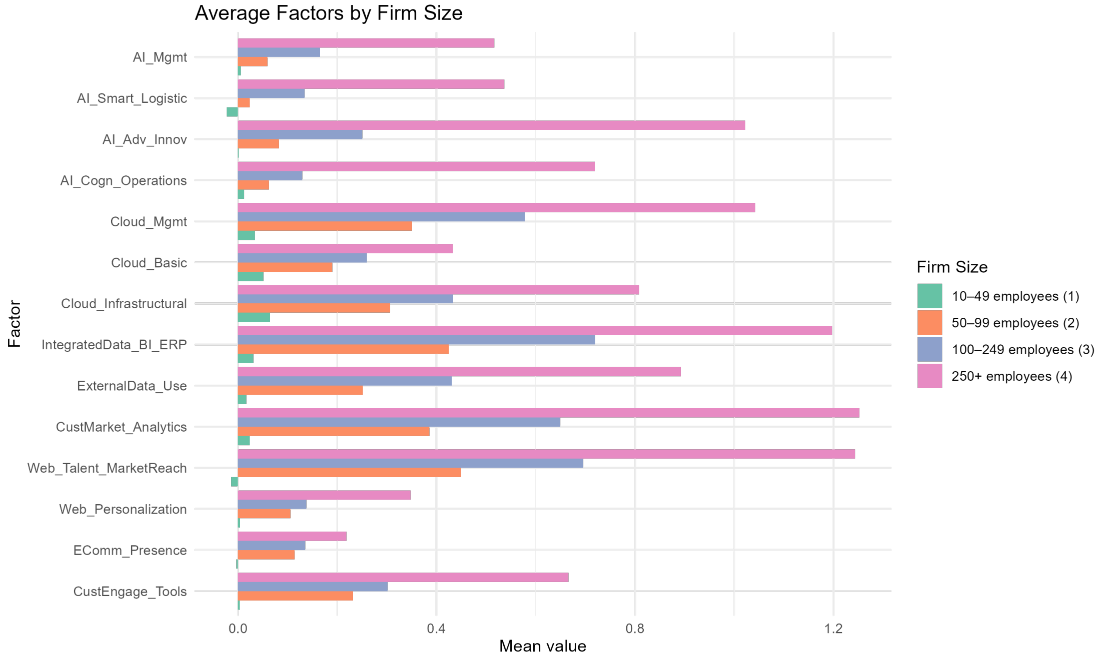

Figura 4.16: Average EFA factors grouped by size

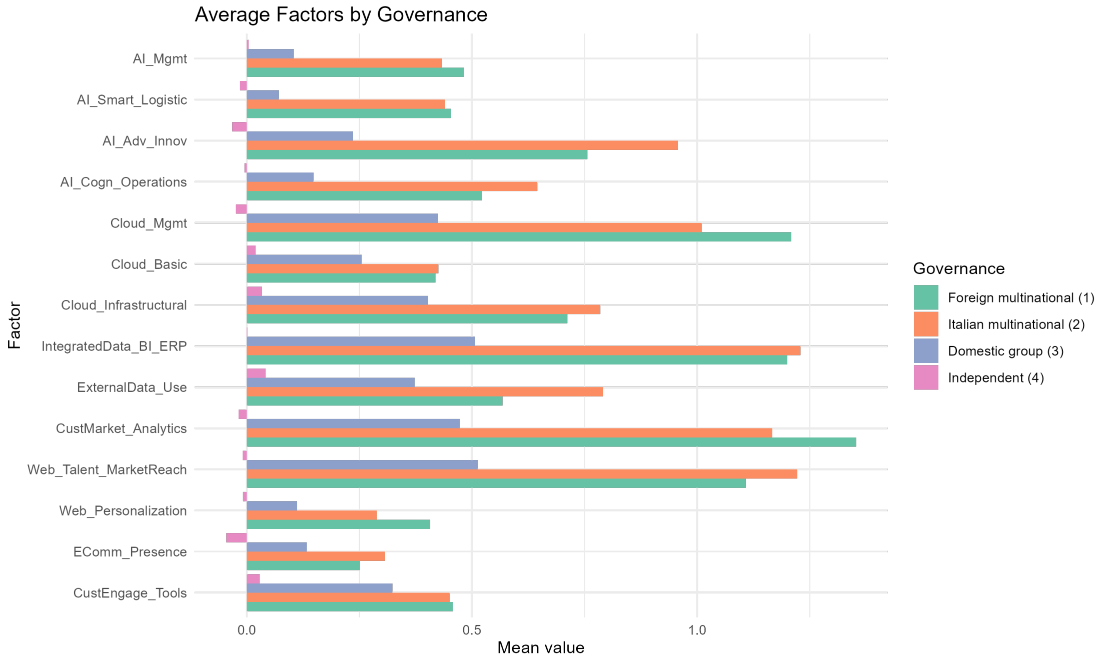

Figura 4.17: Average EFA factors grouped by governance

Second OLS Outcome in Rstudio

For performing the second model the following code chunk is run in RStudio:

```
# 1. Standardization  all MR factors
dati_with_factors_KMO_VM <- dati_with_factors_KMO_VM %>%
  dplyr::mutate(across(dplyr::starts_with("MR"),
                       ~ as.numeric(scale(.)),
                       .names = "{.col}_z"))

# 2. Verify: mean appr. 0,  standard deviation appr. 1
means <-sapply(dplyr::select(dati_with_factors_KMO_VM, dplyr::
    ends_with("_z")), mean, na.rm = TRUE)
sds   <- sapply(dplyr::select(dati_with_factors_KMO_VM, dplyr::
    ends_with("_z")), sd, na.rm = TRUE)

print("Means of standardized factors:")
print(means)

print("Standard deviations of standardized factors:")
print(sds)

# 3. OLS regression  with standardized factors
model_EFA_z <- lm(vagg_add ~ MR1_z + MR2_z + MR3_z + MR4_z + MR5_z
     + MR6_z +
MR7_z + MR8_z + MR9_z + MR10_z + MR11_z +
MR12_z + MR13_z + MR14_z,
                  weights = PESO,
                  data = dati_with_factors_KMO_VM)

summary(model_EFA_z)
```

In contrast, for the second OLS model based on the factors extracted through EFA, standardisation was required. The output of the second OLS model, with the 14 EFA factors as covariates, is reported in Table 4.17 below. The regression model confirms that digital and data-related capabilities are statistically associated with firm performance.

