Working Papers | 113

16.1.2024

Digitalization and Resilience

DATA ASSETS AND FIRM PRODUCTIVITY GROWTH DURING THE COVID-19 PANDEMIC

Heli Koski (corresponding author)
ETLA Economic Research, Finland
heli.koski@etla.fi

Paolo Fornaro
Labour Institute for Economic Research LABORE, Finland
paolo.fornaro@labore.fi

Suggested citation:
Koski, Heli & Fornaro, Paolo (16.1.2024).
"Digitalization and Resilience: Data Assets and Firm Productivity Growth During the COVID-19 Pandemic".
ETLA Working Papers No 113.
http://pub.etla.fi/ETLA-Working-Papers-113.pdf

Abstract

This study investigates the impact of firm-level investments in data assets on productivity growth during the COVID-19 pandemic, utilizing matched employer-employee data of 13,609 Finnish firms for 2015–2020.

Our estimation results indicate that firms with greater pre-pandemic investments in software and database assets and ICT experienced significantly higher labor productivity growth in the first year of the pandemic. Notably, these positive effects are predominantly observed in the service sector, while manufacturing companies did not exhibit statistically significant impacts. Furthermore, our analysis highlights that large service companies with greater investments in data assets demonstrated higher labor productivity growth than their counterparts.

We also identify a noteworthy complementarity between a firm's investments in ICT and databases and employees' skills, as measured by education level. Interestingly, our empirical findings underscore that firms investing more in data, databases and ICT were statistically significantly more likely to belong to the productivity frontier of their industry.

Tiivistelmä

Investoinnit dataan ja yritysten tuottavuuskasvu COVID-19-pandemian aikana

Tämä tutkimus tarkastelee yritysten datavarantoihin tekemien investointien vaikutusta tuottavuuskasvuun COVID-19-pandemian aikana hyödyntäen yhdistettyä työntekijä-työnantaja-aineistoa 13 609 suomalaisesta yrityksestä vuosilta 2015–2020.

Aineistoanalyysimme osoittaa, että pandemiaa edeltävinä vuosina enemmän ohjelmistoihin ja tietokantoihin sekä digitalisaatioon investoineiden yritysten työn tuottavuus kasvoi merkittävästi muita yrityksiä enemmän ensimmäisenä pandemiavuonna. Investointien positiiviset tuottavuusvaikutukset näkyvät kuitenkin pelkästään palvelusektorilla, kun taas teollisuusyritysten joukossa emme löytäneet näyttöä tilastollisesti merkitsevistä vaikutuksista. Lisäksi enemmän dataomaisuuteen ennen pandemiaa investoineiden suurten palvelualan yritysten työn tuottavuus kasvoi muita yrityksiä enemmän vuonna 2020.

Analyysimme osoittaa myös, että työntekijöiden koulutustasolla mitatut taidot täydentävät yrityksen investointeja ICT:hen ja tietokantoihin ja lisäsivät työn tuottavuuden kasvua. Enemmän dataan, tietokantoihin sekä ICT:hen ennen pandemiaa investoineet yritykset kuuluivat ensimmäisenä pandemiavuonna muita yrityksiä todennäköisemmin tuottavuuden eturintamaan.

Ph.D. (Econ.) **Heli Koski** is a Research Director at Etla Economic Research.

Ph.D. (Econ.) **Paolo Fornaro** is a Chief Researcher at Labore.

KTT **Heli Koski** on Elinkeinoelämän tutkimuslaitoksen tutkimusjohtaja.

Ph.D. (Econ.) **Paolo Fornaro** on Laboren johtava tutkija.

Acknowledgements: We thank Eric Bartelsman for helpful comments. The authors are grateful for Business Finland for its financial support.

Kiitokset: Kiitämme Eric Bartelsmania hyödyllisistä kommenteista ja Business Finlandia tutkimuksen rahoittamisesta.

Keywords: Data assets, Digitalization, Productivity, Growth, Resilience, Pandemics

Asiasanat: Dataomaisuus, Digitalisaatio, Tuottavuus, Kasvu, Resilienssi, Pandemia

JEL: D22, L25, O33

1. Introduction

This paper presents a novel empirical investigation into the role of data assets in shaping firms' resilience during the unprecedented challenges posed by the COVID-19 pandemic. Despite the increasing recognition of the strategic importance of data, prior studies have not explicitly explored the impact of data assets on firms' productivity growth. Against the backdrop of a significant digital divide in which larger firms often outpace their counterparts in adopting advanced digital technologies [1], our study delves into whether firms that have invested more in digitalization and embraced business models leveraging data assets were better equipped to weather the economic shock induced by the pandemic. Specifically, we investigate how investments in data assets affected firms' productivity growth during this critical period. We hypothesize that firms with a higher share of data capital or greater dependence on data assets were able to gain a productivity advantage amidst the COVID-19 pandemic. This prospective advantage is attributed to the inherent flexibility and remote operability facilitated by data-centric business models, enabling firms to navigate the challenges posed by the crisis.

Building upon the groundbreaking research by Corrado et al. (2005, 2009), which laid the ground for considering intangible factors of production as capital assets, including software and R&D activities, we extend this notion to encompass data as a capital asset (see also Corrado et al., 2022). Similar to traditional long-term factors of production, such as machinery and buildings, data holds value when produced and utilized repeatedly to generate economic benefits (Corrado et al., 2022).[2] When the data meets the asset criteria of repeated contribution to production over more than one year, there is a clear case for considering data and the knowledge acquired from data as assets (Goodridge et al., 2022).

Exploiting data assets can enhance firm productivity by boosting revenues or reducing costs. Data assets may have been crucial in supporting firm productivity during the COVID-19 pandemic. For example, supply chain data enabled firms to identify and mitigate disruptions, optimize supply chains, and develop new products and services. Additionally, by analyzing databases derived from user online behavior, firms gained insights into customer preferences and pandemic-induced shifts, enabling them to identify new markets, enhance customer engagement through personalized marketing, and provide customized customer experiences,

[1] See, e.g., the survey data of the Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises, accessed July 5, 2023.

[2] However, data differ from tangible capital assets; investments in data are rarely final, and with data updates, its role as a factor of production does not remain the same over time, unlike tangible capital assets. Therefore, the value of a specific type of data can also increase over time. For example, adding new information about consumer behavior to a company's existing data can increase the value of previous observations. Hence, the value of data assets may not rise solely because the value of new data investments exceeds the depreciation of existing data, unlike in the case of traditional capital assets.

ultimately driving revenue growth. These sales-enhancing mechanisms could have been particularly valuable during economic shocks like COVID-19, which restricted personal, face-to-face interactions.

Furthermore, digitalization empowers firms to expand their market reach through online platforms, e-commerce, and digital marketing strategies, increasing sales and revenue (Brynjolfsson & McAfee, 2014). Goldfarb and Tucker (2019) highlight several ways digital technologies may reduce business costs. The internet may *lower search costs* by making information more accessible and improving search accuracy through data-driven algorithms. Digital technologies enable processes with *low replication or* marginal costs after the initial development and implementation, allowing for scalability. *Transporting digitized data over the internet incurs almost zero costs*, facilitating cost-effective connections between geographically dispersed entities and, for instance, making supply change management, logistics and real-time inventory management more efficient. Digital tools help businesses track and analyze data from customers and suppliers, *reducing tracking costs* and enabling streamlined operations. Technology and new data sources facilitate easier verification of identities and reputations, *reducing the verification costs* associated with building trust in business transactions.

