// NO.24-053 | 07/2024

DISCUSSION PAPER

// RAPHAELA ANDRES, THOMAS NIEBEL, AND ROBIN SACK

Big Data and Firm-Level Productivity – A Cross-Country Comparison

Big Data and Firm-Level Productivity – A Cross-Country Comparison*

Raphaela Andres †[1,2], Thomas Niebel ‡[1] and Robin Sack§[1]

[1] *ZEW Mannheim*
[2] *Telecom Paris, Institut Polytechnique de Paris*

July 31, 2024

Abstract

Until today, the question of how digitalisation and, in particular, individual digital technologies affect productivity is still the subject of controversial debate. Using administrative firm-level data provided by the Dutch and the German statistical offices, we investigate the economic importance of data, in particular, the effect of the application of *big data analytics* (BDA) on *labour productivity* (LP) at the firm level. We find that a simple binary measure indicating the mere usage of BDA fails to capture the effect of BDA on LP. In contrast, measures of BDA intensity clearly show a positive and statistically significant relationship between BDA and LP, even after controlling for a firm's general digitalisation level.

Keywords: big data analytics, productivity, administrative firm-level data

JEL Classification: L25, O14, O33

*This work was supported by the German Federal Ministry of Education and Research (BMBF) and is part of the research project "IEDS – Incentives and Economics of Data Sharing" (IEDS002). We are grateful for helpful comments from Irene Bertschek, Daniel Erdsiek, Mary O'Mahony, Dominik Rehse and from the participants of the final workshop of the IEDS project in Dortmund. We thank the staff at CBS Microdata Services for supporting access to the Dutch administrative data as well as Kathrin Stief and Kerstin Stockmayer for supporting access to the German administrative data.

†`Raphaela.Andres@zew.de`.

‡`Thomas.Niebel@zew.de`.

§Corresponding author: ZEW – Leibniz Centre for European Economic Research, *email*: `Robin.Sack@zew.de`, P.O. Box 103443, 68034 Mannheim, Germany.

1 Introduction

Robert Solow's famous quote, "You can see the computer age everywhere but in the productivity statistics" (Solow, 1987), has sparked a controversial discussion on the productivity impact of digitalisation in general and individual digital technologies in particular. This so-called *productivity paradox* describes the slowdown in productivity growth during the 1970s and 1980s as well as during the mid-2000s despite the rapid development and diffusion of information and communication technologies (ICTs) into the economy (Borowiecki et al., 2021). Besides, digital technologies are often seen as so-called general purpose technologies (GPT) (Brynjolfsson and Hitt, 2000) that bear the potential of transforming business processes (Bresnahan and Trajtenberg, 1995; Cardona et al., 2013). However, it can take years or even decades for these effects to fully materialise, potentially explaining why their productivity impacts remain debated (Crafts, 2018; Goldin et al., 2024; Van Ark, 2016). Furthermore, diffusion of ICTs is not happening synchronously across and within industries but depends on firm characteristics, such as size, age, growth, skill-intensity, export-orientation, location, and whether the firm is foreign-owned (Cho et al., 2023; Haller and Siedschlag, 2011), as well as on management or owner characteristics, e.g. age, education, and experience (Andrews et al., 2018; McElheran et al., 2024). Education and on-the-job training as well as decentralised bargaining power within a firm further foster digital technologies' diffusion while it is hampered by too much labour flexibility (i.e. temporary contracts) (Cirillo et al., 2023). Moreover, complementarities exist in the adoption of certain so-called next-generation digital technologies, meaning that firms dealing with these new cutting-edge technologies do not adopt them separately but rather in bundles (Cho et al., 2023; McElheran et al., 2024).

As data storage, data availability, and tools for data analysis become increasingly available and scalable even for smaller firms (Anderton et al., 2020; Borowiecki et al., 2021; Cho et al., 2023; Jorgenson, 2005), big data analyses (BDA), that deliver real-time insights into huge amounts of data, spread at a fast pace and gain importance by supposedly enhancing firm performance (Brynjolfsson and McElheran, 2016).

So far, research has primarily focused on the contribution of BDA to support better decision making (Brynjolfsson and McElheran, 2019) and innovation success (Gierten et al., 2021; Niebel et al., 2019; Wu et al., 2019). However, despite its increasing application, there is hardly any evidence considering the causal impact of BDA on productivity at the firm level, which is of great interest for managers and policy makers. In contrast, there are already empirical insights into the influence of other (recent) digital technologies like broadband or cloud computing on productivity (e.g. Duso et al., 2021; Duso and Schiersch, 2022). This paper extends prior research on the relation between big data use and firm performance by focusing on productivity instead of innovation output as the central measure of firm performance. We conduct micro-econometric analyses of the relationship between BDA and labour productivity using administrative micro-level data from the Dutch and the German statistical offices. In our analysis, we initially find a positive and statistically significant relationship between BDA (measured as a binary variable) and productivity for both countries. However, after controlling for the general digitalisation level of the firm, this effect vanishes. By focusing on various intensity measures instead of the binary measure for BDA, we see a positive and significant association between BDA intensity and LP even after controlling for the general digitalisation level of the firm. This relationship is, however, less pronounced for Germany.

The rest of the paper is structured as follows: Section 2 summarises the relevant literature on the productivity impact of digital technologies as well as the research regarding the implications of BDA. Section 3 continues by introducing the data we use and the additional variables we construct. Afterwards, Section 4 lays out the empirical approaches that we pursue which is followed by the description of our results in Section 5. Section 6 discusses the results, while Section 7 summarises the most important aspects, and outlines open issues to be addressed by future research.

2 Related Literature

On the aggregate level, there is no clear evidence of digital technologies having a positive impact on productivity growth (Andrews et al., 2019). To comprehend the effectiveness, mechanisms, and conditions under which ICT investments yield results, granular firm-level data are essential (Biagi, 2013). This paper aims at filling this gap by exploiting administrative firm-level data from two different countries.

In general, digital technologies are supposed to reduce interaction costs between suppliers and customers and the costs that arise in the context of searching, finding and comparing information, which likely enhances productivity outcomes (Gal et al., 2019; Goldfarb and Tucker, 2019). The literature on low search costs induced by digital technology use and the implications for economic processes and markets is based on the theoretical work by Diamond (1971), Stigler (1961) and Varian (1980). Moreover, ICTs are supposed to serve as enabling technologies that stimulate further innovations (Cardona et al., 2013).

Research on the impact of digitalisation reports a positive association between certain digital technologies, e.g. cloud computing (DeStefano et al., 2023; Duso and Schiersch, 2022) or broadband (Bertschek et al., 2015; Duso et al., 2021), and productivity (Gal et al., 2019). To tap into their full potential, complementary investments in employees' skills (e.g. training) and factors like a firm's intangible capital (e.g. software and data) may be required (Brynjolfsson and McAfee, 2011; Brynjolfsson et al., 2021b; DeStefano et al., 2023; Van Ark, 2016). As shown in the past, firms that invest most in intangible assets exhibit the strongest productivity growth (Crouzet and Eberly, 2018). This might be explained by the fact that intangibles can bridge the gap between the introduction of a new technology and observable advances in productivity (Mohnen et al., 2019). However, up to now, there is hardly any evidence on the impact of the latest wave of technologies, especially of BDA, on firm-level productivity and this paper is the first to present cross-country evidence for the Netherlands and Germany.

Previous research on the impacts of BDA has primarily focused on related topics like companies' innovation success, which may affect firm productivity (Hall, 2011; Rosenberg, 2006). Gierten et al. (2021) find that firms, which use BDA, have a higher tendency to develop innovations regarding processes, products, marketing and the organisation of the firm. In particular, the analysis of user-related data, e.g. from social media, correlates most frequently with a higher propensity to innovate. However, they do not claim causality for their results. Similarly, using IV estimations, Bertschek and Kesler (2022) find causal evidence that firms' adoption of a Facebook page and the feedback obtained from users via this channel positively and significantly affect product innovations. Besides, using a knowledge production function framework with German firm-level data, Niebel et al. (2019) find suggestive evidence of a positive connection between the use of BDA and innovation output as well as the respective innovation's market success. Applying IV estimations, Wu et al. (2020) show that data analytics capabilities are more likely to benefit companies that engage in process improvements and creating new technologies by combining existing ones, while focusing less on disruptive innovations. They conclude that data analytics methods may rather complement specific types of innovation by allowing firms to use and recombine a wide range of existing knowledge. Conti et al. (2024) show that positive effects of BDA on the innovation process exist for both small and large firms. However, the origins of performance improvements associated with these innovation gains vary depending on firm size. While smaller firms use BDA to achieve product innovations to boost their sales, larger firms make use of BDA to foster process innovations and to decrease costs, and the impact increases with firm size.

Related literature regarding artificial intelligence (AI) and innovations reports a positive impact of AI use on process innovations, cost savings and innovation output (Rammer et al., 2022). Although AI does not equal BDA, being capable of analysing extremely large amounts of data is indispensable in the AI context as well. Grashof and Kopka (2023) find that the impact of AI differs between firm sizes and between AI applications versus AI techniques. They argue that small and medium-sized enterprises (SMEs) mostly benefit from AI techniques, while larger firms rather benefit from radical innovation outputs due to AI applications. Moreover, Cho et al. (2023) show that technologies that enable handling or require the availability of large amounts of data, such as AI and big data, exhibit statistical complementarities in their deployment.

Literature on the role of using big data for lifting productivity at the firm level is supposed to be of great interest for managers and policy makers. As such, Müller et al. (2018) apply two-stage least squares IV estimates and report a positive impact of BDA assets on firm productivity, with the effect varying substantially over different industries. According to their findings, firms in IT-intensive industries as well as firms in industries characterised by a higher degree of competitiveness are more likely to show performance gains. Moreover, for the case of start-ups, Rodepeter et al. (2023) point towards the importance of the time horizon for investigating productivity effects of BDA. According to their findings, BDA is associated with a negative impact on adopting start-ups' competitive performance in the short-term. However, conditional on survival, these firms benefit from a higher long-term performance regarding, for instance, growth. Related, Cerqueira et al. (2023) focus on the timing of BDA adoption. Making use of a propensity score estimator to re-weight firms to account for differences in the probability of adoption, they show that first-movers regarding BDA are able to gain a productivity advantage over those who never adopt and over later adopters. According to Wu et al. (2019), the productivity effects of analysing large amounts of data also depend on the internal innovation structure of firms: Those firms having decentralised innovation structures exhibit not only greater demand for analytics skills in their labour force but also seem to reap greater benefits with respect to productivity from these capabilities.

Brynjolfsson and McElheran (2019) find that data-driven decision making (DDD)[1], which is closely linked to BDA particularly focused on supporting management decisions, leads to increased productivity in U.S. manufacturing firms. In particular, they report significant advantages of first movers in adopting DDD, driving productivity gains. They rely i.a. on IV estimations and timing falsification to claim causality of their results. In a later study, Brynjolfsson et al. (2021a) find that complementary inputs, i.e. IT capital, educated workers or an appropriate workplace design, are necessary for predictive analytics[2] to unfold their productivity pay-off. This may serve as one explanation for why beneficiary effects of predictive analytics do not seem to occur for all firms equally. In a recent large-scale field experiment, Bar-Gill et al. (2024) study eBay's staggered introduction of the Seller Hub, a dashboard providing lots of data insights for sellers on their platform. They find that having access to such data analytics tools leads to increases in DDD, which, in turn, leads to increases in revenues of small businesses. The authors claim that SMEs are often constrained in terms of their financial capacity, which leads to lower adoption rates regarding advanced data analytics tools. However, the results also show that the size of the impact of this data analytics tool depends on complementary managerial practices.

Related to this paper, Borowiecki et al. (2021) are among the first to investigate the causal impact of BDA on labour productivity growth. Focusing on non-frontier firms in the Netherlands, they do not find a statistically significant relationship. Other analyses of the impact of BDA on productivity either stay on the descriptive level or focus only on a subset of companies (e.g. publicly listed firms) (Wu et al., 2020).

These drawbacks are mainly due to the fact that, in general, the econometric identification of causal productivity effects of digital technologies is hampered by the presence of endogeneity. Concerns arise from potential *reverse causality*, i.e. more productive firms having a higher tendency towards adopting cutting edge technologies, such as BDA, and from unobserved common confounding factors affecting both digitalisation (e.g. likelihood of conducting BDA) and productivity (Borowiecki et al., 2021). The latter refers, for instance, to better managed firms adopting new technologies more frequently while also being more productive, which may inflate estimates of digital technologies' productivity impact (Andrews et al., 2018; Bloom et al., 2012).

[1] Brynjolfsson and McElheran (2019) define DDD as "collection of and reliance on data for managerial activities", p. 2.

[2] The authors define predictive analytics as "a set of techniques—from data mining to statistical modeling, including, in some firms, machine learning and "AI"—used to analyze historical and current data in order to make predictions about future or unknown events.", p. 218f.

Apart from the literature on BDA and productivity, there is also initial research on AI and productivity. As mentioned before, AI does not equal BDA, but being capable of handling and analysing extremely large amounts of data is essential for AI as well. Czarnitzki et al. (2023) find a positive and significant relationship between AI and productivity based on a cross-section of German firms. In addition, the paper by Calvino and Fontanelli (2023) finds similar results for the majority of selected OECD countries in their sample. Furthermore, they also find an association between AI and LP that is more pronounced for larger firms. Besides, there is also first evidence on how generative AI affects a worker's productivity (Brynjolfsson et al., 2023).