Variable	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	67973.70	776.00	87.599	$< 2e{-}16^{***}$
CustEngage_Tools (MR1_z)	-4766.40	968.10	-4.924	8.58e-07***
EComm_Presence (MR2_z)	1209.10	746.20	1.620	0.10517
Web_Personalization (MR3_z)	-588.40	821.90	-0.716	0.47403
Web_Talent_MarketReach (MR4_z)	8056.20	959.70	8.394	$< 2e{-}16^{***}$
CustMarket_Analytics (MR5_z)	3895.80	1095.60	3.556	0.00038***
ExternalData_Use (MR6_z)	994.40	978.00	1.017	0.30928
IntegratedData_BI_ERP (MR7_z)	9580.70	833.60	11.493	$< 2e{-}16^{***}$
Cloud_Infrastructural (MR8_z)	2821.70	908.30	3.107	0.00190**
Cloud_Basic (MR9_z)	196.90	828.10	0.238	0.81207
Cloud_Mgmt (MR10_z)	169.90	1135.10	0.150	0.88099
AI_Cogn_Operations (MR11_z)	-2883.80	1062.50	-2.714	0.00665**
AI_Adv_Innov (MR12_z)	8831.10	1342.90	6.576	4.99e-11***
AI_Smart_Logistic (MR13_z)	-2136.10	1103.00	-1.937	0.05281·
AI_Mgmt (MR14_z)	241.80	1175.70	0.206	0.83707
Signif. codes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$				

Tabella 4.17: Linear regression results for value added per employee using the 14 standardised EFA factors.

However, looking to Table 4.19, the explained variance remains modest and lightly higher of the one obtained with the first model ($R^2 = 0.03752$, Adjusted $R^2 = 0.03663$). The overall significance of the model ($F(14, 15159) = 42.21$, $p < 2.2 \times 10^{-16}$) suggests that the predictors, when considered jointly, provide meaningful explanatory power despite the limited magnitude of fit.

Statistic	Value	Notes
Min	-7,024,776	—
1Q	-52,480	First quartile
Median	-11,795	—
3Q	44,408	Third quartile
Max	14,225,617	—

Tabella 4.18: Weighted residuals of the second regression model.

Statistic	Value	Notes
Residual Standard Error	303,100	df = 15,159
Multiple R^2	0.03752	—
Adjusted R^2	0.03663	—
F-statistic	42.21	on 14 and 15,159 df
p-value	$< 2.2 \times 10^{-16}$	—

Tabella 4.19: Fit statistics of the OLS regression model using the 14 standardised EFA factors.

Among the factors, the use of integrated ERP and BI data (MR7) emerges as the most strongly associated factor ($\hat{\beta} \approx 9{,}580.70$, $p < 2e^{-16}$), highlighting the central role of ERP and BI systems in consolidating information within unified databases.

This finding is consistent with the distribution of factors across structural variables: large, multinational, and high-tech manufacturing enterprises (characterised by higher average value added per employee) tend to achieve higher scores for MR7, related to the use of ERP and BI adopting one single relational database. Assuming technologies remained unchanged since the end of 2022 (thus removing the time-mismatch limitation), this result suggests that internal data integration has the strongest impact on value creation. However, this effect cannot be quantified in economic units due to the standardised nature of the factor scores.
The use of websites for talent management (e.g., posting job vacancies) and for market purposes (e.g., multi-language websites), represented by MR4, is also significantly and positively associated with productivity ($p < 2e^{-16}$). The multi-languages availability in the website, enhance international market access and the online job vacancies posting increase the probability of attract qualified workforce, that can increase the value added per employee and the LP on the medium term. This factor captures basic but strategic digital infrastructures that directly expand a firm's market opportunities and human capital base. As mentioned in the first regression with *fac1*, sophisticated digital tools are strongly associated with performance in the long run. However, this does not imply that there is no value in more basic technologies. On the contrary, in the short term it is likely that relatively low-cost and more mature digital tools exhibit stronger associations with productivity. Technologies like websites for market and talent reach, provide immediate and tangible benefits with limited adjustment costs, while the positive effects of more complex systems may only materialise after longer periods of integration and organisational adaptation.
Interestingly, MR11, representing the adoption of cognitive AI (e.g., text mining, voice recognition, and text generation) for primary business activities, shows a negative and significant association, while the adoption of self-learning AI for R&D purposes is positive but slightly less significant ($p < 0.00665$). Other AI adoptions are not significant in the model, even though MR13's $p - value$ is very close to the 0.05 threshold. Notably, while AI for logistics (MR13) has a negative coefficient, AI for managerial purposes (MR14) exhibits a positive one. Assuming that AI investments had already been implemented by 2022, negative coefficients may reflect transitional differences or short-term adjustment costs. This interpretation is consistent with existing literature, which shows that the positive effects of AI often materialise with a time lag. For instance, investing in cognitive AI tools may require substantial capital expenditure, with benefits likely to emerge only in the following years. However, **SQ2.4** related to the potential beneficial associations between AI and productivity. The answer is that it depends on the application domain, as the results are heterogeneous. Self-learning AI adopted in R&D exhibits a positive correlation with the dependent variable, whereas cognitive AI applied to primary activities shows a significant negative relationship. The other two

domains do not appear statistically relevant in the model. These findings warrant further investigation using a panel structure (e.g., data for 2023–2025) to test the hypothesis that AI used for innovation generates positive effects earlier, while other applications require more time to deliver benefits and may initially display lagged or even negative impacts.

Customer- and market-related data use, including CRM adoption (MR5) is another key driver ($p < 0.00038$), confirming that the systematic analysis of information (e.g., purchasing behaviour, web data, and social media) is positively associated with performance.

Similarly, infrastructural cloud services (MR8) display a significant and positive association with the dependent variable y ($p < 0.00190$), underscoring the importance of technological infrastructures (e.g., databases, file storage, and ICT security) in enhancing efficiency and resilience. The results found is addressing **SQ2.2**. In aswering it, a distinction must be made: the use of advanced infrastructural cloud services, such as databases, ICT security, computational power, and file storage, shows a positive and significant association with productivity, whereas advanced management-oriented cloud services do not display a significant relationship. For completeness, it should be noted that management cloud services are closely linked to ERP and CRM systems, which themselves show positive associations with productivity. This suggests that management cloud adoption may exert its influence indirectly, through complementarities with these other technologies. At first sight, this may appear to contrast with the results obtained for *fac2*. The difference, however, lies in the interpretation: while *fac2* captures a relative orientation of firms towards infrastructural rather than decision-support technologies, implying lower performance when such tools dominate, MR8 isolates the specific effect of cloud infrastructures and security. These technologies, when adopted alongside other complementary tools, appear to contribute positively to value creation. This interpretation is further supported by the evidence that more decision-support and data-driven technologies, such as ERP and BI (MR7), achieved higher coefficients. In general, the positive relationship between the adoption of infrastructural cloud solutions (i.e., ICT security, databases, platforms, and computational power) and firm performance indicates that sophisticated cloud adoption is more strongly associated with value added than basic cloud adoption. However, more management-based cloud services (e.g., ERP or CRM cloud solutions) do not exhibit any positive association. The factor related to customer engagement tools (e.g., social media) (MR1) shows significant negative effects. Social media engagement tools often require significant time, management, and specialized staff. For some enterprises, the presence on social media might increase their visibility but not directly impacting on productivity, efficiency and value addded per employee. Furthermore another aspect has to be underlines, the sectoral bias: industries with high social media engagement (e.g., retail, tourism) may naturally show lower productivity levels

compared to capital- or knowledge-intensive industries.
Conversely, customer engagement tools and cognitive AI for primary business activities show negative relationships with productivity. With this finding, it has been answered to **SQ2.3**: customer engagement technologies have a negative association probably due to the fact that these tools usually may help to increase the visibility but not directly the value. Contrarily, e-commerce presence is not showing statistically significant associations..
To conclude, same model tuning strategies explained for the first OLS model have been adopted, in this case achieving $R^2 = 0.09486$. While for the first OLS model with MCA variables, deleting the outliers did not change the relevant associations, in this case, doing it showed other positive and relevant associations in addition to the already found ones. Particularly, it resulted a positive significant association between value added per employee and e-commerce presence (MR2), external (geo-spatial) data use (MR6) and AI used for administrative or accountability purposes (MR14). A further negative relevant correlation can be see with use of website for offering personalization services (MR3). It should be emphasized that the removal of outliers can result in the loss of observations that may carry relevant and valuable information. For this reason it has been decided to comment and show just the model with outliers.