We employ Statistics Canada's framework (2019) and classify data assets into three primary categories: data, databases, and data science. Specifically, 'data' pertains to digitally formatted observations capable of storage, transmission, and processing, from which insights can be gleaned. A 'database' represents a structured data store organized for retrieval and manipulation. Instead, 'data science' encompasses the activities wherein individuals employ a spectrum of techniques, methods, algorithms, and systems to extract knowledge and insights from data.

Conventional accounting principles fall short of capturing the value of intangible capital, particularly data assets. Data investments' inherent elusiveness and absence from financial reports render traditional asset valuation methods inapplicable. We use a cost-based approach to assess the value of firms' data capital, similar to the analyses in Statistics Canada (2019) and Corrado et al. (2021). We measure the production costs associated with data-related activities by relying on the wages of occupational groups related to data, databases, and data science. We further consider investments in extracting knowledge from data within the broader context of software and databases[3], the asset class also identified in national accounts, following the approach of Goodridge et al. (2022). To comprehensively explore the impact of digitalization, we construct a variable measuring the wage share of ICT (information and communication technology) workers as a

[3] The national accounts categorize a single asset class, "software and databases", without segregating the two components due to the challenges associated with precisely differentiating between employees involved in software development and database capital formation.

proportion of total wages. This variable serves as an additional explanatory factor in our analysis of firm-level productivity growth.

As discussed above, our empirical analysis builds upon previous research that has sought to define the occupational groups essential for data asset creation and to assess the proportion of time spent on in-house data production by individuals in these occupations (e.g., Statistics Canada, 2019). These prior studies, such as Goodridge et al. (2022), have multiplied time-use-adjusted employment by estimates of average wages for each occupational group to derive aggregate-level estimates of total investments in different data assets. Our study, in contrast, utilizes matched employer-employee data, enabling us to calculate time-adjusted wage sums for each employee in the specified data occupations and aggregate them to the firm level. To our knowledge, our study is the first to directly measure firm-level investments in data assets and their impact on productivity growth.

We utilize matched employer-employee data of 13,609 Finnish companies for the years 2015–2020 to investigate whether and how firms' investment in data assets and digitalization affected their productivity performance amidst the unprecedented challenges posed by the COVID-19 pandemic. Our empirical analysis reveals that firms that had made relatively higher investments in software and database assets, as well as in digitalization more generally, during the pre-pandemic period experienced significantly higher labor productivity growth during the first year of the pandemic. Notably, these positive effects are predominantly observed in the service sector, while manufacturing companies did not exhibit statistically significant impacts. Furthermore, our analysis highlights that large service companies with greater investments in data assets demonstrated higher labor productivity growth than their counterparts, in 2020. Our empirical analysis further indicates that investments in data assets, databases and ICT were positively and statistically significantly associated with a firm's propensity to be among the top 5% of firms in terms of productivity in each industry.

The rest of the paper is organized as follows. Section 2 discusses the previous literature relating to our study, particularly focusing on the empirical efforts to measure the impacts of digitalization on firm productivity. Section 3 introduces the econometric approach we use in our empirical exploration. Section 4 presents the estimation results. Section 5 concludes.

2. Literature review

Despite a vast body of research on the link between ICT, digitalization, and productivity, firm-level evidence on the role of data assets in firm productivity remains limited. Given the extensive literature on this topic, we focus our review on recent empirical studies that closely align with our research objectives. For more comprehensive literature surveys covering earlier research, please refer to Biagi (2013).

The dearth of empirical evidence on the contribution of data assets to productivity stems primarily from the absence of information on data assets on firm balance sheets. Goodridge et al. (2022) address this challenge at the aggregate level by estimating the contribution of data capital deepening to productivity growth for a set of European countries. Utilizing Labor Force Survey data for the 13 EU countries, they estimate national investments in data assets and find that the annual contribution of software and data capital deepening to productivity growth more than tripled from 0.03% to 0.10% between 2011 and 2016. Relatedly, a stream of literature aims to estimate the data economy's value or size. For instance, Statistics Canada (2019) and Bondt & Mushkudiani (2019) estimate the data value in the business sector by calculating the labor inputs used in data-based work based on employment and wage data extracted from the registry data. Ker and Mazzini (2020) use data on investments in data processing, such as IT hardware, software, and services, to estimate the size of the data economy in the US business sector. Their calculations show that the usage of data processing products increased from around 26 billion US dollars in 2012 to over 36 billion US dollars in 2017, accounting for 0.25 percent of all intermediate input usage.

Economic literature further indicates that the value of data and its evolution over time is highly context dependent. Farbood & Veldkamp (2021) model the depreciation of data used for forecasting and the data economy. Their theoretical model indicates that when data used for forecasting is scarce, its growth increases firms' productivity and transactions, generating more data and further enhancing productivity and creating new data. This “data feedback loop” leads to increasing returns to forecast data when it is scarce. However, in the long run, large datasets used for forecasting (e.g., company forecasts regarding demand or costs) exhibit diminishing returns similar to tangible capital assets. The reason, however, differs from tangible capital assets: when dealing with large datasets, the forecast error approaches zero, and therefore, adding more data does not significantly improve the forecast. Empirical research findings support this theory (for example, Bajari et al., 2018, regarding Amazon's demand forecasting).

However, increasing the amount of data may yield economic benefits in certain applications. When forecast data is used as input in research and development (R&D), more data allows for better-targeted or more useful innovations, resulting in greater value-added data. Farbood & Veldkamp (2021) emphasize that data used in R&D should be measured separately from other data, similar to how macroeconomists distinguish between traditional capital investments and R&D investments.

Data is pivotal in robotics, serving as the foundation for training, monitoring, and improving robot performance. The growing prominence of automation and industrial robotics warrants a concise examination of recent studies investigating the effects of robotics on productivity and employment. While a comprehensive review of the literature is beyond the scope of this work, we aim to provide the reader with a selection of notable examples. A seminal study on the aggregate impact of industrial robots is that of Graetz and Michaels (2018). Utilizing data from the International Federation of Robotics (IFR), the authors assess the effects of robot adoption across various countries and industries from 1993-2007. A key finding of their analysis is the substantial decline in the price of industrial robots during the study period, which has been associated with a surge in robot adoption. Furthermore, the authors estimate a significant positive impact of robotics on labor productivity, with a point estimate of 0.36 percentage points to annual growth, albeit with diminishing returns. The study also examines the employment implications of robotics, finding a nonsignificant average effect on total hours worked while witnessing a shift in employment from low-skill to middle-skill and high-skill occupations. Additionally, the analysis highlights a positive effect on wages and total factor productivity and a negative impact on output prices.