3 Data Sources and Variable Definitions

3.1 Dutch Data

Our main explanatory variable of interest is the firm-level information on *big data analytics* usage. This data, along with important control variables on general ICT intensity, come from the Dutch version of the "Survey on ICT Usage and E-Commerce in Enterprises"[3], which is administered by Eurostat and conducted by the Dutch Statistical Office CBS[4]. The survey contains information on the use and the importance of different ICTs. The key *big data*-variable is constructed by combining the answers to several questions regarding the use of big data to a binary indicator.[5] The key components encompass questions on whether a firm indicates running big data analyses and which kind of big data sources it uses. Big data can be retrieved from smart devices or sensors ("internal data source"), focus on geolocation data or data generated from social media as well as other big data sources ("external data source"). The binary indicator equals one if a firm indicates to use at least one of the different data types. Different from the general Eurostat survey, CBS included questions on the big data usage of firms not only in the compulsory years of 2016, 2018, and 2020, but additionally in 2017. Therefore, our Dutch analyses include an additional year compared to the German analyses.

[3] `www.doi.org/10.21242/52911.2020.00.00.1.1.0`.

[4] CBS, Centraal Bureau voor de Statistiek `https://www.cbs.nl/en-gb/`.

[5] The Community Survey on ICT Usage and E-Commerce in Enterprises defines big data as being "generated from activities that are carried out electronically and from machine-to-machine communications". In particular, three distinct characteristics - significant volume, variety of different formats of complex data, and velocity, which refers to the high frequency of data generation - describe big data. Besides, big data analysis contains the use of techniques, technologies and software tools for the analysis of big data obtained from enterprises' own data sources or other data sources. The questions on BDA always refer to the usage in year $t-1$. For more information on the exact framing of the questions, see Figure A.1 in the Appendix.

The firm performance metrics, which are the independent variables of interest, as well as additional firm characteristics such as the number of employees stem from the *Production Statistics* provided by CBS. Furthermore, information on gross investments to model stocks and flows of different assets are added, which come from the *Investment Statistics* provided by CBS.[6]

Finally, we merge external data obtained from the EU-KLEMS capital and national accounts (Bontadini et al., 2023) that include deflators and depreciation rates (for all assets), each at the two-digit industry level.

The final Dutch data set consists of the merge of the four above-mentioned data sets. Due to merging the data sets, the resulting sample is not representative of the population of Dutch firms and we control for industry sectors and firm size in our estimations. Overall, our final data set includes data for 23 industries over the period 2016 to 2020 (without 2019), resulting in 25,122[7] firm-year observations (see Table A.2 for the number of observations by industry).

Table 3.1 provides detailed summary statistics of all the variables used in our empirical analysis. It shows that on average 32 percent of the firms are using some type of BDA, with own data and social media data being the most common data sources for BDA. In particular, large firms, that have at least 250 employees, seem to be more likely to use BDA. While almost 44 percent used BDA already back in 2016, this share increased to more than half of the large firms in 2017 and then remained on this level, amounting to more than 56 percent in 2020 (see Figure 3.1). In contrast, small firms with less than 50 employees started on a comparably low level with hardly every fifth firm applying BDA. This share increased until 2020 to more than 26 percent, corresponding to an increase of roughly one fourth. The average Dutch firm in our sample has 185 employees and a LP of about 92.5 thousand euro. Over 80 percent of the firms are either small or medium-sized enterprises and are predominantly (68 percent) located in the services sector. Table A.3 in the Appendix provides a distinction between firms that use BDA and those who are not using BDA. It clearly shows that BDA-using firms are substantially larger in terms of value added (about 10 million euro vs. about 31 million euro) and employment (125 vs. 309 employees). They are also more digitalised with the share of employees using a computing device with internet access being 9 percentage points larger for the BDA-using firms.

[6] `https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research`.

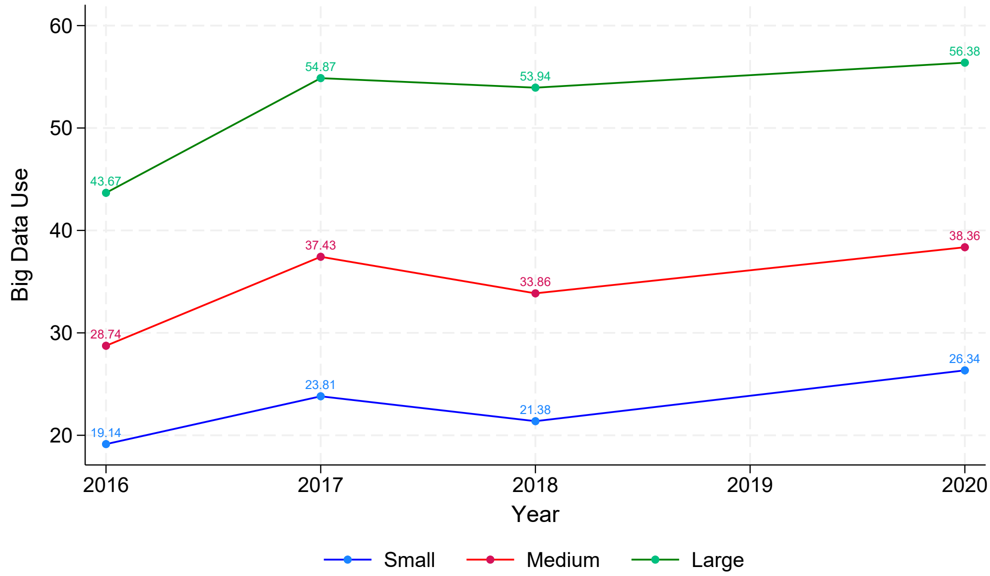
[7] The control variable *Sh_internet* is not available in 2017. We use the average of 2016 and 2018. As we have an unbalanced panel, this reduces the number of observations in regressions with this control variable to 23,905.

Table 3.1: Summary Statistics Dutch Estimation Sample

	N	Mean	Median	SD	p10	p90
BDA own	25122	.17	0	.37	0	1
BDA geo	25122	.12	0	.32	0	1
BDA social	25122	.17	0	.38	0	1
BDA other	25122	.13	0	.33	0	1
BDA	25122	.32	0	.47	0	1
BDA intensity	25122	.58	0	1	0	2
BDA all	25122	.025	0	.15	0	0
BDA 3 types	25122	.069	0	.25	0	0
BDA intensity std	25122	.003	-.58	1	-.58	1.5
BDA int. == 0	25122	.68	1	.47	0	1
BDA int. == 1	25122	.16	0	.37	0	1
BDA int. == 2	25122	.094	0	.29	0	0
BDA int. == 3	25122	.045	0	.21	0	0
BDA int. == 4	25122	.025	0	.15	0	0
Sh_internet	23905	70	81	33	18	100
VA	25122	16,632,101	4,115,975	93,729,434	663,508	29,751,635
L	25122	185	60	552	11	385
K	25122	27,976,096	1,358,032	319299705	129,064	23,461,162
LP	25122	92,579	66,558	102,669	32,670	159,095
Small firm	25122	.45	0	.5	0	1
Medium firm	25122	.4	0	.49	0	1
Large firm	25122	.16	0	.37	0	1
Manufacturing	25122	.24	0	.43	0	1
Services	25122	.68	1	.47	0	1

NOTE: See Table A.1 for a detailed description of the variables.

Figure 3.1: Big Data Analytics Use over Time by Firm Size Class in the Netherlands



NOTE: The graph shows the evolution of the adoption rate of BDA by firm size class over time.

3.2 German Data

Similar to the Dutch data set, the German data set is constructed out of several different data sources. To ensure comparability between the German and the Dutch analyses, the data sources are as comparable as possible. As such, the information if big data analytics is employed by a firm stems from the German counterpart of Eurostat's "Survey on ICT Usage and E-Commerce in Enterprises". Additional information, such as deflators and depreciation rates at the industry-level, stem from the EU-KLEMS capital and national accounts (Bontadini et al., 2023). As these data sources are either identical or administered by the same organization, the variable definitions are consistent across countries and ensure high data comparability.

The firm-level data stem from the AFiD-Panel Service Firms[8] and the AFiD-Panel Manufacturing Firms[9] provided by the German statistical offices. The data sets contain a variety of firm characteristics and performance metrics.

Amongst others, information on the number of employees, on gross investments as well as on value added are contained. Value added serves as output variable for the estimation of the production function

[8] www.doi.org/10.21242/47415.2020.00.01.1.1.0.
[9] www.doi.org/10.21242/42221.2021.00.01.1.1.0.

(see Section 4).

Our final panel data set is constructed by linking the four above-mentioned data sources covering the years between 2016 and 2020 (without 2017 and 2019 as questions regarding BDA in Germany are only asked biennially). It is composed of 18 manufacturing and services sector industries, resulting in 8,612 firm-year observations (see Appendix A.8 for the number of observations by industry).

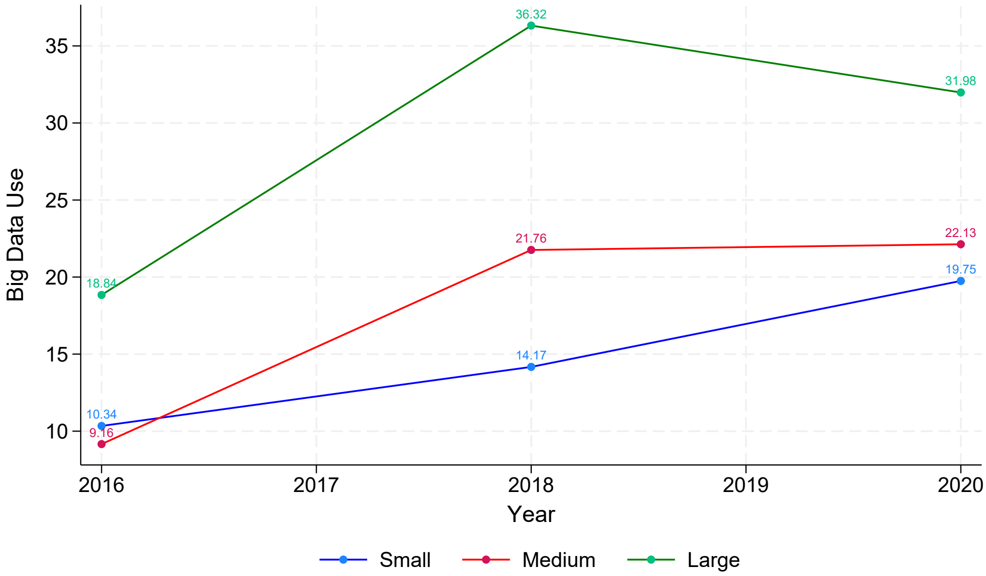
Table 3.2: Summary Statistics German Estimation Sample

	N	Mean	Median	SD	p10	p90
BDA own	8594	.085	0	.28	0	0
BDA geo	8551	.09	0	.29	0	0
BDA social	8547	.098	0	.3	0	0
BDA other	8502	.064	0	.24	0	0
BDA	8612	.22	0	.41	0	1
BDA intensity	8612	.33	0	.74	0	1
BDA all	8612	.0082	0	.09	0	0
BDA 3 types	8612	.026	0	.16	0	0
BDA intensity std	8612	.12	-.39	1.1	-.39	1.4
bd_int== 0	8612	.78	1	.41	0	1
bd_int== 1	8612	.13	0	.34	0	1
bd_int== 2	8612	.057	0	.23	0	0
bd_int== 3	8612	.018	0	.13	0	0
bd_int== 4	8612	.0082	0	.09	0	0
Sh_internet	8612	58	57	35	10	100
VA	8612	33,499,338	6,671,497	302653488	1,087,510	50,242,972
L	8612	372	121	2,159	27	674
K	8612	42,515,103	3,269,941	731814302	140,390	44,430,174
LP	8612	75,455	55,530	214,922	21,749	115,252
Small firm	8612	.22	0	.42	0	1
Medium firm	8612	.44	0	.5	0	1
Large firm	8612	.33	0	.47	0	1
Manufacturing	8612	.57	1	.49	0	1
Services	8612	.43	0	.49	0	1

NOTE: See Table A.1 for a detailed description of the variables.

The summary statistics in Table 3.2 offer a detailed description of the variables used in our empirical analysis. On average, 22 percent of the German firms in our sample indicate to use at least some type of BDA, with social media (9.8 percent) and geolocation data (9 percent) being named most frequently. In contrast to the Netherlands, even among large German firms ($\geq$ 250 employees) adoption of BDA was comparably low in 2016, amounting to less than 19 percent (see Figure 3.2). After a jump to around 36

Figure 3.2: Big Data Analytics Use over Time by Firm Size Class in Germany



NOTE: The graph shows the evolution of the adoption rate of BDA by firm size class over time.

percent in 2018, the share of large firms using BDA went back to 32 percent in 2020. Small firms (< 50 employees) in Germany also started with a comparably small adoption rate of just above 10 percent. Over the period of observation, this share increased steadily to almost 20 percent in 2020, which is comparable to the value of medium-sized firms. The average firm in the German sample has 372 employees and shows a LP of around 75.5 thousand euro. Roughly two thirds of the firms are classified as either small- or medium-sized and more than half of the firms (57 percent) belong to the manufacturing sector. A detailed distinction between firms using BDA and those who do not can be found in Table A.9 in the Appendix. In general, it can be shown that firms that use BDA are, on average, considerably larger regarding the number of people employed (794 vs. 254 employees) and value added (about 20 million euro vs. about 83 million euro). Moreover, BDA-using firms exhibit a 14 percentage points higher digitalisation level as indicated by the share of employees using a computing device with internet access.