Chapter 5

Conclusions

5.1 Summary of Objectives and Approaches

This thesis investigated the relationship between the adoption of digital technologies and enterprise productivity in Italy, measured as value added per employee. The analysis was conducted on data collected by the ICT Survey 2023, the SBS Frame 2022 and the Statistical Register of Enterprise Groups (complemented by FATS for the multinational governance types). The merged framework used contains 97 variables and 15,187 rows (enterprises), subsequently reduced to 15,174 usable observations due to negligible weights. The idea of focusing on the year 2023 is related to the fact that the ICT survey had more interesting and technology-specific variables. Furthermore, the uniqueness of this cross-sectional dataset lies in the fact that it does not only include information on technologies and the typical structural indicators (e.g, turnover or size) collected in ICT surveys regarding enterprises, but also incorporates data on value added and governance. The central research questions were twofold: first, to characterise patterns of ICT adoption among Italian firms, and second, to examine which dimensions of digitalisation are statistically associated with productivity performance, measured through value added per employee. To address these questions, two different weighted dimensionality reduction techniques were applied. Multiple Correspondence Analysis was used to extract underlying dimensions of technological adoption from categorical survey data, while Exploratory Factor Analyses were employed to identify latent constructs within four subsets of technologies, particularly in the following areas: website usage, social media, mobile applications, e-sales, cloud computing, data analytics, and artificial intelligence. For the EFA, two different rotation methods for the extraction of the factors have been adopted: one looking to hidden correlation between factors and technology adoption (`oblimin`) and one trying to obtain orthogonal dimensions prioritizing independency (`varimax`). Subsequently, two Ordinary Least Squares regression models were estimated: one employing the two factors obtained via MCA, and the other using the fourteen latent constructs derived from EFA (`rotation=varimax`). This dual approach allowed for a robust triangulation of findings, balancing parsimony (MCA) with granularity (EFA). In fact, the two dimensionality reduction approaches adopted (MCA and EFA) and their respective OLS regressions results should not be interpreted in isolation but rather as complementary. The MCA provides a synthetic overview of ICT adoption, highlighting broad co-adoption patterns and latent dimensions that explain the majority of variance in firms' behaviours. The EFA, in turn, disentangles these dimensions into more specific and interpretable factors, making it possible to assess which particular technologies drive the associations observed and which show weaker or divergent effects. Taken together, the two techniques offer both a general framework and a detailed lens, thus enabling a more robust interpretation of how ICT adoption relates to productivity.

5.2 Main Empirical Findings

5.2.1 Evidence from MCA

The MCA highlighted a highly polarised technological landscape among Italian enterprises. The first dimension (*Dim1*) explained 88.75% of total inertia and clearly separated firms with high digital sophistication and intensity from those with minimal adoption. The second dimension (*Dim2*), while accounting for a smaller share of variability (8.84%), captured the modality of technological choice, distinguishing firms oriented toward in-house, decision-support and data-driven applications (e.g., BI, AI, CRM) from those relying predominantly on infrastructural and outsourced tools (e.g., cloud services purchased for storage or ICT security). The overall inertia was low (0.0663), suggesting relative homogeneity across most firms, with advanced adoption concentrated in a minority of large, multinational, and high-tech enterprises. Plotting the coordinates of these two dimensions for each of the sixteen active variables analyzed it emerges how some technologies are often used together like:

- e-sales and ERP;
- cloud computing purchased for basic business functionalities coupled with the one for ICT security;
- data sharing and analysis;
- CRM and BI software;

- cloud computing used for management softwares (CRM and ERP) cloud for platform development and computational power.

Furthermore other eight supplementary variables, not considered in the dimensions computation, have been used for interpreting the results. Structural aspects like value added, size, governance and technological intensity are associated with the adoption of certain technologies. Particularly, bigger and multinational firms are likely to adopt more sophisticated digital tools (e.g, AI and advanced softwares). Firms with excellent productive performance tend to adopt AI technologies. However, from the scatterplot it is evident how, advanced or intermediate AI (adopting more than two of this kind of technologies) represent more rare and isolated phenomena. High-tech services (e.g, ICT and IS, TLC and R&D activities) tend to prefer more data-driven technologies like CRM and BI while high-tech manufacturing are close to ERP and e-sales adoption.

5.2.2 Findings from OLS Regression with MCA Factors

Regression analysis confirmed the strong association between sophisticated technology adoption and productivity (`vagg_add`). Both MCA factors were statistically significant and show a positive relationship with value added per employee. Factor 1 (*fac1*), underlines that advanced and integrated technological adoption is positively correlated with higher value added per employee. Factor 2 (*fac2*), capturing the modality of adoption, also showed a positive association, but its interpretation is more nuanced. Higher values of *fac2*, corresponding to data-driven and decision-support systems, were positively associated with productivity. Conversely, lower values of *fac2*, indicating reliance on infrastructural technologies alone, correlated with lower performance, confirming the limited role of such tools when not complemented by advanced applications.

5.2.3 Evidence from EFA

The EFA was conducted on four main survey sections, following an extensive variable selection process based on the outputs of the Bartlett and KMO tests. Using `oblimin` rotation, 11 factors were extracted, which displayed stronger intercorrelations than the 14 factors obtained with varimax rotation. Since the latter are less correlated (due to the orthogonality of the technique) they were employed in the OLS regression. Overall, the factors obtained under both rotations are comparable, confirming the consistency of the latent patterns identified. Compared to the two broad MCA dimensions, the EFA factors are more specific, allowing for more detailed and granular insights. These factors reveal underlying constructs in the data, grouping together different clusters of adopted technologies. Specifically, clustering variables around underlying thematic constructs, they capture different use of: (i) websites (content engagement, personalization, or market/talent reach), (ii) data analyzed (customer behaviour, geospatial information, or business process data), (iii) cloud services purchased (managerial, infrastructural, or basic functionalities), and (iv) AI adoptions (cognitive AI for core business activities, self-learning AI for innovation, and automated AI for logistics).

5.2.4 Findings from OLS Regression with EFA Factors

The regression model using fourteen EFA-derived factors provided further granularity in comparison to the MCA one, enabling a more robust results' interpretation. Though explanatory power remained modest (Adjusted $R^2 \approx 0.036$), the global model was nonetheless highly significant ($F(14, 15159) = 42.21,\ p < 2.2 \times 10^{-16}$). Among the co-variates, five emerged as positively associated with productivity (measured by value added per employee):

- Integrated enterprise data management (MR7) that confirmed the pivotal role of ERP and BI, when the data are collected on a shared database, in consolidating data flows, streamlining operations, and enabling informed decision-making.

