ISSN: 2282-4189

LUISS

Research Center
for European Analysis
and Policy

Uncertainty and Investments in Data and R&D

Siavash Mohades and Maria Savona

Working Paper 9/2025

November 6, 2025

Uncertainty and Investments in Data and R&D*

Siavash Mohades[1†] Maria Savona[2,3]

[1]Maastricht University, [2]Luiss Guido Carli, [3]University of Sussex

Abstract

This paper investigates whether investments in data influence firms' R&D and whether the two are productivity-enhancing complements. We conceptualise and test whether investments in data reduce market uncertainty, thereby mitigating the inherent uncertainty of R&D and enhancing research and innovation investment. Using Italian firm-level data from 2002 to 2024 and exploiting the GDPR as an instrument, we identify a positive causal effect of data on R&D investment. Moreover, we find that data and R&D are complementary in enhancing both short- and long-term productivity. Our analyses also identify a positive role of R&D for productivity only when firms are data-intensive.

Keywords: Data, R&D, Digitalisation, Innovation, Productivity, Uncertainty

JEL Codes: D22, D25, D82, O31, O33

*This project is funded by Progetti di Rilevante Interesse Nazionale (PRIN) titled "The Economics of Data" awarded to Luiss Guido Carli University and was previously published as the CESifo Working Paper 12230. We also thank (in alphabetical order) Carol Corrado, Diane Coyle, Marco Grazzi, Jonathan Haskel, Lenard Lieb, Francesca Lotti, Valentina Meliciani, Fabio Pieri, Mercedes Teruel, and Tania Treibich for their valuable comments, which significantly improved this paper. Our gratitude also goes to our colleagues at Maastricht University and Luiss Guido Carli, as well as conference participants in ENGAGE, ENTeG, Productivity Research Conference and Global INTAN-Invest Conferences.

†Corresponding author. s.mohadesforooshani@maastrichtuniversity.nl

1 Introduction

Digital transformation (digitalisation) and data-intensive processes (datafication) underpin the 4th Industrial Revolution[1] and have increased the importance of the role of data as a core production factor, delivering substantial firm-level gains in efficiency and innovation (Björkdahl, 2020). Digitalisation is the transformation of information into digital formats that are easier to access, transfer, and analyse (Aronica, Piacentino and Picone, 2022). Datafication refers to the increasingly prominent collection, storage, and software-based analysis of information (Bounie, Dubus and Waelbroeck, 2021), that contribute to firm's strategic knowledge, service optimisation, and value generation (Mayer-Schönberger and Cukier, 2013; Klein et al., 2022). Digitalisation and datafication might also affect firms' investment strategies in Research and Development (R&D). Investing in data might represent a trade-off with respect to investing in R&D from a financial perspective. Or, by improving information access, investing in data might reduce uncertainty and thereby complement and stimulate investments in R&D. Furthermore, the interaction of higher data intensity and R&D can affect firms' productivity performance by mitigating the intrinsic uncertainty's costs of innovation.

This article addresses two related questions. First, we investigate whether investment in data leads to higher R&D expenditure, based on the conjecture that enhanced information reduces uncertainty and might lead firms to "afford" higher investment in R&D, and potentially make firms more efficient (Brynjolfsson, Hitt and Kim, 2011). The accumulation of databases as a new production factor analysed through Machine Learning (ML), generates information and analytical intelligence that can be used to reduce the intrinsic uncertainty linked to innovation (Savona, 2019). Unlike previous Information and Communication Technologies (ICT) waves, data-driven technologies enable firms to acquire information on a larger scale and faster speed, thereby reducing uncertainties in inventory management, marketing, and production inputs (Frankel and Kamenica, 2019). Reduced uncertainty can, in turn, lower costs, stimulate firm growth, and increase market concentration (Eeckhout and Veldkamp, 2023; Brynjolfsson, Jin and Wang, 2023; Hagiu and Wright, 2023). We argue that firms face a portfolio of different types of uncertainty: while data investment may not directly reduce R&D uncertainty, which stems from the inherent unpredictability of innovation, it can diminish overall market uncertainty (Groh and Pfäuti, 2023) and improve risk diversification, thereby leaving more scope for firms to bear R&D-related risks.

The causal relationship between investments in data and R&D is not straightforward. R&D generates data that can support subsequent innovation, but it also relies on data-intensive processes, raising endogeneity concerns. To address this, we exploit the General Data Protection Regulation (GDPR) in 2016[2] as a source of exogenous variation in data investments. Because GDPR was mandated at the EU level, its adoption was external to firm-level investment decisions. The regulation constrains how firms collect and store data to protect individuals' privacy, which may increase compliance costs and alter business models, for example, by requiring new value chain functions, changes in customer service, or adjustments in data-handling practices (Lindgren,

[1]Also referred to as Industrie 4.0, see Kagermann, Wahlster and Helbig (2013); Bj¨orkdahl (2020)

[2]Implemented in Italy as Decreto Legislativo 10 Agosto 2018, n. 101

2016; Aridor, Che and Salz, 2023). At the same time, GDPR can strengthen incentives to purchase and store data by reducing the risk of future judicial costs under less transparent regimes (EU Commission, 2020).

Second, we study whether investments in data and R&D, as two distinct classes of intangible assets[3] are complementary to enhance labour productivity, building on literature on intangible assets Corrado et al. (2021). Prior research has primarily documented such complementarity at the country level (Nonnis, Bounfour and Kim, 2023), as well as at the firm and industry levels (Ennen and Richter, 2010; Delbecque, Bounfour and Barreneche, 2015; Añón Higón, Gómez and Vargas, 2017), primarily across managerial and organisational capital and innovation.[4] Furthermore, evidence shows that the adoption and use of Artificial Intelligence (AI), as a data-intensive activity, generates strong complementarity and learning spillovers with existing expertise in data management, networking, and communication (Igna and Venturini, 2023).

The impact of intangible assets on productivity and economic growth is not new in the economic literature (Arrow, 1972; Machlup, 1962; Romer, 1990; Hall, 2011; Goodridge and Haskel, 2023), though intangible assets change shape alongside different waves of technical change. Studies by Akerman, Gaarder and Mogstad (2015), Truant, Broccardo and Dana (2021), and Cirillo et al. (2023) document the positive effects of digital technology adoption on labour productivity, mainly through cost-reduction mechanisms, albeit with large heterogeneities at the firm level (Nucci, Puccioni and Ricchi, 2022). While seminal research has highlighted that, among intangible investments, R&D activities exert the most substantial impact on firm performance (Griliches, 1984; Hall, Mairesse and Mohnen, 2010; Shefer and Frenkel, 2005; Siliverstovs et al., 2016; Mohnen, 2019), the latest wave of digitalisation might entail different forms of complementarity with R&D . For R&D to generate long-term benefits, firms must invest in complementary inputs (Teece, 1986), including tangible assets such as laboratory machinery, as well as intangible assets such as organisational capital, human capital, and Information and Communication Technologies (ICT) (Calvino et al., 2022).

This article addresses two questions: "Does investing in data affect R&D investments? Secondly, "Is there a multiplicative effect on labour productivity of the joint investments in data and R&D, with respect to sole investments in data, sole investments in R&D?"

The article aims to contribute to the literature on innovation and firm dynamics by studying the relationship between investments in data and analytics, R&D and productivity. We build on Hall (2011) and explicitly consider the high heterogeneity in intangible assets, as well as the conjecture related to trade-offs in types of firm risks. We empirically test the presence of a causal relationship between investments in data and R&D. We do so by applying an instrumental variable (IV) approach to a database of firms provided by the Bank of Italy's Survey of Industrial and Service Firms. Second, we test the potential complementarity (presence of a multiplier effect) of investments in data and R&D to labour productivity. Specifically, we examine whether sole or combined

[3]This class of assets comprises a collection of resources for producing and marketing new or improved products and processes. It encompasses both internally generated resources, such as innovative knowledge, designs, blueprints, brand equity, internal software, and construction projects, as well as externally acquired resources like technology licenses, patents, copyrights, and economic competencies gained through the procurement of management and consulting services (Arrighetti, Landini and Lasagni, 2014).

[4]The extent of complementarity depends on the type of asset. For example, a firm might choose to invest in data analytics to provide tailored products rather than spending resources on extensive marketing campaigns, implying that acquiring one intangible asset (data and software) can substitute for another (marketing) (Grabowska and Saniuk, 2021; Piccarozzi, Aquilani and Gatti, 2018).

data and R&D investments affect firm labour productivity performance. Our findings initially reveal that more investments in data have led to more R&D investments, specifically through lower uncertainty. Second, we find that not only do firms experience within-year gains in labour productivity by combining investments in data and R&D, but they may also continue to gain in the longer term, as previously shown by Peters et al. (2017). The data used in our analysis comes from the Bank of Italy's Survey of Industrial and Service Firms, which includes information on more than 11,000 unique firms between 2002 and 2024. This survey offers high-quality data on firms' investments across different classes of intangible assets and their structural characteristics. The reliability of this survey, ensured through regular verification and direct contact between Bank of Italy economists and firms in each region, has made it a trusted source in the literature (Pozzi and Schivardi, 2016; Manaresi and Pierri, 2024; Cingano and Pinotti, 2013). This high data quality enables us to investigate questions concerning production and investment in intangibles with greater analytical precision and confidence.

The remainder of the article is structured as follows. First, we introduce a conceptual background in Section 2. In Section 3, we detail our empirical strategy. We then motivate the choice of Italy as the study context and describe the dataset of interest, the Survey of Industrial and Service Firms, and the construction of the variables relevant to our empirical specifications in Section 4. Section 5 presents the results. Section 6 provides a battery of robustness checks. Section 7 concludes by summarising findings and discussing their implications.

2 Conceptual Background

This section presents the theoretical framework to address the research questions laid out in the previous section: the causal impact of investments in data on the financial resources devoted to R&D and the presence of a complementarity effect of such investments on labour productivity performance of firms. By putting both of these conjectures together, we present our research agenda graphically:

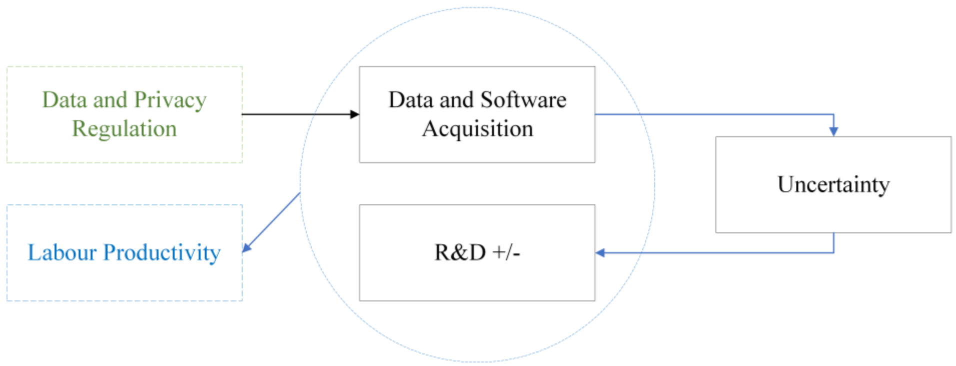


Figure 1: Conceptual Framework of this Study

We now discuss each of these trajectories in detail.

From Data to R&D

Building on the real options theory, we propose a conceptual channel through which investing in data and software can lead to higher investments in R&D. We argue that by increasing firms' information set through deeper analytics and more data, firms face less market uncertainty, can sustain growth through lower costs, gain market shares, and further invest in R&D (Del Monte and Papagni, 2003; Shefer and Frenkel, 2005). This theory predicts that better information about future states of the economy reduces the option value of waiting and accelerates investment in irreversible projects like R&D (Dixit and Pindyck, 1994).

Throughout our study, we treat data and analytics as a key input to stock information, that is a gateway to reduce uncertainty (CFOTech, 2023; Wondrium, 2016). From a firm's perspective, uncertainty represents a hidden cost, leading to delays in investment and hiring (Cuervo-Cazurra and Annique Un, 2010; Bloom et al., 2022). Even though business cycles affect firms' uncertainty, data and analytics can partially help reduce these uncertainties (Simon, 2013). Also, the literature highlights that reducing uncertainty increases firm size and might lead to specialisation upgrading, both of which increase markups (Eeckhout and Veldkamp, 2023) and a firm's value (Salvi et al., 2021). Firms that fail to undertake mitigating uncertainty and risk strategies are more likely to underperform (Wu et al., 2017; Tysiac, 2015). Therefore, having more data translates into a more extensive information set[5], which could reduce overall uncertainty across a portfolio of activities and motivate the firm to engage in other high-uncertainty activities such as R&D.

However, there might be direct channels that connect data and R&D. The extant literature has shown that data can improve R&D outcomes by providing novel decision-making innovation possibilities (Niebel, Rasel and Viete, 2019), R&D management and its technologies across various industries (Blackburn et al., 2017) and is also expected to enhance consumer intimacy, accelerate time-to-market, and promote open innovation in R&D (Farrington and Alizadeh, 2017).

Complementarity of Data and R&D

A firm intensive in data and analytics, and performing innovative activities like R&D, might combine these intangible assets to gain higher productivity. To ascertain the presence of a cumulative, self-reinforcing effect of investing jointly in data and R&D, we examine the possible complementarity of these investments in enhancing labour productivity.

So far, we discussed that investing in data can potentially lead to more R&D, and through more data-driven decision-making and innovative production processes, the firm may become increasingly efficient, with labour productivity growing. Previously, Müller, Fay and Vom Brocke (2018) documented that big data and analytics assets are associated with a 3-7% improvement in firm productivity, but the effect varies by industry. Bakhshi, Bravo-Biosca and Mateos-Garcia (2014) found that more extensive data use improves firms' productivity, and its combination with other organisational attributes such as organisational structure and innovation contributes

[5]It has also been shown that models with big data beliefs are updated more predicatbly Van Oijen (2017) and those with uncertainty are solved more precisely (Ning and You, 2019).

positively to firm performance. Longitudinal studies have demonstrated positive returns on R&D investment, with company-financed research yielding higher returns and basic research providing a productivity premium (Lichtenberg and Siegel, 1991). These findings underscore the critical role of data and software in enhancing R&D capabilities and driving innovation and productivity. By combining these insights with those previously provided, we demonstrate that there may be a loop in which firms become data-intensive, become further involved in innovative activities, and consequently become more productive.