4 Empirical Specification

We start the analysis with a value-added Cobb-Douglas production function (Equation 1), which includes the production inputs labour (L) and capital (K). The dependent variable refers to real gross value added (Y).

$$Y = A * L^{\beta_l} * K^{\beta_k} \tag{1}$$

By dividing both sides by L, we obtain labour productivity ($\frac{Y}{L}$, referred to as LP from now on) as new dependent variable (Equation 2).

$$\frac{Y}{L} = LP = A * L^{\beta_{l-1}} * K^{\beta_k} \tag{2}$$

We assume variable returns to scale (VRS), as it is less restrictive than constant returns to scale (CRS). Consequently, the sum of the coefficients does not necessarily have to add up to one (Equation 3).

$$\beta_l + \beta_k \neq 1 \tag{3}$$

Lastly, we take Equation 2 in logarithmic terms and include a dummy variable for the use of BDA as well as a variable accounting for firms' general digitalisation level ($Sh\_internet_{it}$). The resulting production function to be estimated takes on the representation of Equation 4.

$$lp_{it} = \beta_0 + (\beta_l - 1) * l_{it} + \beta_k * k_{it} + \beta_{Sh\_internet} * Sh\_internet_{it} + \beta_{BDA} * BDA_{it} + \underbrace{\varepsilon_{it}}_{TFP_{it}+\epsilon_{it}} \tag{4}$$

One apparent threat to the application of the augmented Cobb-Douglas production function is its susceptibility to the adverse influence of endogeneity. This is captured by the observed error term ε_{it} that contains not only the true independent and identically distributed error term (ϵ_{it}) but also the unobserved total factor productivity (TFP_{it}).

To quantify the impact of BDA on productivity, we start with the augmented Cobb-Douglas production function with labour productivity lp_{it} by firm i in period t as the dependent variable described in Equation 4 as starting point. Labour l_{it} and capital k_{it}[10] represent the production inputs, $Sh\_internet_{it}$ is a variable that controls for the general digitalisation level of each firm in each period[11], and BDA_{it} is the main independent variable of interest. The latter is a binary indicator variable, equal to 1 if a

[10]See Section A.1 in the Appendix for a detailed description of the calculation of the capital stocks.

[11]Here, we use the share of employees in each firm-year observation that uses computing devices (PCs, notebooks, tablets or smartphones) with internet access for daily work.

firm uses BDA in the respective year t, and 0 otherwise. We further add a vector of controls $\mathbf{X_{it}}$ that accounts for various confounding factors, such as firm size and the industry in which the firm is operating as well as year dummies controlling for time fixed effects (see Equation 5) and estimate it using OLS.

$$lp_{it} = \beta_0 + (\beta_l - 1) * l_{it} + \beta_k * k_{it} + \beta_{Sh\_internet} * Sh\_internet_{it} + \beta_{BDA} * BDA_{it} + \beta_{\mathbf{X}} * \mathbf{X_{it}} + \varepsilon_{it} \quad (5)$$

5 Econometric Results

5.1 Results for the Dutch Sample

5.1.1 Big Data Analytics as Binary Measure

Table 5.1: OLS Regressions Netherlands: BDA as Binary Variable

BDA=1	0.0814***	0.0859***	0.0477***	0.0143
	(0.009)	(0.009)	(0.009)	(0.009)
ln(L)	-0.1635***	-0.1637***	-0.1891***	-0.1929***
	(0.005)	(0.005)	(0.005)	(0.005)
ln(K)	0.1129***	0.1130***	0.1691***	0.1746***
	(0.003)	(0.003)	(0.004)	(0.004)
Sh_internet				0.0048***
				(0.000)
Constant	10.1816***	10.2165***	9.2644***	9.0089***
	(0.037)	(0.038)	(0.051)	(0.051)
Year DVs	No	Yes	Yes	Yes
Industry DVs	No	No	Yes	Yes
Observations	25122	25122	25122	23905
Adjusted R-squared	0.066	0.068	0.173	0.218

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Column (1) of Table 5.1 shows the results of a basic OLS regression of lp_{it} on capital k_{it} and labour l_{it} without controlling for the general digitalisation level, $Sh\_internet_{it}$, industry or year fixed effects. The coefficient of BDA_{it} indicates that using BDA-tools is associated with an 8.14 percent increase in lp.

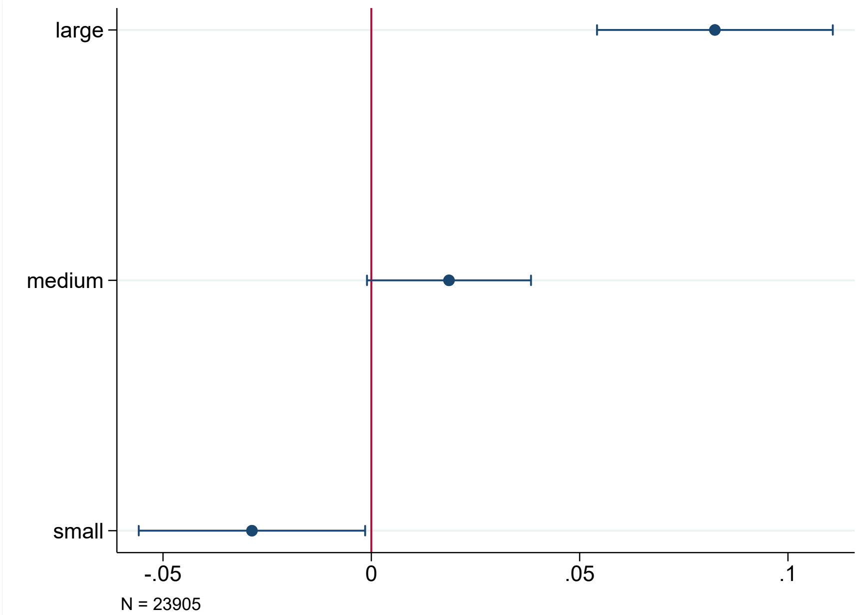
Adding year-dummies to control for time fixed effects in column (2) hardly affects this result. In column (3), we additionally add industry dummies to account for industry fixed effects. This results in a bisection of the coefficient on BDA, which remains highly statistically significant at the 1 percent significance level. However, after controlling for the general digitalisation level ($Sh\_internet_{it}$) of a firm in column (4), which corresponds to the representation of the production function as depicted in Equation 5, the BDA-coefficient turns statistically insignificant. This indicates that previously, the positive effect of the level of digitalisation of a firm was captured by the BDA-coefficient.
The remainder of this section digs deeper into the effect of BDA on lp by exploiting our fine-grained data in more detail.

Table A.4 in the Appendix breaks up the aggregated BDA-variable into its sub-components. The analysis reveals a statistically significant and positive association between BDA and labour productivity for firms using big data from internal sources (own). A similar association exists for firms using other big data sources ($other$) not covered by internal, social media, or geolocation data. In contrast, no such relation can be observed when looking at geolocation (geo) or social media ($social$) data. This indicates that the aggregated BDA-indicator hides the differential effect that different big data sources exert.

Figure 5.1 depicts the marginal effect of BDA on lp with respect to firm size after adding an interaction term between firm size and BDA to Equation 5[12]. In line with the literature, which finds firms that are i.a. larger to be more successful in the adoption and usage of digital technologies (Haller and Siedschlag, 2011), we show that the marginal effect of BDA increases with firm size and is significantly different from zero for the group of large firms.

[12] The coefficient corresponds to the sum of the coefficients of the binary BDA-variable and of the interaction term between BDA and $firm\ size\ class$. For detailed results, see Table A.6.

Figure 5.1: Big Data Analytics in the Netherlands by Firm Size

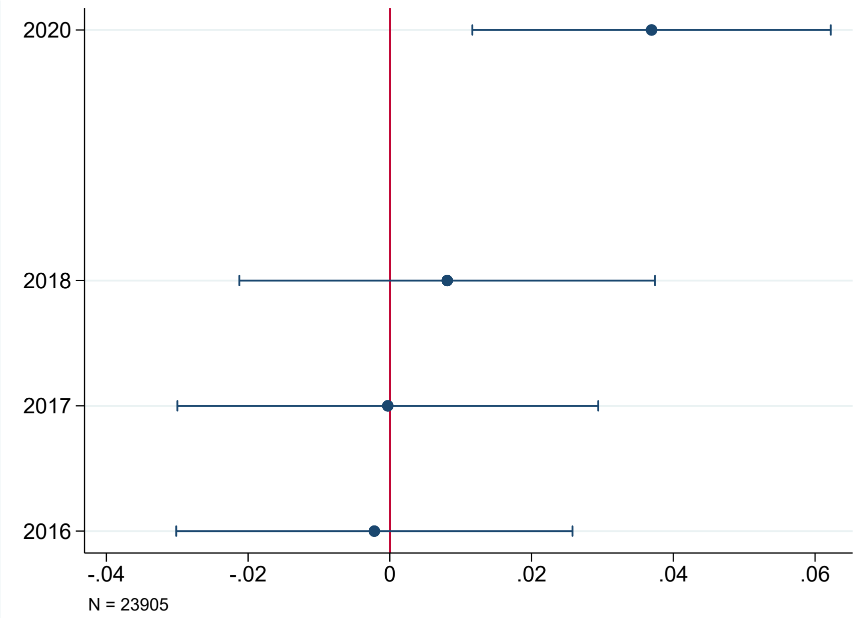


Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of BDA when including both BDA and its interaction with the respective $firm\ size\ class$-dummy in the regression. Also see Table A.6 for exact regression results.

Similarly, Figure 5.2 depicts the marginal effect of BDA on lp over time after adding an interaction term between survey $year$ and BDA to Equation 5[13]. In line with findings on the diffusion of digital technologies, we find that the effect of BDA on lp increases over time and becomes statistically significantly different from zero in 2020, the most recent data point available. In Figure A.2 in the Appendix, we also show the results for including both BDA and the interaction terms between BDA and each industry.

[13] The coefficient corresponds to the sum of the coefficients of the binary BDA-variable and of the interaction term between BDA and the survey $year$-variable. For detailed regression results, see Table A.6.

Figure 5.2: Big Data Analytics in the Netherlands by Year



NOTES: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of BDA when including both BDA and its interaction with the respective survey *year*-dummy in the regression. Also see Table A.6 for exact regression results.

Besides, the above results also hold when we subdivide the sample according to the different *years* of observation as well as according to the different *firm size classes* (see Table A.5 in the Appendix, column 1-4 for years and column 5-9 for firm size classes).

5.1.2 Big Data Analytics Measured as Intensity

In Table 5.2, we extend our analysis presented before by various BDA intensity measures that explicitly take into account the different forms of BDA a firm applies, thereby, indicating how prevailing BDA is in a particular firm in a certain year. The results in column (1) correspond to column (4) in Table 5.1 and serve as benchmark. This means that we control for the general digitalisation level as well as for year and industry fixed effects in all regressions presented here. In column (2), we replace the binary BDA-indicator by a categorical variable ranging from 0 to 4, with a value of 0 representing a firm that does not use any kind of BDA and 4 for a firm that uses internal, geolocation, social media and further big data sources. *BDA intensity* exhibits a positive and highly statistically significant coefficient, indicating that more intensive use of BDA is associated with increased lp. Column (3) confirms this result. Here, we use a binary variable, indicating whether a firm is not only using BDA but belongs to the group of the most

Table 5.2: OLS Regressions Netherlands: BDA Intensity

	(1)	(2)	(3)	(4)	(5)	(6)
BDA=1	0.0143					
	(0.009)					
BDA intensity		0.0133***				
		(0.004)				
BDA all=1			0.0927***			
			(0.028)			
BDA 3 types=1				0.0685***		
				(0.016)		
BDA intensity std					0.0132***	
					(0.004)	
BDA intensity=1						0.0082
						(0.011)
BDA intensity=2						-0.0106
						(0.014)
BDA intensity=3						0.0532***
						(0.019)
BDA intensity=4						0.0976***
						(0.028)
ln(L)	-0.1929***	-0.1939***	-0.1929***	-0.1936***	-0.1939***	-0.1936***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(K)	0.1746***	0.1742***	0.1746***	0.1742***	0.1742***	0.1742***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Sh_internet	0.0048***	0.0048***	0.0048***	0.0048***	0.0048***	0.0048***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	9.0089***	9.0195***	9.0130***	9.0208***	9.0271***	9.0203***
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
Year DVs	Yes	Yes	Yes	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23905	23905	23905	23905	23905	23905
Adjusted R-squared	0.218	0.218	0.218	0.218	0.218	0.218

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

intensive users, referring to the usage of all four of the aforementioned big data sources for their analyses. Again, the coefficient is positive and highly statistically significant. In column (4), we extend this binary variable now indicating whether a firm uses at least 3 of the big data sources. Although the coefficient decreases by roughly one fourth in size, it is still highly statistically significant. In column (5), we add a standardised BDA-measure, which provides for each firm a normalised measure of the intensity of BDA use[14]. As before, the coefficient of interest remains highly statistically significant. Lastly, in column (6), we include binary variables for each level of BDA *intensity*. In accordance with prior results, the coefficients for the variables that indicate more intensive use of BDA, that is firms which use three or four different big data sources, are positive and highly statistically significant.