- Website usage for talent and market (MR4) reach that shows how even simpler technologies can also benefit enterprises. Particularly in this case, offering a multi-language website or posting job vacancies online, respectively increase a cross-border visibility and the attraction of qualified workforce.

- Self-learning AI for R&D and innovation activities (MR12), underlining the positive power of tools like machine learning. Hypothesizing that this technology was adopted already in 2022, this can be related to two main reasons. First, firms adopting such tools are typically innovation-oriented and already characterised by higher absorptive capacity, which allows them to integrate new technologies more effectively and translate AI-driven insights into efficiency gains. Second, AI adoption in R&D entails lower adjustment costs compared to the integration of cognitive AI in daily operations, since it can be implemented in modular projects without disrupting core activities. As a result, the benefits of AI in R&D might tend to materialise earlier, reinforcing the positive association with value creation despite the general time-lag problem. This aspect requires further investigation, for instance by adopting a panel structure, to assess whether the benefits of AI (e.g., ML, DL or neural network), when applied to new product development or data-analytical research, materialise earlier than those of other AI applications.

- Infrastructural cloud use (MR8) that highlights the importance of robust digital tools like data storage and ICT security in enhancing resilience, efficiency, and scalability. This result, helped us to better interpretate *fac2*. Infrastructural technologies like this one still have a positive correlation on value added per employee if complemented by the adoption of data-driven technologies.

- CRM and customer/market data use (MR5), which reflects the growing strategic importance of customer and market insights for value creation, complemented by customer relationship management software.

Moreover, negative and significant associations with value added are observed for:

- The adoption of customer engagement tools (MR1), which may reflect sectoral bias or the fact that such tools have only indirect effects on productivity.

- The adoption of cognitive AI technologies, such as text mining, image generation, or facial recognition, in primary business activities like production, sales or marketing and (MR11), suggesting that the benefits of advanced technologies may be delayed due to high costs or the need for organisational adjustments.

5.3 Addressing the Research Questions and Hypothesis

In this section, the two initial research questions formulated in the literature review are addressed in light of the empirical evidence obtained through MCA, EFA, and OLS regression analyses. The convergence of MCA and EFA results allows several broader conclusions to be drawn.

RQ1: What ICT patterns govern technological choice decisions in Italy?

Technology adoption in Italy follows a polarised and complementary structure, strongly influenced by structural dimensions like size, governance and sector. Digital sophistication is concentrated in a minority of large and multinational firms, while the majority rely primarily on basic and more infrastructural tools.
The MCA identified two broad dimensions: (i) overall digital intensity, ranging from low to highly advanced adopters, and (ii) the mode of adoption, contrasting in-house, decision-support, and data-driven tools (e.g., BI, CRM, AI) with outsourced infrastructural solutions (e.g., cloud services, ICT security).
The EFA confirmed these dimensions and provided a more granular perspective, revealing latent constructs grouped around: (i) website usage (content engagement, personalisation, market/talent reach), (ii) data analytics (customer behaviour, geospatial, business processes), (iii) cloud adoption purposes (basic, managerial,

infrastructural), and (iv) AI applications (innovation, operational, logistics, managerial support).
Importantly, adoption patterns are not random but shaped by strong complementarities: certain technologies are frequently implemented together, such as ERP with e-sales, basic cloud services with ICT security, data analytics and sharing, BI with CRM. Adoption choices are shaped by sectoral, size, and governance characteristics, with complementarities observed among advanced managerial and customer-oriented technologies. Large, high-tech, and multinational firms are more likely to adopt advanced tools, whereas SMEs and firms in traditional sectors remain largely confined to basic infrastructures.

RQ2: Which technological aspects mostly affect productivity, as measured by value added per employee?

The first regression analysis highlights that productivity gains are strongly associated with advanced, data-driven and decision-support technologies. The results for *fac2* suggest that a predominant reliance on infrastructural technologies is negatively associated with performance. However, in the second OLS model, infrastructural cloud services show a positive association, indicating that when properly complemented, ICT tools such as cloud solutions for file storage or computational power remain valuable contributors to firm performance.
Integrated ERP and BI use (MR7) within one single relational database, emerges as the strongest association, underscoring the central role of enterprise-wide data integration. In addiction to this, AI tools like ML, DL and neural networks used for research and product development purposes are strongly related with value added per employee as well. Contrarily, cognitive-AI adoption, mostly large and multinational enterprise, appear negatively correlated with value added per employee. However, this finding should be interpreted with caution, as the literature shows that AI investments often generate delayed but positive effects on enterprise performance. Additionally, the use of websites for broader market penetration and talent attraction (MR4) is also strongly associated with higher productivity, probably due to its consolidated effects on expanding multinational market presence and facilitating the recruitment of skilled and young individuals. Although simpler than advanced tools such as BI software, this result suggests that even basic technologies can contribute positively to performance when integrated within a broader and more sophisticated digital ecosystem. Having a solid digital foundation appears essential not only for enhancing the effectiveness of advanced tools, but also for enabling basic technologies to benefit from integration with more sophisticated systems.

5.4 Limitations

This study is not without limitations. The technology adoption information collected by the 2023-ICT Survey and used for this thesis lack of information related to ICT training and security, aspects that would have been interesting to insert in the analyses since they have been proved by the literature to play a pivotal role. Furthermore, the cross-sectional design, together with the fact that certain technologies were measured contemporaneously with or even after the outcome variable (value added in 2022), precludes causal inference. Findings should therefore be interpreted as associations rather than causal relationships. Ideally, the dataset would also include value added per employee for subsequent years (e.g., 2025 and beyond), which would allow the assessment of the delayed effects of technology adoption, but such data were not available. A further limitation concerns the dependent variable itself: in some cases, value added per employee takes negative values. This makes impossible to use logarithmic solutions for the model tuning step. As an alternative, turnover per employee (already extractable from the ICT-2023), always positive, could have been used, enabling the application of logarithmic or logit transformations, but at the cost of reducing the effectiveness of the measure as a proxy for productivity. Another limitation relates to the exclusion of two survey sections from the EFA (e-sales and invoicing), as they did not reach sufficient scores on the KMO and Bartlett's tests. This implies that certain aspects could not be explored in the factor analysis. Conversely, the selective approach adopted for the MCA and based on descriptive statistics allowed the inclusion of the most explicative variables, although at the cost of losing additional information. In fact, selecting variables using the correlation matrix, as done for MCA, is useful for reducing redundancy and excluding weakly correlated variables, but it is inherently subjective and limited to bivariate relationships. By contrast, measures such as KMO and Bartlett's test provide more objective and statistically grounded evidence of the overall suitability of the dataset for factor analysis, even though they do not directly point to which variables should be removed. Furthermore, the factors' labelling process was performed and defined by the author, and therefore entails a certain degree of subjectivity. Moreover, the relatively low R^2 values (between 0.03–0.09 even in tuned models) indicate that while digital technologies is significantly associate to performance, they explain only a small share of the variance. This can indicate that there are other elements, not technological, not present in the model that could enable the OLS models to catch higher variance. To support this hypothesis, trying to add structural variables, like size, governance and technology intensity R^2 reach values around 0.13. Indeed, other factors, such as sectoral dynamics, management quality, or macroeconomic conditions, undoubtedly play a significant role. To reinforce this point, even though the high-tech service activities are the one adopting more digital solutions, it has been shown how the highest value added per employee is achieved by high-tech manufacturing. To conclude, outlier exclusion increased explanatory power (R^2) but at the risk of neglecting extreme yet relevant cases, often corresponding to the most technologically advanced or innovative firms.