3 Empirical Strategy

We develop our empirical analyses to study the conceptual framework explained in detail in Section 2. We discussed our hypothesis that investing in data, software, and analytics allows firms to manage and afford more uncertainty, providing the leeway to redirect resources toward high-risk ventures such as R&D. In this section, we aim to estimate a direct causal relationship between investing in data and investing in R&D, and test whether uncertainty mediates this effect. After that, we study the potential complementarity of investing in these intangible assets for labour productivity.

In what follows, we first present our identification strategy and motivation for using GDPR as an instrument for investments in data. Second, we conduct an IV analysis to examine the relationship between investments in data and analytics and R&D activities. We then use a firm-level uncertainty measure to study the potential mediation of uncertainty between data and R&D investments. Lastly, we explore the potential complementarity between investments in data and R&D by studying whether there are gains in productivity from jointly investing in R&D and data.

GDPR as an Instrument for Data

We are aware of multiple ways in which the direct relationship between data and R&D could be exposed to endogeneity issues. First, R&D generates data and in-house software, which may introduce a potential reverse causality. Furthermore, firms' R&D investment plans are affected by uncertainty (Bloom, 2007) and increase uncertainty (Fung, 2006). For our framework to identify the causal direction and to overcome the issues of simultaneity and reverse causality between data investments and R&D, we use the General Data and Privacy Regulation (GDPR) as an exogenous instrument for investments in data.

Firms increasingly collect data on user preferences, online searches, and even geographic movements through cookies and other digital traces. In 2016, the European Union (EU)[6] introduced the GDPR, which protects the personal data of all EU residents regardless of where processing occurs, by requiring anonymisation and limiting the use of Personally Identifiable Information (PII). PII refers to any information that can identify an individual, either alone or when combined with other sources. Within this framework, personal data encompasses identifiers such as IP addresses, cookies, digital fingerprints, and location data (Goddard, 2017).

[6]The application of the GDPR requires member states to adopt and adapt the legislation to their specific institutional and business environments. In Italy, this occurred through the Data Protection and Governance Decree (Decreto Legislativo 10 Agosto 2018, n. 101).

Firms face considerable challenges in complying with these broad and complex rules, including the costs of compliance, the risk of fines and litigation, and the uncertainty arising from extensive definitions of personal data (Layton and Elaluf-Calderwood, 2019; Johnson, 2023; Urban et al., 2018). The literature has previously studied the various ways in which the GDPR can affect firms' venture investments (Jia, Jin and Wagman, 2020), revenues (Goldberg, Johnson and Shriver, 2024; Koski and Valmari, 2020), data sharing (Gal and Aviv, 2020), and the cost of acquiring data, especially for smaller firms (Demirer et al., 2024). Because the GDPR promotes principles such as data minimisation and limited purpose, firms may reduce or postpone data acquisition when potential penalties, up to 4% of global annual turnover, outweigh the expected benefits. At the same time, by introducing a clear and harmonised regulatory framework, the GDPR may also lower institutional uncertainty and encourage further investment in data and analytics (EU Commission, 2020; Voss and Houser, 2019).[7] Consequently, GDPR is a plausibly exogenous shock to firms' data-related activities, making it a suitable instrument for identifying the causal effect of data investments on innovation.

Our identification strategy is based on how the GDPR has impacted the data intensity of each sector to construct treatment and control groups. Calvino and Criscuolo (2019) stressed the presence of sector-level heterogeneity in data intensity. In our identification strategy, more data-intensive sectors are assumed to be more exposed and affected by the inytroduction of GDPR, and therefore are identified as the treatment group. To identify the sectors most affected by GDPR, we rely on insights from EUKLEMS-INTANProd data on firms' investments in databases, as introduced in Section 4.

More data, more R&D?

In this section, we assess a potential causal relationship between investments in data and R&D as also described in Sections 1 & 2. We build on the intuitions put forward by Angrist, Imbens and Rubin (1996); Angrist (2022) and construct a two-step control function model similar to a 2SLS. The first step is to predict the values of investments in data using a GDPR dummy, fixed effects ($\mu_i + \lambda_t$) and a small set of control variables ($X_{i,t}$). These controls include sales growth and employee growth (Grazzi, Jacoby and Treibich, 2016; Farboodi et al., 2019), lagged by one year, and the firm's age (Fortune and Shelton, 2014). Formally:

$$\frac{\text{Data}_{i,t}}{L_{i,t}} = \gamma \text{GDPR}_{s,2016} + \mu_i + \lambda_t + X_{i,t}\Lambda + u_{i,t} \tag{1}$$

Where L represents the units of labour and μ_i and λ_t are firm and time-fixed effects, respectively. This estimation yields the predicted values for the data as $\frac{\widehat{\text{Data}_{i,t}}}{L_{i,t}}$. We then insert the predicted values in our main regression equation featuring the sector-fixed effects (ψ_s) and run separate regressions for $\tau = 0, 1, 2, 3, 4, 5$ to assess the impact of investments in data on R&D over time as follows:

$$\frac{\text{R\&D}_{i,t}}{L_{i,t}} = \alpha \frac{\widehat{\text{Data}_{i,t-\tau}}}{L_{i,t-\tau}} + X_{i,t}\beta + \psi_s + \mu_i + \lambda_t + \varepsilon_{i,t} \tag{2}$$

[7] Beyond economics, scholars have examined GDPR's effects on web privacy (Degeling et al., 2018), content sharing (Lefrere et al., 2020), legal anonymisation (Bolognini and Bistolfi, 2017), and technology narratives (Li, Yu and He, 2019).

In the second-stage regression, we extend our analysis and control for firms' lagged high-tech investments, as they provide leeway for more investments in other intangibles (Schivardi and Schmitz, 2019). Furthermore, following IMF (2015); Altomonte et al. (2022), we control for financial constraints, as they can curb firms' investments in R&D. We also control for whether large firms or those in the North drive the majority of R&D investments.

Does uncertainty mediate the relationship between Data and R&D?

Our conjecture, supported by the theoretical literature illustrated in section 2 is that firms' investments in data infrastructure reduce uncertainty and thereby encourage investments in R&D. To test this mechanism, we investigate whether forecast-error based uncertainty (FCE) previously introduced by Mohades et al. (2025) serves as a mediating channel through which data investments influence R&D expenditures. We focus on FCE as it captures firms' ability to accurately predict future sales, a concrete, operational dimension of uncertainty that data analytics should directly improve. Unlike measures of output volatility or subjective managerial perceptions, forecast errors reflect the informational quality of firms' decision-making processes, precisely the domain where data investments should have their strongest impact. FCE for each firm i at time t is calculated as:

$$\text{Forecast Error}_{i,t} \equiv \text{FCE}_{i,t} = \frac{E_t\text{Sales}_{i,t+1} - \text{Sales}_{i,t+1}}{\text{Sales}_{i,t+1}}$$

Our analysis employs a rolling window approach that allows us to trace how this mediation mechanism evolves over time. This temporal perspective proves essential for uncovering the data-uncertainty-R&D nexus: by examining consecutive time periods, we reveal how technological upgrading and economic conditions affect the effectiveness of data investments in reducing uncertainty and enabling innovation.

For each rolling window spanning years $[t_{\text{start}}, t_{\text{end}}]$, we estimate a two-stage system. In the first stage, we regress forecast-error uncertainty on data investment intensity, controlling for firm characteristics and fixed effects:

$$\text{FCE}_{i,t} = \gamma_1 \frac{\text{Data}_{i,t}}{L_{i,t}} + X_{i,t}\beta + \mu_i + \lambda_t + \psi_s + u_{i,t}, \quad t \in [t_{\text{start}}, t_{\text{end}}] \tag{3}$$

where $\text{FCE}_{i,t}$ denotes the normalised forecast error of sales growth for firm i in year t for $t+1$, $X_{i,t}$ includes lagged sales growth, lagged size growth, firm age, size, location, past profitability, and financial constraints, and μ_i, λ_t, and ψ_s represent firm, year, and sector fixed effects, respectively. The coefficient γ_1 captures the effect of data investment on forecast uncertainty.

In the second stage, we estimate the effect of forecast-error uncertainty on R&D investment using an instrumental variables approach, where uncertainty is instrumented with data investment:

$$\frac{\text{R\&D}_{i,t}}{L_{i,t}} = \beta_2 \widehat{\text{FCE}}_{i,t} + X_{i,t}\delta + \mu_i + \lambda_t + \psi_s + \varepsilon_{i,t}, \quad t \in [t_{\text{start}}, t_{\text{end}}] \tag{4}$$

where $\widehat{\text{FCE}}_{i,t}$ denotes the instrumented forecast error from the first stage. The coefficient β_2 represents the causal effect of uncertainty on R&D intensity, addressing potential endogeneity between these variables.

The indirect effect of data investment on R&D operating through the uncertainty channel is given by the product:

$$\text{Indirect Effect} = \gamma_1 \times \beta_2 \tag{5}$$

To account for the variance in this product, we employ a bootstrap procedure with 200 replications for each window, constructing confidence intervals for the indirect effect. A significantly positive indirect effect indicates that data investment reduces forecast errors ($\gamma_1 < 0$), which in turn enables higher R&D ($\beta_2 < 0$), confirming our theoretical framework of uncertainty-mediated innovation.

Our rolling window approach proves essential for revealing this mediation mechanism. By allowing the relationships between data, uncertainty, and R&D to vary over time, we uncover economically meaningful patterns that static, full-sample analyses would obscure. The data-uncertainty-R&D nexus is not a fixed structural relationship but rather one that evolves alongside technological progress and firms' organisational capabilities. This temporal heterogeneity reflects the fundamental nature of digital transformation: the returns to data investments depend critically on the availability of complementary assets, analytical tools, and human capital that have themselves developed over time.

Estimating Complementarity of Data and R&D Investments for Productivity

Here we investigate whether investing jointly in data and R&D contributes positively to the firm's productivity. To do so, we need to introduce a different setting in line with the framework described in Section 2. Our dependent variable of interest is the production capacity per labour as a proxy for productivity. However, we are interested in both the within-period (year) effect and the lagged effect because the use and combination of these investments may take more than one year to affect firms' performance (Capasso, Treibich and Verspagen, 2015).

Therefore, our estimation equation of interest is as follows:

$$\text{Lab Prod}_{i,t} \equiv \frac{\log(Y^{PC}_{i,t})}{\log(\text{EMP}_{i,t})} = \alpha_1 \frac{\widehat{\text{Data}}_{i,t-\tau}}{L_{i,t-\tau}} + \alpha_2 \frac{\text{R\&D}_{i,t-\tau}}{L_{i,t-\tau}} + \alpha_3 \frac{\widehat{\text{Data}}_{i,t-\tau}}{L_{i,t-\tau}} \times \frac{\text{R\&D}_{i,t-\tau}}{L_{i,t-\tau}} + X_{i,t}\beta + \mu_i + \lambda_t + \psi_s + \varepsilon_{i,t} \tag{6}$$

The functional form above includes the interaction between R&D and data investments and runs separate regressions for $\tau = 0,1,2,3,4,5$. The positive coefficient of the interaction term (α_3) will be a sign of potential complementarity and the presence of a complementarity effect. $X_{i,t}$ includes controls such as all other investments, age, North and large dummies to control for the high productivity of frontier firms and lagged high-tech investments, with insights from Kumbhakar et al. (2012).

4 Data & Variables: The Bank of Italy Survey of Industrial and Service Firms

This section presents the rationale for focusing on Italy as a case study to address the research questions. We describe the dataset, and the key variables constructed for our analysis.

Benchmarking Italy: Intangible Investments and Productivity

In contrast to evidence showing firm-level gains from digitalisation and innovation-based activities, the diffusion of new technologies has coincided with a slowdown in aggregate productivity growth, giving rise to the so-called "productivity puzzle of the 21st century" (Acemoglu et al., 2014; Coyle, 2019). In Italy, this slowdown has been closely associated with lower investments before and during the Global Financial Crisis and the European debt crisis (Busetti, Giordano and Zevi, 2015; Mohades et al., 2025). Moreover, in the first two decades of the century, the Italian economy has exhibited one of the lowest shares of intangible investments among advanced economies (Crass and Peters, 2014), potentially driven by underinvestment in big data and e-commerce rather than in other technologies or cloud computing (Calvino et al., 2022). Figure 2 illustrates that the slow, secular growth of intangible investments relative to all investments has not been matched by an accelarated increase in investments in R&D, data and software in the age of ICT and big data.

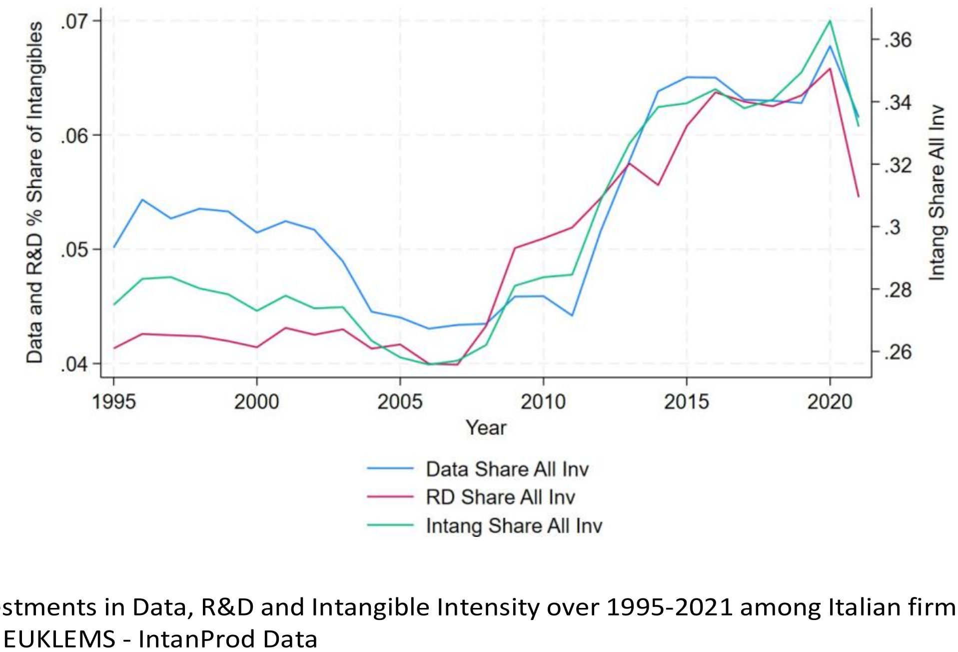


Figure 2: Investments in Data, R&D and Intangible Intensity over 1995-2021 among Italian firms (all sectors).
Data source: EUKLEMS - IntanProd Data

As a result of these trends, Italy has received recommendations from the EU Commission to encourage investments in innovation and new technologies. Tax incentives motivated by these recommendations have contributed to increased investments in R&D (Briguglio et al., 2019), particularly following the introduction of Impresa 4.0 in 2014 (ISTAT, 2018). Previous studies have shown that intangible assets can be at least as productive as tangible assets for Italian manufacturing firms (Bontempi and Mairesse, 2008), and that adopting high-end digital technologies increases the likelihood of growth (Cassetta, Meleo and Pini, 2016). Such assets are therefore a promising source of growth and productivity gains, as they generally require lower costs and less external financing (Marrocu, Paci and Pontis, 2012). In an economy that has faced decades of low productivity growth and limited innovation and investments in high-tech sectors, ensuring a productivity-increasing role for data and its combination with other types of investments can introduce two primary trajectories for research and policy design: it motivates policy and academic attention to these activities, and it leaves space for further exploration of other potential drivers of productivity decline, such as population ageing or demographic shifts (Maestas, Mullen and Powell, 2023; Kotschy and Bloom, 2023).