In line with our findings in Section 5.1, results following the inclusion of both BDA intensity and an interaction term between BDA intensity and $firm\ size\ class$ (see Table A.7 in the Appendix for detailed regression results) show a statistically significantly positive marginal effect of BDA intensity for large firms[15]. This effect can be observed for both measures of BDA intensity, one indicating whether a firm uses at least three different BDA-types (see Figure A.3) and the other using the standardised BDA intensity measure (see Figure A.4). Besides, after interacting BDA intensity with the respective survey *year* (see Table A.7 in the Appendix for detailed regression results), we receive positive and statistically significant marginal effects for BDA intensity for the years 2018 and 2020[16], again for both measures of BDA intensity (see Figure A.5 and Figure A.6).

[14]See Section A.2 in the Appendix for a detailed description.

[15]The coefficient corresponds to the sum of the coefficients of the respective BDA *intensity*-variable and of the interaction term between BDA *intensity* and $firm\ size\ class$. For detailed results, see Table A.7.

[16]The coefficient corresponds to the sum of the coefficients of the respective BDA *intensity*-variable and of the interaction term between BDA *intensity* and the survey *year*-variable. For detailed results, see Table A.7.

5.2 Results for the German Sample

5.2.1 Big Data Analytics as Binary Measure

Table 5.3: OLS Regressions Germany: BDA as Binary Variable

	(1)	(2)	(3)	(4)
BDA=1	0.0587***	0.0713***	0.0737***	0.0240
	(0.021)	(0.022)	(0.021)	(0.020)
ln(L)	-0.0660***	-0.0684***	-0.0485***	-0.0469***
	(0.010)	(0.010)	(0.010)	(0.010)
ln(K)	0.0918***	0.0919***	0.0720***	0.0678***
	(0.005)	(0.005)	(0.005)	(0.005)
Sh_internet				0.0058***
				(0.000)
Constant	9.8298***	9.9023***	9.9069***	9.8143***
	(0.055)	(0.058)	(0.065)	(0.063)
Year DVs	No	Yes	Yes	Yes
Industry DVs	No	No	Yes	Yes
Observations	8612	8612	8612	8612
Adjusted R-squared	0.111	0.114	0.242	0.283

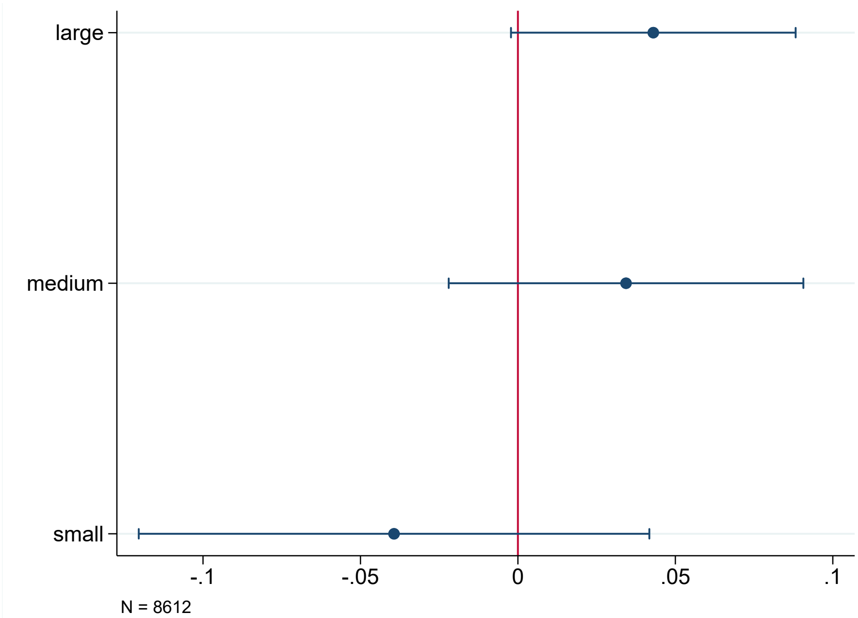
NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 5.3 presents the baseline results for estimating the production function for Germany via OLS, using BDA as a binary variable as main independent variable of interest. Similar to the findings for the Dutch data, the coefficient on BDA in column (1) is positive and statistically significant at the 1 percent significance level when only capital k and labour l are included in the production function. After adding time dummies in column (2), the BDA-coefficient remains highly statistically significant and increases in size. Even after accounting for industry fixed effects in column (3), the coefficient on BDA remains mainly unchanged. However, similar to the case of the Dutch data, the coefficient of the binary BDA-measure turns statistically insignificant once we control for the general digitalisation level of the firm in column (4).

As in the Dutch case, we try to disentangle the effect of the aggregated binary BDA-variable into its sub-components in Table A.10 in the Appendix. Again, we observe a statistically significant positive relationship between BDA and lp for firms that apply BDA using either internal data sources (own) or other big data sources ($other$), which are neither covered by internal nor geolocation nor social media data. The coefficients for own and $other$ are highly statistically significant and roughly twice as large as in the Dutch data. This further supports the assumption that the aggregated BDA-variable might hide some of the effect heterogeneity of the different big data sources that can be used for BDA.

Figure 5.3 shows the marginal effect of BDA according to $firm\ size\ class$ on lp and follows from the inclusion of an interaction term between BDA and $firm\ size\ class$-dummies in Equation 5. Similarly to the Dutch results, we see an increasing marginal effect of BDA along firm size[17]. However, in contrast to the results shown in Section 5.1, the coefficient remains statistically insignificant, although the confidence interval decreases for higher size classes.

Figure 5.3: Big Data Analytics in Germany by Firm Size



Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of BDA when including both BDA and its interaction with the respective $firm\ size\ class$-dummy in the regression. Also see Table A.12 for exact regression results.

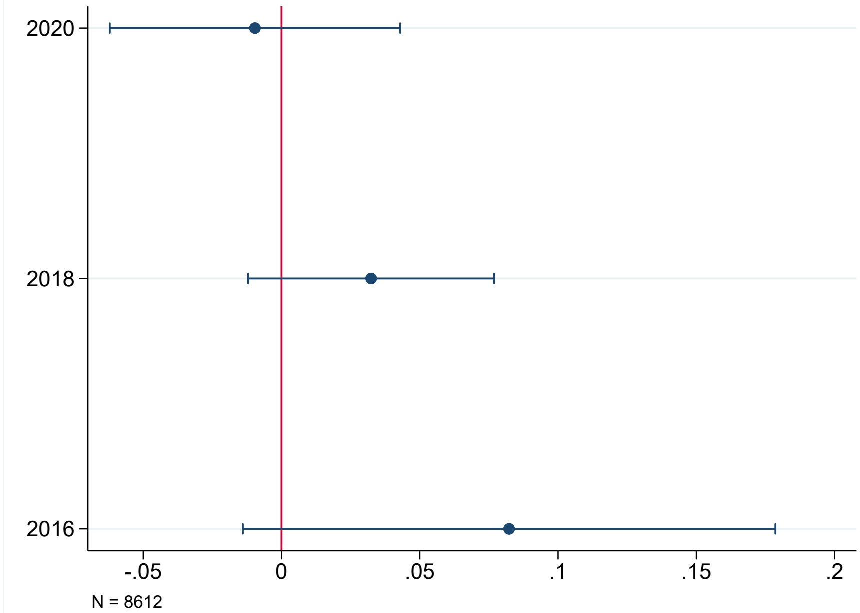
Also, we do not find a statistically significant time trend after including BDA alongside an interaction term between BDA and the survey $year$-dummies in Equation 5[18], which may be owed to the comparably

[17] The coefficient corresponds to the sum of the coefficients of the binary BDA-variable and of the interaction term between BDA and $firm\ size\ class$. For detailed results, see Table A.12.

[18] The coefficient corresponds to the sum of the coefficients of the binary BDA-variable and of the interaction term between BDA and the survey $year$-variable. For detailed regression results, see Table A.12.

more limited observation count for Germany. The marginal effect of BDA with respect to respective survey *year* is depicted in Figure 5.4. Moreover, results for the marginal effects after the inclusion of interaction terms between BDA and each *industry* are displayed in Figure A.7.

Figure 5.4: Big Data Analytics in Germany by Year



NOTES: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of BDA when including both BDA and its interaction with the respective survey *year*-dummy in the regression. Also see Table A.12 for exact regression results.

As in the Dutch case, we replicate the above heterogeneity analyses by subsample regressions instead of interaction terms (see Table A.11 in the Appendix). In contrast to the Dutch results, no statistically significant effects of BDA on lp can be observed.

5.2.2 Big Data Analytics Measured as Intensity

In Table 5.4, we proceed as in Section 5.1.2 and replace the binary BDA-indicator (column 1) by measures that capture the intensity of BDA at the firm-year level (column 2-5). Column (2) shows the results when applying a categorical variable measuring the BDA intensity by summing up the number of different BDA-types a firm uses. Similar to the results for the Netherlands, the coefficient of BDA *intensity* is statistically significant, though only at the 10 percent significance level. In column (3), the dependent variable is again of binary nature, equalling 1 if a firm makes use of all four types of BDA asked in the questionnaire and zero if it uses less than all or none of them. In contrast to the Dutch results, the

Table 5.4: OLS Regressions Germany: BDA Intensity

	(1)	(2)	(3)	(4)	(5)	(6)
BDA=1	0.0240					
	(0.020)					
BDA intensity		0.0209*				
		(0.011)				
BDA all=1			0.1023			
			(0.091)			
BDA 3 types=1				0.0875*		
				(0.046)		
BDA intensity std					0.0158**	
					(0.007)	
BDA intensity=1						0.0085
						(0.024)
BDA intensity=2						0.0332
						(0.039)
BDA intensity=3						0.0842
						(0.051)
BDA intensity=4						0.1114
						(0.091)
ln(L)	-0.0469***	-0.0481***	-0.0461***	-0.0470***	-0.0487***	-0.0480***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
ln(K)	0.0678***	0.0677***	0.0678***	0.0678***	0.0677***	0.0677***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Sh_internet	0.0058***	0.0058***	0.0058***	0.0058***	0.0058***	0.0058***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	9.8143***	9.8215***	9.8099***	9.8153***	9.8311***	9.8215***
	(0.063)	(0.063)	(0.062)	(0.062)	(0.063)	(0.063)
Year DVs	Yes	Yes	Yes	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8612	8612	8612	8612	8612	8612
Adjusted R-squared	0.283	0.283	0.283	0.283	0.283	0.283

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

coefficient is statistically insignificant for Germany, which may be due to the considerably lower number of observations in the German data. Using a somewhat broader binary measure in column (4), that equals 1 if a firm uses at least three different big data sources, the coefficient is again positive and statistically significant at the 10 percent significance level. The coefficient of the standardised BDA intensity-measure in column (5) shows the expected sign and is statistically significant at the 5 percent significance level. Lastly, in column (6), dummy variables for each level of $BDA\ intensity$ are included. In contrast to the results found for the Dutch data, even the coefficients for $BDA\ intensity\ = 3$ and $BDA\ intensity = 4$ remain statistically insignificant for the German data, which may be attributed to the lower observation count and, thus, less variance in the German data.

Interestingly, after including an interaction term between BDA intensity and $firm\ size\ class$ (see Table A.13 in the Appendix for detailed results), the pattern mirrors the one we observe for the Netherlands. In particular, the marginal effect of BDA on lp turns statistically significant for the group of large firms. In the German case, we already observe a statistically significant coefficient for the interaction with medium-sized firms These observations hold for both measures of BDA intensity, meaning whether a firm uses at least three different BDA-types (see Figure A.8) as well as for the standardised BDA intensity measure (see Figure A.9).

To sum it up, the coefficients of the different BDA intensity-measures for Germany are on average larger in size than the results found for the Netherlands in Section 5.1.2. However, in contrast to the Dutch results, the regression coefficients for BDA are only statistically significant at lower statistical significance level and are partly not statistically significant at all in the German case. Looking at the interaction of the BDA-dummy with size classes and years, the effects we measure are stronger for the Dutch case. In contrast, the results regarding the interaction of the BDA intensity-measures with size classes are stronger for the German data.

5.3 Summary Results

Summing up the results for Germany and the Netherlands, we do not find a statistically significant effect of the binary BDA indicator on LP after controlling for the general level of digitalisation as presented in Table 5.5. This suggests that the statistically significant effect observed before rather captures the impact of the general level of digitalisation than the effect of BDA. However, when examining potential effect heterogeneity, we find, at least for the Netherlands, that especially large firms seem to benefit from the use of BDA. Besides, as diffusion and adoption of the technology may take some time, we see that the use of BDA significantly impacts LP in the year 2020, the last year in which the BDA information is observed.

Table 5.5: Summary of the Results: Big Data Analytics as Binary Measure

	NL	**DE**
All (w/o general level of digitalisation)	✓	✓
All (controlling for level of digitalisation)	–	–
Year	2020	–
Firm size	large	–

NOTE: ✓ means a statistically significant finding, – means no significant finding. If only single years or size classes are indicated, these are the only significant estimates.

As the aggregate binary indicator fails to differentiate between varying degrees of BDA use, we looked closer at different measures of intensity of BDA use in Section 5.1.2 and Section 5.2.2 (see Table 5.6 for an overview of the results). Here, we show that even after controlling for the general level of digitalisation, we find a statistically significant positive effect of intensive BDA use on LP for both, the variable indicating whether a firm uses at least three different big data types and the one representing a standardised BDA intensity measure. Again, we also look at potential effect heterogeneity. In the case of the Netherlands, we find that both intensity measures exhibit a statistically significant positive impact on LP for large firms and when interacting the intensity measure with the years 2018 and 2020. In the German data, we also find this statistically significant positive effect for both measures for large firms. However, interacting intensity measures with the year of observation hardly yields any statistically significant effects.