5.5 Policy and Managerial Implications

The findings hold relevant implications for both policymakers and managers. For policymakers, the evidence supports the design of targeted interventions aimed at reducing the significant digital divide across firm size, geography, and governance structures. Enterprises with fewer than 50 employees, characterised by independent governance and operating in low-tech sectors, which together represent 61% of the Italian industrial context, appear systematically disadvantaged in the adoption of advanced technologies. Moreover, the analysis of 2023 data shows that only 5% of enterprises adopt AI tools, and, of the aforementioned, merely 3% report the use of more than four distinct AI technologies. Among the remaining 95% of non-adopters, just 4% declared an intention to invest in AI, but refrained mainly due to a lack of internal competencies or the high costs involved. All these evidences together underscore the need for targeted financial incentives, training programmes, and infrastructure investments.
These findings suggest that managers should prioritise investments in integrated data management systems and decision-support technologies, while recognising that infrastructural tools are necessary enablers rather than productivity drivers in themselves. Cloud adoption is important, but its benefits are maximised when embedded within broader digital strategies, for example, when combined with applications such as CRM and BI. Another emerged aspect it is the value of analyzing customer-data and its positive association with productivity. By contrast, AI adoption requires careful evaluation in light of an organisation's readiness and the specific domains in which it is applied. Effective digitalisation therefore requires not only technology adoption but also parallel investments in skills, organisational readiness, and strategic alignment, ensuring that advanced tools are embedded into broader value-creation processes.

5.6 Future Research Directions

While the present analysis provides novel insights into the relationship between digitalisation and enterprise productivity, several avenues for further investigation remain open. A first step would be to extend the analysis to longitudinal data once available. A panel structure would allow for disentangling causality from correlation, overcoming the temporal mismatch between outcome variables (2022) and explanatory variables (2023) that currently constrains interpretation. A relevant constraint emerges from the fact that the ICT survey questions (and consequently the variables in the Istat dataset) change from year to year, making it difficult to consistently track the temporal evolution of certain variables. Second, more granular sectoral analyses could enrich the findings. The present study has identified systematic differences across technological intensity, size, and governance categories; however, the dynamics within specific industries (e.g., pharmaceuticals, ICT services, traditional manufacturing) remain underexplored. Industry-level case studies could therefore complement the quantitative evidence presented here. Third, the integration of complementary dimensions, such as workforce skills, organisational practices, and innovation strategies, which are not included in the current survey, could substantially enhance explanatory power. The relatively low R^2 values observed in the regression models underline the importance of non-technological factors, suggesting that productivity results from the interaction between digital adoption and broader organisational capabilities. Future analyses could also benefit from integrate additional alternative clustering techniques to better capture heterogeneous digitalisation patterns, as well as from the development of a new digital index based on the factors extracted. In addition, the heterogeneous results found for artificial intelligence indicate the need for deeper investigation into its implementation contexts. Future research should explore not only whether firms adopt AI, but also which organisational complementarities (e.g., workforce skills) are required, and what risks or unintended effects may arise. Lastly, an interesting avenue for further research would be to compare the factors identified through the two dimensionality reduction techniques (MCA and EFA) across other European countries (e.g., Germany, France, Spain). Such a cross-country investigation would provide a broader perspective on how Italy is positioned relative to its main counterparts in terms of digital adoption and its association with productivity. Overall, these extensions would not only strengthen causal inference and robustness but also provide a more comprehensive understanding of the mechanisms through which digitalisation translates into economic performance.

References

[1] Eurostat. *Digitalisation in Europe 2025 - Interactive Publication*. Accessed: 2025-06-30. 2025. URL: https://ec.europa.eu/eurostat/web/interactive-publications/digitalisation-2025 (cit. a p. 1).

[2] M. L. Giorgetti and L. Zirulia, cur. *Digital Transition and the European Industrial Policy*. a cura di. Milano: Milano University Press, 2025. DOI: 10.54103/milanoup.180. URL: https://doi.org/10.54103/milanoup.180 (cit. alle pp. 1, 21, 24, 27, 29, 30).

[3] M. Bragagnolo. «The role of supply chains». In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 39–42. ISBN: 9791255101659. DOI: 10.54103/milanoup.180.c268. URL: http://dx.doi.org/10.54103/milanoup.180.c268 (cit. a p. 1).

[4] Oxford English Dictionary. *Digitalization*. https://www.oed.com/dictionary/digitalization_n1?tl=true&tab=factsheet. Accessed: 2025-07-07. 2024 (cit. a p. 1).

[5] P. C. Verhoef, T. Broekhuizen, Y. Bart, A. Bhattacharya, J. Qi Dong, N. Fabian and M. Haenlein. «Digital transformation: A multidisciplinary reflection and research agenda». In: *Journal of Business Research* 122 (gen. 2021), pp. 889–901. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2019.09.022. URL: http://dx.doi.org/10.1016/j.jbusres.2019.09.022 (cit. a p. 2).

[6] T. Vallone. «The digitalization of the supply chain». In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 43–48. ISBN: 9791255101659. DOI: 10.54103/milanoup.180.c269. URL: http://dx.doi.org/10.54103/milanoup.180.c269 (cit. alle pp. 2, 27).

[7] M. M. A. P. Mick, J. L. Kovaleski and D. M. G. Chiroli. «Sustainable Digital Transformation Roadmaps for SMEs: A Systematic Literature Review». In: *Sustainability* 16.19 (ott. 2024), p. 8551. ISSN: 2071-1050. DOI: 10.3390/su16198551. URL: http://dx.doi.org/10.3390/su16198551 (cit. alle pp. 2, 28).

[8] Y. K. Dwivedi et al. «Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action». In: *International Journal of Information Management* 63 (apr. 2022), p. 102456. ISSN: 0268-4012. DOI: 10.1016/j.ijinfomgt.2021.102456. URL: http://dx.doi.org/10.1016/j.ijinfomgt.2021.102456 (cit. alle pp. 2, 28).

[9] R. Martínez-Peláez et al. «Sustainable Digital Transformation for SMEs: A Comprehensive Framework for Informed Decision-Making». In: *Sustainability* 16.11 (mag. 2024), p. 4447. ISSN: 2071-1050. DOI: 10.3390/su16114447. URL: http://dx.doi.org/10.3390/su16114447 (cit. alle pp. 2, 27, 28).

[10] F. Bonfiglio. «A New Age in Digital Trust». In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 103–115. ISBN: 9791255101659. DOI: 10.54103/milanoup.180.c275. URL: http://dx.doi.org/10.54103/milanoup.180.c275 (cit. alle pp. 2, 31).