The impact of industrial robots on productivity and employment has also been investigated by Acemoglu et al. (2020), looking at French firms, and Acemoglu and Restrepo (2020) for the US. In the first study, the authors examine the adoption of robots among French firms between 2010 and 2015 and distinguish the effects of robotics between market-level and firm-level effects. At the market level, the analysis results highlight an overall increase in productivity matched by a decline in employment and labor shares. However, firms that adopt robots show an increase in employment, probably due to the reallocation of resources to cost-cutting firms. Focusing on the US economy from 1990 to 2007, Acemoglu and Restrepo (2020) analyze the impact of robotization on wages and employment at the commuting zone level. The paper's main finding is that adopting industrial robots is associated with a significant decline in employment (precisely, a drop of 0.37 points in the employment-population ratio) and wages (a drop of 0.73 percent). However, these negative effects are attenuated, while remaining significant when considering trade across commuting zones.

Bearing close relevance to our work, Abidi et al. (2022) focus on the relationship between digitalization and business resilience during the COVID-19 pandemic. Their study, encompassing a sample of firms from Middle Eastern and Central Asian economies, reveals that digital technologies have served as a source of resilience against the pandemic shock. Firms with a higher degree of digitalization experienced a less pronounced decline in sales during the crisis. Oikonomou et al. (2023) further explore the mitigating effects of digital technologies on the labor market repercussions of the COVID-19 pandemic. Utilizing a comprehensive dataset of US establishments, they examine how ICT technologies have alleviated the pandemic's impact on labor outcomes, specifically the unemployment rate. Their findings indicate a positive association between digital technology adoption and labor market resilience, primarily attributed to the increased prevalence of remote work arrangements and the creation of digitally intensive jobs. The authors employ the historical share of routine workers as an instrumental variable to establish causal inference.

Our methodological approach to examining the impact of data on productivity closely aligns with that of Gal et al. (2019). Their study tries to gauge a causal relationship by looking at the relationship between industry-level digitalization levels and firm-level productivity growth, which should attenuate endogeneity concerns (because the degree of digitalization is less affected by firm-level unobservable characteristics). Notably, Gal et al. (2019) investigates the differential effects of various digital technologies on productivity growth and explore the interplay between digitalization and other productivity-enhancing factors, such as skill availability. Their findings reveal a strong, positive connection between industry-level adoption of digital technologies and firm-level productivity growth. Additionally, they demonstrate that this relationship is amplified by the availability of skills within the industry, emphasizing the complementary role of education in facilitating digital technology adoption.

Cette et al. (2022a, 2022b) delve into the multifaceted effects of ICT on productivity and labor share. Their first study employs a growth accounting methodology to assess the contribution of ICT capital accumulation, encompassing hardware and software, to labor productivity growth. Their findings reveal a significant positive impact, indicating that ICT capital accumulation enhances labor productivity. Additionally, they investigate the impact of robot adoption on aggregate growth across various countries, including Germany, Japan, and Italy, during specific periods. Their analysis demonstrates that robot adoption has been a driving force behind aggregate growth in these countries. In their second study, Cette et al. (2022b) focus on the French economy, examining the dual impact of employing ICT specialists and adopting digital technologies on labor productivity and on the labor share. Their findings reveal a positive and statistically significant effect of ICT specialists and digital technologies on labor productivity, indicating that these factors contribute to enhanced workforce efficiency. However, they also uncover a negative impact on labor share, suggesting that the benefits of ICT may not be fully captured by labor compensation.

Gordon and Sayed (2020) examine the differential impact of ICT adoption on productivity growth in the United States and a group of ten EU countries. They critically assess the methodological limitations of the growth accounting approach and opt for industry-level data (KLEMS) to provide a more nuanced analysis. Their findings reveal that ICT-intensive service industries predominantly drove the productivity resurgence of the US economy between 1995 and 2005. In contrast, the EU did not experience comparable productivity growth due to ICT. Notably, both economies have witnessed a decline in productivity gains from ICT since 2005.

While not directly studying the impact of digital technologies on productivity, Niebel et al. (2019) offer a related analysis focusing on the effect of big data usage on innovation propensity. Using a sample of German firms, they find a positive, significant association between the use of big data and the propensity to innovate and the intensity of innovation. The tight link between innovation and productivity growth makes this work relevant to us. DeStefano et al. (2023) focus on cloud computing technology and study its effect on multiple-level outcomes while using zip-code expansion of broadband internet as an instrument. Interestingly, they find that the impact of cloud computing depends on whether the firm is a young/entrant or an incumbent. For the first group, the authors find that adopting cloud computing leads to an increase in sales and employment (and an insignificant impact on labor productivity). In contrast, for the incumbents, they find a weaker effect on growth, together with a tendency to become more geographically dispersed.

Further solidifying the positive impact of digitalization on productivity, Czarnitzki et al. (2023) examine the specific effects of adopting artificial intelligence (AI) technology. Their study, encompassing a sample of German firms, reveals that integrating AI into a firm's operations is associated with a notable increase in productivity. This finding underscores the transformative potential of AI in enhancing operational efficiency and driving productivity gains.

In concluding our literature review, we turn to a study by Anderton et al. (2023) that offers a more nuanced perspective on the relationship between digitalization and productivity. Utilizing an extensive panel dataset of over 2 million firms, their findings reveal an average positive impact of digitalization on TFP growth. However, the observed impact is modest and, notably, essentially absent for a significant portion of firms. The authors highlight that digitalization benefits among laggard firms are limited to the top 30% most productive businesses within that group. Additionally, they observe a reinforcing interaction between digitalization and a high degree of intangible asset intensity. These findings underscore that digitalization should not be viewed as a universal solution for stimulating productivity and emphasize the need for well-conceived strategies that consider firm-level heterogeneity.

3. Empirical approach

3.1 Analytical framework

Our empirical specification closely parallels the analytical framework developed by Gal et al. (2019). Denoting the log change of labor productivity for firm *f*, belonging to industry *i* in year *t* (in practice, the change between 2019 and 2020), as $= \Delta LP_{f,t}$, our regression model can be written as:

$$\Delta LP_{f,t} = \alpha_1 \Delta LP_{frontier,i,t} + \alpha_2 Gap_{f,i,t-1} + \beta Data_f + \gamma X_{f,t} + \delta_i + \epsilon_t. \quad (1)$$

In the same fashion as for $\Delta LP_{f,t}$, $\Delta LP_{frontier,i,t}$ represents the log-change of labor productivity, but in this case, it is the average productivity growth for frontier firms of industry *i* at time *t,* where we define frontier as the top-5% firms in terms of productivity. On the other hand, $Gap_{f,i,t-1}$ denotes the gap between the log-productivity of firm *f* and that of frontier firms in the same industry. Economic theory predicts α_1 to be positive but less than 1 (implying positive spillovers from the frontier to the rest of the economy) and α_2 to be positive (catch-up effects). Our main coefficient of interest is β, which captures the relation between the share of a firm's data assets of total assets (which we denote as $Data_f$) and productivity growth at the firm level. $X_{f,t}$ is a vector of firm-level control, specifically firm size and age. Finally. δ_i are industry fixed effects.

