// NO.23-061 | 12/2023

DISCUSSION PAPER

// ELISA RODEPETER CHRISTOPH GSCHNAIDTNER,
AND HANNA HOTTENROTT

Big Data and Start-up Performance

Big Data and Start-up Performance

Elisa Rodepeter[1,2], Christoph Gschnaidtner[2], and Hanna Hottenrott[1,2]
elisa.rodepeter@zew.de, christoph.gschnaidtner@tum.de, hanna.hottenrott@tum.de

[1] ZEW – Leibniz Centre for European Economic Research, L7 1, 68161 Mannheim, Germany
[2] Technical University of Munich, TUM School of Management, Dept. of Economics & Policy, Arcisstraße 21, 80333 Munich, Germany

This version: December 20, 2023

Abstract

Big Data (BD) is becoming widely available and manageable. This raises the question of whether Big Data Analytics (BDA) in companies leads to better decision-making and hence performance. Based on a large, representative set of start-ups in Germany, we study the adoption of BDA among small and young ventures and analyze its economic effects using various short- and longer-term performance measures. We investigate the effect of adopting BDA on the new ventures' cost structure, sales, profits, survival rate, growth, and their probability of receiving Venture Capital (VC) financing while taking into account factors that drive BDA adoption. Our findings, however, show that using BDA does not lead to an immediate competitive advantage in terms of the classical short-term performance measures. BDA adoption is rather associated with greater sales/profit uncertainty, higher (personnel) costs, and a higher probability of failure. Yet, the increased risk of adopting BDA is compensated by a prospect of higher long-term performance conditional on survival. BDA-adopting start-ups perform better than comparable ones when considering longer-term performance measures such as their growth and their ability to secure VC.

Keywords: Big data, Innovation, Productivity, Start-ups, Survival, Venture capital

Acknowledgments: We thank Sandra Gottschalk and Marius Berger for their help in data access and preparation. We further want to thank Maikel Pellens for his highly appreciated feedback. We are also indebted to seminar participants at the CBE seminar at NHH Bergen, the IAB Research Seminar, the RWI Research Seminar as well as to conference participants at REGIS Summer School 2023, the DRUID23 annual conference, the Summer 2023 BGPE workshop, the CISS Summer School 2023, the ZEW Conference of the Dynamics of Entrepreneurship (CoDE23), the 6th RISE Workshop 2023, and Claudia Steinwender for valuable comments.

1. Introduction

The relevance of Big Data (BD) and Big Data Analytics (BDA) has increased significantly in recent years. While firms generate data all along the value chain, the declining costs of all aspects of computing, such as storage, processing, and bandwidth, facilitated the development of new data processing methods (Gandomi & Haider 2015, Sena et al. 2019). Companies have adopted these data analytics methods to improve products and services, internal processes, or customer satisfaction (Manyika et al. 2011, Bajari et al. 2019, Suoniemi et al. 2020, Rammer & Es-Sadki 2022). BDA is therefore considered to be the 'next frontier for innovation, competition, and productivity' (Manyika et al. 2011), 'the next big thing in innovation' (Gobble 2015), or 'the management revolution' (McAfee & Brynjolfsson 2012). Data has repeatedly been called 'the oil of the 21st century' or even the 'world's most valuable resource'.[1] There seems to be a consensus that innovations based on BDA are key to future sustainable economic development and growth (Farboodi et al. 2019, Aghion et al. 2021).

However, insights about the value-enhancing effects of BDA on both the level of the whole economy as well as on the individual firm level are still limited, and considering the high expectations, the question emerges as to whether BDA can actually meet these. Anecdotal evidence and case studies show how individual companies successfully incorporate BDA into their existing business models or how BDA can even give rise to completely new business models. For example, UPS implemented a data-driven route optimization (Sena et al. 2019). Banks and insurance companies increasingly use advanced BDA techniques for fraud detection, risk management, or the pre-processing of claims (Paravisini & Schoar 2013). Larger corporations like Amazon, Alphabet, and Microsoft incorporate BDA at the

[1] For example, *The Economist* denoted in 2017 an entire article on this specific topic, see `https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data`.