[11] M. Florio. «A European Alternative to Tech Giants in the Public Interest». In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 93–101. ISBN: 9791255101659. DOI: 10.54103/milanoup.180.c274. URL: http://dx.doi.org/10.54103/milanoup.180.c274 (cit. alle pp. 2, 30).

[12] M. Berlinguer. «EU Cloud Policy requires an innovative governance». In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 81–91. ISBN: 9791255101659. DOI: 10.54103/milanoup.180.c273. URL: http://dx.doi.org/10.54103/milanoup.180.c273 (cit. alle pp. 2, 30).

[13] OECD. *Empowering SMEs for Digital Transformation and Innovation: The Italian Way*. Accessed: 2025-07-26. 2024. URL: https://www.oecd.org/en/blogs/2024/06/empowering-smes-for-digital-transformation-and-innovation-the-italian-way.html (cit. a p. 3).

[14] ISTAT. *Imprese e ICT – Anno 2024*. Tavole statistiche in formato XLSX. 2024. URL: https://www.istat.it/comunicato-stampa/imprese-e-ict-anno-2024/ (visitato il 24/09/2025) (cit. alle pp. 3, 5–12, 14).

[15] Istituto Nazionale di Statistica (ISTAT). *Statistiche Report - Imprese e ICT - Anno 2024*. Rapp. tecn. Accessed: 2025-07-09. ISTAT, gen. 2025. URL: https://www.istat.it/wp-content/uploads/2025/01/Statreport_ICT2024-1.pdf (cit. alle pp. 4, 37).

[16] Istituto Nazionale di Statistica (ISTAT). *Statistiche Report - Imprese e ICT - Anno 2023*. Rapp. tecn. Accessed: 2025-07-24. ISTAT, gen. 2024. URL: https://www.istat.it/it/files/2023/12/report-imprese_2023.pdf (cit. alle pp. 12, 35, 36).

[17] *Rapporto annuale sulla produttività 2025*. Rapp. tecn. Rapporto congiunto su produttività, eterogeneità d'impresa e trasformazione digitale. Roma: CNEL; ISTAT; Banca d'Italia, 2025 (cit. alle pp. 15, 16).

[18] F. Nucci, C. Puccioni and O. Ricchi. «Digital technologies and productivity: A firm-level investigation». In: *Economic Modelling* 128 (nov. 2023), p. 106524. ISSN: 0264-9993. DOI: `10.1016/j.econmod.2023.106524`. URL: `http://dx.doi.org/10.1016/j.econmod.2023.106524` (cit. a p. 16).

[19] M. Borowiecki, J. Pareliussen, D. Glocker, E. J. Kim, M. Polder and I. Rud. *The impact of digitalisation on productivity: Firm-level evidence from the Netherlands*. OECD, set. 2021. DOI: `10.1787/e800ee1d-en`. URL: `http://dx.doi.org/10.1787/e800ee1d-en` (cit. a p. 16).

[20] B. (editor) Anghel et al. *Digitalisation and productivity: A report by the ESCB expert group on productivity, innovation and technological change*. 339. European Central Bank, 2024 (cit. alle pp. 16, 17).

[21] P. Koutroumpis. «The economic impact of broadband: Evidence from OECD countries». In: *Technological Forecasting and Social Change* 148 (nov. 2019), p. 119719. ISSN: 0040-1625. DOI: `10.1016/j.techfore.2019.119719`. URL: `http://dx.doi.org/10.1016/j.techfore.2019.119719` (cit. a p. 17).

[22] Carlo Cambini, Elena Grinza and Lorien Sabatino. «Ultra-fast broadband access and productivity: Evidence from Italian firms». In: *International Journal of Industrial Organization* 86 (2023), p. 102901. DOI: `10.1016/j.ijindorg.2022.102901` (cit. a p. 17).

[23] Wolfgang Briglauer, Carlo Cambini, Klaus Gugler and Lorien Sabatino. «Economic benefits of new broadband network coverage and service adoption: evidence from OECD member states». In: *Industrial and Corporate Change* (2025), pp. 1–26. DOI: `10.1093/icc/dtae043` (cit. a p. 17).

[24] L. Hurbean, A. Dodescu, S. Briciu et al. «The Impact of Business Intelligence and Analytics on Managerial Effectiveness and Work Performance». In: *Journal of Business Research or related (exact source to be verified)* (2023). Available at ResearchGate: `https://www.researchgate.net/publication/373785865_The_Impact_of_Business_Intelligence_and_Analytics_on_Managerial_Effectiveness_and_Work_Performance` (cit. a p. 18).

[25] Eurostat. *Glossary: Data analytics*. `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Data_analytics`. Accessed: YYYY-MM-DD. 2024 (cit. a p. 18).

[26] S. Bianchini and L. Michalkova. *Data Analytics in SMEs: Trends and Policies*. Rapp. tecn. OECD, 2019. DOI: `10.1787/18151965`. URL: `https://www.oecd-ilibrary.org/science-and-technology/data-analytics-in-smes-trends-and-policies_18151965` (cit. a p. 18).

[27] Belgin Tuncay. «Effects of Business Intelligence Techniques on Enterprise Productivity». In: *International Journal of Business and Management Studies* 2.1 (2010), pp. 105–114. URL: https://www.researchgate.net/publication/46471303_Effects_of_Business_Intelligence_Techniques_on_Enterprise_Productivity (cit. a p. 18).

[28] T. Kgakatsi, K. Molefi and K. Kelepile. «The Impact of Big Data on SME Performance: A Systematic Review». In: *Businesses* 4.1 (2024), p. 38. DOI: 10.3390/businesses4010038. URL: https://www.mdpi.com/2673-7116/4/1/38 (cit. a p. 18).

[29] Jude Kisang Asongwe. «The Significance of Data Analysis and Data Analytics in Enterprise Resource Planning and Management». In: (2023). Available at ResearchGate: https://www.researchgate.net/publication/374367814_The_significance_of_Data_Analysis_and_Data_Analytics_in_Enterprise_Resource_Planning_and_Management (cit. a p. 18).

[30] Eurostat. *Large enterprises used more e-business applications.* https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20240516-2. Accessed: 2025-07-30. 2024 (cit. a p. 19).

[31] Zdzislaw Polkowski, Joanna Mazurkiewicz and Henryk Krawczyk. «The Integration of BI, ERP and CRM Systems». In: *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS).* IEEE, 2016, pp. 1417–1421. DOI: 10.15439/2016F478. URL: https://ieeexplore.ieee.org/document/7733340 (cit. a p. 19).

[32] Eurostat. *Cloud computing – statistics on the use by enterprises.* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cloud_computing_-_statistics_on_the_use_by_enterprises. Accessed: 2025-07-30. 2024 (cit. a p. 19).

[33] M. Willie. «Revolutionizing Employee Productivity through Cloud-Enabled Innovation». In: *Journal of Human Resource and Sustainability Studies* 12.2 (2024), pp. 45–58. DOI: 10.4236/jhrss.2024.122004. URL: https://www.scirp.org/journal/paperinformation.aspx?paperid=136993 (cit. a p. 20).