Though our empirical methodology primarily aligns with the analytical framework established by Gal et al. (2019), notable deviations emerge in the metrics employed and the focus of the analysis. Firstly, while Gal et al. (2019) relied on industry-level digitalization indices derived from survey data, our study delves into the granular role of firm-level investments in data assets in shaping firm productivity. Given the inherent characteristic of data assets being primarily created and utilized in-house, one approach for measuring capital formation is to evaluate the total costs incurred in generating three distinct data asset types. Our methodology entails identifying workers involved in the production of data assets and is based on the wage share of workers associated with data-producing occupations in the registry data. The advantage of this sum-of-costs approach lies in its ability to capture the share of data assets relative to all assets rather than relying on subjective assessments of digital technology usage derived from surveys. Additionally, we extend the analysis by estimating alternative specifications for the impact of digitalization captured by the wage shares of ICT occupations. For simplicity, our model development below focuses on data assets, while we estimate separate models for digitalization or ICT more generally. Moreover, using registry data instead of survey data enables us to extend our analysis of the impact on firm-level productivity to a broader population of companies.

Gal et al. (2019) specification employs the productivity frontier data, defined as the top 5% in each industry in a firm's country. It excludes firms at the productivity frontier from the sample to mitigate potential endogeneity concerns. However, this exclusion might lead to underestimating the impact exerted by digital technologies or data assets, as these frontier firms may be the first to utilize and invest in data-driven technologies and data assets effectively. We, instead, use the industry-specific labor productivity data concerning *the international productivity frontier* (excluding Finland) to calculate the variables $\Delta LP_{frontier,i,t}$ and $Gap_{f,i,t-1}$ (see next section for more detailed discussion). This removes the endogeneity bias and enables us to keep the total sample of companies, including the Finnish productivity frontier, in the analysis.

While our basic specification controls for firm-level characteristics and industry dummies, it is still susceptible to endogeneity concerns regarding the estimation of the effects of data assets on labor productivity growth. For instance, firms with brighter productivity prospects may be more inclined to invest in data assets, leading to an upward bias in the estimated relationship between data investments and productivity growth.

Additionally, better-managed firms, which are more likely to experience positive productivity dynamics, may also be more likely to invest in data, further confounding the causal interpretation of the results.

To further mitigate endogeneity concerns and attempt to establish causal links, we rely on an instrumental variable (IV) approach. Specifically, we draw inspiration from the works of Gal et al. (2019) and Borowiecki et al. (2021) and instrument the firm-level indicators for data assets using the industry-level (2-digit industry levels, using the NACE Rev 2 classification) data assets where we exclude the firm of interest from the industry average. As a secondary instrument, we introduce a binary variable, $foreign_f$, which takes the value of one if a firm is under foreign ownership and 0 otherwise. The main assumption about the validity of the first instrument is that industry-level investments in data assets or digitalization intensity influence individual firm-level productivity performance through technological spillover effects that impact a firm's s investments in data assets, but not directly via other channels. When firms in an industry invest more in data assets, they accumulate a knowledge base for effectively using these technologies. This knowledge can spread to other firms in the industry, for example, through employee mobility, reducing overall investment costs for individual firms. Additionally, the broader adoption of data use can create competitive pressures that drive firms to keep up with the latest technological advances. This assumption also aligns with the approach employed by Borowiecki et al. (2021).

Furthermore, we hypothesize that foreign ownership affects firms' investments in new technologies and skilled labor force but not directly on labor productivity growth.[4] Building on extant research, we posit that foreign ownership positively influences firm productivity, driven by heightened investments in data assets and a more skilled workforce.[5] Empirical findings consistently indicate distinctions between foreign-owned and domestic firms, with the former tending to employ a workforce with higher education and skill levels (Koch & Smolka, 2019). The research conducted by Koch and Smolka (2019) underscores that the superior productivity levels observed in firms with foreign ownership firms stem from their ability to integrate

[4] We also estimated the baseline specification (1) incorporating foreign ownership as an additional control variable directly influencing firm-level productivity growth. The estimated coefficient for foreign ownership was statistically insignificant, suggesting that foreign ownership does not have a direct impact on the dependent variable, implying its potential suitability as an instrumental variable.

[5] We also estimated the models using foreign ownership as the explanatory variable for productivity growth, i.e., assuming that it might directly impact productivity growth. Our data did not detect a direct relationship: the estimated coefficients of the foreign ownership dummy variable were statistically insignificant in all estimated models.

advanced technologies with appropriately trained personnel, resulting in a heightened skill intensity within their production processes.

The IV specification is implemented using a typical two-stage approach, where the first-stage regression is:

$$Data_f = \alpha_1 \Delta LP_{frontier,i,t} + \alpha_2 Gap_{f,i,t-1} + \theta Data_{-f,i} + \zeta foreign_f + \gamma X_{f,i,t} + \delta_i + \epsilon_t, \quad (2)$$

where $Data_{-f,i}$ represent the share of data assets of all assets at the industry level, while excluding firm f. We expect the coefficient associated with $Data_{-f,i}$ to be positive, because of possible spillover effects. The second-stage regression is then:

$$\Delta LP_{f,t} = \alpha_1 \Delta LP_{frontier,i,t} + \alpha_2 Gap_{f,i,t-1} + \beta \widehat{Data}_f + \gamma X_{f,i,t} + \delta_i + \epsilon_t, \quad (3)$$

where $\widehat{Data}_f$ are the fitted values obtained from regression model (2).

3.2 Data

Our empirical analysis draws primarily upon firm-level registry data from Statistics Finland. The central variable of interest is the change in (log) labor productivity between 2019 and 2020, during the initial year of the COVID-19 pandemic. Labor productivity is measured as the ratio of value added to employment, with employment expressed in full-time equivalents. These variables are sourced from the Financial Statement Panel of Statistics Finland. Observations with negative value added are excluded from the analysis. Nominal value added is deflated using industry-specific deflators at the 2-digit level whenever available; otherwise, 1-digit level deflators are employed. In the final sample, we have observations from 14,320 Finnish companies.

We construct the international productivity frontier variables using the CompNet data.[6] We extract the log labor productivity (real value added based) of the top 5 % of companies at the 2-digit level industry for 2019 and 2020 from eight Western Europe countries available in the dataset (i.e., Sweden, Belgium, Denmark, France, Italy, Spain, Croatia and Portugal). We then calculate the average productivity frontier of each industry among the countries and subtract a firm's log labor productivity in 2019 from the corresponding international productivity frontier value to form the variable $Gap_{f,i,t-1}$. We generate the variable

[6] More detailed information on the CompNet data is available on "User Guide for the 9th Vintage of the CompNet Dataset", https://www.comp-net.org/data/9th-vintage/. The user must be aware that small differences in data collection rules and procedures across countries may exist and are out of CompNet's control. Nevertheless, comparability issues appear to be limited.

$\Delta LP_{frontier,i,t}$ by calculating the 2020-2019 difference in the log labor productivity at the international industry productivity frontiers.