Data and Variables

We use a multi-year version of the firm-level dataset "Survey of Industrial and Service Firms"[8] conducted and collected annually by the Bank of Italy since the early 1980's. This survey is administered annually to the manufacturing and services firms with at least 20 employees. Since the 1990s, the Survey of Industrial and Service Firms has included questions on firms' actual and expected values on multiple variables, including turnover, different classes of investments, and capacity utilisation. Following the Great Recession of 2007-08, the survey was expanded to include questions on firms' financing conditions and constraints. Furthermore, the survey comprises year-by-year questions on various interesting economic research topics, such as the Russo-Ukrainian War or the energy crisis.[9]

We limit our sample to start from 2002 as the number of participating firms surges from 1,500 to around 4,000 this year by including firms in the services sector and expanding to more industries within manufacturing. We observe 12,094 unique firms in our sample between 2002 and 2024, with 94,330 corresponding observations or, on average, 7.80 annual survey responses per firm. We use the data on firms' characteristics, such as location, industry, number of employees, sales, capacity utilisation, profitability, year of establishment, investments in data, investments in R&D and other classes of investments. Firms answer directly the amount they spend on data and software acquisition as well as their R&D expenditures - independently from the other[10]. However, measuring data investments in other samples can prove challenging for multiple reasons.

First, data is typically an intermediate and experience good, meaning its value is not observable before consumption (information asymmetry in the seller's favour). Furthermore, the intermediate hand in the market holds substantial market power (Ichihashi, 2021). Second, data is non-rival (Determann, 2018; Jones and Tonetti, 2020). As a result, there may be social benefits to data being used widely across firms, even considering privacy regulation. Fearing creative destruction, firms may choose to accumulate their data, leading to the inefficient use of non-rival data. In addition, data depreciation is not trivial and differs from the normal wear and tear of other

[8]" Indagine sugli investimenti delle imprese manifatturiere" (Bank of Italy, 2024)

[9]The dataset is available through an online platform that supports code execution in R and STATA. The output is restricted to text (.txt format), preventing us from analysing variables' distributional properties, such as histograms or Kernel densities. Additionally, direct access to the data and individual observations is limited due to confidentiality restrictions, preventing us from viewing the percentiles, minimum, and maximum values of the variables.

[10]The explicit terms used for these variables are: "Total expenditure on software & databases and mineral explorations" and "Expenditure on R&D; design and test products"

assets (OECD, 2020; Li, Nirei and Yamana, 2019). Lastly, a large proportion of data is collected through free services (e.g., Search Engines, Cookies, Online Advertising Services, and Social Media). Therefore, recording and monetising free transactions of data has made its measurement as an input for digital firms particulalry complicated. In our dataset, firms report the values for investments in data directly. Therefore, we expect these values only to include the direct expenditure of firms on data, and hence, represent a conservative (lower bound) estimate of firms' acquisitions of data and software.

We construct a measure of firm performance, specifically labour productivity, using variables provided in the Survey of Industrial and Service Firms. Leveraging information on firms' sales and reported capacity utilisation, we calculate each firm's production capacity per unit of labour. This approach allows us to avoid relying on (in any case unavailable) production data. It ensures that our productivity measure is less affected by demand-driven fluctuations in business cycles that would arise if we used sales per labour directly. Formally, the production capacity ($Y^{PC}_{i,t}$) of firm i in year t is computed from reported total sales ($Y^{S}_{i,t}$) and capacity utilisation ($0 < CU_{i,t} < 1$) as:

$$CU_{i,t} \equiv \frac{Y^{S}_{i,t}}{Y^{PC}_{i,t}} \quad \therefore \quad Y^{PC}_{i,t} = \frac{Y^{S}_{i,t}}{CU_{i,t}} \tag{7}$$

We then choose to calculate labour productivity as production capacity per labour as follows:

$$\text{Lab Prod}_{i,t} = \frac{\log(Y^{PC}_{i,t})}{\log(\text{EMP}_{i,t})} \tag{8}$$

In our analyses, we control for various firm characteristics as follows. Initially, we use the log of age, which is calculated as the natural logarithm of the difference between the observation year and the establishment year. Firms located in central and northern Italy take the dummy North = 1. We also calculate growth of the number of employees (Emp Growth) and sales growth (Sales Growth), using lagged values as a control in our analysis, in line with the existing literature. The profitability dummy controls for the firms' recent performance and potential internal funding of investments, constructed based on a categorical variable regarding the size of profits and losses. We categorise firms that reported large and moderate profits as profitable, those without profit or losses, and significant losses as non-profitable. We also construct a dummy for observations with higher than 25% of their investments in high-tech and assign them as High Tech firms. Lastly, we follow Mohades et al. (2025) and construct a financial constraints dummy based on the questions included after 2009. Firm-year observations that entirely, partially or by self-selection[11] did not receive their required funds, are categorised as financially constrained.

The descriptive statistics of the variables taken directly from the survey and those that we constructed are presented in Table 1:

[11]These firms did not apply for funds as they were sure their request would be rejected.

Variable	Mean	SD	N
R&D Investments	1324.63	22539.94	61,426
Data and Software Investments	1031.42	32556.09	94,330
Nb. of Employees	362.43	2414.94	94,330
All other Inv	5105.56	74390.08	94,330
Sales Growth	-0.01	0.26	74,094
Employees Growth	0.01	0.12	74,107
Age	35.50	25.07	93,781
Financial Constraints	0.06	0.24	45,350
Profit	0.68	0.47	87,211
North	0.66	0.48	94,330
High Tech	0.32	0.47	23,270
Lab Prod (log)	0.30	0.27	54,776
Data per Labour	0.76	0.34	57,305
R&D per Labour	1.05	0.40	20,964
Observations per Sector			
Manufacturing			67,662
Services			26,668
All Sample			94,330

Table 1: Descriptive Statistics

Overall, more than 65% of our firm-year observations are from firms located in the centre and north of Italy. More than 68% have recorded profits, whereas only 32% have made significant high-tech investments. Furthermore, our firms are, on average, more than 35 years old.

EUKLEMS-INTANProd Data

To accurately estimate the effect of data investments on R&D, using data regulation as an instrument, we first construct treatment and control groups based on firms' data sensitivity, i.e., the extent to which they are likely to respond to data privacy regulations, which depends on their propensity to invest in data. At the sector level, the EUKLEMS-INTANProd project provides estimates of sectoral investments in intangible assets across European economies (Bontadini et al., 2023). We match these sector-level data to our NACE classification (Table A.1) and focus on the estimates for Italy from 1995 to 2019. This allows us to identify the sectors most sensitive to data regulation, which we use to define our treatment and control groups. We use the GDPR as our baseline specification from 2016 and define the GDPR dummy applied to highly data-intensive sectors post-2016. This choice takes the anticipation effect of EU-level legislation.

After performing trend validity and t-tests for after and before 2016 on both raw and detrended investments in data, we note that similar to the existing literature, sectors SS7 (Other industries, Energy), SS8 (Wholesale and retail commerce), SS10 (Transport and communication) and SS11 (Real estate activities and IT) have been increasingly investing in databases and hence are more "data intensive". These findings align with the literature (Haddara, Salazar and Langseth, 2023; Park and Kim, 2021; Nepelski, Stancik et al., 2011; Giannopoulos and Moschovou, 2023). Firms in these sectors are grouped as treatment and those present in other sectors as our control, as they should not be (or be less) sensitive to the introduction of a regulation that affects data acquisitions. Our identification strategy yields a treatment group of 10,513 observations and a control group of 21,565 observations after 2016.

5 Results

This section presents the results from the empirical settings developed in Section 3. First, we display the results of our first-stage regression. Next, we present the results on the relationship between investing in data and R&D expenditures and discuss how uncertainty may act as a mediator. We emphasise the findings on the potential complementarity between R&D and data investments to labour productivity. We also demonstrate that these gains vary across both short-term and long-term horizons. Lastly, we show that the marginal impact of investments in data and R&D depends heavily on the levels of investment in the other, further reinforcing evidence of their complementarity.

First-stage Regression: GDPR and Investments in Data

Table 2 provides results on our first-stage regression (Equation 1) introduced in Section 3. We find that controlling for firm and time-fixed effects, GDPR is positively correlated with investments in data:

	Dependent Variable: Data Inv per Labour
Treatment	0.039***
	(0.006)
Sales Growth$_{t-1}$	0.029***
	(0.006)
Age	0.016***
	(0.005)
Emp Growth$_{t-1}$	0.001
	(0.014)
Constant	0.684***
	(0.0189)
Firms	6,497
Observations	34,941
Time FE	✓
Firm FE	✓

Table 2: Regression Results for the first-stage regression in Equation 1

*** Significant at 99% CI, ** at 95% CI, * at 90% CI. Treated Sectors: Energy and extraction - Wholesale and retail commerce - Real Estate and IT - Transport and communications

This first-stage regression shows that GDPR might have offered lower judicial costs and higher transparency, thereby reducing the risk of future privacy-related concerns when holding clients' or other businesses' data, in line with EU Commission (2020). Furthermore, this result can be of interest in studying firms' heterogeneity in perceiving privacy regulation, as suggested by Taufick (2021). It has been shown that smaller firms may be negatively affected by regulations that affect the business environment (Schiffer and Weder, 2001). A more transparent legal system may benefit larger firms as they can internalise data risks before regulation and implement better data governance practices and integrate all their processes within the framework at lower costs and with less uncertainty about the data possess (Sabatino and Sapi, 2022).

Second stage regression - Data and R&D

Table 3 presents the results of the second-stage regression in Equation 2. We find support for our hypotheses introduced in Sections 2 & 3: more investments in data and analytics lead to higher expenditures in R&D.

Dependent Variable: R&D Inv per Labour Instrument for Data Investments: GDPR Implementation 2016						
	(1)	(2)	(3)	(4)	(5)	(6)
Data inv per Labour	1.642***	1.815***	1.802***	1.802***	-0.939	-0.944
	(0.421)	(0.429)	(0.430)	(0.430)	(0.670)	(0.670)
Sales Growth$_{t-1}$	-0.061***	-0.065***	-0.066***	-0.066***	0.016	0.016
	(0.015)	(0.015)	(0.015)	(0.015)	(0.024)	(0.024)
Log of Age	-0.031	-0.036*	-0.037*	-0.037*	0.007	0.007
	(0.016)	(0.016)	(0.016)	(0.016)	(0.036)	(0.036)
Emp Growth$_{t-1}$	-0.037	-0.039	-0.034	-0.034	-0.090*	-0.090*
	(0.023)	(0.023)	(0.023)	(0.023)	(0.035)	(0.035)
Profit$_{t-1}$	0.009	0.009	0.010	0.010	0.005	0.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.010)	(0.010)
Large			-0.039*	-0.039*	-0.021	-0.022
			(0.015)	(0.015)	(0.025)	(0.025)
North			-0.156	-0.156	-0.728***	-0.728***
			(0.093)	(0.093)	(0.151)	(0.152)
Fin Cons				-0.001		0.007
				(0.009)		(0.016)
High Tech$_{t-1}$					0.011	0.011
					(0.008)	(0.008)
Constant	-0.108	-0.068	-0.062	-0.062	2.471***	2.476***
	(0.293)	(0.325)	(0.337)	(0.337)	(0.506)	(0.506)
Sector FE		✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Within R^2	0.016	0.018	0.018	0.018	0.038	0.039
Firms	3,633	3,633	3,626	3,626	1,903	1,903
Observations	14,054	14,054	13,985	13,985	5,163	5,163

Table 3: Regression Results For the second-stage regression in Equation 2

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval. Treated Sectors: Energy and extraction - Wholesale and retail commerce - Real Estate and IT - Transport and communications

We also find a weak negative relationship between sales growth and R&D investments. However, profits in the previous year do not contribute to ongoing R&D investments. Larger firms and those operating in northern Italy are not necessarily more active in R&D investments, and those with larger proportions of high-tech-related investments were not detected to be more actively engaging in research and development, similar to Beaudry and Breschi (2003). However, when we control for the share of high-tech related investments (out of all investments), the positive effect of data on R&D disappears.