[34] Tomaso Duso and Alexander Schiersch. «The Impact of Cloud Computing on Firm Productivity: Causal Evidence from Germany». In: *Information Economics and Policy* 62 (2025), p. 100536. DOI: 10.1016/j.infoecopol.2025.100536. URL: https://doi.org/10.1016/j.infoecopol.2025.100536 (cit. a p. 20).

[35] P. Schulze Brock, A. Katsinis, J. Lagüera González, L. Di Bella, L. Odenthal, M. Hell, B. Lozar and B. Secades Casino. *Annual Report on European SMEs 2024/2025: SME Performance Review*. JRC142263. Luxembourg: Publications Office of the European Union, 2025. DOI: 10.2760/7714438. URL: https://data.europa.eu/doi/10.2760/7714438 (visitato il 07/07/2025) (cit. a p. 20).

[36] European Commission. *SME Definition*. https://single-market-economy.ec.europa.eu/smes/sme-fundamentals/sme-definition_en. Accessed: 2025-07-08. 2020 (cit. a p. 21).

[37] M. Bianchini and M. Lasheras Sancho. *SME digitalisation for competitiveness: The 2025 OECD D4SME Survey*. OECD, apr. 2025. DOI: 10.1787/197e3077-en. URL: http://dx.doi.org/10.1787/197e3077-en (cit. alle pp. 21, 22, 28, 29).

[38] A. Da Rocha, L. Neves da Fonseca and C. Kogut. «Small firm internationalization using digital platforms: an assessment and future research directions». In: *International Marketing Review* 41.5 (ago. 2024), pp. 981–1015. ISSN: 0265-1335. DOI: 10.1108/imr-10-2023-0280. URL: http://dx.doi.org/10.1108/IMR-10-2023-0280 (cit. a p. 22).

[39] Eurostat. *Usage of AI technologies increasing in EU enterprises*. Accessed: 2025-07-07. 2025. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20250123-3 (cit. a p. 22).

[40] R. Bommasani et al. *On the Opportunities and Risks of Foundation Models*. 2021. DOI: 10.48550/ARXIV.2108.07258. URL: https://arxiv.org/abs/2108.07258 (cit. a p. 23).

[41] L. Abrardi, C. Cambini and L. Rondi. «Artificial intelligence, firms and consumer behavior: A survey». In: *Journal of Economic Surveys* (2021). DOI: 10.1111/joes.12455. URL: https://doi.org/10.1111/joes.12455 (cit. alle pp. 23–25).

[42] I. M. Cockburn, R. Henderson and S. Stern. «The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis». In: *The Economics of Artificial Intelligence*. University of Chicago Press, 2019, pp. 115–148. ISBN: 9780226613475. DOI: 10.7208/chicago/9780226613475.003.0004. URL: http://dx.doi.org/10.7208/chicago/9780226613475.003.0004 (cit. a p. 24).

[43] Stefano Da Empoli, Matteo Cassoli, Alessandro D'Amato, Maria Rosaria Della Porta and Domenico Salerno. *L'impresa dell'IA: Come le PMI italiane possono diventare più competitive grazie all'intelligenza artificiale*. Technical Report. Accessed: 2025-09-19. Istituto per la Competitività (I-Com), 2025.

URL: https://www.i-com.it/wp-content/uploads/2025/06/Studio-I-Com-TeamSystem_LIMPRESA-DELLIA-1.pdf (cit. alle pp. 24, 77).

[44] Competition and Markets Authority. *AI Foundation Models Update paper*. https://www.gov.uk/government/publications/ai-foundation-models-update-paper. Accessed: 2025-07-07. 2024 (cit. a p. 25).

[45] I. van de Poel and L. M. M. Royakkers. *Ethics, Technology, and Engineering: An Introduction*. Chapter 3. Wiley-Blackwell, 2011. ISBN: 978-1-4443-3094-6 (cit. a p. 25).

[46] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter and L. Floridi. «The ethics of algorithms: Mapping the debate». In: *Big Data & Society* 3.2 (dic. 2016). ISSN: 2053-9517. DOI: 10.1177/2053951716679679. URL: http://dx.doi.org/10.1177/2053951716679679 (cit. a p. 25).

[47] B. Friedman and H. Nissenbaum. «Bias in computer systems». In: *ACM Transactions on Information Systems* 14.3 (lug. 1996), pp. 330–347. ISSN: 1558-2868. DOI: 10.1145/230538.230561. URL: http://dx.doi.org/10.1145/230538.230561 (cit. a p. 25).

[48] J. M. Durán and G. Pozzi. «Trust and Trustworthiness in AI». In: *Philosophy & Technology* 38.1 (feb. 2025). ISSN: 2210-5441. DOI: 10.1007/s13347-025-00843-2. URL: http://dx.doi.org/10.1007/s13347-025-00843-2 (cit. alle pp. 25, 26).

[49] D. Danks and A. J. London. «Algorithmic Bias in Autonomous Systems». In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. IJCAI-2017. International Joint Conferences on Artificial Intelligence Organization, ago. 2017, pp. 4691–4697. DOI: 10.24963/ijcai.2017/654. URL: http://dx.doi.org/10.24963/ijcai.2017/654 (cit. a p. 26).

[50] L. Haefner and R. Sternberg. «Spatial implications of digitization: State of the field and research agenda». In: *Geography Compass* 14.12 (set. 2020). ISSN: 1749-8198. DOI: 10.1111/gec3.12544. URL: http://dx.doi.org/10.1111/gec3.12544 (cit. a p. 26).

[51] N.J. Foss. «The Impact of the Covid-19 Pandemic on Firms' Organizational Designs». In: *Journal of Management Studies* 58.1 (ott. 2020), pp. 270–274. ISSN: 1467-6486. DOI: 10.1111/joms.12643. URL: http://dx.doi.org/10.1111/joms.12643 (cit. a p. 26).

[52] J. Camhi, K. Sneader and B. Sternfels. *How COVID-19 has pushed companies over the technology tipping point—and transformed business forever*. McKinsey & Company. 5 Ott. 2020. URL: https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever (visitato il 07/07/2025) (cit. a p. 26).

[53] M. Apedo-Amah et al. *Unmasking the Impact of COVID-19 on Businesses: Firm Level Evidence from Across the World. World Bank Group Policy Research Working Paper No. 9434*. World Bank Group. Ott. 2020. URL: https://documents1.worldbank.org/curated/en/399751602248069405/pdf/Unmasking-the-Impact-of-COVID-19-on-Businesses-Firm-Level-Evidence-from-Across-the-World.pdf (visitato il 07/07/2025) (cit. a p. 26).

[54] W. Dyba and E. Di Maria. «Assessment and support of the digitalization of businesses in Europe during and after the COVID-19 pandemic». In: *Regional Science Policy & Practice* 16.1 (gen. 2024), p. 12717. ISSN: 1757-7802. DOI: 10.1111/rsp3.12717. URL: http://dx.doi.org/10.1111/rsp3.12717 (cit. alle pp. 26, 29).

[55] Eurostat. *Impact of COVID-19 on e-sales of enterprises*. European Commission. 15 Dic. 2022. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Impact%5C_of%5C_COVID-19%5C_on%5C_e-sales%5C_of%5C_enterprises (visitato il 07/07/2025) (cit. alle pp. 26, 27).