Statistics Canada's (2019) framework centers on selecting specific occupational groups essential for building data assets. One challenge with this approach is that within the three key areas where firms invest in data assets — data, databases, and data science — workers may only spend part of their work time producing data assets. To account for this, time use factors are employed to adjust the estimated compensation of workers to reflect the amount of time they spend working on in-house data production. Table 1 illustrates the correspondence between the occupations identified by Statistics Canada and their alignment with the occupational classification employed production aligns with those used by Statistics Canada, using the averages of their estimated time use factors. Additionally, we distinguish the occupational group related to software and databases to explore how the software and database formation contribute to productivity growth. Here, we follow the recommended time use factor of 50 % that is applied in most countries and used by Goodridge et al. (2022). To further explore the impact of a company's investments in digitalization, we define a group of ICT occupations. This framework draws upon the work of Calvino et al. (2018) and has been adapted to the Finnish context, as documented in Ali-Yrkkö et al. (2020).

The comprehensive registry data at our disposal enables us to link firm-level information with individual worker-level data. We construct a variable that captures a firm's investments in data assets by employing worker-level salary data containing occupation information. Our initial step involves determining whether an individual's occupation can be classified as data related. Following the classification of worker occupations, we assign each worker to their respective firm. The individual employee-level salary data allows us to calculate the proportion of salaries associated with data generation for each firm and year using the assumed time use factors. When calculating the proportion of total costs incurred in data formation relative to the total costs of all capital formation, non-wage costs (e.g., non-wage labor costs and overheads) can be disregarded under the assumption that the multiplier for non-wage costs remains constant for both data assets and other capital, as the proportion remains unaffected by the multiplication. Subsequently, we average these wage proportions over the period spanning from 2015 to 2019. These proportions are utilized in our empirical estimations as indicators of a firm's investments in data assets.

– TABLE 1 HERE –

We extracted additional control variables from the Business Register, another register-based data source maintained by Statistics Finland. These variables include firm age, number of employees, and industry affiliation. Based on the number of employees, we categorize firms into size classes and exclude those with fewer than ten employees, following the approach adopted by Gal et al. (2019). In our estimations, we control for firm age and the logarithm of the number of employees. Also, we include industry dummies to account for various market-specific factors that could influence fluctuations in firm-level productivity between 2019 and 2020, such as industry-specific variations in the impact of the COVID-19 pandemic.

– TABLE 2a & 2b HERE –

Table 2a provides the summary statistics for the variables employed in our analysis. Among various data-related assets, firms have prominently directed their investments towards software and database assets, constituting an average allocation of 1.8% of resources. In stark contrast, investments in data science or data analytics between 2015 and 2019 accounted for a mere 0.04% of total assets on average.

Table 2b delves into the nuanced distinctions in firms' average investments in diverse data assets between large and small and medium-sized enterprises. Our t-tests reveal statistically significant investment pattern differences ($p<0.05$). Large firms exhibit a more pronounced focus on generating databases and software and database assets than SMEs. Additionally, large firms allocate a greater proportion of their total wages to ICT workers, indicating more substantial investments in digitalization. These observations suggest that large firms tend to prioritize internal development and maintenance of software and databases, whereas SMEs are inclined to outsource these activities. The divergence in investment patterns between large firms and SMEs may be attributed to various factors, including disparities in IT infrastructure maturity or variations in the availability of resources and specialized expertise, particularly for developing and maintaining intricate software and databases.

Table 2c extends our examination to compare the average investments in different data assets and digitalization between service and manufacturing companies. Additionally, we present the results of t-tests comparing the magnitude of investments in these distinct company categories. On average, service companies have made statistically significantly larger investments in all asset types than manufacturing companies. Service companies exhibit statistically significant, larger investments on average across all asset types compared to manufacturing companies. These findings underscore the inherently data-intensive nature of operations and business models within the service sector, coupled with a heightened complementarity of ICT with other assets, distinguishing them from their manufacturing counterparts.

4. Estimation results

We employed the two-stage least squares (2SLS) regression approach to analyze the determinants of firm-level productivity growth. In the first stage of the 2SLS estimations, we used industry-level average investments in data assets (excluding each firm's own investments) and foreign ownership as instrumental variables. Our findings from the first stage suggest that industry-level investments in data assets have a positive and statistically significant impact on a firm's investments in these assets. Furthermore, foreign ownership exhibits a positive and statistically significant coefficient in most first-stage estimations, though not all. Both the F-test and Kleibergen–Paap tests decisively reject their null hypotheses at the 99% confidence level, attesting to the strength of our instrumental variables. The Hansen-J test for overidentifying restrictions is never rejected, providing additional confidence in the validity of our instruments. In sum, the results from the first-stage regression instill confidence in the strength and validity of the instrumental variables, underpinning the reliability of our subsequent 2SLS analysis of firm-level productivity growth determinants.

Our estimation results (Tables 3-7) reveal that the estimated coefficients $\widehat{\alpha_1}$ and $\widehat{\alpha_2}$ representing the effects of the average international productivity growth of frontier firms within the industry and the gap between a firm's and the international frontier's productivity growth, respectively, corroborate the predicted magnitude and statistical significance proposed by economic theory. The spillover effects stemming from the international frontier's productivity growth are notable: approximately 60 percent of their productivity growth is transmitted to the average Finnish company. This finding resonates with recent research of Lind and Ramondo (2022), which highlights that around 90 percent of production in Finland draws upon knowledge disseminated from global sources. The catching up effect is also evident. One percent increase in the gap between a firm's labor productivity growth and that of the international frontier is associated with an approximately 0.08% increase in the firm's labor productivity growth.

The estimation results presented in Table 3 demonstrate that firms that made relatively greater investments in software and database assets and ICT during the period 2015-2019 experienced significantly higher productivity growth during the initial year of the pandemic than their counterparts. Specifically, a one-percentage-point increase in a firm's allocation of investments to software and database assets (ICT), relative to its investments in all assets, is associated with a productivity growth increase of approximately 0.5% (0.1%). The empirical findings indicate that companies with comparatively higher pre-pandemic allocations to investments in data assets, databases, and data science did not demonstrate statistically significant differences in productivity growth compared to their industry counterparts. This observation aligns with the hypothesis that these investments may have been forward-looking, anticipating future productivity enhancements that were not fully materialized during the year 2020.

When the coefficients of the data asset variables were estimated separately for large firms and small and medium-sized enterprises (SMEs), we found that relatively more resources invested in databases and software and ICT have notably contributed to the productivity growth of firms in both size classes (Table 4). The empirical results further indicate that the productivity growth of both SMEs and larger enterprises that directed relatively greater resources towards other categories of data-related assets did not exhibit statistically significant deviations ($p < 0.05$) from their counterparts.

In Table 5, we report the regressions' results similar to specifications (1) – (3), this time adding an interaction term between the digitalization investments and the share of employees with high education to study the possible complementarity between digitalization and skills. The interaction term between the two variables appears positive and statistically significant for the databases and ICT but not for the other data assets. These results provide evidence of a certain level of complementarity between investments in digitalization and skilled labor, highlighting their joint impact on productivity growth. This corroborates findings emphasized in previous studies, such as Gal et al. (2019).