Table 5.6: Summary of the Results: Big Data Analytics Measured as Intensity

	NL BDA 3 types	**DE** BDA 3 types	**NL** BDA intensity std.	**DE** BDA intensity std.
All (controlling for level of digitalisation)	✓	✓	✓	✓
Year	2018, 2020	–	2018, 2020	2016
Firm size	large	large	large	large

NOTE: ✓ means a statistically significant finding, – means no significant finding. If only single years or size classes are indicated, these are the only significant estimates.

6 Discussion

The results obtained in the course of our analysis serve as first evidence for the productivity-enhancing effect of BDA at the firm level. However, they have to be interpreted carefully. First, our OLS regressions may be subject to endogeneity issues. This is despite our efforts to control for industry and year fixed effects, other potential confounders, and interaction effects, as well as the use of alternative measures for our variable of interest (BDA). In particular, we cannot control for total factor productivity, which is, thus, contained in the error term and may affect both LP and our BDA-measure as well as other input factors, possibly introducing an omitted variable bias (OVB). Especially in the case of Germany, the limited panel structure restricts our ability to apply more advanced regression techniques. This prevents us from estimating productivity using methods proposed by Ackerberg et al. (2015), Levinsohn and Petrin (2003), and Olley and Pakes (1992). Further analyses may be hampered by the shift of focus of the "Survey on ICT Usage and E-Commerce in Enterprises" in subsequent waves from BDA towards the related topic of AI. Although BDA and AI show complementarities in their deployment (Cho et al., 2023), being capable of handling large amounts of data can be seen as a prerequisite for technologies, such as AI, that have recently raised public interest. This critically necessitates a profound understanding of how BDA disseminates and how it affects individual firms as well as the economy as a whole.

Second, due to the binary nature of the BDA items in the "Survey on ICT Usage and E-Commerce in Enterprises", the data we use do not provide us with information regarding the intensity of use of BDA within a firm. They only tell us whether a firm uses any kind of BDA, irrespective of the extent. We try to overcome this limitation in the data by constructing a measure of BDA intensity that takes the variety of BDA input data as a proxy for how established and widespread BDA use is within a firm.

In order to at least mitigate potential endogeneity issues, we also applied various fixed-effects (FE) regressions. However, these were most of the times (NL) respectively always (DE) insignificant. As this might partly be driven by the short and unbalanced panel structure, we refrain from drawing any definite conclusion from the FE results.

With respect to the generalisability of the results, we find it particularly interesting that despite the considerable differences between Germany and the Netherlands with respect to the size of the economy, the industry structure, and the general digitalisation level, we observe very similar results. Furthermore, existing evidence for Portugal largely confirms our findings for Germany and the Netherlands. In line with our results that the impact of BDA varies along various dimensions - i.a. firm size, time, and intensity of use -, Conti et al. (2024) find that BDA only has a positive impact on firm performance in terms of value added for larger firms. This makes us tentatively confident regarding the external validity of our results.

7 Conclusion

This paper examines the relationship between big data analytics (BDA) and labour productivity (LP) using firm-level data from Dutch and German statistical offices. Applying pooled OLS regressions, we find a statistically significant and positive relationship between BDA (measured as a binary variable) and productivity. After controlling for the general digitalisation level of the firm, the effect vanishes. However, we observe substantial heterogeneity across industries, firm sizes, and years. For example, we still find positive coefficients for BDA (after controlling for the general digitalisation level) for large firms and the year 2020 (only for the Netherlands).

As a binary measure for the adoption of BDA might insufficiently reflect the real benefits of exploiting data, we extend our analysis to various intensity measures for BDA. For the Netherlands, we see a significant association between BDA intensity and LP. For Germany, this relationship is less pronounced but still measurable.

Future research should examine the external validity of our findings by testing whether this relationship also holds in other settings. Moreover, having panel data on a higher frequency over a longer period of time would benefit the credibility of the analysis.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). "Identification properties of recent production function estimators". In: *Econometrica* 83.6, pp. 2411–2451.

Anderton, R., V. Jarvis, V. Labhard, J. Morgan, F. Petroulakis, and L. Vivian (2020). *Virtually everywhere? Digitalisation and the euro area and EU economies: Degree, effects, and key issues.* ECB Occasional Paper Series 244. European Central Bank.

Andrews, D., C. Criscuolo, and P. N. Gal (2019). *The best versus the rest: divergence across firms during the global productivity slowdown.* CEP Discussion Papers dp1645. Centre for Economic Performance, LSE. URL: `https://ideas.repec.org/p/cep/cepdps/dp1645.html`.

Andrews, D., G. Nicoletti, and C. Timiliotis (2018). *Digital technology diffusion: A matter of capabilities, incentives or both?* OECD Economics Department Working Papers 1476. OECD. URL: `https://doi.org/10.1787/7c542c16-en`.

Bar-Gill, S., E. Brynjolfsson, and N. Hak (2024). "Helping small businesses become more data-driven: A field experiment on eBay". In: *Management Science.* URL: `https://doi.org/10.1287/mnsc.2021.02026`.

Bertschek, I., W. Briglauer, K. Hüschelrath, B. Kauf, and T. Niebel (2015). "The economic impacts of broadband internet: A survey". In: *Review of Network Economics* 14.4, pp. 201–227. URL: `https://doi.org/10.1515/rne-2016-0032`.

Bertschek, I. and R. Kesler (2022). "Let the user speak: Is feedback on Facebook a source of firms' innovation?" In: *Information Economics and Policy* 60, p. 100991.

Biagi, F. (2013). *ICT and productivity: A review of the literature.* Institute for Prospective Technological Studies Digital Economy Working Paper 2013/09. European Commission, Joint Research Centre (JRC). URL: `https://ideas.repec.org/p/ipt/decwpa/2013-09.html`.

Bloom, N., R. Sadun, and J. Van Reenen (2012). "Americans do IT better: US multinationals and the productivity miracle". In: *American Economic Review* 102.1, pp. 167–201.

Bontadini, F., C. Corrado, J. Haskel, M. Iommi, and C. Jona-Lasinio (2023). *EUKLEMS & INTANProd: industry productivity accounts with intangibles - Sources of growth and productivity trends: methods and main measurement challenges.* URL: `https://euklems-intanprod-llee.luiss.it/wp-content/uploads/2023/02/EUKLEMS_INTANProd_D2.3.1.pdf`.

Borowiecki, M., J. Pareliussen, D. Glocker, E. J. Kim, M. Polder, and I. Rud (2021). *The impact of digitalisation on productivity: Firm-level evidence from the Netherlands*. OECD Economics Department Working Papers 1680. OECD. URL: https://doi.org/10.1787/e800ee1d-en.

Bresnahan, T. F. and M. Trajtenberg (1995). "General purpose technologies 'engines of growth'?" In: *Journal of Econometrics* 65.1, pp. 83–108.

Brynjolfsson, E. and L. M. Hitt (2000). "Beyond computation: Information technology, organizational transformation and business performance". In: *The Journal of Economic Perspectives* 14.4, pp. 23–48.

Brynjolfsson, E., W. Jin, and K. McElheran (2021a). "The power of prediction: predictive analytics, workplace complements, and business performance". In: *Business Economics* 56, pp. 217–239.

Brynjolfsson, E., D. Li, and L. R. Raymond (2023). *Generative AI at work*. Working Paper 31161. National Bureau of Economic Research. URL: http://www.nber.org/papers/w31161.

Brynjolfsson, E. and A. McAfee (2011). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Lexington, Mass: Digital Frontier Press.

Brynjolfsson, E. and K. McElheran (2016). "The rapid adoption of data-driven decision-making". In: *American Economic Review* 106.5, pp. 133–139.

— (2019). *Data in action: data-driven decision making and predictive analytics in US manufacturing*. Rotman School of Management Working Paper 3422397. Rotman School of Management.

Brynjolfsson, E., D. Rock, and C. Syverson (2021b). "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies". In: *American Economic Journal: Macroeconomics* 13.1, pp. 333–72.

Calvino, F. and L. Fontanelli (2023). *A portrait of AI adopters across countries*. OECD Science, Technology and Industry Working Papers 2023/02. OECD Publishing. URL: https://doi.org/10.1787/0fb79bb9-en.

Cardona, M., T. Kretschmer, and T. Strobel (2013). "ICT and productivity: Conclusions from the empirical literature". In: *Information Economics and Policy* 25.3, pp. 109–125.

Cerqueira, C., F. Alexandre, and M. Portela (2023). *Digitalization: the edge of first movers*. NIPE Working Papers 11/2023. NIPE-Universidade do Minho. URL: https://ideas.repec.org/p/nip/nipewp/11-2023.html.

Cho, J., T. DeStefano, H. Kim, I. Kim, and J. H. Paik (2023). "What's driving the diffusion of next-generation digital technologies?" In: *Technovation* 119, p. 102477.

Cirillo, V., L. Fanti, A. Mina, and A. Ricci (2023). "The adoption of digital technologies: Investment, skills, work organisation". In: *Structural Change and Economic Dynamics* 66, pp. 89–105.

Conti, R., M. G. de Matos, and G. Valentini (2024). "Big data analytics, firm size, and performance". In: *Strategy Science* 9.2, pp. 135–151.

Crafts, N. (2018). "The productivity slowdown: Is it the 'New Normal'?" In: *Oxford Review of Economic Policy* 34.3, pp. 443–460.

Crouzet, N. and J. Eberly (2018). "Intangibles, investment, and efficiency". In: *AEA Papers and Proceedings* 108, pp. 426–31.

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

DeStefano, T., R. Kneller, and J. Timmis (2023). "Cloud computing and firm growth". In: *The Review of Economics and Statistics*, pp. 1–47. URL: https://doi.org/10.1162/rest_a_01393.

Dhyne, E., J. Konings, J. Van den bosch, and S. Vanormelingen (2021). "The return on information technology: Who benefits most?" In: *Information Systems Research* 32.1, pp. 194–211.

Diamond, P. A. (1971). "A model of price adjustment". In: *Journal of economic theory* 3.2, pp. 156–168.

Duso, T., M. Nardotto, and A. Schiersch (2021). *Broadband and productivity: Structural estimates for Germany.* Discussion Papers of DIW Berlin 1988. Berlin: DIW Berlin, German Institute for Economic Research. URL: https://ideas.repec.org/p/diw/diwwpp/dp1988.html.

Duso, T. and A. Schiersch (2022). *Let's switch to the cloud: Cloud adaption and its effect on IT investment and productivity.* Discussion Papers of DIW Berlin 2017. DIW Berlin, German Institute for Economic Research. URL: https://ideas.repec.org/p/diw/diwwpp/dp2017.html.

Gal, P., G. Nicoletti, T. Renault, S. Sorbe, and C. Timiliotis (2019). *Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries.* OECD Economics Department Working Papers 1533. OECD. URL: https://doi.org/10.1787/5080f4b6-en.

Gierten, D., S. Viete, R. Andres, and T. Niebel (2021). *Firms going digital: Tapping into the potential of data for innovation.* OECD Digital Economy Papers 320. OECD. URL: https://doi.org/10.1787/ee8340c1-en.

Goldfarb, A. and C. Tucker (2019). “Digital economics”. In: *Journal of economic literature* 57.1, pp. 3–43.

Goldin, I., P. Koutroumpis, F. Lafond, and J. Winkler (2024). “Why is productivity slowing down?” In: *Journal of Economic Literature* 62.1, pp. 196–268.

Grashof, N. and A. Kopka (2023). “Artificial intelligence and radical innovation: an opportunity for all companies?” In: *Small Business Economics* 61.2, pp. 771–797.

Hall, B. H. (2011). *Innovation and productivity*. Working Paper 17178. National Bureau of Economic Research. URL: `http://www.nber.org/papers/w17178`.

Haller, S. A. and I. Siedschlag (2011). “Determinants of ICT adoption: Evidence from firm-level data”. In: *Applied Economics* 43.26, pp. 3775–3788.

Jorgenson, D. W. (2005). “Accounting for growth in the information age”. In: *Handbook of Economic Growth*. Ed. by Philippe Aghion and Steven N. Durlauf. Vol. 1, Part A. Elsevier, pp. 743–815.

Levinsohn, J. and A. Petrin (2003). “Estimating production functions using inputs to control for unobservables”. In: *The review of economic studies* 70.2, pp. 317–341.

McElheran, K., J. F. Li, E. Brynjolfsson, Z. Kroff, E. Dinlersoz, L. Foster, and N. Zolas (2024). “AI adoption in America: Who, what, and where”. In: *Journal of Economics & Management Strategy* 33.2, pp. 375–415. URL: `https://doi.org/10.1111/jems.12576`.

Mohnen, P., M. Polder, and G. Van Leeuwen (2019). “ICT, R&D, and organizational innovation: exploring complementarities in investment and production”. In: *Measuring and Accounting for Innovation in the 21st Century*. University of Chicago Press.

Müller, O., M. Fay, and J. vom Brocke (2018). “The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics”. In: *Journal of Management Information Systems* 35.2, pp. 488–509.

Niebel, T., F. Rasel, and S. Viete (2019). “BIG data – BIG gains? Understanding the link between big data analytics and innovation”. In: *Economics of Innovation and New Technology* 28.3, pp. 296–316.

Olley, G. S. and A. Pakes (1992). *The dynamics of productivity in the telecommunications equipment industry*. Working Paper 3977. National Bureau of Economic Research. URL: `http://www.nber.org/papers/w3977`.

Rammer, C., G. P. Fernández, and D. Czarnitzki (2022). “Artificial intelligence and industrial innovation: Evidence from German firm-level data”. In: *Research Policy* 51.7, p. 104555.