So far, our analysis has only been concerned about "within-year" effects of investments in data on R&D activity by setting $\tau = 0$. As discussed previously, our analysis is also concerned with longer-term effects. By setting $\tau > 0$, we present our results in Equation 2 in Figure 3:

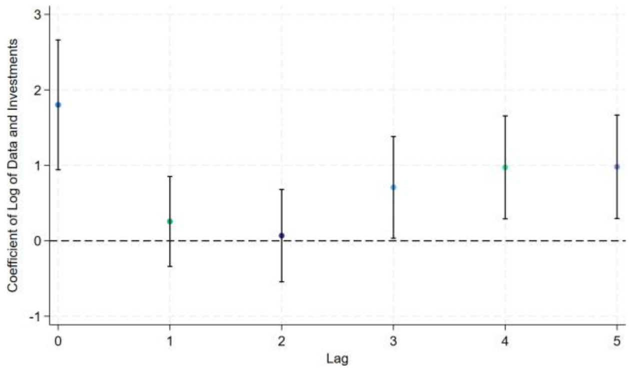


Figure 3: The effect of data on R&D over time for: $\tau = 0, 1, 2, 3, 4, 5$ in Equation 2

We document that data investments' simultaneous effect on R&D is significantly larger than zero. Furthermore, it is only considerably positive within the year and weakly in the long term. This result should motivate policymakers to encourage more investments in data and provide more straightforward data sharing via B2B data transfer, as it can increase firms' engagement in innovative activities in the short term. The simultaneous effect of these investments can also suggest that investments in data need to be persistent enough to influence firms' innovation activities continuously. The positive coefficient of data for $\tau \in [3, 5]$ can be an interesting puzzle, as it contradicts the rapid depreciation of data. This confirms that in the current digital technologies paradigm, it is the scale of investments in data that matters, and that obsolescence is not a major issue when investing in data. The regression results for $\tau > 0$ can be found in Table A.9 of Appendix E.

By combining our results from this section and Section 5, we can establish the empirical results of the relationship between GDPR, investing in data and R&D. So far, we have shown that post-GDPR years come with firms' larger investments in data and analytics, and more data and analytics have increased firms 'within-year R&D investments.

Uncertainty as a Mediator

We estimate the mediation analysis previously discussed for two sets of rolling windows: 2-year windows (providing high temporal resolution) and 5-year windows (offering more stable estimates with larger sample sizes). The 5-year windows are particularly valuable for identifying persistent mechanisms, as they smooth out transitory fluctuations and capture medium-term relationships between intangible investments, uncertainty, and innovation. Our focus on 5-year windows reflects their superior statistical properties: they provide sufficient observations for reliable estimation while maintaining temporal granularity to detect structural changes in the mediation relationship.

Our rolling window analysis reveals a clear and strengthening mediation mechanism linking data investment to R&D through forecast-error reduction. Below, Figure 4 illustrates the evolution of the indirect effect over time. The pattern that emerges demonstrates how the data-uncertainty-R&D channel has intensified as firms' analytical capabilities and the broader technological ecosystem have matured.

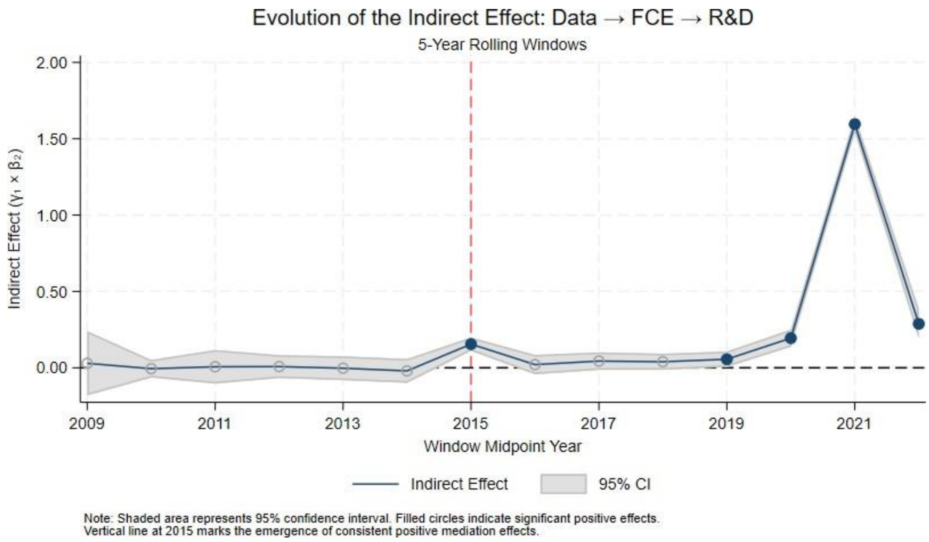


Figure 4: Evolution of the Indirect Effect: Data → Forecast Error → R&D (5-Year Windows). This figure plots the indirect mediation effect ($\gamma_1 \times \beta_2$) from 5-year rolling windows. The shaded area represents the 95% confidence interval based on bootstrapped standard errors. The vertical line indicates the 2013-2017 window, marking the emergence of consistently positive and significant mediation effects.

Table A.11 in Appendix F presents the complete results across all 5-year windows. The most striking finding is the emergence and strengthening of positive mediation effects beginning in 2013. The 2013-2017 window shows an indirect effect of 0.16 ($p < 0.05$), indicating that data investments had begun to meaningfully reduce forecast uncertainty and thereby enable higher R&D expenditures. This effect grows steadily: the 2014-2018 window yields an indirect effect of 0.21 ($p < 0.05$), and the 2015-2019 window produces 0.23 ($p < 0.01$). The number of observations across these windows remains substantial (ranging from 7,400 to 8,600 firm-years), ensuring that these effects reflect genuine economic relationships rather than sample artefacts.

Most remarkably, the 2019-2023 window exhibits an indirect effect of 1.60 ($p < 0.01$), which represents a near-tenfold amplification relative to earlier periods. This dramatic intensification demonstrates that data investments have become increasingly powerful in reducing forecast uncertainty and unlocking innovation capacity. Even the most recent complete window (2020-2024) maintains a substantial indirect effect of 0.29 ($p < 0.05$), confirming that the mechanism persists and remains meaningful.

The temporal pattern we document carries important implications for understanding digital transformation and its impact on core investments in innovative and uncertain research. The strengthening of the mediation effect from 2013 onwards coincides precisely with the increasingly prominent paradigm of big data analytics from early 21st century, machine learning applications, and cloud computing infrastructure (Brynjolfsson, Hitt and Kim, 2011; McAfee et al., 2012). Over this period, firms transitioned from merely accumulating data to actively leveraging advanced analytics to extract actionable insights. Our results suggest that this technological evolution has had profound consequences for innovation capacity: as data analytics capabilities improved, firms gained the ability to reduce forecast uncertainty more effectively, thereby creating the conditions for increased R&D investment.

The exceptionally large indirect effect observed in the 2019-2023 window ($\gamma_1 \times \beta_2$ = 1.60) provides compelling evidence of data's role during periods of heightened uncertainty and under more data regulation. This window encompasses the COVID-19 pandemic, which dramatically disrupted markets and rendered traditional forecasting methods obsolete. Firms with strong data analytics capabilities are able to leverage on a crucial advantage: they could track real-time changes in consumer behaviour, identify emerging patterns in markets, and adjust forecasts dynamically as conditions evolved (Sharma, Adhikary and Borah, 2020). Our findings indicate that this superior forecasting ability translated directly into greater capacity to undertake R&D projects even during the crisis. Rather than retreating from innovation due to uncertainty, data-intensive firms could reduce their forecast errors and maintain or expand their innovation activities. The recent growth in the indirect effect suggests that as more firms develop analytical capabilities, the marginal returns to data access for innovation may be increasing rather than diminishing.

The contrast between 2-year and 5-year windows (with 2-year results reported in Appendix Table A.10) further illustrates the value of a rolling window approach. Although 2-year windows capture short-term dynamics and provide evidence of mediation in several periods, they exhibit greater volatility due to smaller sample sizes and the influence of transitory shocks. The 5-year windows smooth these fluctuations while preserving sufficient temporal resolution to identify the structural shift that occurred during the 2010s. This balance between stability and temporal specificity makes the 5-year results our preferred specification for drawing substantive conclusions about the mediation mechanism.

In summary, our rolling window mediation analysis establishes that data investments enable R&D activities by reducing forecast uncertainty, and that this mechanism has strengthened substantially over time. For Italian firms, which face ongoing productivity challenges, these findings highlight the strategic importance of data investments not only for operational efficiency but as fundamental enablers of innovation capacity. The evidence that this relationship has intensified in recent years suggests that the innovation returns to data infrastructure may be entering a period of acceleration, making continued investment in these capabilities increasingly critical for competitive success. Our results provide direct empirical support for this channel: by improving forecast accuracy, data investments reduce the perceived downside risk from innovation failures, making firms more willing to allocate resources to high-uncertainty R&D activities. This represents a shift in firms' risk portfolios as they face less market uncertainty and can therefore accommodate more innovation uncertainty within their overall risk capacity. Lastly, our analysis reveals that crises years with higher uncertainty may amplify the relationship between data and R&D by allowing more margins for data's role as information.

Complementarity Effects

The estimation of Equation 6 provides key insights into the potential complementarity of data and R&D for labour productivity. Our findings in Table 4 show that data investments do not contribute to productivity, whereas we document an adverse effect of R&D investments on productivity. Importantly, their interaction displays a positive contribution to productivity in specifications of (1) to (6), confirming our hypothesis on complementarity in Sections 2 & 3. Furthermore, older firms tend to display lower levels of productivity, while large firms exhibit higher productivity. Investments in high-tech-related resources do not display simultaneous productivity gains when we account for investments in R&D and data. Crass and Peters (2014) found that intangible assets are usually complementary, but they obtained mixed results when studying the combinations of R&D and other intangible assets. It is important to note that these effects reported in Table 4 are within the year when $\tau = 0$: a combination of data and R&D in any given year leads to significantly positive labour productivity gains, which may even offset the negative impact of pure R&D.

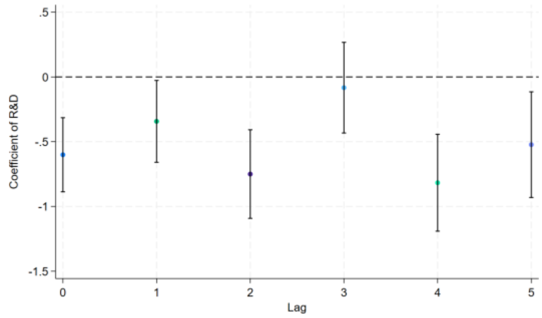
Our results identify an essential complementary intangible asset to R&D. Data acts as information, reducing firms' uncertainty about markets. Firms can afford to bear more innovation uncertainty and engage more in R&D by freeing up space in a portfolio of various uncertainties. Combining these two assets increases efficiency and thus improves labour productivity by reducing the simultaneous negative impact of R&D on productivity.

	Dependent Variable: Labour Productivity						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Data inv per Labour	-0.180	-0.128	-0.134	-0.230	-0.246	-0.137	0.644
	(0.220)	(0.222)	(0.222)	(0.225)	(0.225)	(0.227)	(0.444)
R&D inv per Labour	-0.616***	-0.628***	-0.616***	-0.607***	-0.594***	-0.601***	-0.353
	(0.139)	(0.139)	(0.139)	(0.143)	(0.143)	(0.143)	(0.267)
Data inv per Labour	0.821***	0.835***	0.820***	0.800***	0.783***	0.792***	0.454
×R&D inv per Labour	(0.186)	(0.186)	(0.186)	(0.191)	(0.191)	(0.191)	(0.356)
All other Investments				0.006***	0.006***	0.006***	0.003**
				(0.001)	(0.001)	(0.001)	(0.001)
Log of Age						-0.015***	-0.018
						(0.004)	(0.010)
North						-0.001	-0.043
						(0.023)	(0.040)
Large						0.037***	0.037***
						(0.004)	(0.008)
High Tech$_{t-1}$							-0.001
							(0.002)
Constant	0.477**	0.444**	0.399*	0.524**	0.613***	0.554**	-0.008
	(0.165)	(0.166)	(0.188)	(0.168)	(0.170)	(0.173)	(0.336)
Sector FE		✓		✓	✓	✓	✓
Within R^2	0.009	0.012	0.014	0.023	0.026	0.039	0.044
Firms	2,809	2,809	2,809	2,310	2,310	2,302	1,356
Observations	12,295	12,295	12,295	9,804	9,804	9,747	3,747

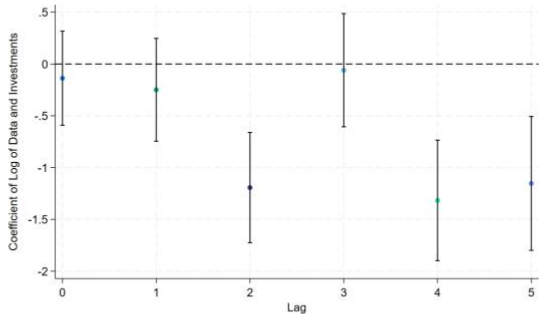
Table 4: Regression Results of Equation 6 with Random Effects and $\tau = 0$

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval. The interaction variables are jointly significantly different from zero. Treated Sectors: Energy and extraction - Wholesale and retail commerce - Real Estate and IT - Transport and communications

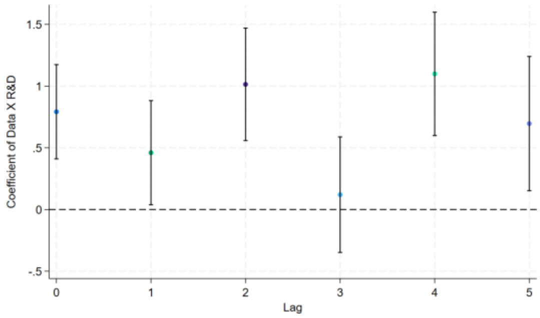
The regression results for $\tau > 0$ are provided in Appendix D. Below and in Figure 5, we provide a graphical representation of the coefficients of lagged investments. These graphs provide insight into the effect of investments in each of the data and R&D and their complementarity effect over time:



(a) The Coefficients of R&D



(b) The Coefficients of Data



(c) The Coefficients of Data × R&D

Figure 5: Coefficients for: $\tau = 0, 1, 2, 3, 4, 5$ in Equation 6

We can observe that the returns of R&D may only be documented in very long horizons, similar to the findings of Aydin, Alrajhi and Jouini (2018); Guellec and Van Pottelsberghe de la Potterie (2004). On the other hand, investments in data do not exhibit simultaneous productivity costs. However, combining these two investments yields a persistently positive impact on labour productivity. Therefore, the multiplier effect is present and statistically significant, and is reasonably large compared to the sole impact and does not increase or decrease to a large extent over time. However, this analysis shows that analysing solely within-year effects might be incomplete as the firms' labour productivity continues to be positively affected over time as they combine investments in data and R&D.