Tables 6 and 7 present results from our model estimations, elucidating the influence of data assets on productivity growth for manufacturing and service firms. Notably, Table 7 provides a more granular examination, explicitly focusing on differentiating effects between small and medium-sized enterprises (SMEs) and large companies. Our findings highlight that the positive and statistically significant contribution of ICT and databases and software to productivity growth is observed exclusively in the service sector, encompassing both SMEs and large companies. In contrast, none of the estimated coefficients for data or digital asset variables demonstrate positive and statistically significant effects among manufacturing companies. Additionally, our data reveals that a one-percentage-point increase in investments in data assets resulted in, on average, 5.7% higher labor productivity growth among large service companies. Moreover, we identify a marginally significant positive impact of investments in data science on the productivity growth of large companies in the service sector.

We complement our analysis by exploring whether the data assets and digitalization relate to the firm’s propensity to be in the productivity frontier, defined as the top 5% of firms in terms of productivity in each industry. We estimate a simple probit model with the dummy variable getting a value of one if a firm belongs to the domestic productivity frontier of its industry during the first year of the pandemic as the dependent variable and use the data asset variables, firm age, log number of employees and industry dummies as explanatory variables. Our findings indicate that firms that invested more in data, databases and ICT were statistically significantly more likely to be among the firms in the productivity frontier (Table 8). In contrast, investments in software and databases and data science were not statistically significantly related to a firm's propensity to be in the productivity frontier.

5. Conclusions

The global disruption caused by the COVID-19 pandemic posed unprecedented challenges to businesses, impacting supply chains, altering consumer behavior, and heightening economic uncertainty. Our empirical analysis using data from 13,609 Finnish companies for the years 2015-2020 reveals that firms that had made relatively higher investments in software and database assets, and in digitalization more generally, during the pre-pandemic period experienced significantly higher labor productivity growth during the first year of the pandemic. These positive effects were predominantly observed in the service sector, while manufacturing companies did not exhibit statistically significant impacts. Furthermore, our analysis highlights that large service companies with greater investments in data assets demonstrated higher labor productivity growth than their counterparts in 2020.

Our investigation reveals that leading up to the onset of the pandemic, firms within the service sector surpassed their manufacturing counterparts in dedicating resources to investments in data assets and digitalization. Furthermore, within the service sector, entities that made more substantial investments in these assets, particularly larger service companies, witnessed a significant increase in labor productivity growth in the initial year of the pandemic. The efficacy of these investments can be attributed to the versatile and adaptable characteristics of data assets, facilitating their application across a spectrum of business processes. This adaptability proved especially advantageous during the pandemic, where restrictions on face-to-face interactions underscored the amplified value derived from such investments.

Nevertheless, in 2020, the discernible productivity gains stemming from investments in data science were not evident within Finnish firms. This lack of observable impact may be attributed to the adjustment costs associated with efficiently utilizing data assets. These adjustment costs, which companies may not have fully mitigated, impede realizing the full potential of their investments in data science. Such costs encompass adopting complementary technologies like artificial intelligence and machine learning, providing comprehensive training for personnel, and implementing organizational and work practice modifications. An alternative explanation for this phenomenon is the imperfect nature of the proxy measures employed to capture firm-level investments in data science. Many software and application developers, categorized within the domain of software and database creation, are also actively involved in data analysis.

The methodology applied to gauge data investments offers valuable insights; however, it is imperative to acknowledge its inherent limitations. The reliance on occupation-based classification, assumed time use factors, and exclusive focus on wage costs serves as a proxy for a firm's investments in data assets. While presently representing perhaps the most robust register-based measure for firm-level investments in data assets, this approach may overlook certain critical aspects of data investments. A prospective avenue for enhancing accuracy and reliability could involve adopting a more comprehensive approach incorporating direct data expenditure measures. Such an approach could be pursued in the future, contingent upon the collection of firm-level registry data would be reformed to encompass the necessary information for refining the measurement of data investments.

Our findings, however, offer evidence of the productivity potential arising from companies' investment in data assets and adopting business models reliant on data utilization. We also identify a noteworthy complementarity between a firm's investments in ICT and databases and employees' skills, as measured by education level. Despite the nascent stage of data utilization in 2020, our results emphasize that firms investing more in data, databases, and ICT were significantly more likely to belong to the productivity frontier of their industry. The swift diffusion of generative AI across sectors underscores the increasing trend of data usage among companies. Encouraging and supporting investments in data, along with complementary assets like education and enhanced digital skills for employees, is crucial. Tailoring strategies to sector- and firm-specific needs is essential for maximizing aggregate productivity benefits from data and digitalization, fostering resilience in the face of unforeseen challenges.

References

Abidi, N., El Herradi, M., & Sakha, S. (2022). *Digitalization and resilience: firm-level evidence during the COVID-19 pandemic*. International Monetary Fund.

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020, May). Competing with robots: Firm-level evidence from France. In *AEA papers and proceedings* (Vol. 110, pp. 383-388). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of political economy*, *128*(6), 2188-2244.

Ali-Yrkkö, J., Koski, H., Kässi, O., Pajarinen, M., Valkonen, T. Hokkanen, M., Hyvönen, N., Koivusalo, E., Laaksonen, J., Laitinen, J. & Nyström, Enni (2020). The Size of the Digital Economy in Finland and Its Impact on Taxation. ETLA Report No 106.

Anderton, R., Reimers, P., & Botelho, V. (2023). Digitalisation and productivity: gamechanger or sideshow?

BEA (2019). Treatment of data in National Accounts. Paper Prepared for the United States Bureau of Economic Analysis.

Bajari, P., Chernozhukov, V., Hortaçsu, A. & Suzuki, J. (2019). The impact of big data on firm performance: An empirical investigation. *AEA Papers and Proceedings* 109, 33-37.

Biagi, F. (2013). ICT and Productivity: A Review of the Literature.

Bondt & Mushkudiani (2019). Estimating the Value of Data in the Netherlands.

Borowiecki, M., Pareliussen, J., Glocker, D., Kim, E. J., Polder, M., & Rud, I. (2021). The impact of digitalisation on productivity: Firm-level evidence from the Netherlands.

Calvino, F., Criscuolo, C., Marcolin, L. and Squicciarini, M. (2018). A taxonomy of digital intensive sectors. *OECD Working Paper* 2018/14, OECD, Paris, France.

Cette, G., Devillard, A., & Spiezia, V. (2022a). Growth factors in developed countries: a 1960–2019 growth accounting decomposition. *Comparative Economic Studies*, 1-27.

Cette, G., Nevoux, S., & Py, L. (2022b). The impact of ICTs and digitalization on productivity and labor share: evidence from French firms. *Economics of innovation and new technology*, *31*(8), 669-692.

Coyle, D. & Li, W. (2021). The Data Economy: Market Size & Global Trade. SSRN 3973028.

Corrado, C., Hulten, C. & Sichel, D. (2005). Measuring Capital and Technology: An Expanded Framework. In C. Corrado, J. Haltiwanger, and D. Sichel (eds), *Measuring Capital in the New Economy, Studies in Income and Wealth No. 65*, University of Chicago Press, Chicago, IL, 11–46.

Corrado, C., Hulten, C. & Sichel, D. (2009). Intangible Capital and U.S. Economic Growth," *Review of Income and Wealth*, 55, 661–85.

Corrado, C., Haskel, J., Iommi, M. & Jona-Lasinio, C. (2022). Measuring data as an asset: Framework, methods and preliminary estimates. *OECD Economics Department Working Papers*, No. 1731, OECD Publishing, Paris, https://doi.org/10.1787/b840fb01-en.