core of their business models (Bajari et al. 2019). But also newcomers such as *OpenAI* or multi-million (or even multi-billion) funded start-ups like *6sense* and *databricks* are extremely successful in building their entire services around BDA.Beyond these anecdotes, empirical evidence on the impact of BDA is still scarce: Brynjolfsson et al. (2011) provide larger-scale evidence on the effect of data-driven decision-making (DDD), an integral element of BDA, on company performance. Combining data of 179 large, publicly listed U.S. firms over the period from 2005 to 2009 with 2008 survey data on their BDA usage, they show that DDD leads to an increase in productivity and market value.[2] However, Brynjolfsson et al. (2011) do not find any effect on accounting measures such as return on assets, return on equity, and asset utilization. The results also show that the (well-established) positive effect of investment in IT capital on the firms' market values is even more pronounced when paired with BDA. This is in line with empirical findings by Müller et al. (2018), who document a significant positive effect of BDA on firm productivity in highly competitive and IT-intensive sectors drawing on a sample of 814 publicly traded U.S. companies from 2008 to 2014. Both studies show that the positive impact of BDA is greatest when combined with the necessary (IT-)infrastructure. This notion is supported by a more recent study of Brynjolfsson et al. (2021), who show for 30,000 U.S.-based manufacturing firms that the use of prediction automation[3] can increase a plant's productivity if, and only if, paired with the right assets. Relevant complementary assets include IT capital, the stock of skilled workers, flow-efficient production, and managerial skills.

While these studies shed light on some aspects of the effects of BDA in the business context, the evidence is based on established companies in the United States or specific

[2] Similar effects, i.e. the rapid diffusion of DDD being consistent with higher productivity, are also documented by Brynjolfsson and McElheran (2016) for manufacturing plants.

[3] Prediction automation is only a subset of BDA analyzing historical and current data to predict future or unknown values, events, or probabilities.

sectors. This leaves the question of whether the findings can be generalized to smaller and younger firms (or companies outside the U.S.). Whether the value-enhancing effects of BDA also hold for younger firms is a critical question, as they may lack complementary assets or scale. Liability of newness and resulting resource constraints might hamper acquiring complementary skills and technologies that seem to be essential (Brynjolfsson et al. 2021). On the other hand, their very newness, implying low adjustment costs and the absence of path dependencies, might be of advantage when adopting BDA as part of their business model - potentially multiplying the returns to BDA.

With our study, we set out to investigate whether start-ups - defined as independent businesses that are younger than seven years - can economically benefit from the use of BDA or whether the costs and risks outweigh the advantages, reserving the highest returns to larger, established cooperations. By doing so, we contribute to three strands of literature: Firstly, we add to the literature on the adoption and impact of BDA on firm performance by focussing on start-ups. Secondly, we draw upon the theory of the Resource-Based View (RBV) and add to our understanding of BDA as a critical resource. Thirdly, we contribute to the literature on VC funding by showing that it plays a crucial role in financing start-ups adopting new and high-risk technologies.

We base our analyses on a unique and extensive data set of German start-ups observed over the period from 2010 to 2018. The panel includes, among others, information on various founder and firm characteristics, economic performance, as well as financing information, allowing us to analyze the effects of BDA over a longer and, at the same time, more recent period (as compared to Brynjolfsson et al. 2011, Brynjolfsson and McElheran 2016, Müller et al. 2018, Brynjolfsson et al. 2021). We extend the existing body of empirical literature on the economic impact of BDA adoption to younger firms. With our detailed data, we analyze the drivers of BDA adoption as well as its implications for the development of start-ups, including both short-term and longer-term outcomes. We account for selection into BDA adoption using entropy balancing based on important factors such as founder human capital, firm characteristics, and sector attributes. Comparing BDA adopters to a similar set of non-adopters, we show that adoption comes with higher personnel costs and higher variance in sales and profits, as well as a lower survival probability. While these findings indicate the costliness and riskiness of BDA adoption, they also suggest that the profile may appeal to investors. In line with this, we find that BDA-adopting companies have a significantly higher likelihood to attract VC, while there is no difference in bank financing. BDA adopting start-ups - conditional on survival - also show grow greater employee growth.

Aghion et al. (2021) describe innovative founders as Schumpeterian Entrepreneurs with a high probability of failure but high growth potential if their start-ups survive. Our findings confirm this idea by first showing that start-ups adopting BDA do not outperform their competitors in the short run but rather encounter more entrepreneurial risk. However, at the same time, we also show that this higher risk can pay off in the longer run – a classical high-risk, high-growth pattern.

2. Theoretical Background and Hypotheses

2.1. Big Data as a Resource

The recent uptake of BDA did not come unexpectedly. Various frameworks have been proposed to explain how the adoption of newly developed tools, such as BDA, and the implementation of the respective and equally novel processes can generate value within the firm (Barney 1991, Grant 1996, Winter 2003). The RBV suggests that a firm's competitive advantage is strongly determined by its resources. These include the firms' tangible and intangible resources as well as human capital. Differences in strategies and firm performances hence stem from differences in these resources, their combination, and their utilization (Barney 1991).