Marginal Effects and Complementarity

The impact of investments in intangibles on productivity is heterogeneous at various levels of intangible-asset intensity (Corrado et al., 2021; Roth, Sen and Rammer, 2023). These gains or losses may be strictly related to the levels of intangible investments: firms that invest more in intangible assets benefit proportionally more from a productivity perspective (Kaus, Slavtchev and Zimmermann, 2024). This is the case among Italian firms as well: Italian firms have experienced productivity gains using the Industry 4.0-related technologies, but size-related heterogeneity among firms is significant Cefis, Scrofani and Tubiana (2025). To understand these complexities, we estimate the marginal effects of investments in data and R&D to account for heterogeneity in their complementarity. Because the model includes the interaction term $\frac{\widehat{Data}_{i,t-\tau}}{L_{i,t-\tau}} \times \frac{R\&D_{i,t-\tau}}{L_{i,t-\tau}}$, the marginal effects of each variable are conditional on the level of the other. Recall the estimation equation from our model:

$$\text{Lab Prod}_{i,t} \equiv \frac{\log(Y^{PC}_{i,t-\tau})}{\log(\text{EMP}_{i,t-\tau})} = \alpha_1 \frac{\widehat{\text{Data}}_{i,t-\tau}}{L_{i,t-\tau}} + \alpha_2 \frac{\text{R\&D}_{i,t-\tau}}{L_{i,t-\tau}} + \alpha_3 \frac{\widehat{\text{Data}}_{i,t-\tau}}{L_{i,t-\tau}} \times \frac{\text{R\&D}_{i,t-\tau}}{L_{i,t-\tau}} + X_{i,t}\beta + \mu_i + \lambda_t + \psi_s + \varepsilon_{i,t}$$

For simplicity, in this section we set $\tau = 0$. Henceforth, the partial derivative of Lab $\text{Prod}_{i,t}$ with respect to $\frac{\widehat{\text{Data}}_{i,t}}{L_{i,t}}$ is given by:

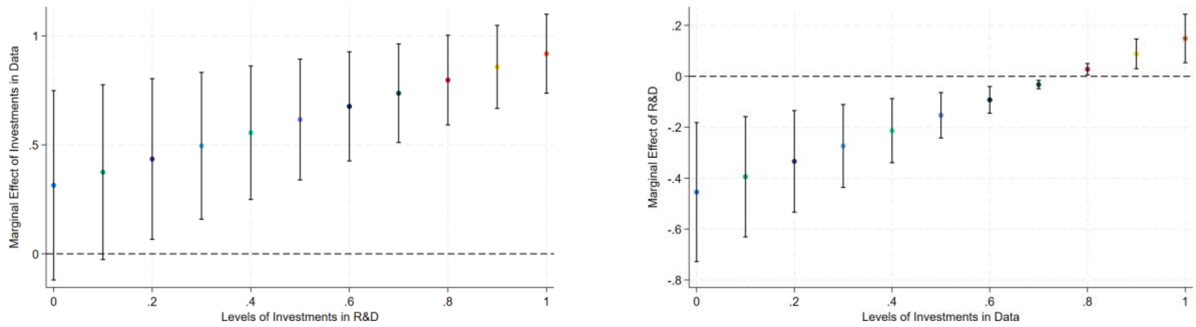
$$\frac{\partial \text{Lab Prod}_{i,t}}{\partial \frac{\widehat{\text{Data}}_{i,t}}{L_{i,t}}} = \alpha^1 + \alpha_3 \cdot \frac{\text{R\&D}_{i,t}}{L_{i,t}}.$$

This expression shows that the effect of data per labour is related linearly to investments in R&D per labour.

Similarly, the partial derivative of Lab $\text{Prod}_{i,t}$ with respect to $\frac{\text{R\&D}_{i,t}}{L_{i,t}}$ is:

$$\frac{\partial \text{Lab Prod}_{i,t}}{\partial \frac{\text{R\&D}_{i,t}}{L_{i,t}}} = \alpha_2 + \alpha_3 \cdot \frac{\widehat{\text{Data}}_{i,t}}{L_{i,t}}.$$

Here, the effect of R&D depends on investments in data. Below, we show the marginal effects introduced so far on labour productivity:



(a) Marginal Effects of Investments in Data

(b) Marginal Effects of Investments in R&D

Figure 6: Marginal Effects of Investments in Data and Investments in R&D on Labour Productivity

At low levels of R&D, the marginal effect is small and statistically insignificant; however, as investments in data increase, the effect becomes stronger and highly significant. At low levels of data investments, the marginal impact of R&D is negative and significant. As firms invest more in data, this negative effect diminishes and switches sign from negative to positive. High investments in data, therefore, are even more critical than previously thought, as they can offset the short-term costs of innovation activities. These results highlight a complementarity between data and R&D, where the positive effect of one variable is amplified when the other is high, confirming previous findings of (Cong et al., 2022).

6 Robustness Checks

As robustness checks, we analyse whether excluding fixed effects can affect our analysis. We then investigate whether our division of firms into control and treatment groups is robust to simply divide the sample by service and manufacturing. We also demonstrate that our results remain unaffected if we use the Italian timeline of

GDPR (2018 instead of 2016) as a reference point. We also estimate our model for the multiplier effect without using our identification strategy to predict the data investment values by directly applying the GDPR estimate Equation 6 to the investments in data, thereby addressing any potential endogeneity among our independent variables.

Our baseline specification in studying the impact of data investments on R&D in Section 3 includes time and firm-fixed effects. Furthermore, our result can be an outcome of a positive trend in both data investments and R&D. In Appendix C.1.2, we present the results without time and firm-fixed effects. We can observe that, even though the magnitude of the impact of data investments on R&D has increased, the sign and significance of this coefficient remain robust. Appendix C.2.1 attempts to scrutinise our results of the complementarity effect analysis presented in 3 by removing the time and fixed effects. We find that our take on the complementarity between these investments is robust to the removal of some of the multi-dimensional fixed effects.

Our analysis also accounts for the bias introduced by using the rule of thumb in Section 3 and how it may be driven primarily by a few sectors. Therefore, we modify our analysis by using the manufacturing sector as the control group and services as the treatment, as services have been mainly impacted and are more data-intensive compared to manufacturing firms across the entire sample. The results of regressions are provided in Appendix C.1.3 for $\tau = 0$. We find that changing treatment and control groups to broader sectors does not affect our main takeaways from the analyses.

We also suspect that the actual adoption of the GDPR was not expected by firms until 2018, as an Italian framework was introduced two years later than the original 2016 deadline. Therefore, firms might not necessarily adjust their behaviour pre-shock. We extend our analysis in Appendix C.1.4 by using the Italian announcement of the GDPR in 2018 as an instrument, rather than the European-wide regulation in 2016. We find that our instrumental variable analysis remains robust to this specification as well, and more investments in data have had a simultaneous positive effect on R&D investments.

Although both investments in data and R&D appear as independent variables in the analysis in Section 3, one may raise the issue that addressing the endogeneity of data to R&D can underestimate their role in labour productivity, and the analysis would be safe without using the predicted values of data from the first-stage regression. Therefore, we use raw values of data per labour 1 into Equation 6 for $\tau = 0$. In this way, we check whether our baseline specification underestimates the impact of independent variables on labour productivity. The results of this application are provided in Appendix C.2.2. We show that investments in data display positive contributions to labour productivity, whereas the evidence on R&D's contribution and the multiplier effect is weak and has a vague direction.

7 Conclusion

This article has addressed the complex relationship between investment in data and R&D. We asked whether, in a context of pervasive and prominent digitalisation and datafication of production, investing in data entails a substantial reduction of market uncertainty and whether this leads firms to be able to 'afford' more of the intrinsic uncertainty associated to investments in R&D, traditionally considered more risky.

The overarching aim of this work is therefore that of unpacking the link between digitalisation on the one hand and activities of invention and innovation on the other hand. The question is relevant, as firms might face a financial trade-off - i.e. digitalising means diverting resources from and divesting in R&D - or they might benefit from a virtuous circle between digitalisation and research, mediated by a reduction in uncertainty - i.e. digitalising means reducing market uncertainty and hence being able to afford more innovation uncertainty.

This overarching question can be operationalised by testing whether there is substitution or complementarity between investments in data and R&D. Should there is a complementary effect, - i.e. investing in data leads to more R&D - we also ask whether this joint investment effort is driving a higher productivity performance, compared to devoting effort only on data or only on R&D.

Based on an IV approach, which exploits the introduction of the GDPR regulation in 2016, we shed light on the complex relationship between data and R&D investments, as well as their joint impact on labour productivity, using firm-level survey data from the Bank of Italy.

Our findings show that investments in data complement those in R&D and support therefore the potential for radical innovation activities. In particular, we are able to show that investments in data and analytic intelligence lead to increased R&D activity within firms via lower uncertainty. Our empirical findings also indicate a complementarity effect of data and R&D investments on labour productivity performance. This suggests that firms benefit from investing in data in the short and long term, emphasising the importance of diversified investment strategies in the era of data-driven decision-making. We also find that introducing the GDPR may have positively affected firms' data investments. This result suggests that regulation may have increased transparency of the legal framework, encouraging larger firms to invest more in data.

Our findings contribute to several strands of literature. First, we provide novel evidence on the mechanisms through which data investments allow more uncertain R&D, demonstrating that forecast-error reduction serves as a key channel. This extends prior work on data and performance (Brynjolfsson, Hitt and Kim, 2011) by identifying a specific pathway linking data to R&D expenditures. Second, our results add to the literature on uncertainty and investment (Bloom, 2007), showing that data-driven uncertainty reduction can enable rather than merely respond to innovation activities. Third, we contribute to understanding the dynamics of digital transformation by documenting how the returns to data investments have evolved as complementary capabilities to traditional R&D investments. This is an important contribution to the literature on intangible assets. Intangible assets are very different in nature: exploring their complementarity from the perspective of reducing market and innovation uncertainty might shed new light on the theoretical mechanisms that underpin the incentive structure of firms in devising investment strategies in intangibles.

These findings carry significant implications for academic research and policy-making, that looks at the complex interplay between datafication, R&D, productivity, and sector-specific dynamics.

Firstly, our research highlights the positive impact of the GDPR on firms' data investments. This suggests that a clear and transparent legal framework can encourage firms, especially larger ones, to invest more in data and analytics. Policymakers can leverage this insight to design regulations that encourage data investments, thereby driving innovation and productivity growth. Secondly, the significant positive correlation between data and R&D investments and labour productivity underscores the importance of a comprehensive investment strategy. It would therefore be critical to promote policies that encourage firms to identify and seek specific complementarities. The one between data and R&D has been found to be particularly relevant, as these investments appear to have a multiplier effect on productivity. Third, despite we do not focus on this here, a potentially significant policy target might be the creation of incentives for B2B data sharing, which would enable the distribution of the benefits of data investments and prevent detrimental market concentration in data-intensive markets, as well as barriers to entry for small players (Goos and Savona, 2024; Savona, 2024). Overall, our findings seem relevant in the case of Italy, where both data intensity and R&D expenditures have historically lagged behind other advanced economies (Crass and Peters, 2014).

Future research could delve deeper into the mechanisms through which data regulation impacts firms' investment decisions. Management quality, human capital, IT infrastructure and other firm characteristics can be the source of significant heterogeneity in the perception of privacy regulation and in the strategies of optimisation of data use in R&D to better productivity performance.

References

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2014. "Return of the Solow Paradox? IT, Productivity, and Employment in US Manufacturing." American Economic Review, 104(5): 394–99.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. 2015. "The skill complementarity of broadband internet." The Quarterly Journal of Economics, 130(4): 1781–1824.

Altomonte, Carlo, Peter Bauer, Alberto Maria Gilardi, and Chiara Soriolo. 2022. "Intangible assets, industry performance and finance during crises." Baffi Carefin Centre Research Paper, , (2022-173).

Angrist, Joshua D. 2022. "Empirical strategies in economics: Illuminating the path from cause to effect." Econometrica, 90(6): 2509–2539.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of causal effects using instrumental variables." Journal of the American statistical Association, 91(434): 444–455.

Añón Higón, Dolores, Jaime Gómez, and Pilar Vargas. 2017. "Complementarities in innovation strategy: do intangibles play a role in enhancing firm performance?" Industrial and Corporate Change, 26(5): 865–886.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz. 2023. "The effect of privacy regulation on the data industry: empirical evidence from GDPR." The RAND Journal of Economics, 54(4): 695–730.

Aronica, Martina, Davide Piacentino, and Pasquale Massimo Picone. 2022. "Big data analytics and internationalisation in Italian firms." Developments in Entrepreneurial Finance and Technology, , ed. David Audretsch, Maksim Belitski, Nada Rejeb and Rosa Caiazza, Chapter 4, 67–81. Edward Elgar Publishing.

Arrighetti, Alessandro, Fabio Landini, and Andrea Lasagni. 2014. "Intangible assets and firm heterogeneity: Evidence from Italy." Research Policy, 43(1): 202–213.

Arrow, Kenneth Joseph. 1972. Economic welfare and the allocation of resources for invention. Springer.

Aydin, Necati, Ahmed Naser Alrajhi, and Jamel Habib Jouini. 2018. "Estimating The Impact Of R&D Spending On Total Factor Roductivity For OECD Countries: Pooled Mean Group Approach." The Journal of Developing Areas, 52(2): 159–168.

Bakhshi, Hasan, Albert Bravo-Biosca, and Juan Mateos-Garcia. 2014. "The analytical firm: Estimating the effect of data and online analytics on firm performance." Nesta Working Paper.

Bank of Italy. 2024. "Survey on Industrial and Service Firms, 2002-2023." Data retrieved from Bank of Italy, Source: Bank of Italy.

Beaudry, Catherine, and Stefano Breschi. 2003. "Are firms in clusters really more innovative?" Economics of innovation and new technology, 12(4): 325–342.

Björkdahl, Joakim. 2020. "Strategies for digitalization in manufacturing firms." California management review, 62(4): 17–36.