Czarnitzki, D., Fernández, G. P., & Rammer, C. (2023). Artificial intelligence and firm-level productivity. *Journal of Economic Behavior & Organization*, *211*, 188-205.

DeStefano, T., Kneller, R., & Timmis, J. (2023). Cloud computing and firm growth. *Review of Economics and Statistics*, 1-47.

Farboodi, M. & Veldkamp, L. (2021). A model of the data economy. *National Bureau of Economic Research Working Paper Series* No. 28427.

Gal, P., Nicoletti, G., von Rüden, C., Sorbe, S., & Renault, T. (2019). Digitalization and productivity: in search of the holy grail-firm-level empirical evidence from European countries. *International Productivity Monitor*, *37*(Fall), 39-71.

Goldfarb, A. & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57, 3–43.

Goodridge, P., Haskel, J. & Edquist, H. (2022). We see data everywhere except in the productivity statistics. *Review of Income and Wealth* 68, 4.

Gordon, R. J., & Sayed, H. (2020). *Transatlantic technologies: The role of ICT in the evolution of US And European productivity growth* (No. w27425). National Bureau of Economic Research.

Graetz, G., & Michaels, G. (2018). Robots at work. *Review of Economics and Statistics*, *100*(5), 753-768.

Jones, C.I. & Tonetti, C. (2018). Nonrivalry and the economics of data. *American Economic Review* 110, 2819-59.

Ker, D. & Mazzini, E. (2020). Perspectives on the value of data & data flows. *OECD Digital Economy Papers*, No. 299, OECD Publishing. Paris.

Koch, M. & Smolka, M. (2019). Foreign ownership and skill-biased technological change. *Journal of International Economics* 118, 84-104.

Lind, N. & Ramondo, N. (2022). Global Innovation and Knowledge Diffusion. *National Bureau of Economic Research Working Paper Series* No. 30590.

Niebel, T., Rasel, F., & Viete, S. (2019). BIG data–BIG gains? Understanding the link between big data analytics and innovation. *Economics of Innovation and New Technology*, *28*(3), 296-316.

OECD (2022). Going Digital Toolbox Note: Measuring the economic value of data. *Directorate for science, technology & innovation committee on Digital Economy Policy.*

Oikonomou, M., Pierri, N., & Timmer, Y. (2023). IT shields: Technology adoption and economic resilience during the COVID-19 pandemic. *Labour Economics*, *81*, 102330.

Statistics Canada (2019). The value of data in Canada: Experimental estimates.

Table 1. The classification of data-related occupations and ICT

Classification of occupations, Statistics Canada (2019)	**Classification of occupations, Statistics Finland**	**Assumed time use factor (% of data production)**
Data: occupational groups		
Financial and investment analysts	2413 Financial analysts	15
Customer and information services supervisors Other customer and information services representatives	121 Business services and administration managers	45
Data entry clerks	4132 Data entry clerks	100
Survey interviewers and statistical clerks	4227 Survey and market research interviewers	95
Mathematicians, statisticians and actuaries	2120 Mathematicians, actuaries and statisticians	25
Economists and economic policy researchers and analysts	2631 Economists	25
Databases: occupational groups		
Computer and information systems managers	133 Information and communications technology managers	45
Database analysts and data administrators	252 Database and network professionals	95
Information systems testing technicians	351 ICT operations and user support	40
Data science: occupational groups		
Financial and investment analysts	2413 Financial analysts	55
Statistical officers and related research support officers	3314 Statistical, mathematical and related associates	95
Mathematicians, statisticians and actuaries	2120 Mathematicians, actuaries and statisticians	55
Economists and economic policy researchers and analysts	2631 Economists	55
Goodridge et al. (2022) classification	**Software and databases** 251 Software and applications developers and analysts 252 Database and network professionals	50
Classification of occupations: ICT 251 Software and applications developers and analysts 252 Database and network professionals 133 Information and communications technology service managers 351 Information and communications technology operations and user support 2153 Telecommunications engineers 7422 Information and communications technology installers and servicers		

Table 2a. Descriptive statistics, calculated for all firms in the sample, for the year 2020.

Dependent variable: Log change of LP	Data	Database	Software & database	Data science	ICT
Data (%)	13,609	0.69	2.27	0	62.75
Database (%)	13,609	0.89	3.49	0	81
Database & software (%)	13,609	1.76	6.36	0	50
Data science (%)	13,609	0.04	0.51	0.00	22.89
ICT (%)	13,609	5.25	17.1	0	100
Age	13,609	21.73	16.70	2	120
Number of employees	13,609	59	200	10	6306
High education share (%)	13,609	25.60	25.89	0	100
Instrumental variables					
Foreign ownership	13,609	0.11	0.32	0	1
Industry-level: Data	13,609	0.68	0.78	0	5.70
Industry-level: Database	13,609	1.22	2.32	0	16.44
Industry-level: Database & Software	13,609	2.23	6.16	0	30.16
Industry-level: Data science	13,609	0.055	0.1	0	1.18
Industry-level: ICT	13,609	6.39	16.22	0	75.1

Table 2b. T-tests of data asset variables by size

Variable	Large: mean	SME: mean	Difference (SME-Large)	p-value
Data	0.51	0.70	0.18	p<0.1
Database	1.52	0.87	-0.65	p<0.001
Software & database	2.61	1.73	-0.87	p<0.01
Data science	0.052	0.039	-0.013	p>0.1
ICT	7.63	5.16	-2.47	p<0.001
Number of observations	487	13122		

Table 2c. T-tests of data asset variables: service vs. manufacturing

Variable	Service: mean	Manuf: mean	Difference	p-value
Data	0.844	0.507	0.019	p<0.001
Database	1.643	0.616	-1.027	p<0.001
Software & database	3.331	0.671	-2.66	p<0.001
Data science	0.066	0.045	-0.215	p<0.001
ICT	9.148	2.078	-7.07	p<0.001
Number of observations	8233	2903		

Table 3. The estimation results of the 2SLS (2-instrument) models for productivity growth.

Dependent variable: Log change of LP	Data	Database	Software & database	Data science	ICT
Asset	0.003 (0.003)	0.0023 (0.004)	0.005*** (0.001)	0.01 (0.044)	0.001*** (0.001)
d_LP	0.565*** (0.107)	0.564*** (0.107)	0.602*** (0.108)	0.561*** (0.107)	0.58*** (0.107)
Gap from frontier	0.082*** (0.006)	0.081*** (0.00589)	0.085*** (0.006)	0.081*** (0.006)	0.083*** (0.006)
Number of obs	13,609	13,609	13,609	13,609	13,609
R-squared	0.07	0.07	0.07	0.07	0.07
1st stage F	198.31***	118.49***	689***	11.44***	774.9***
Kleibergen-Paap	306.4***	149.41***	326***	22.43***	331***
Hansen J	0.172	0.155	0.18	0.275	0.14

All models include a full set of 1-digit industry dummies, the (log) of the number of employees and the age of the firm, as controls. *, **, *** indicate statistical significance at 10, 5 and 1 % level, respectively.