[56] C. A. Carnevale Maffè. «Digital supply chains and the industrial metaverse». In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 49–60. ISBN: 9791255101659. DOI: 10.54103/milanoup.180.c270. URL: http://dx.doi.org/10.54103/milanoup.180.c270 (cit. a p. 27).

[57] European Commission. *Europe's Digital Decade: Digital Targets for 2030*. Accessed on 2025-07-07. 2024. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030%5C_en (visitato il 07/07/2025) (cit. a p. 29).

[58] Eurostat. *Digitalisation in Europe – 2025 edition*. Accessed on 2025-07-07. 2024. URL: https://ec.europa.eu/eurostat/web/interactive-publications/digitalisation-2025 (visitato il 07/07/2025) (cit. a p. 29).

[59] Eurostat. *Digital Intensity by size class of enterprise (isoc_e_dii) - Reference: Metadata*. Accessed on 2025-07-07. 2024. URL: https://ec.europa.eu/eurostat/cache/metadata/en/isoc%5C_e%5C_dii%5C_esmsip2.htm (visitato il 07/07/2025) (cit. a p. 29).

[60] European Commission. *Italy 2025 Digital Decade Country Report*. Accessed: 2025-07-07. 2025. URL: https://digital-strategy.ec.europa.eu/en/factpages/italy-2025-digital-decade-country-report (cit. a p. 29).

[61] Content European Commission: Directorate-General for Communications Networks, Technology, OMDIA and Point Topic. *Broadband coverage in Europe 2024 – Mapping progress towards the coverage objectives of the digital decade – Final report*. Final report. Accessed: 2025-07-07. Publications Office of the European Union, 2025. DOI: 10.2759/441597. URL: https://data.europa.eu/doi/10.2759/441597 (cit. a p. 29).

[62] European Commission. *The Digital Europe Programme*. Accessed: 2025-07-07. 2025. URL: https://digital-strategy.ec.europa.eu/en/activities/digital-programme (cit. a p. 29).

[63] Friso Bostoen. «Understanding the Digital Markets Act». In: *Antitrust Bulletin* 68.2 (2023). Tilburg Law School Research Paper, Available at SSRN: https://ssrn.com/abstract=4440819, pp. 263–306. DOI: 10.2139/ssrn.4440819 (cit. a p. 29).

[64] European Commission. *Digital Markets Act: ensuring fair and open digital markets*. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en. Accessed: 2025-07-16. 2022 (cit. a p. 29).

[65] Manuel Wörsdörfer. «The Digital Markets Act and E.U. Competition Policy: A Critical Ordoliberal Evaluation». In: *Philosophy of Management* 22.3 (2022), pp. 1–23. DOI: 10.1007/s40926-022-00213-4 (cit. a p. 29).

[66] M. Almada and N. Petit. «The EU AI Act: Between Product Safety and Fundamental Rights». In: *SSRN Electronic Journal* (2022). ISSN: 1556-5068. DOI: 10.2139/ssrn.4308072. URL: http://dx.doi.org/10.2139/ssrn.4308072 (cit. a p. 29).

[67] M. Kop. *EU Artificial Intelligence Act: The European Approach to AI*. https://www.researchgate.net/publication/354852375_EU_Artificial_Intelligence_Act_The_European_Approach_to_AI. Accessed: 2025-07-07. 2021 (cit. a p. 29).

[68] D. Krause. «The EU AI Act and the Future of AI Governance: Implications for U.S. Firms and Policymakers». en. In: *Unpublished* (2025). DOI: 10.13140/RG.2.2.35108.18569. URL: https://rgdoi.net/10.13140/RG.2.2.35108.18569 (cit. a p. 29).

[69] Charlotte Siegmann and Markus Anderljung. *The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market*. 2022. DOI: 10.48550/ARXIV.2208.12645. URL: https://arxiv.org/abs/2208.12645 (cit. a p. 29).

[70] Governo Italiano Presidenza del Consiglio dei Ministri. *Piano Nazionale di Ripresa e Resilienza (PNRR)*. PDF report. Accessed 2025-07-07. 2021. URL: `https://www.governo.it/sites/governo.it/files/PNRR%5C_0.pdf` (cit. a p. 29).

[71] Adriana Lotti. «NRRP and Digital Adoption: What Effects on the Country's System?» In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 75–80. ISBN: 9791255101659. DOI: `10.54103/milanoup.180.c272`. URL: `http://dx.doi.org/10.54103/milanoup.180.c272` (cit. a p. 29).

[72] Maria Pia Abbracchio. «Foreword». In: *Digital transition and the European industrial policy*. Milano University Press, dic. 2024, pp. 7–9. ISBN: 9791255101659. DOI: `10.54103/milanoup.180.c264`. URL: `http://dx.doi.org/10.54103/milanoup.180.c264` (cit. a p. 30).

[73] ISTAT. *Sistema Informativo Frame Territoriale*. 2024. URL: `https://www.istat.it/scheda-qualita/sistema-informativo-frame-territoriale/` (cit. a p. 41).

[74] ISTAT. *Registro Statistico dei Gruppi d'Impresa*. 2024. URL: `https://www.istat.it/scheda-qualita/registro-statistico-dei-gruppi-dimpresa/` (cit. alle pp. 41, 42).

[75] Eurostat. *Glossary: Foreign Affiliates Statistics (FATS)*. `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Foreign_affiliates_statistics_(FATS)`. Accessed: YYYY-MM-DD (cit. a p. 42).

[76] StataCorp. *mca — Multiple and joint correspondence analysis*. Accessed: 2025-09-15. College Station, TX, USA, 2023. URL: `https://www.stata.com/manuals/mvmca.pdf` (cit. alle pp. 45, 46, 62).

[77] Andy Field, Jeremy Miles and Zoe Field. «Discovering Statistics Using R». In: 1st. London: SAGE Publications, 2012. Cap. Exploratory Factor Analysis, pp. 749–811. ISBN: 978-1-4462-0045-2 (cit. alle pp. 45, 63, 64).

[78] European Commission. *List of NACE codes*. Accessed: 2025-09-10. 2010. URL: `https://ec.europa.eu/competition/mergers/cases/index/nace_all.html` (cit. alle pp. 54, 55).

[79] Eurostat. *High-tech industry and Knowledge–intensive services: Annex 3 – High-tech aggregation by NACE Rev.2*. Rapp. tecn. Accessed: 2025-09-19. Eurostat, 2023. URL: `https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf` (cit. a p. 55).

[80] Eurostat. *Businesses in the manufacturing sector*. Data extracted in January 2025. 2025. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Businesses_in_the_manufacturing_sector (cit. alle pp. 93, 94).

[81] DataCamp. *Multicollinearity in Regression: How to See and Fix Issues*. Accessed: 2025-08-02. 2024. URL: https://www.datacamp.com/tutorial/multicollinearity (cit. a p. 97).

[82] Author(s) of the chapter. «Linear Regression». In: *People Analytics for the Workforce: A Guide to Data-Driven HR*. Springer, 2023. Cap. 10, pp. -–-. DOI: 10.1007/978-3-031-28674-2_10. URL: https://link.springer.com/chapter/10.1007/978-3-031-28674-2_10 (cit. a p. 97).