Björkdahl, Joakim. 2020. "Strategies for Digitalization in Manufacturing Firms." California Management Review, 62(4): 17–36.

Blackburn, Michael, Jeffrey Alexander, J David Legan, and Diego Klabjan. 2017. "Big Data and the Future of R&D Management: The rise of big data and big data analytics will have significant implications for R&D and innovation management in the next decade." Research-Technology Management, 60(5): 43–51.

Bloom, Nicholas, Steven J Davis, Lucia S Foster, Scott W Ohlmacher, and Itay Saporta-Eksten. 2022. "Investment and Subjective Uncertainty." National Bureau of Economic Research Working Paper 30654.

Bloom, Nick. 2007. "Uncertainty and the Dynamics of R&D." American Economic Review, 97(2): 250–255.

Bolognini, Luca, and Camilla Bistolfi. 2017. "Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation." Computer law & security review, 33(2): 171–181.

Bontadini, Filippo, Carol Corrado, Jonathan Haskel, Massimiliano Iommi, and Cecilia Jona-Lasinio. 2023. "EUKLEMS & INTANProd: Industry Productivity Accounts with Intangibles." Directorate General for Economic and Financial Affairs (DG-ECFIN) of the European Commission Deliverable D2.3.1. Procurement procedure ECFIN/2020/O.P./0001 - Provision of Industry level growth and productivity data with special focus on intangible assets - 2020/S 114-275561.

Bontempi, Maria Elena, and Jacques Mairesse. 2008. "Intangible capital and productivity: an exploration on a panel of Italian manufacturing firms." National Bureau of Economic Research.

Bounie, David, Antoine Dubus, and Patrick Waelbroeck. 2021. "Selling strategic information in digital competitive markets." The RAND Journal of Economics, 52(2): 283–313.

Briguglio, Damiano, Lazaros Dimitriadis, Virginia Maestri, and Gianluca Papa. 2019. "Private Investment in Italy." European Economy Discussion Papers.

Brynjolfsson, Erik, Lorin M Hitt, and Heekyung Hellen Kim. 2011. "Strength in numbers: How does data-driven decisionmaking affect firm performance?" Available at SSRN 1819486.

Brynjolfsson, Erik, Wang Jin, and Xiupeng Wang. 2023. "Information technology, firm size, and industrial concentration." National Bureau of Economic Research.

Busetti, Fabio, Claire Giordano, and Giordano Zevi. 2015. "Main Drivers of the Recent Decline in Italy's Non-Construction Investment." Bank of Italy Occasional Paper, , (276).

Calvino, Flavio, and Chiara Criscuolo. 2019. "Business dynamics and digitalisation."

Calvino, Flavio, Stefano DeSantis, Isabelle Desnoyers-James, Sara Formai, Ilaria Goretti, Silvia Lombardi, Francesco Manaresi, and Giulio Perani. 2022. "Closing the Italian digital gap: The role of skills, intangibles and policies."

Capasso, Marco, Tania Treibich, and Bart Verspagen. 2015. "The medium-term effect of R&D on firm growth." Small Business Economics, 45: 39–62.

Cassetta, Ernesto, Linda Meleo, and Marco Pini. 2016. "The role of digitalization in the internationalization of Italian manufacturing firms." L'industria, , (2): 305–328.

Cefis, Elena, Stefania Scrofani, and Matteo Tubiana. 2025. "Innovation policies and firms' productivity: The Italian Industry 4.0 Plan for digital transformation." The Journal of Technology Transfer, 1–33.

CFOTech. 2023. "Data reduces uncertainty and drives better decisions ." Shannon Williams. Available at: https://cfotech.com.au/story/data-reduces-uncertainty-and-drives-better-decisions, Accessed: 31 July 2023.

Cingano, Federico, and Paolo Pinotti. 2013. "Politicians at work: The private returns and social costs of political connections." Journal of the European Economic Association, 11(2): 433–465.

Cirillo, Valeria, Lucrezia Fanti, Andrea Mina, and Andrea Ricci. 2023. "New digital technologies and firm performance in the Italian economy." Industry and Innovation, 30(1): 159–188.

Cong, Lin William, Wenshi Wei, Danxia Xie, and Longtian Zhang. 2022. "Endogenous growth under multiple uses of data." Journal of Economic Dynamics and Control, 141: 104395.

Corrado, Carol, Chiara Criscuolo, Jonathan Haskel, Alexander Himbert, and Cecilia Jona-Lasinio. 2021. "New evidence on intangibles, diffusion and productivity."

Coyle, Diane. 2019. "Do-it-yourself Digital: the Production Boundary, the Productivity Puzzle and Economic Welfare." Economica, 86: 750–774.

Crass, Dirk, and Bettina Peters. 2014. "Intangible assets and firm-level productivity." ZEW-Centre for European Economic Research Discussion Paper, , (14-120).

Cuervo-Cazurra, Alvaro, and C Annique Un. 2010. "Why some firms never invest in formal R&D." Strategic Management Journal, 31(7): 759–779.

Degeling, Martin, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. 2018. "We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy." arXiv preprint arXiv:1808.05096.

Delbecque, Vincent, Ahmed Bounfour, and Andrés Barreneche. 2015. "Intangibles and Value Creation at the Industrial Level: Delineating Their Complementarities." Intangibles, Market Failure and Innovation Performance, , ed. Ahmed Bounfour and Tsutomu Miyagawa, 27–56. Cham:Springer International Publishing.

Del Monte, Alfredo, and Erasmo Papagni. 2003. “R&D and the growth of firms: empirical analysis of a panel of Italian firms.” Research policy, 32(6): 1003–1014.

Demirer, Mert, Diego J Jiménez Hernández, Dean Li, and Sida Peng. 2024. “Data, privacy laws and firm production: Evidence from the GDPR.” National Bureau of Economic Research.

Determann, Lothar. 2018. “No one owns data.” Hastings LJ, 70: 1.

Dixit, Avinash K, and Robert S Pindyck. 1994. Investment under uncertainty. Princeton university press.

Eeckhout, Jan, and Laura Veldkamp. 2023. “Data and markups: A macro-finance perspective.” Working paper.

Ennen, Edgar, and Ansgar Richter. 2010. “The whole is more than the sum of its parts—or is it? A review of the empirical literature on complementarities in organizations.” Journal of Management, 36(1): 207–233.

EU Commission. 2020. “Two years of the GDPR: Questions and answers.” Accessed: 21 December 2023.

Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. 2019. “Big data and firm dynamics.” Vol. 109, 38–42, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Farrington, Ted, and Azar Alizadeh. 2017. “On the Impact of Digitalization on R&D: R&D practitioners reflect on the range and type of digitalization’s likely effects on R&D management.” Research-Technology Management, 60(5): 24–30.

Fortune, Annetta, and Lois M Shelton. 2014. “Age Matters: Disentangling the Effect of R&D Investment in the Global Chemical Products Industry.” Business Management Dynamics, 3(11).

Frankel, Alexander, and Emir Kamenica. 2019. “Quantifying Information and Uncertainty.” American Economic Review, 109(10): 3650–80.

Fung, Michael K. 2006. “”R&D, knowledge spillovers and stock volatility”.” Accounting and Finance, 46(1): 107 – 124. Cited by: 15.

Gal, Michal S, and Oshrit Aviv. 2020. “The competitive effects of the GDPR.” Journal of Competition Law & Economics, 16(3): 349–391.

Giannopoulos, Athanasios G, and Tatiana P Moschovou. 2023. “Estimating the Value of Information Technology in the Productivity of the Transport Sector.” Future Transportation, 3(2): 601–614.

Goddard, Michelle. 2017. “The EU General Data Protection Regulation (GDPR): European regulation that has a global impact.” International Journal of Market Research, 59(6): 703–705.

Goldberg, Samuel G, Garrett A Johnson, and Scott K Shriver. 2024. “Regulating privacy online: An economic evaluation of the GDPR.” American Economic Journal: Economic Policy, 16(1): 325–358.

Goodridge, Peter, and Jonathan Haskel. 2023. "Accounting for the slowdown in UK innovation and productivity." Economica, 90(359): 780–812.

Goos, Maarten, and Maria Savona. 2024. "The governance of artificial intelligence: Harnessing opportunities and mitigating challenges." Research Policy, 53(3): 104928.

Grabowska, Sandra, and Sebastian Saniuk. 2021. Modern marketing for customized products under conditions of fourth industrial revolution. Routledge: New York, NY, USA.

Grazzi, Marco, Nadia Jacoby, and Tania Treibich. 2016. "Dynamics of investment and firm performance: Comparative evidence from manufacturing industries." Empirical Economics, 51: 125–179.

Griliches, Zvi. 1984. "Introduction to" R & D, Patents, and Productivity"." In R&D, Patents, and Productivity. 1–20. University of Chicago Press.

Groh, Carl-Christian, and Oliver Pfäuti. 2023. "Monetary Policy and Uncertainty in the Data Economy." In Monetary Policy and Uncertainty in the Data Economy: Groh, Carl-Christian— uPfäuti, Oliver. [Sl]: SSRN.

Guellec, Dominique, and Bruno Van Pottelsberghe de la Potterie. 2004. "From R&D to productivity growth: Do the institutional settings and the source of funds of R&D matter?" Oxford bulletin of economics and statistics, 66(3): 353–378.

Haddara, Moutaz, Arra Salazar, and Marius Langseth. 2023. "Exploring the Impact of GDPR on Big Data Analytics Operations in the E-Commerce Industry." Procedia Computer Science, 219: 767–777.

Hagiu, Andrei, and Julian Wright. 2023. "Data-enabled learning, network effects, and competitive advantage." The RAND Journal of Economics, 54(4): 638–667.

Hall, Bronwyn H. 2011. "Innovation and productivity." National bureau of economic research.

Hall, Bronwyn H, Jacques Mairesse, and Pierre Mohnen. 2010. "Measuring the Returns to R&D." In Handbook of the Economics of Innovation. Vol. 2, 1033–1082. Elsevier.

Ichihashi, Shota. 2021. "Competing data intermediaries." The RAND Journal of Economics, 52(3): 515–537.

Igna, Ioana, and Francesco Venturini. 2023. "The determinants of AI innovation across European firms." Research Policy, 52(2): 104661.

IMF. 2015. Private Investment: What's the Holdup? Washington, D.C.:International Monetary Fund. Chapter 4 in World Economic Outlook: Uneven Growth—Short- and Long-Term Factors, April 2015.

ISTAT. 2018. "Report on competitiveness." Italian National Institute of Statistics (ISTAT).

Jia, Jian, Ginger Zhe Jin, and Liad Wagman. 2020. "GDPR and the Localness of Venture Investment." Available at SSRN 3436535.

Johnson, Garrett A. 2023. "Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond." The Economics of Privacy. University of Chicago Press.

Jones, Charles I, and Christopher Tonetti. 2020. "Nonrivalry and the Economics of Data." American Economic Review, 110(9): 2819–2858.

Kagermann, Henning, Wolfgang Wahlster, and Johannes Helbig. 2013. "Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0." Industry-Science Research Alliance, Germany. Final report of the Industrie 4.0 Working Group.

Kaus, Wolfhard, Viktor Slavtchev, and Markus Zimmermann. 2024. "Intangible capital and productivity: Firm-level evidence from German manufacturing." Oxford Economic Papers, 76(4): 970–996.

Klein, Tobias J, Madina Kurmangaliyeva, Jens Prufer, and Patricia Prufer. 2022. "How important are user-generated data for search result quality? Experimental evidence." In How Important Are User-Generated Data for Search Result Quality? Experimental Evidence: Klein, Tobias J.— uKurmangaliyeva, Madina— uPrufer, Jens— uPrufer, Patricia. [Sl]: SSRN.

Koski, Heli, and Nelli Valmari. 2020. "Short-term Impacts of the GDPR on Firm Performance." ETLA Working Papers.

Kotschy, Rainer, and David E Bloom. 2023. "Population aging and economic growth: From demographic dividend to demographic drag?"

Kumbhakar, Subal C, Raquel Ortega-Argilés, Lesley Potters, Marco Vivarelli, and Peter Voigt. 2012. "Corporate R&D and firm efficiency: evidence from Europe's top R&D investors." Journal of Productivity Analysis, 37: 125–140.

Layton, Roslyn, and Silvia Elaluf-Calderwood. 2019. "A Social Economic Analysis of the Impact of GDPR on Security and Privacy Practices." 1–6.

Lefrere, Vincent, Logan Warberg, Cristobal Cheyre, Veronica Marotta, Alessandro Acquisti, et al. 2020. "The impact of the GDPR on content providers."

Lichtenberg, Frank R, and Donald Siegel. 1991. "The impact of R&D investment on productivity–New evidence using linked R&D–LRD data." Economic inquiry, 29(2): 203–229.

Li, He, Lu Yu, and Wu He. 2019. "The impact of GDPR on global technology development."

Lindgren, Peter. 2016. "GDPR regulation impact on different business models and businesses." Journal of Multi Business Model Innovation and Technology, 4(3): 241–254.

Li, Wendy CY, Makoto Nirei, and Kazufumi Yamana. 2019. Value of data: there's no such thing as a free lunch in the digital economy. RIETI.

Machlup, Fritz. 1962. The production and distribution of knowledge in the United States. Vol. 278, Princeton university press.

Maestas, Nicole, Kathleen J. Mullen, and David Powell. 2023. "The Effect of Population Aging on Economic Growth, the Labor Force, and Productivity." American Economic Journal: Macroeconomics, 15(2): 306–332.

Manaresi, Francesco, and Nicola Pierri. 2024. "The Asymmetric Effect of Credit Supply on Firm-Level Productivity Growth." Journal of Money, Credit and Banking, 56(4): 677–704.

Marrocu, Emanuela, Raffaele Paci, and Marco Pontis. 2012. "Intangible capital and firms' productivity." Industrial and corporate change, 21(2): 377–402.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

McAfee, Andrew, Erik Brynjolfsson, Thomas H Davenport, DJ Patil, and Dominic Barton. 2012. "Big data: the management revolution." Harvard business review, 90(10): 60–68.

Mohades, Siavash, Giulia Piccillo, Tania Treibich, and Maria Savona. 2025. "Firms' Investment and Capacity Utilisation: The Role of Financial Constraints and Uncertainty." CESifo CESifo Working Paper 12108.