In the digital era, (big) data and the insights gained through BDA constitute such resources that may provide - alone or in combination with other resources - a competitive advantage. In light of the RBV framework, BD and BDA are, however, likely of higher value if complemented by particular other resources and potentially worthless if these are lacking. Firms need to build and shape their resources and use of resources, for example, by linking BDA and (data-related) human capital to enhance human decision-making (Helfat 2023). The dynamic RBV extends the RBV framework by stressing that resources must adapt to changes in a firm's products and markets but also to changes in the respective industries and environment (Sena et al. 2019). This includes technological advances, such as the emergence of BDA (Raguseo et al. 2020). Thus, the (dynamic) RBV may see BDA as an additional resource that can unfold value-increasing effects on other resources. The latter makes it imperative for firms to react to and might ultimately even explain the apparent rise of BDA among a large number of firms. The empirical literature confirms the notion of the RBV and highlights the importance of data-related tangible and intangible resources. This includes data itself, infrastructure and technologies for data storage and processing, general resources such as money and time (Gupta & George 2016), human capital in the form of technical and managerial skills (Tambe 2014, Akhtar et al. 2019, Wang et al. 2019), and fully intangible resources such as data-driven culture (Gupta & George 2016, Akhtar et al. 2019, Wang et al. 2019), and organizational learning (Gupta & George 2016, Yu et al. 2019, Wang et al. 2019).

Drawing from these ideas, recent studies have taken a closer look at the relationship between BDA and firm performance. Aydiner et al. (2019) show that business process performance mediates the relationship between BDA and firm performance.[4] BDA may also increase the firm's dynamic capabilities, which serve as a further mediator between BDA and firm performance (Côrte-Real et al. 2017, Wamba et al. 2017). These dynamic capabilities include the development of both incremental and radical innovations (Mikalef et al. 2019, Ciampi et al. 2021), as well as co-innovations (Lozada et al. 2019). While directly showing the explanatory power of the RBV framework in the case of BDA, these studies focus on specific links and mechanisms within the individual firm. Although less obvious, we can also derive implications of the RBV in studies focusing on the effect of BDA at a macro level. Brynjolfsson et al. (2011) show the necessity of pairing BDA with IT capital. The complementary resources identified by Brynjolfsson et al. (2021) bear a striking resemblance to the complementaries considered crucial in the RBV (Gupta & George 2016).

In the context of start-ups, the RBV helps to explain the trade-off between scarce financial resources to invest in and to experiment with new technologies versus the advantage of low adjustment costs when including BDA in an existing resource structure. Ciampi et al. (2021), for example, stress the importance of entrepreneurial orientation as a moderator of BDA and business performance. In addition, unlike incumbents that are often strapped to a technological status quo, start-ups are less path-dependent and, hence, less prone to the 'curse of knowledge' (Henderson 2006). Yet, at the same time, start-ups might simply lack the (financial) resources to follow through with adopting a novel technology such as BDA, which may require significant upfront investments. BDA capabilities, including competencies in data management, physical and virtual data infrastructure, and talent capability,

[4] Here, business process performance is a measure in terms of better coordination of regional, national, and global activities, helping firms to achieve economies of scale, increased labor productivity, adequate responses to customers' requests, and more efficient and effective meetings.

are all costly and may reduce the incentives to adopt BDA as well as the entrepreneurial risk associated with its adoption Akter et al. (2016).

2.2. Hypotheses

To answer the question of whether the adoption of BDA is a source of competitive advantage for young ventures, this paper takes a three-dimensional approach. We distinguish between the start-ups' short-term performance, their riskiness, and, eventually, their longer-term performance. While theory suggests that using BDA comes with potential benefits in terms of efficiency, product quality, and customer satisfaction, the RBV argues, and the empirical literature confirms that for BDA to have a positive impact on firm performance, it needs to be paired with complementary resources and capabilities (Gupta & George 2016, Mikalef et al. 2019, Wamba et al. 2017). Building these usually takes some time and investments, resulting in lagged performance effects and higher costs (Tambe 2014). To capture the positive short-term performance effects of BDA, we use information on sales. In addition (and in line with Tambe 2014), we explicitly show the dynamics between sales and cost drivers to understand the development of the start-ups' profits. Potential benefits are most likely to (eventually) materialize in higher sales but come with increased costs (rather than cost advantages), for example, due to highly skilled personnel or investments in hardware (Tambe 2014, Gupta & George 2016). As a result, we hypothesize that new ventures that use BDA do not immediately materialize better performance:

Hypothesis 1. *In the short run, the adoption BDA does not result in increased firm performance as determined by classical accounting performance measures, such as sales, costs, and profits.*

Considering lagged performance effects and substantial investments in complementary resources, the adoption of BDA is likely to result in higher uncertainty or entrepreneurial risk.We use two measures of entrepreneurial risk to test this assumption: The first one is volatility in the classical performance measures (compare Alvarez et al. 2005), which comprises both income stream uncertainty (Miller & Bromiley 1990) and volatility in costs. The second measure captures risk by relying on firm survival. Especially in the context of start-ups, survival can be considered as a lower bound for company performance (Brynjolfsson et al. 2021). Moreover, a measure capturing downside loss might be preferred to classical volatility, as entrepreneurs at the beginning of their endeavors are mostly concerned with the former (Janney & Dess 2006).