Rasel, F. (2016). “Combining information technology and decentralized workplace organization: SMEs versus larger firms”. In: *International Journal of the Economics of Business* 23.2, pp. 199–241.

Rodepeter, E., C. Gschnaidtner, and H. Hottenrott (2023). *Big data and start-up performance.* ZEW Discussion Papers 23-061. ZEW - Leibniz Centre for European Economic Research. URL: `https://ideas.repec.org/p/zbw/zewdip/283582.html`.

Rosenberg, N. (2006). “Innovation and Economic Growth”. In: *Innovation and Growth in Tourism.* OECD, pp. 43–52. URL: `https://doi.org/10.1787/9789264025028-4-en`.

Solow, R. M. (1987). “We’d better watch out”. In: *New York Times Book Review.* July 12, 1987, p. 36.

Stigler, G. J. (1961). “The economics of information”. In: *Journal of political economy* 69.3, pp. 213–225.

Van Ark, B. (2016). “The productivity paradox of the new digital economy”. In: *International Productivity Monitor* 31, pp. 3–18.

Varian, H. R. (1980). “A model of sales”. In: *The American economic review* 70.4, pp. 651–659.

Wu, L., L. Hitt, and B. Lou (2020). “Data analytics, innovation, and firm productivity”. In: *Management Science* 66.5, pp. 2017–2039.

Wu, L., B. Lou, and L. Hitt (2019). “Data analytics supports decentralized innovation”. In: *Management Science* 65.10, pp. 4863–4877.

A Appendix

A.1 Calculation of Capital Stocks

Capital stocks at the firm level are constructed following the perpetual inventory method (PIM) (Equation 6) and are based on the deflated gross investments by the individual firms. The PIM approach follows Dhyne et al. (2021)

$$K_{it} = (1 - \delta_t) * K_{it-1} + i_{it}. \tag{6}$$

K_{it} refers to the real capital stock of firm i in period t, i_{it} to the real gross investments of i in t and δ to the depreciation rate for capital. Nominal values are deflated using the information from the EU-KLEMS data. In particular, we use information on the gross fixed capital formation deflators by year at the two-digit industry level. Besides, the EU-KLEMS data also provide information on yearly depreciation rates.

As shown in Equation 6, each period t's capital stocks are calculated by depreciating the previous period's $(t - 1)$ capital stocks using the corresponding depreciation rate for capital δ_t and adding the gross investments of the respective time period t.

A.2 Standardisation of the BDA-Intensity Variable

In Sections 5.1.2 and 5.2.2 we use different approaches for measuring the intensity of BDA-usage. First, we use a categorical variable *BDA intensity* ranging from 0 to 4, with a value of 0 meaning that a firm that does not use any kind of *BDA* and 4 for that a firm uses all 4 types of big data sources. Second, the binary variable *BDA all* equals 1 if all 4 of the big data sources are in use. Third, we created a binary variable now indicating whether a firm uses at least 3 of the big data sources. Lastly, we add a standardised *BDA*-measure *BDA intensity std* similar to e.g. Rasel (2016):

$$BDA\ intensity\ std = S\Big(S(BDA_{own}) + S(BDA_{geo}) + S(BDA_{social}) + S(BDA_{other})\Big) \tag{7}$$

$$with\ \ S = \frac{x - \mu_x}{SD_x} \tag{8}$$

A.3 Detailed Variable Description

Figure A.1: Definition of Big Data Analytics in the Eurostat ICT Survey 2018

Module G: Big data analysis

(Scope: enterprises with computers)
- Optional

Big data are generated from activities that are carried out electronically and from machine-to-machine communications (e.g. data produced from social media activities, from production processes, etc.)

Big data typically have characteristics such as:

- Significant **volume** referring to vast amounts of data generated over time.

- **Variety** referring to the different format of complex data, either structured or unstructured (e.g. text, video, images, voice, docs, sensor data, activity logs, click streams, coordinates, etc.).

- **Velocity** referring to the high speed at which data is generated, becomes available and changes over time.

Big data analysis refers to the use of techniques, technologies and software tools for analysing **big data** extracted from your own enterprise's data sources or other data sources.

G1. *[21]	**During 2017, did your enterprise analyse big data from any of the following data sources?** *(Please refer to the definition of big data above; include big data analysis conducted by external service providers)* ***- Optional***	Yes	No
	a) Enterprise's own data from smart devices or sensors (e.g. Machine to Machine -M2M- communications, digital sensors, Radio frequency identification tags RFID[22], etc.) *(in the context of big data)*	☐	☐
	b) Geolocation data from the use of portable devices (e.g. portable devices using mobile telephone networks, wireless connections or GPS) *(in the context of big data)*	☐	☐
	c) Data generated from social media (e.g. social networks, blogs, multimedia content sharing websites, etc.) *(in the context of big data)*	☐	☐
	d) Other big data sources not specified above	☐	☐

Source: Eurostat Community survey on ICT usage and E-commerce in Enterprises. https://circabc.europa.eu/ui/group/89577311-0f9b-4fc0-b8c2-2aaa7d3ccb91/library/b44e7e01-e75e-4ab5-be1c-dae7a1d80ef2?p=1&n=-1&sort=name_DESC.

Table A.1: List of variables

Variable Name	Description
BDA own	BDA - own data
BDA geo	BDA - geolocation data
BDA social	BDA - social media data
BDA other	BDA - other data
BDA	Big Data Analysis (own or geo or social or other)
BDA intensity	BDA intensity: Sum of all BDA types
BDA all	BDA intensity: Firm has all BDA types
BDA 3 types	BDA intensity: Firm has at least 3 BDA types
BDA intensity std	BDA intensity: Standardize sum of all BDA types
Sh_internet	Share employees having computer with internet
VA	Real value added
L	Labour - FTE
K	Tangible capital stock
LP	Real value added/labour
ln(LP)	ln(labour productivity)
ln(K)	ln(tangible capital stock)
ln(L)	ln(labour - FTE)
Small firm	Small firm (< 50 FTE)
Medium firm	Medium firm FTE >= 50 & FTE < 250
Large firm	Large firm (> 250 FTE)
Manufacturing	Manufacturing sector (10-33)
Services	Services sector (45-95 with gaps)

NOTE: See Figure A.1 for a detailed definition of the BDA variable.

A.4 Additional Tables: Netherlands

Table A.2: Number of Observations by Industry Group and respective NACE 2 Codes for the Dutch Estimation Sample

	N	**Percentage**	**NACE Codes**
Food/Beverages	606	2.41	10, 11
Textiles/Clothing	234	0.93	13, 14, 15
Wood/Paper	403	1.60	16, 17
Chemicals/Pharmaceuticals	500	1.99	20, 21
Rubber/Plastics	370	1.47	22
Glass/Ceramics/Concrete	288	1.15	23
Metals	830	3.30	24, 25
Machinery/Equipment	998	3.97	28, 33
Electronics/Electrical	505	2.01	26, 27
Vehicles	482	1.92	29, 30
Furniture/Other Manufacturing	611	2.43	31, 32
Energy/Oil	248	0.99	19, 35
Water Supply/Waste/Recycling	340	1.35	36, 37, 38, 39
Construction	1482	5.90	41, 42, 43
Trade	5134	20.44	45, 46, 47
Transportation/Postal Services	2140	8.52	49, 50, 51, 52, 53, 79
Accommodation	702	2.79	55, 56
Printing/Publishing/Media	567	2.26	18, 58, 59, 60
IT-Services/Telecommunications	2070	8.24	61, 62, 63
Real Estate	172	0.68	68
Consulting/Advertising	1891	7.53	69, 70, 73, 75, 77
Technical Engineering/R&D	1158	4.61	71, 72
Other Producer Services	3391	13.50	74, 78, 80, 81, 82, 95
Total	25122	100.00	

NOTE: These industry groups are included as dummy variables in our regression analysis.

Table A.3: Summary Statistics Dutch Estimation Sample: No BDA vs BDA

	No BDA			BDA		
	N	Mean	Median	N	Mean	Median
BDA own	16978	0	0	8144	.51	1
BDA geo	16978	0	0	8144	.37	0
BDA social	16978	0	0	8144	.52	1
BDA other	16978	0	0	8144	.39	0
BDA	16978	0	0	8144	1	1
BDA intensity	16978	0	0	8144	1.8	2
BDA all	16978	0	0	8144	.076	0
BDA 3 types	16978	0	0	8144	.21	0
BDA intensity std	16978	-.58	-.58	8144	1.2	1.3
BDA int. == 0	16978	1	1	8144	0	0
BDA int. == 1	16978	0	0	8144	.5	0
BDA int. == 2	16978	0	0	8144	.29	0
BDA int. == 3	16978	0	0	8144	.14	0
BDA int. == 4	16978	0	0	8144	.076	0
Sh_internet	16080	67	80	7825	76	90
VA	16978	9,627,990	3,284,053	8144	31,233,746	7,315,336
L	16978	125	48	8144	309	105
K	16978	10,364,262	1,029,120	8144	64,691,925	2,677,297
LP	16978	90,935	65,425	8144	96,006	69,096
Small firm	16978	.51	1	8144	.32	0
Medium firm	16978	.38	0	8144	.43	0
Large firm	16978	.11	0	8144	.26	0
Manufacturing	16978	.26	0	8144	.19	0
Services	16978	.66	1	8144	.72	1

NOTE: See Table A.1 for a detailed description of the variables.

Table A.4: OLS Regressions Netherlands: BDA Types

	(1) All	(2) Own	(3) Geo	(4) Social	(5) Other
BDA=1	0.0143				
	(0.009)				
BDA own=1		0.0318***			
		(0.011)			
BDA geo=1			0.0017		
			(0.012)		
BDA social=1				0.0145	
				(0.011)	
BDA other=1					0.0560***
					(0.012)
ln(L)	-0.1929***	-0.1932***	-0.1920***	-0.1925***	-0.1946***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(K)	0.1746***	0.1740***	0.1750***	0.1749***	0.1746***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Sh_internet	0.0048***	0.0048***	0.0048***	0.0048***	0.0048***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	9.0089***	9.0181***	9.0027***	9.0052***	9.0183***
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
Year DVs	Yes	Yes	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.218	0.218	0.218	0.218	0.218
Observations	23905	23905	23905	23905	23905

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.5: OLS Regressions Netherlands: Split Samples by Year and Firm Size Class

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	2016	2017	2018	2020	All	Small	Medium	Large	SME
BDA=1	-0.0092	0.0087	0.0007	0.0371**	0.0143	-0.0138	0.0098	0.0425**	0.0018
	(0.017)	(0.019)	(0.019)	(0.016)	(0.009)	(0.017)	(0.012)	(0.017)	(0.010)
ln(L)	-0.1826***	-0.2031***	-0.1936***	-0.2028***	-0.1929***	-0.1813***	-0.1594***	-0.1587***	-0.1923***
	(0.011)	(0.011)	(0.010)	(0.010)	(0.005)	(0.013)	(0.014)	(0.015)	(0.007)
ln(K)	0.1663***	0.1706***	0.1858***	0.1783***	0.1746***	0.1626***	0.1632***	0.1512***	0.1697***
	(0.008)	(0.008)	(0.008)	(0.007)	(0.004)	(0.007)	(0.006)	(0.007)	(0.004)
Sh_internet	0.0052***	0.0059***	0.0048***	0.0040***	0.0048***	0.0029***	0.0052***	0.0060***	0.0042***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	9.1434***	9.0455***	8.7870***	8.9151***	9.0089***	9.1304***	8.9917***	9.2213***	9.0631***
	(0.101)	(0.107)	(0.104)	(0.093)	(0.051)	(0.101)	(0.091)	(0.129)	(0.060)
Year DVs	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5568	4671	5808	7858	23905	10286	9716	3903	20002

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.6: OLS Regressions Netherlands: BDA Interaction with Size Classes or Years

	(1)	(2)	(3)
BDA=1	0.0143	-0.0022	-0.0287*
	(0.009)	(0.017)	(0.017)
ln(L)	-0.1929***	-0.1929***	-0.2015***
	(0.005)	(0.005)	(0.009)
ln(K)	0.1746***	0.1747***	0.1729***
	(0.004)	(0.004)	(0.004)
Sh_internet	0.0048***	0.0048***	0.0048***
	(0.000)	(0.000)	(0.000)
BDA=1 X Year=2017		0.0019	
		(0.025)	
BDA=1 X Year=2018		0.0103	
		(0.024)	
BDA=1 X Year=2020		0.0391*	
		(0.023)	
Medium			-0.0061
			(0.016)
Large			0.0087
			(0.029)
BDA=1 X Medium			0.0473**
			(0.020)
BDA=1 X Large			0.1111***
			(0.024)
Constant	9.0089***	9.0115***	9.0671***
	(0.051)	(0.051)	(0.055)
Year DVs	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes
Observations	23905	23905	23905
Adjusted R-squared	0.218	0.218	0.219