Mohnen, Pierre. 2019. R&D, innovation and productivity. Springer.

Müller, Oliver, Maria Fay, and Jan Vom Brocke. 2018. "The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics." Journal of management information systems, 35(2): 488–509.

Nepelski, Daniel, Juraj Stancik, et al. 2011. "The Top World R&D-investing Companies from the ICT Sector: A Company-level Analysis." JRC Scientific and Technical Report IPTS.

Niebel, Thomas, Fabienne Rasel, and Steffen Viete. 2019. "BIG data–BIG gains? Understanding the link between big data analytics and innovation." Economics of Innovation and New Technology, 28(3): 296–316.

Ning, Chao, and Fengqi You. 2019. "Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming." Computers & Chemical Engineering, 125: 434–448.

Nonnis, Alberto, Ahmed Bounfour, and Keungoui Kim. 2023. "Knowledge spillovers and intangible complementarities: Empirical case of European countries." Research Policy, 52(1): 104611.

Nucci, Francesco, Chiara Puccioni, and Ottavio Ricchi. 2022. "Digital Technologies and Productivity: A Firm-Level Investigation for Italy." Government of the Italian Republic (Italy), Ministry of Economy and Finance, Department of the Treasury Working Paper, , (3).

OECD. 2020. “A roadmap toward a common framework for measuring the digital economy.” Report for the G20 Digital Economy Task Force, available at.

Park, Chankook, and Minkyu Kim. 2021. “Characteristics influencing digital technology choice in digitalization projects of energy industry.” Environmental and Climate Technologies, 25(1): 356–366.

Peters, Bettina, Mark J Roberts, Van Anh Vuong, and Helmut Fryges. 2017. “Estimating dynamic R&D choice: an analysis of costs and long-run benefits.” The RAND Journal of Economics, 48(2): 409–437.

Piccarozzi, Michela, Barbara Aquilani, and Corrado Gatti. 2018. “Industry 4.0 in management studies: A systematic literature review.” Sustainability, 10(10): 3821.

Pozzi, Andrea, and Fabiano Schivardi. 2016. “Demand or productivity: What determines firm growth?” The RAND Journal of Economics, 47(3): 608–630.

Romer, Paul M. 1990. “Endogenous technological change.” Journal of political Economy, 98(5, Part 2): S71–S102.

Roth, Felix, Ali Sen, and Christian Rammer. 2023. “The role of intangibles in firm-level productivity–evidence from Germany.” Industry and Innovation, 30(2): 263–285.

Sabatino, Lorien, and Geza Sapi. 2022. “Online privacy and market structure: Theory and evidence.” Information Economics and Policy, 60: 100985.

Salvi, Antonio, Filippo Vitolla, Michele Rubino, Anastasia Giakoumelou, and Nicola Raimo. 2021. “Online information on digitalisation processes and its impact on firm value.” Journal of Business Research, 124: 437–444.

Savona, Maria. 2019. “The Value of Data: Towards a Framework to Redistribute It.” Science Policy Research Unit, University of Sussex SWPS 2019-21. Research Note.

Savona, Maria. 2024. “Data Governance: Main Challenges.” Vol. 25, 28–31, CESifo.

Schiffer, Mirjam, and Beatrice Weder. 2001. Firm size and the business environment: Worldwide survey results. Vol. 43, World Bank Publications.

Schivardi, Fabiano, and Tom Schmitz. 2019. “The IT Revolution and Southern Europe’s Two Lost Decades.” Journal of the European Economic Association, 18(5): 2441–2486.

Sharma, Amalesh, Anirban Adhikary, and Sourav Bikash Borah. 2020. “Covid19 s impact on supply chain decisions: Strategic insights from NASDAQ 100 firms using Twitter data.” Journal of business research, 117: 443–449.

Shefer, Daniel, and Amnon Frenkel. 2005. “R&D, firm size and innovation: an empirical analysis.” Technovation, 25(1): 25–32.

Siliverstovs, Boriss, et al. 2016. "R&D and non-linear productivity growth." Research policy, 45(3): 634–646.

Simon, Phil. 2013. Too big to ignore: the business case for big data. Vol. 72, John Wiley & Sons.

Taufick, Roberto D. 2021. "The underdeterrence, underperformance response to privacy, data protection laws." Technology in Society, 67: 101752.

Teece, David J. 1986. "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy." Research Policy, 15(6): 285–305.

Truant, Elisa, Laura Broccardo, and Léo-Paul Dana. 2021. "Digitalisation boosts company performance: an overview of Italian listed companies." Technological Forecasting & Social Change, 173: 121173.

Tysiac, Ken. 2015. "Data analytics helps auditors gain deep insight." Journal of Accountancy, 219(4): 52.

Urban, Tobias, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2018. "The unwanted sharing economy: An analysis of cookie syncing and user transparency under GDPR." arXiv preprint arXiv:1811.08660.

Van Oijen, Marcel. 2017. "Bayesian methods for quantifying and reducing uncertainty and error in forest models." Current Forestry Reports, 3: 269–280.

Voss, W Gregory, and Kimberly A Houser. 2019. "Personal data and the GDPR: providing a competitive advantage for US companies." American Business Law Journal, 56(2): 287–344.

Wondrium. 2016. "How Gathering Data can Reduce Uncertainty ." Connel Fullenkamp. Available at: https://www.wondriumdaily.com/reducing-uncertainty-creating-information, Accessed: 31 July 2023.

Wu, Kuo-Jui, Ching-Jong Liao, Ming-Lang Tseng, Ming K Lim, Jiayao Hu, and Kimhua Tan. 2017. "Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties." Journal of Cleaner Production, 142: 663–676.

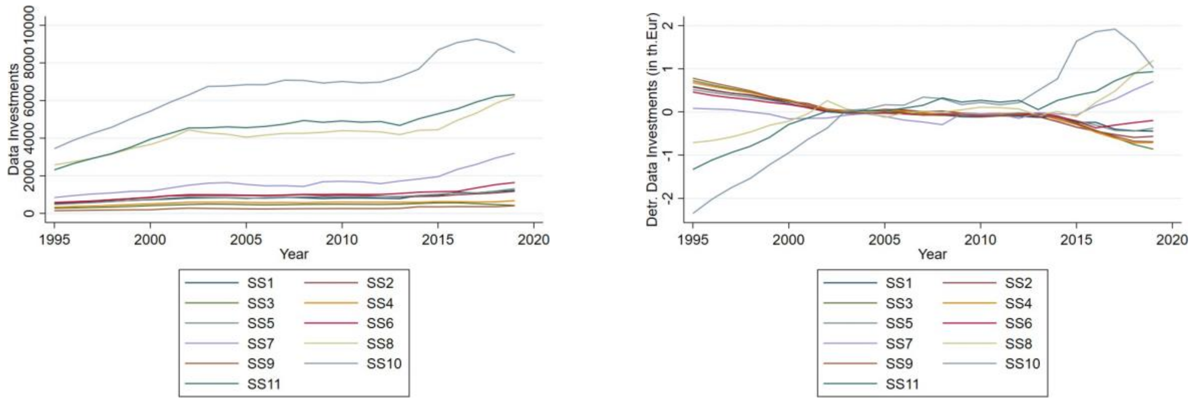
A Additional Descriptive Tables

A.1 Sectoral Division of INVIND Data

Variable	Values	Description	ATECO 2002	ATECO 2007	NACE
settor11	SS1	Food industries, beverages, and tobacco products	DA	10, 11, 12	C10-C12
	SS2	Textiles, clothing, and hide, leather, and footwear products	DB, DC	13, 14, 15	C13-C15
	SS3	Coke manufacturing, chemical industry, rubber, and plastics	DF, DG, DH	19, 20, 21, 22	C19-C21
	SS4	Processing of non-metallic minerals	DI	23	C22-C23
	SS5	Metal engineering industry	DJ, DK, DL, DM	24, 25, 26, 27, 28, 29, 30, 33	C24-C25
	SS6	Other manufacturing industries	DD, DE, DN	16, 17, 18, 31, 32	C16-C18,C26-C33,E
	SS7	Other industries excluding construction	CA, CB, CE	05, 06, 07, 08, 09, 35, 36, 37, 38, 39	D
	SS8	Wholesale and retail commerce	G	45, 46, 47	G45-G46
	SS9	Hotels and restaurants	H	55, 56	I
	SS10	Transport and communications	I	49, 50, 51, 52, 53, 58, 59, 60, 61, 62, 63	H,J
	SS11	Real estate activities, IT, etc.	K	68, 69, 70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82	L,M,N

Table A.1: Sector categorisation for the *settor11* variable in the INVIND dataset and its comparability to ATECO by ISTAT.

B EUKLEMS Figures



(a) Investments in Databases and Software

(b) Detrended Investments in Databases and Software

Figure A.1: Sectors' Investments in Data, Source: EUKLEMS-INTANProd

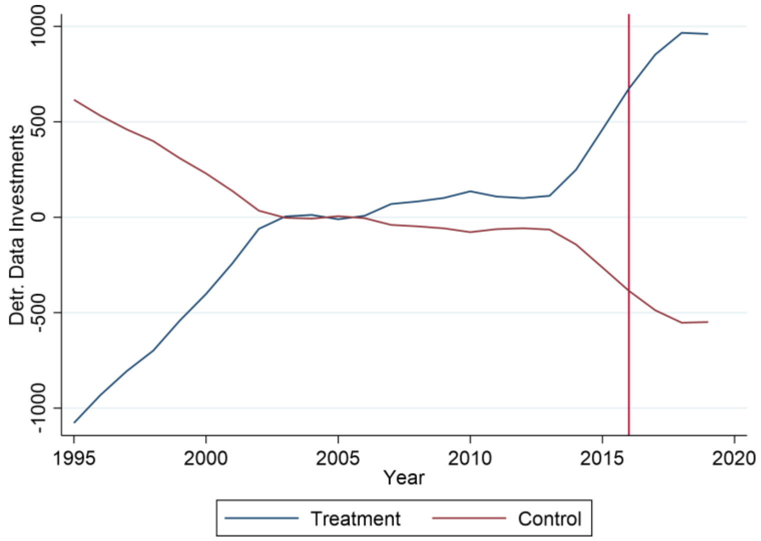


Figure A.2: Detrended Investments in Databases and Software in Control and Treatment Groups, Source: EUKLEMS INTANProd for EU Economies (1995-2019). The Line represents GDPR in 2016

C Robustness Checks - Regression Tables

C.1 More Data, More R&D?

C.1.1 Removing firm and time fixed effects

	Dependent Variable: Log of R&D Investments Instrument for Data Investments: GDPR Implementation 2016					
Data inv per Labour	1.620***	4.372***	4.372***	4.367***	-0.544	-0.552
	(0.341)	(0.360)	(0.361)	(0.361)	(0.603)	(0.603)
Sales Growth$_{t-1}$	-0.054***	-0.124***	-0.124***	-0.124***	-0.001	-0.001
	(0.013)	(0.014)	(0.014)	(0.014)	(0.022)	(0.022)
Log of Age	-0.017	-0.082***	-0.080***	-0.081***	-0.034*	-0.034*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.016)	(0.016)
Emp Growth$_{t-1}$	-0.023	-0.038	-0.029	-0.029	-0.062	-0.061
	(0.022)	(0.022)	(0.022)	(0.022)	(0.033)	(0.033)
Profit$_{t-1}$	0.019**	0.021***	0.022***	0.022***	0.016	0.016
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)
Large			-0.044***	-0.044***	-0.072***	-0.072***
			(0.010)	(0.010)	(0.014)	(0.014)
North			0.017	0.017	0.030	0.030
			(0.013)	(0.013)	(0.018)	(0.018)
Fin Cons				-0.005		0.011
				(0.009)		(0.015)
High Tech$_{t-1}$					0.014*	0.014*
					(0.007)	(0.007)
Constant	-0.192	-2.087***	-2.084***	-2.079***	1.505***	1.508***
	(0.236)	(0.247)	(0.248)	(0.248)	(0.416)	(0.416)
Sector FE		✓	✓	✓	✓	✓
Firms	3,633	3,633	3,626	3,626	1,903	1,903
Observations	14,054	14,054	13,985	13,985	5,163	5,163

Treated Sectors: Energy and extraction - Wholesale and retail commerce - Real Estate and IT - Transport and communications

Table A.2: Regression Results for the second stage regression in Equation 2 without firm and time FE

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

C.1.2 Removing instrumenting data with GDPR

	Dependent Variable: Log of R&D Investments Instrument for Data Investments: GDPR Implementation 2016					
Data inv per Labour	0.055***	0.056***	0.056***	0.056***	0.069***	0.069***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.018)	(0.018)
Sales Growth$_{t-1}$	-0.014	-0.014	-0.014	-0.014	-0.001	-0.002
	(0.012)	(0.012)	(0.012)	(0.012)	(0.019)	(0.019)
Log of Age	0.003	0.002	-0.000	-0.000	-0.028	-0.028
	(0.016)	(0.016)	(0.016)	(0.016)	(0.036)	(0.036)
Emp Growth$_{t-1}$	-0.001	0.001	0.002	0.002	-0.057	-0.057
	(0.026)	(0.026)	(0.026)	(0.026)	(0.039)	(0.039)
Profit$_{t-1}$	0.008	0.008	0.009	0.009	0.013	0.013
	(0.007)	(0.007)	(0.007)	(0.007)	(0.011)	(0.011)
Large			-0.023	-0.023	0.007	0.008
			(0.017)	(0.017)	(0.028)	(0.028)
North			-0.039	-0.039	-0.338	-0.338
			(0.096)	(0.096)	(0.202)	(0.202)
Fin Cons				0.001		-0.014
				(0.010)		(0.018)
High Tech$_{t-1}$					0.013	0.013
					(0.008)	(0.008)
Constant	0.967***	1.067***	1.121***	1.121***	1.554***	1.551***
	(0.055)	(0.138)	(0.158)	(0.158)	(0.230)	(0.230)
Sector FE		✓	✓	✓	✓	✓
Firms	3,011	3,011	3,006	3,006	1,591	1,591
N	11,015	11,015	10,959	10,959	4,125	4,125

Table A.3: Regression Results for the second stage regression in Equation 2 without instrumenting Data Investments