HYPOTHESIS 2. *Start-ups adopting BDA are associated with higher entrepreneurial risk, reflected by higher volatility in their performance measures and a lower probability of survival.*

Considering these short-term measures does, however, not account for longer-term outcomes. Benefits may take time to materialize as a result of initial adoption and learning costs. Further, those measures are neither adjusted for risk nor capture the value of intangible assets (Brynjolfsson et al. 2011). As all these factors play a crucial role when considering how BDA affects firm performance, it might be helpful to additionally consider different, more forward-looking concepts. We propose two such measures within this paper. The first one, following, e.g., Gilbert et al. (2006) and Chatterji et al. (2019), is employee growth.[5] We assume that start-ups with a more positive outlook and profit expectations for the longer term will and can hire more employees, even though they do not necessarily outperform their competitors in terms of sales or profits (yet). Another measure for expected

[5] Further examples that suggest the use of employee growth as a measure for new venture performance are Feeser & Willard (1990), Chandler & Hanks (1993), Reid & Smith (2000), Lange et al. (2007), or Nielsen (2015).

future firm performance commonly used in the literature is the market value (Brynjolfsson et al. 2011, Hitt et al. 2002). As most newly founded ventures are usually privately owned and not publicly listed, stock market data is unavailable. We, therefore, infer the future market value of the firm through its attractiveness to VC investors. This is based on the idea that VC investors discount start-ups' future performance and investments, thus serving as a testimony of their potential (future) market values.[6] Whereas traditional lenders like banks do not profit from any upside potential, VC investors can monetize the higher performance of their equity investments in case of an exit event and, thus, are willing to take on additional risk. There exists an extensive body of literature on how VC investors screen and evaluate start-ups. Selection factors include intellectual property, technological capability like reflected in BDA, the products or services, the management team, and the market and industry outlook (Gompers et al. 2020, Kaplan & Strömberg 2004, Hellmann & Puri 2000). Assuming in line with prior literature that BDA leads to a competitive advantage in the longer term, we hypothesize:

Hypothesis 3. *In the longer run, start-ups adopting BDA derive returns of this activity as reflected in a higher likelihood of securing VC and in higher employee growth.*

In the following, we test these hypotheses in the context of new businesses started in Germany between 2010 and 2015 and active in various service sectors and manufacturing.

[6] In the last decades, VC has become an increasingly important source of financing for start-ups in the US (Lerner & Nanda 2020) but also in Europe (Bertoni et al. 2015, Berger et al. 2021). Although the VC market in Europe is much smaller, it is receiving increasing attention from policymakers, for example, as part of the European Commission's EU 2020 strategy or the Green New Deal (Berger et al. 2021, Wallace 2020).

3. Data and Methodology

3.1. Data Sources and Variables

The analysis draws from several data sources. The main source is the IAB/ ZEW Start-up Panel. The panel comprises a large, representative sample of start-ups in Germany and tracks the founding process, business activities, and the development of newly founded companies with the help of annual computer-aided telephone interviews. The questionnaire further collects a large set of founder and company-specific information. The survey sample of start-ups is drawn as a stratified random sample from the Mannheim Enterprise Panel, which represents the entire universe of companies registered in Germany and recorded by Creditreform, the largest credit rating agency in Germany. The final sample consists of about 6,000 young companies that are interviewed once per year up to eight times.

When entering the panel, the surveyed companies are not older than three years and remain in the sample until a maximum age of six years. The collected information thus captures the first phase of the firms' life cycle.[7] We link this data with the ZEW Transaction Data Base on VC transactions. It contains transactional data for equity deals from the Bureau van Dijk's Zephyr Data Base and data provided by Majunke Consulting.[8] As a third data source, we track start-up survival based on information provided by Creditreform that allows us to observe exits until the end of 2021.