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

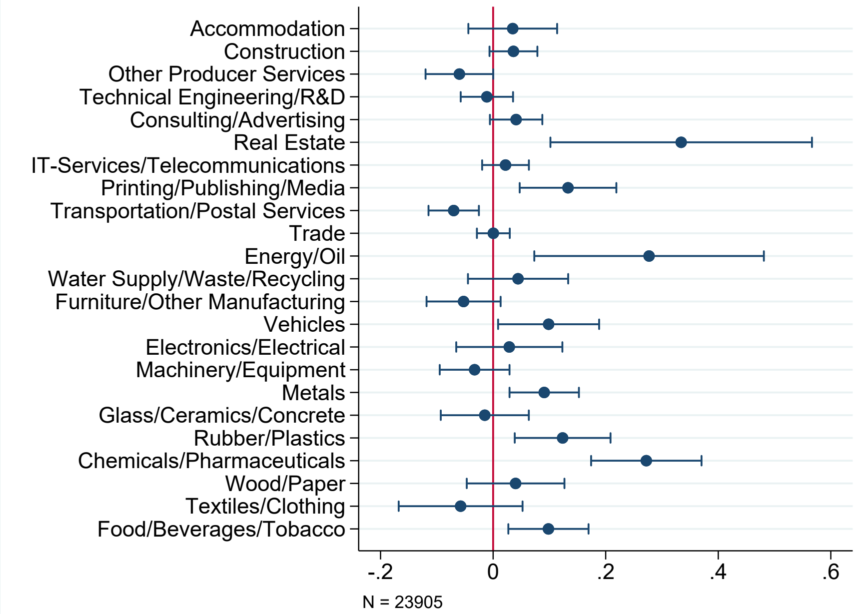
Table A.7: OLS Regressions Netherlands: BDA Intensity Interaction with Size Classes or Years

	(1)	(2)	(3)	(4)
BDA 3 types=1	0.0520	0.0515		
	(0.037)	(0.037)		
BDA 3 types=1 X Year=2017	-0.0244			
	(0.049)			
BDA 3 types=1 X Year=2018	0.0645			
	(0.050)			
BDA 3 types=1 X Year=2020	0.0191			
	(0.045)			
BDA 3 types=1 X medium		-0.0319		
		(0.043)		
BDA 3 types=1 X large		0.0841*		
		(0.045)		
BDA intensity std			0.0047	-0.0081
			(0.009)	(0.009)
Year=2017 X BDA intensity std			0.0022	
			(0.012)	
Year=2018 X BDA intensity std			0.0131	
			(0.012)	
Year=2020 X BDA intensity std			0.0134	
			(0.011)	
Medium X BDA intensity std				0.0148
				(0.010)
Large X BDA intensity std				0.0528***
				(0.011)
Medium		0.0104		0.0121
		(0.016)		(0.016)
Large		0.0447		0.0455
		(0.028)		(0.028)
Constant	9.0218***	9.0635***	9.0249***	9.0726***
	(0.051)	(0.055)	(0.051)	(0.055)
All controls	Yes	Yes	Yes	Yes
Observations	23905	23905	23905	23905
Adjusted R-squared	0.218	0.219	0.218	0.219

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

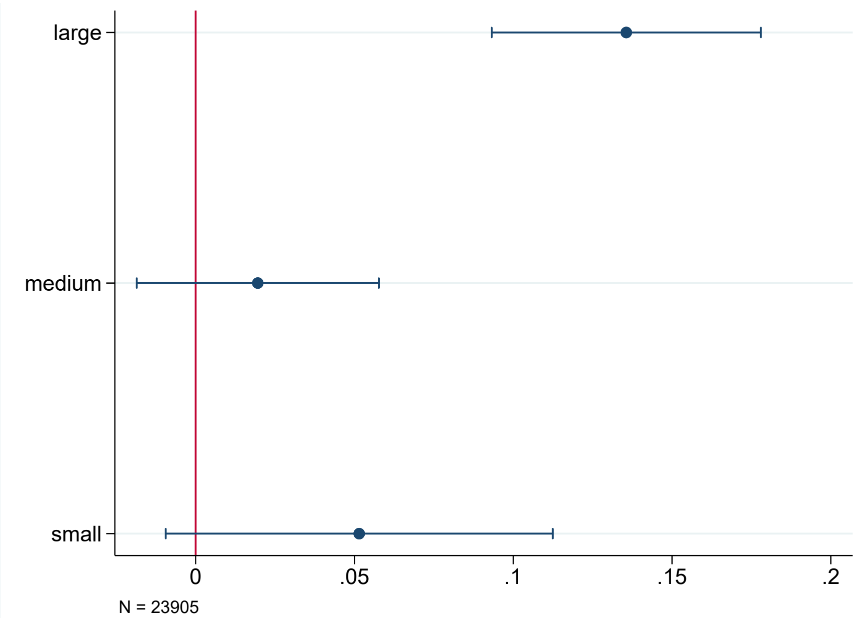
A.5 Additional Graphs: Netherlands

Figure A.2: Big Data Analytics by Industry



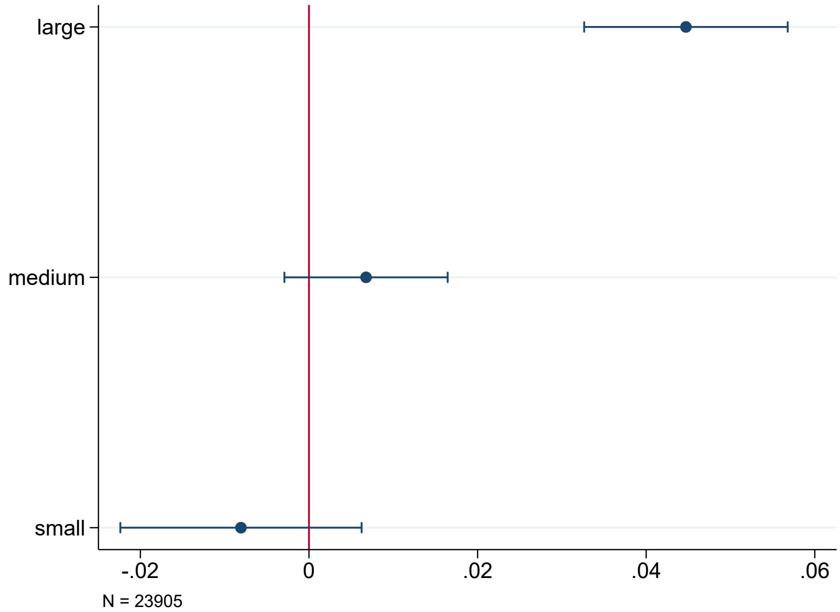
Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* when including both *BDA* and its interaction with the respective *industry*-dummy in the regression (see Table A.2 for the industry distribution).

Figure A.3: Big Data Analytics Intensity (Using at least 3 BDA Types) by Firm Size



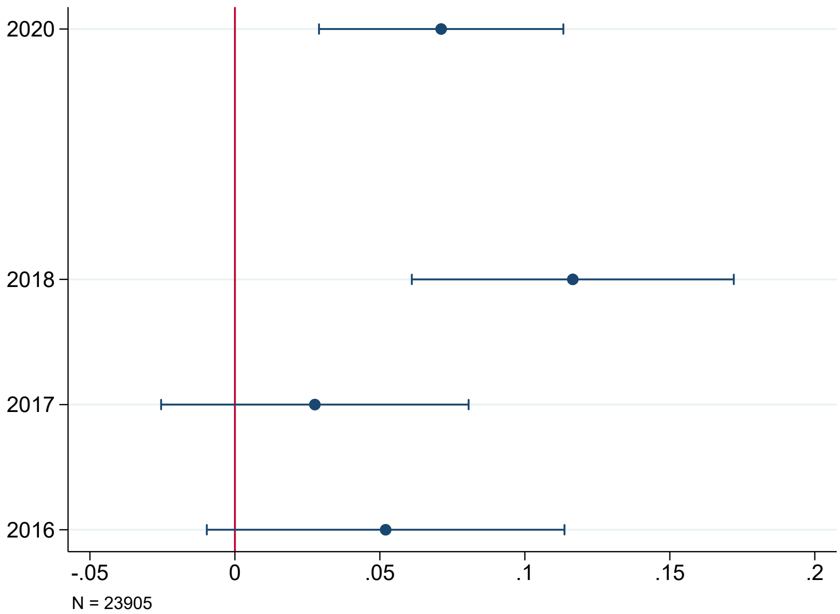
Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective *firm size class*-dummy in the regression. Also see Table A.7 for exact regression results.

Figure A.4: Big Data Analytics Intensity (Standardised Measure) by Firm Size



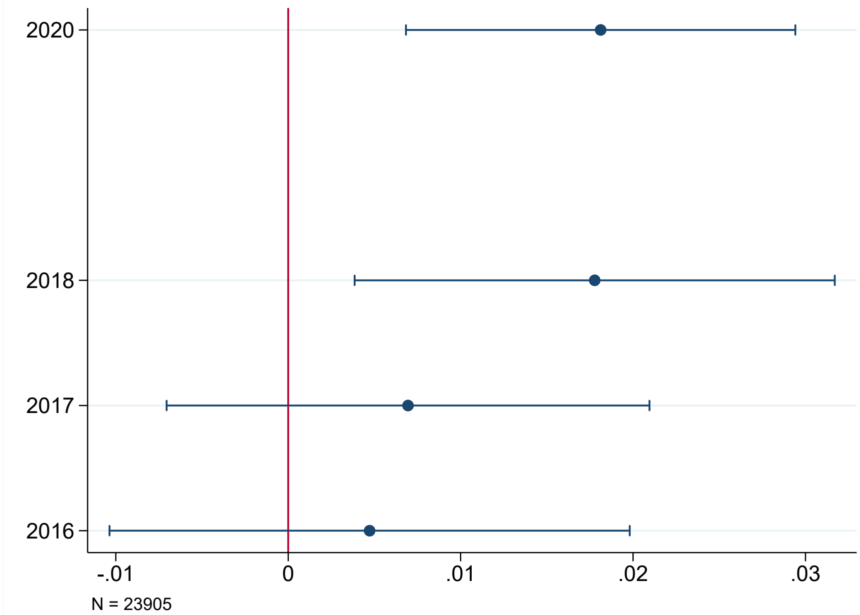
Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective *firm size class*-dummy in the regression. Also see Table A.7 for exact regression results.

Figure A.5: Big Data Analytics Intensity (Using at least 3 BDA Types) by Year



Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective survey *year*-dummy in the regression. Also see Table A.7 for exact regression results.

Figure A.6: Big Data Analytics Intensity (Standardised Measure) by Year



Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective survey *year*-dummy in the regression. Also see Table A.7 for exact regression results.

A.6 Additional Tables: Germany

Table A.8: Number of Observations by Industry Group and respective NACE 2 Codes for the German Estimation Sample

	N	**Percentage**	**NACE Codes**
Food/Beverages/Tobacco	676	7.85	10, 11, 12
Textiles/Clothing	374	4.34	13, 14, 15
Wood/Paper	335	3.89	16, 17
Oil	17	0.20	19
Chemicals/Pharmaceuticals	284	3.30	20, 21
Rubber/Plastics	263	3.05	22
Glass/Ceramics/Concrete	198	2.30	23
Metals	667	7.75	24, 25
Electronics/Electrical	529	6.14	26, 27
Machinery/Equipment	684	7.94	28, 33
Vehicles	513	5.96	29, 30
Furniture/Other Manufacturing	278	3.23	31, 32
Transportation/Postal Services	1014	11.77	49, 50, 51, 52, 53, 79
Printing/Publishing/Media	573	6.65	18, 58, 59, 60
IT-Services/Telecommunications	798	9.27	61, 62, 63
Consulting/Advertising	379	4.40	69, 70, 73, 75, 77
Technical Engineering/R&D	300	3.48	71, 72
Other Producer Services	730	8.48	74, 78, 80, 81, 82
Total	8612	100.00	

Note: These industry groups are included as dummy variables in our regression analysis.

Table A.9: Summary Statistics German Estimation Sample: No BDA vs BDA

	No BDA			**BDA**		
	N	Mean	Median	N	Mean	Median
BDA own	6735	0	0	1859	.39	0
BDA geo	6695	0	0	1856	.41	0
BDA social	6690	0	0	1857	.45	0
BDA other	6657	0	0	1845	.3	0
BDA	6739	0	0	1873	1	1
BDA intensity	6739	0	0	1873	1.5	1
BDA all	6739	0	0	1873	.038	0
BDA 3 types	6739	0	0	1873	.12	0
BDA intensity std	6739	-.39	-.39	1873	2	1.2
bd_int== 0	6739	1	1	1873	0	0
bd_int== 1	6739	0	0	1873	.62	1
bd_int== 2	6739	0	0	1873	.26	0
bd_int== 3	6739	0	0	1873	.081	0
bd_int== 4	6739	0	0	1873	.038	0
Sh_internet	6739	55	50	1873	69	80
VA	6739	19,835,483	5,890,682	1873	82,661,497	11,356,903
L	6739	254	109	1873	794	209
K	6739	19,586,731	2,855,043	1873	125010722	5,955,122
LP	6739	73,181	53,983	1873	83,635	61,845
Small firm	6739	.24	0	1873	.16	0
Medium firm	6739	.46	0	1873	.39	0
Large firm	6739	.3	0	1873	.46	0
Manufacturing	6739	.6	1	1873	.48	0
Services	6739	.4	0	1873	.52	1

NOTE: See Table A.1 for a detailed description of the variables.