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

C.1.3 Manufacturing vs Services

	Dependent Variable: Log of R&D Investments Instrument for Data Investments: GDPR Implementation 2016					
Data per labour	1.943***	1.995***	1.999***	1.983***	-0.611	-0.620
	(0.524)	(0.527)	(0.529)	(0.530)	(0.833)	(0.834)
Sales Growth$_{t-1}$	-0.066***	-0.066***	-0.068***	-0.068***	-0.039	-0.039
	(0.018)	(0.018)	(0.018)	(0.018)	(0.050)	(0.050)
Age	0.051**	0.052**	0.046*	0.046*	-0.257**	-0.258**
	(0.019)	(0.019)	(0.020)	(0.020)	(0.090)	(0.090)
Emp Growth$_{t-1}$	-0.072*	-0.072*	-0.071*	-0.071*	-0.121	-0.120
	(0.031)	(0.031)	(0.031)	(0.031)	(0.067)	(0.067)
Profit$_{t-1}$	0.025**	0.025**	0.027**	0.026**	0.043	0.042
	(0.009)	(0.009)	(0.009)	(0.009)	(0.031)	(0.031)
Large			0.008	0.008	0.072	0.072
			(0.026)	(0.026)	(0.057)	(0.057)
North			0.300	0.300		
			(0.184)	(0.184)		
Fin Cons				-0.013		0.018
				(0.012)		(0.047)
High Tech$_{t-1}$					0.024	0.025
					(0.025)	(0.025)
Constant	-0.618	-0.471	-0.692	-0.678	2.287***	2.294***
	(0.392)	(0.448)	(0.471)	(0.471)	(0.623)	(0.625)
Sector FE		✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Firms	2,930	2,930	2,930	2,930	1,001	1,001
Observations	8,193	8,193	8,151	8,151	1,358	1,358
Treated Sectors: All in Services						

Table A.4: Regression Results for Equation 2 for Manufacturing and Services sectors as control and treatments, respectively

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

C.1.4 GDPR 2018 instead of GDPR 2016

	Dependent Variable: Log of R&D Investments					
	Instrument for Data Investments: GDPR Announcement 2018					
Data inv per Labour	0.799	0.912*	0.914*	0.915*	-0.973	-0.976
	(0.414)	(0.420)	(0.420)	(0.420)	(0.516)	(0.516)
Sales Growth$_{t-1}$	-0.045**	-0.048**	-0.048**	-0.049**	0.018	0.018
	(0.016)	(0.016)	(0.016)	(0.016)	(0.022)	(0.022)
Log of Age	-0.009	-0.012	-0.013	-0.013	0.007	0.007
	(0.017)	(0.017)	(0.017)	(0.017)	(0.035)	(0.035)
Emp Growth$_{t-1}$	-0.034	-0.035	-0.032	-0.032	-0.095**	-0.095**
	(0.024)	(0.024)	(0.024)	(0.024)	(0.035)	(0.035)
Profit$_{t-1}$	0.011	0.011	0.012	0.011	0.006	0.006
	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
Large			-0.037*	-0.037*	-0.019	-0.020
			(0.016)	(0.016)	(0.025)	(0.025)
North			-0.150	-0.150	-0.727***	-0.727***
			(0.098)	(0.098)	(0.151)	(0.151)
Fin Cons				-0.003		0.006
				(0.009)		(0.016)
High Tech$_{t-1}$					0.011	0.011
					(0.008)	(0.008)
Constant	0.445	0.536	0.684*	0.684*	2.487***	2.490***
	(0.286)	(0.317)	(0.326)	(0.326)	(0.404)	(0.404)
Sector FE		✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
r2	0.016	0.017	0.018	0.018	0.038	0.038
Firms	3,553	3,553	3,551	3,551	1,902	1,902
Observations	13,273	13,273	13,230	13,230	5,173	5,173

Treated Sectors: Energy and extraction - Wholesale and retail commerce - Real Estate and IT - Transport and communications

Table A.5: Regression Results for Equation 2 for GDPR 2016 as Instrument

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

C.2 Complementarity Analysis

C.2.1 Removing Firm and Time Fixed Effects

	Dependent Variable: Labour Productivity						
Data per Labour	-0.091	-0.151	-0.151	-0.367	-0.418	-0.185	0.728
	(0.216)	(0.216)	(0.216)	(0.229)	(0.228)	(0.230)	(0.440)
Log of R&D Investments	-0.629***	-0.618***	-0.618***	-0.649***	-0.626***	-0.546***	-0.272
	(0.137)	(0.137)	(0.137)	(0.147)	(0.147)	(0.145)	(0.266)
×Log of R&D Investments							
	(0.183)	(0.183)	(0.183)	(0.196)	(0.196)	(0.194)	(0.355)
Log of Other Investments				0.014***	0.014***	0.013***	0.013***
				(0.001)	(0.001)	(0.001)	(0.001)
Age						-0.015***	-0.011
						(0.003)	(0.006)
North						0.096***	0.085***
						(0.008)	(0.011)
Large						0.075***	0.098***
						(0.004)	(0.006)
High Tech$_{t-1}$							-0.004
							(0.002)
Constant	0.381*	0.502**	0.502**	0.543**	0.659***	0.446**	-0.261
	(0.162)	(0.162)	(0.162)	(0.171)	(0.171)	(0.170)	(0.326)
Sector FE		✓		✓	✓	✓	✓
Firms	2,809	2,809	2,809	2,310	2,310	2,302	1,356
Observations	12,295	12,295	12,295	9,804	9,804	9,747	3,747

Table A.6: Regression Results: Complementarity analysis in Equation 6 without time and firm fixed effects

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

C.2.2 Removing IV from Complementarity Analyses

	Dependent Variable: Labour Productivity						
	Data Investments instrumented by GDPR 2016						
Data per Labour	0.015	0.014	0.014	0.017	0.017	0.013	0.015
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.016)
Log of R&D Investments	0.002	0.002	0.002	0.001	0.001	0.001	-0.003
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.012)
Log of Data Investments	-0.007	-0.008	-0.007	-0.010	-0.010	-0.008	-0.010
×R&D inv per Labour	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.014)
Log of Other Investments				0.007***	0.007***	0.006***	0.005***
				(0.001)	(0.001)	(0.001)	(0.001)
Age						-0.006*	0.001
						(0.003)	(0.007)
North						0.001	-0.057
						(0.022)	(0.059)
Large						0.051***	0.039***
						(0.004)	(0.008)
High Tech$_{t-1}$							0.001
							(0.002)
Constant	0.347***	0.346***	0.430***	0.334***	0.452***	0.432***	0.524***
	(0.007)	(0.007)	(0.051)	(0.009)	(0.060)	(0.063)	(0.087)
Sector FE		✓		✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓
Firms	3,192	3,192	3,192	2,659	2,659	2,651	1,497
N	13,050	13,050	13,050	10,589	10,589	10,514	3,958

Treated Sectors: Energy and extraction - Wholesale and retail commerce - Real Estate and IT - Transport and communications

Table A.7: Regression Results: Complementarity analysis in Equation 6

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

D Lagged Productivity Gains

Dependent Variable: Labour Productivity							
Lag (τ)	0	1	2	3	4	5	6
Data inv per Labour	0.314	-0.253	-0.679*	-1.248***	-1.256***	-0.552	0.099
	(0.222)	(0.244)	(0.272)	(0.276)	(0.293)	(0.331)	(0.337)
R&D inv per Labour	-0.455**	-0.399*	-0.510**	-0.603***	-0.725***	-0.190	0.188
	(0.139)	(0.155)	(0.174)	(0.175)	(0.187)	(0.207)	(0.214)
Data inv per Labour	0.603**	0.538**	0.701**	0.819***	0.983***	0.255	-0.241
$\times$R&D inv per Labour	(0.188)	(0.209)	(0.235)	(0.235)	(0.252)	(0.280)	(0.289)
All other Investments	0.005***	0.005***	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age	-0.019***	-0.002	-0.007	-0.013*	-0.009	-0.008	-0.004
	(0.004)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
North	0.003	0.004	0.003	-0.039	-0.026	-0.039	-0.042
	(0.024)	(0.025)	(0.028)	(0.034)	(0.034)	(0.033)	(0.039)
Large	0.032***	0.032***	0.034***	0.040***	0.025***	0.023***	0.018*
	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)
Constant	0.245	0.646***	0.949***	1.419***	1.390***	0.889***	0.427
	(0.167)	(0.185)	(0.206)	(0.208)	(0.222)	(0.252)	(0.256)
Sector FE	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓
R2	0.045	0.028	0.025	0.032	0.022	0.018	0.022
Firms	2,251	1,985	1,860	1,736	1,613	1,498	1,354
Observations	9,206 7,745	6,884	6,176	5,501	4,861	4,110	

Table A.8: Regression Results: Complementarity analysis in Equation 6 for $\tau > 0$ with fixed effects

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

E Lagged Data and R&D

	Dependent Variable: Log of R&D Investments Instrument for Data Investments: GDPR Announcement 2018					
Lag (τ)	0	1	2	3	4	5
Log of Data Investments	1.649***	0.098	-0.026	0.554	0.725*	0.756*
	(0.437)	(0.296)	(0.305)	(0.335)	(0.343)	(0.349)
	No Lags					
Sales Growth$_{t-1}$	-0.070***	-0.024*	-0.026*	-0.031*	-0.029*	-0.031*
	(0.016)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
Log of Age	-0.026	0.020	0.012	-0.007	-0.020	-0.012
	(0.017)	(0.019)	(0.021)	(0.023)	(0.021)	(0.020)
Emp Growth$_{t-1}$	-0.032	-0.032	-0.053	-0.063	-0.067*	-0.067
	(0.024)	(0.027)	(0.029)	(0.032)	(0.034)	(0.035)
Large	-0.037*	-0.043*	-0.037*	-0.038	-0.033	-0.015
	(0.016)	(0.017)	(0.019)	(0.020)	(0.020)	(0.021)
North	-0.149	-0.315**	-0.396***	-0.565***	-0.635***	-0.742***
	(0.098)	(0.105)	(0.115)	(0.137)	(0.141)	(0.150)
Profit$_{t-1}$	0.011	0.016*	0.011	0.010	0.010	0.010
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Constant	0.177	1.118***	1.293***	1.052***	1.010**	0.998**
	(0.336)	(0.270)	(0.283)	(0.314)	(0.327)	(0.344)
Sector FE	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
r2	0.019	0.020	0.022	0.022	0.023	0.023
Firms	3,551	3,081	2,697	2,381	2,341	2,176
Observations	13,230	11,512	10,078	8,958	8,410	7,887

Table A.9: Regression Results: Causal analysis in Equation 2 for $\tau > 0$ with random effects

*** Significant with 99% Confidence Interval, ** Significant with 95% Confidence Interval, * Significant with 90% Confidence Interval

F Rolling Window Mediation Analysis

Table A.10: Rolling Window Mediation Analysis: Forecast Error Uncertainty (2-Year Windows)

Window			Coefficients		Indirect Effect	
Start	End	N	γ_1	β_2	$\gamma_1 \times \beta_2$	Std. Error
2012	2013	2,929	0.015	−1.288	−0.020	0.097
2013	2014	3,087	−0.084	3.956***	−0.334***	0.068
2014	2015	3,044	0.065	0.098	0.006	0.085
2015	2016	3,018	−0.066	−2.331	0.153*	0.083
2016	2017	3,142	0.005	0.614	0.003	0.051
2017	2018	3,037	0.005	−0.910	−0.005	0.050
2018	2019	2,770	0.008	1.441	0.012	0.087
2019	2020	2,578	−0.036	−0.625	0.022	0.041
2020	2021	2,527	−0.038	−1.078	0.041	0.083
2021	2022	2,873	−0.059	−1.186	0.070	0.100
2022	2023	3,159	−0.027	8.631***	−0.232***	0.061

Notes: This table reports results from rolling 2-year window mediation analysis. γ_1 is the coefficient from the first-stage regression of forecast error (FCE) on data investment per labour. β_2 is the coefficient from the second-stage IV regression of R&D per labour on instrumented FCE. The indirect effect is the product $\gamma_1 \times \beta_2$, estimated via bootstrap with 200 replications. All regressions include firm fixed effects, year fixed effects, sector fixed effects, and controls for lagged sales growth, lagged size growth, age, size, location, past profitability, and financial constraints. *** p<0.01, ** p<0.05, * p<0.10.

Table A.11: Rolling Window Mediation Analysis: Forecast Error Uncertainty (5-Year Windows)

Window			Coefficients		Indirect Effect	
Start	End	N	γ_1	β_2	$\gamma_1 \times \beta_2$	Std. Error
2007	2011	7,018	−0.010	−2.940	0.030	0.107
2008	2012	6,891	−0.010	0.563	−0.006	0.029
2009	2013	6,998	−0.026	−0.260	0.007	0.056
2010	2014	7,210	−0.016	−0.484	0.008	0.038
2011	2015	7,365	−0.007	0.367	−0.003	0.040
2012	2016	7,494	−0.012	1.758	−0.020	0.040
2013	2017	7,726	−0.019	−8.122***	0.155***	0.022
2014	2018	7,602	−0.015	−1.443	0.021	0.032
2015	2019	7,409	−0.104***	−0.423	0.044	0.029
2016	2020	7,136	−0.072**	−0.552	0.040	0.027
2017	2021	6,918	−0.070**	−0.801	0.056**	0.026
2018	2022	6,867	−0.053	−3.631***	0.194***	0.029
2019	2023	7,040	−0.085**	−18.732***	1.596***	0.027
2020	2024	5,686	−0.034	−8.571***	0.288**	0.050

Notes: This table reports results from rolling 5-year window mediation analysis. γ_1 is the coefficient from the first-stage regression of forecast error (FCE) on data investment per labour. β_2 is the coefficient from the second-stage IV regression of R&D per labour on instrumented FCE, using data investment as an instrument. The indirect effect is the product $\gamma_1 \times \beta_2$, estimated via bootstrap with 200 replications. All regressions include firm fixed effects, year fixed effects, sector fixed effects, and controls for lagged sales growth, lagged size growth, age, size, location, past profitability, and financial constraints. *** p<0.01, ** p<0.05, * p<0.10.