Table 4. The estimation results of the 2SLS (2-instrument) models for productivity growth: large vs SMEs.

Dependent variable: Log change of LP	Data	Database	Software & database	Data science	ICT
Asset: SME	0.003 (0.003)	0.002 (0.004)	0.005*** (0.001)	0.008 (0.044)	0.001** (0.0006)
Asset: Large company	0.057* (0.035)	0.005 (0.005)	0.006*** (0.002)	0.33 (0.253)	0.002*** (0.0007)
d_LP_frontier	0.56*** (0.107)	0.566*** (0.107)	0.602*** (0.107)	0.553*** (0.107)	0.582*** (0.107)
Gap from frontier	0.082*** (0.006)	0.0813*** (0.006)	0.085*** (0.006)	0.0813*** (0.006)	0.0835*** (0.006)
Number of obs	13,609	13,609	13,609	13,609	13,609
R-squared	0.07	0.07	0.07	0.07	0.07
1st stage F	32.66***	63.96***	195***	13.54***	316***
Kleibergen-Paap	343***	152.8***	331***	22.56***	333.8***
Hansen J	0.179	0.149	0.186	0.29	0.145

All models include a full set of 1-digit industry dummies, the (log) of the number of employees and the age of the firm, as controls. *, **, *** indicate statistical significance at 10, 5 and 1 % level, respectively.

Table 5. The estimation results of the 2SLS (2 instruments) models for productivity growth: interaction with high education employment share.

Dependent variable: Log change of LP	Data	Database	Software & database	Data science	ICT
Asset:	0.013** (0.006)	-0.0125** (0.005)	0.0009 (0.003)	-0.4 (0.608)	-0.002* (0.001)
Asset X High edu share	-0.0002** (0.0000)	0.0002*** (0.000)	0.00006* (0.00003)	0.005 (0.006)	0.00004*** (0.000)
d_LP_frontier	0.55*** (0.108)	0.57*** (0.107)	0.604*** (0.108)	0.58*** (0.108)	0.59*** (0.107)
Gap from frontier	0.084*** (0.006)	0.0826*** (0.006)	0.085*** (0.006)	0.082*** (0.006)	0.0844*** (0.006)
High edu share	0.0005*** (0.000)	0.0002 (0.0002)	0.000 (0.000)	0.0003** (0.0002)	0.000 (0.000)
Number of obs	13,609	13,609	13,609	13,609	13,609
R-squared	0.07	0.07	0.07	0.05	0.07
1st stage F	131***	69***	444.76***	6.17***	491***
Kleibergen-Paap	294***	131***	316***	8.63**	331.3***
Hansen J	0.03	0.14	0.12	0.13	0.21

All models include a full set of 1-digit industry dummies, the (log) of the number of employees and the age of the firm, as controls. *, **, *** indicate statistical significance at 10, 5 and 1 % level, respectively.

Table 6. The estimation results of 2SLS models (2 instruments) for productivity growth, manufacturing vs service firms.

Dependent variable: Log change of LP	Data	Database	Software & database	Data science	ICT
2SLS results					
Asset*manufacturing:	-0.0169 (0.051)	0.034 (0.024)	-0.027** (0.014)	0.65 (0.78)	-0.0001 (0.005)
Asset*service	0.005 (0.0035)	0.001 (0.004)	0.006*** (0.001)	0.0229 (0.043)	0.002 *** (0.006)
d_LP_frontier	0.564*** (0.107)	0.564*** (0.107)	0.6*** (0.108)	0.56*** (0.11)	0.63*** (0.11)
Gap from frontier	0.0815*** (0.006)	0.0812*** (0.006)	0.085*** (0.006)	0.081*** (0.006)	0.075*** (0.007)
Number of obs	11,136	11,136	11,136	11,136	11,136
R-squared	0.07	0.06	0.07	0.03	0.07
1st stage F	11.79***	5.23***	17.06***	3***	14.61***
Kleibergen-Paap	32.78***	12.87***	40.14***	8.53**	23.55***
Hansen J	0.24	0.1	0.39	0.16	0.18

All models include a full set of 1-digit industry dummies, the (log) of the number of employees and the age of the firm, as controls. *, **, *** indicate statistical significance at 10, 5 and 1 % level, respectively.

Table 7: The estimation results of 2SLS models (2 instruments) for productivity growth, manufacturing vs service firms, and large vs SMEs.

Dependent variable: Log change of LP	Data	Database	Software & database	Data science	ICT
2SLS results					
Asset*Manu*SME	-0.016 (0.05)	0.033 (0.025)	-0.028* (0.014)	0.66 (0.84)	-0.0001 (0.000)
Asset*Manu*Large	0.009 (0.098)	0.029 (0.046)	-0.013 (0.0379)	0.82 (0.88)	0.000 (0.12)
Asset*Serv*SME	0.005 (0.0035)	0.001 (0.004)	0.005*** (0.001)	0.02 (0.04)	0.002*** (0.000)
Asset*Serv*Large	0.057** (0.027)	0.0064 (0.005)	0.007*** (0.001)	0.36* (0.22)	0.003*** (0.000)
d_LP_frontier	0.564*** (0.107)	0.564*** (0.107)	0.6*** (0.108)	0.56*** (0.11)	0.63*** (0.11)
Gap from frontier	0.0815*** (0.006)	0.0812*** (0.006)	0.085*** (0.006)	0.081*** (0.006)	0.075*** (0.007)
Number of obs	11,136	11,136	11,136	11,136	11,136
R-squared	0.06	0.06	0.07	0.03	0.07
1st stage F	23.98***	185.06***	16.5***	3.52***	28.92***
Kleibergen-Paap	32.5***	18.79***	43***	7.08**	25.8***
Hansen J	0.27	0.09	0.41	0.18	0.2

All models include a full set of 1-digit industry dummies, the (log) of the number of employees and the age of the firm, as controls. *, **, *** indicate statistical significance at 10, 5 and 1 % level, respectively.

Table 8. Probit model estimates for a firm's propensity to be among the industry's top 5 % companies in terms of labor productivity levels. The models include a 1-digit industry dummies, the log of the number of employees and the age of the firm, as additional controls.

Dependent variable: Frontier	Nobs	Coef.	Std.Error	z-value	P-value	95% Confidence Interval
Data	13,609	0.033	0.007	4.73	0.000	[0.001, 0.02]
Database	13,609	0.011	0.005	2.28	0.023	[0.000, 0.019]
Software & database	13,609	0.007	0.005	1.57	0.11	[-0.002, 0.017]
Data science	13,609	0.024	0.022	1.09	0.27	[-0.019, 0.068]
ICT	13,609	0.005	0.002	2.29	0.022	[0.001, 0.008]

The models include a 1-digit industry dummies, the log of the number of employees and the age of the firm, as additional controls.

Elinkeinoelämän tutkimuslaitos

ETLA Economic Research

ISSN-L 2323-2420
ISSN 2323-2420 (print)
ISSN 2323-2439 (pdf)

Publisher: Taloustieto Oy

Tel. +358-9-609 900
www.etla.fi
firstname.lastname@etla.fi

Arkadiankatu 23 B
FIN-00100 Helsinki