Table A.10: OLS Regressions Germany: BDA Types

	(1) All	(2) Own	(3) Geo	(4) Social	(5) Other
BDA=1	0.0240				
	(0.020)				
BDA own=1		0.0806***			
		(0.026)			
BDA geo=1			-0.0319		
			(0.030)		
BDA social=1				0.0114	
				(0.030)	
BDA other=1					0.1140***
					(0.034)
ln(L)	-0.0469***	-0.0489***	-0.0446***	-0.0466***	-0.0512***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
ln(K)	0.0678***	0.0673***	0.0677***	0.0680***	0.0680***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Sh_internet	0.0058***	0.0058***	0.0058***	0.0058***	0.0057***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	9.8143***	9.8318***	9.8069***	9.8098***	9.8373***
	(0.063)	(0.062)	(0.062)	(0.062)	(0.062)
Year DVs	Yes	Yes	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.283	0.283	0.283	0.283	0.285
Observations	8612	8594	8551	8547	8502

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.11: OLS Regressions Germany: Split Samples by Year and Firm Size Class

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2016	2018	2020	All	Small	Medium	Large	SME
BDA=1	0.0545	0.0395	-0.0019	0.0240	0.0029	0.0062	0.0297	0.0072
	(0.060)	(0.028)	(0.033)	(0.020)	(0.051)	(0.035)	(0.028)	(0.028)
ln(L)	-0.0775***	-0.0530***	-0.0212	-0.0469***	-0.1373*	0.0005	-0.0830***	-0.0135
	(0.019)	(0.015)	(0.018)	(0.010)	(0.074)	(0.025)	(0.023)	(0.018)
ln(K)	0.0936***	0.0651***	0.0565***	0.0678***	0.0488***	0.0541***	0.1011***	0.0534***
	(0.012)	(0.007)	(0.008)	(0.005)	(0.008)	(0.007)	(0.012)	(0.005)
Sh_internet	0.0069***	0.0060***	0.0046***	0.0058***	0.0038***	0.0054***	0.0065***	0.0048***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Constant	9.5503***	9.8180***	9.7438***	9.8143***	10.4639***	9.7672***	9.4895***	9.8811***
	(0.149)	(0.084)	(0.100)	(0.063)	(0.251)	(0.143)	(0.163)	(0.086)
Year DVs	No	No	No	Yes	Yes	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2106	3493	3013	8612	1908	3825	2879	5733

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.12: OLS Regressions Germany: BDA Interaction with Size Classes or Years

	(1)	(2)	(3)
BDA=1	0.0240	0.0823	-0.0393
	(0.020)	(0.059)	(0.049)
ln(L)	-0.0469***	-0.0473***	-0.0589***
	(0.010)	(0.010)	(0.018)
ln(K)	0.0678***	0.0677***	0.0676***
	(0.005)	(0.005)	(0.005)
Sh_internet	0.0058***	0.0058***	0.0058***
	(0.000)	(0.000)	(0.000)
BDA=1 X Year=2018		-0.0499	
		(0.063)	
BDA=1 X Year=2020		-0.0919	
		(0.066)	
Medium			0.0324
			(0.029)
Large			0.0363
			(0.050)
BDA=1 X Medium			0.0737
			(0.060)
BDA=1 X Large			0.0823
			(0.056)
Constant	9.8143***	9.8082***	9.8457***
	(0.063)	(0.062)	(0.079)
Year DVs	Yes	Yes	Yes
Industry DVs	Yes	Yes	Yes
Observations	8612	8612	8612
Adjusted R-squared	0.283	0.283	0.283

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

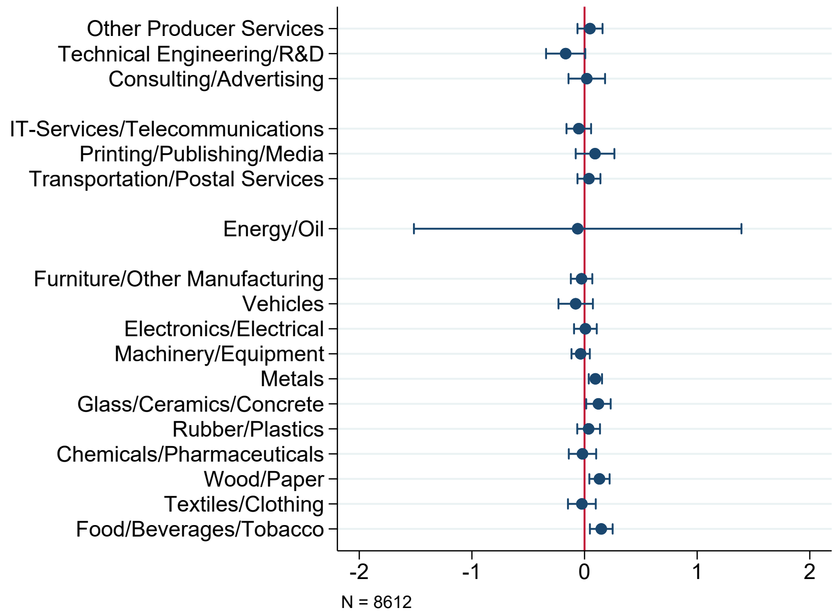
Table A.13: OLS Regressions Germany: BDA Intensity Interaction with Size Classes or Years

	(1)	(2)	(3)	(4)
BDA 3 types=1	0.1607	-0.2346*		
	(0.099)	(0.143)		
BDA 3 types=1 X Year = 2018	-0.1210			
	(0.121)			
BDA 3 types=1 X Year = 2020	-0.0528			
	(0.122)			
BDA 3 types=1 X Medium		0.2749*		
		(0.160)		
BDA 3 types=1 X Large		0.4125***		
		(0.155)		
BDA intensity std			0.0308*	-0.0272
			(0.018)	(0.020)
Year = 2018 X BDA intensity std			-0.0158	
			(0.020)	
Year = 2020 X BDA intensity std			-0.0211	
			(0.021)	
Medium X BDA intensity std				0.0435*
				(0.024)
Large X BDA intensity std				0.0562**
				(0.022)
Medium		0.0452		0.0531*
		(0.029)		(0.029)
Large		0.0525		0.0621
		(0.050)		(0.051)
Constant	9.8144***	9.8493***	9.8331***	9.8607***
	(0.062)	(0.077)	(0.063)	(0.079)
All controls	Yes	Yes	Yes	Yes
Observations	8612	8612	8612	8612
Adjusted R-squared	0.283	0.284	0.283	0.284

NOTES: Dependent Variable: ln(labour productivity). Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

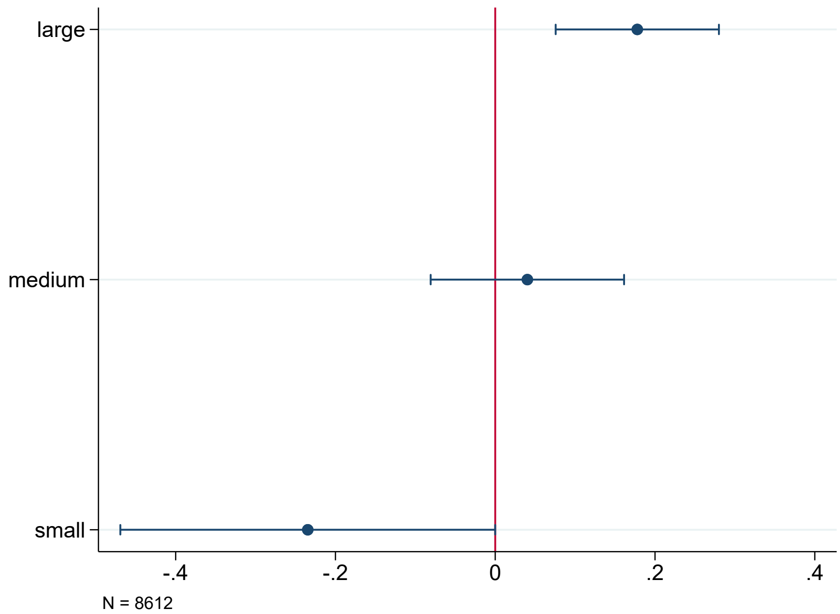
A.7 Additional Graphs: Germany

Figure A.7: Big Data Analytics by Industry



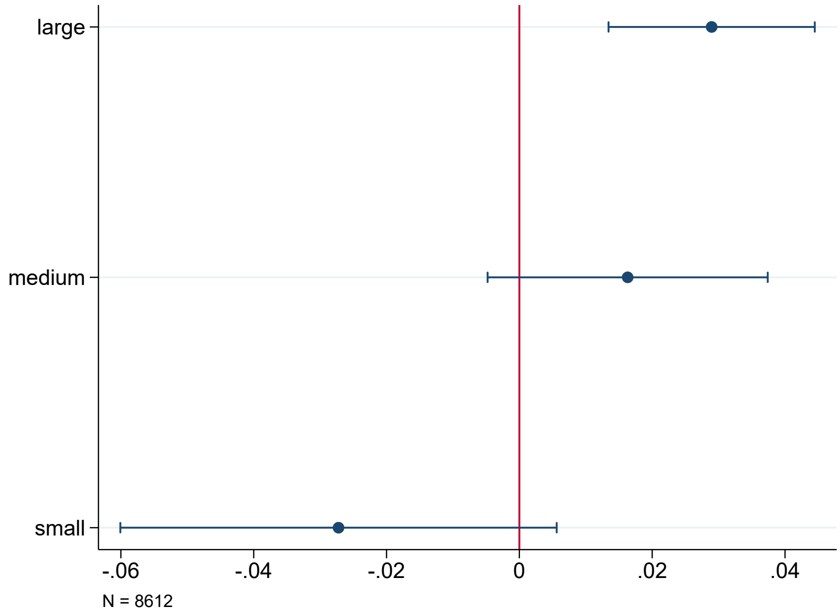
NOTES: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* when including both *BDA* and its interaction with the respective *industry*-dummy in the regression (see Table A.8 for the industry distribution).

Figure A.8: Big Data Analytics Intensity (at least 3 BDA Types) by Firm Size



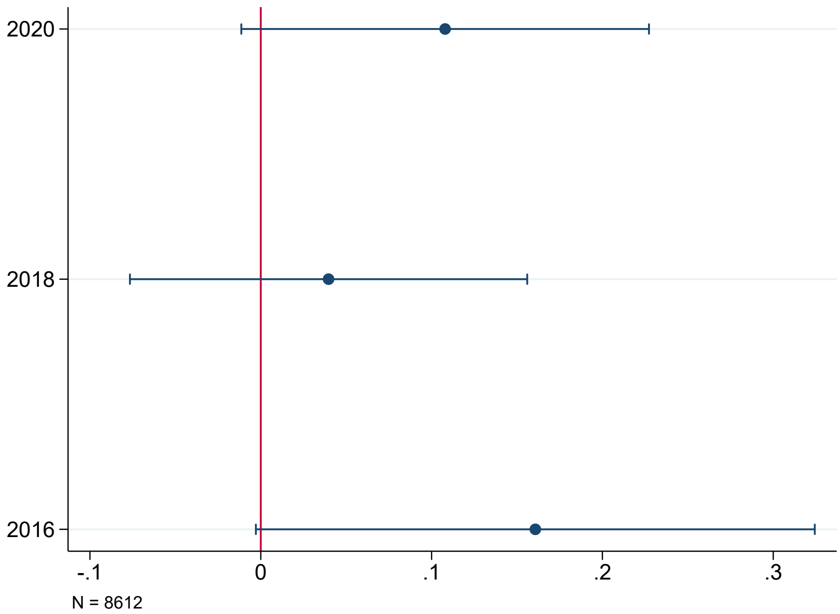
NOTES: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective *firm size class*-dummy in the regression. Also see Table A.13 for exact regression results.

Figure A.9: Big Data Analytics Intensity (Standardised) by Firm Size



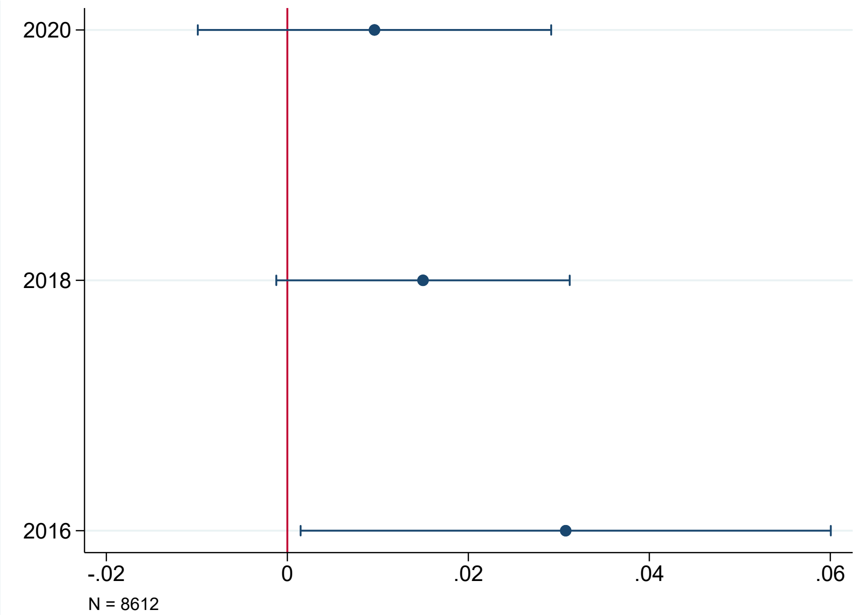
Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective *firm size class*-dummy in the regression. Also see Table A.13 for exact regression results.

Figure A.10: Big Data Analytics Intensity (at least 3 BDA Types) by Year



Notes: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective survey *year*-dummy in the regression. Also see Table A.13 for exact regression results.

Figure A.11: Big Data Analytics Intensity (Standardised) by Year



NOTES: Dependent Variable: ln(labour productivity). The graph depicts the average marginal effect of *BDA* intensity when including both *BDA* intensity and its interaction with the respective survey *year*-dummy in the regression. Also see Table A.13 for exact regression results.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//

IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de