The resulting data set includes unique and comprehensive information covering founder and start-up characteristics, business activities (compare Table A.2 in the appendix for an overview of the number of firms listed in each sector), and information about the start-ups' financing structure. In 2017, the IAB/ ZEW included new questions addressing the topic of digitization and BDA usage, referring to the following definitions of BD and BDA during the interviews:

[7] For more details on the IAB/ ZEW Start-up Panel, see Fryges et al. (2009).

[8] For more details on the ZEW Transaction Data Base Berger et al. (2021).

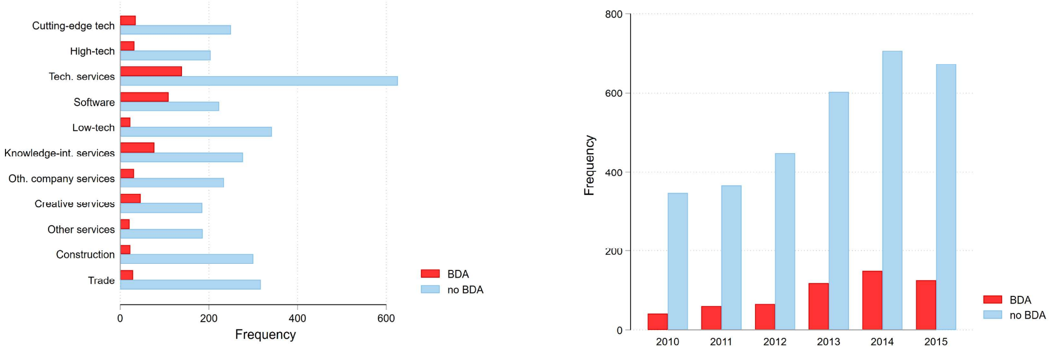

Figure 1 **BDA use by industry sector (left) and start-up cohort (right; indicated by founding year).**

> *BD is rapidly growing volumes of data generated from activities conducted electronically, for example, social media activities, GPS usage, etc. BDA refers to concepts, processes, technologies, and software applications that help process the rapidly growing and diverse volume of data (from corporate or external data sources) for qualitative and quantitative analyses as a basis for management decisions.*[9]

Based on this definition, founders were asked to indicate whether their start-up uses BDA, which approximately 15% of the 3,670 firms in our sample affirmed. While we use the panel structure of the data in the subsequent empirical analyses, it is important to note that, due to the one-off nature of the question, the BDA activity information is assumed to be a time-invariant firm characteristic. Provided that the observed firms are newly founded ventures (the firms are, on average, two years old), the assumption is reasonable that BDA, if used at all, was adopted at or shortly after the start-up was founded.

Overall, the final data set consists of 18,388 firm-year observations of start-ups founded between 2010 and 2015. BDA adopters can be found across all sectors (Figure 1, left), and adoption increased substantially in more recent cohorts (Figure 1, right). Most of the start-ups included in the panel are, by nature, relatively young and small. More than half of the

[9] IAB/ ZEW Start-up Panel Questionnaire, 2017, p. 32; the original definitions and questions were provided in German.

Table 1 Summary Statistics

Variable	Complete Sample				BDA = 0				BDA = 1				T-test
	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.	(p-value)
Total costs (t)	398.98	1,369.25	0.00	91,600.00	373.43	1,380.34	0.00	91,600.00	546.62	1,293.86	0.00	23,500.00	0.00
Investment costs (t)	31.91	166.19	0.00	10,000.00	28.65	121.03	0.00	5,000.00	51.08	321.49	0.00	10,000.00	0.00
R&D costs (t)	24.04	108.30	0.00	35,000.00	17.59	82.78	0.00	2,800.00	62.74	198.09	0.00	3,500.00	0.00
Operating costs (t)	339.80	1,293.59	0.00	91,600.00	323.01	1,323.39	0.00	91,600.00	435.82	1,103.21	0.00	23,000.00	0.00
Material costs (t)	172.75	1,174.81	0.00	90,000.00	171.35	1,224.22	0.00	90,000.00	180.66	844.14	0.00	20,000.00	0.67
Personnel costs (t)	132.80	300.13	0.00	9,500.00	119.10	256.61	0.00	6,300.00	211.82	471.77	0.00	9,500.00	0.00
Other operating costs (t)	33.69	104.76	0.00	7,000.00	32.27	102.93	0.00	7,000.00	41.82	114.40	0.00	2,000.00	0.00
Sales (t)	471.10	1,556.19	0.00	80,000.00	441.33	1,455.69	0.00	80,000.00	643.27	2,034.44	0.00	36,700.00	0.00
Profit (t)	33.05	210.32	-4,500.00	7,000.00	32.70	115.90	-4,500.00	3,600.00	35.02	396.68	-4,000.00	7,000.00	0.79
Profit indicator	2.31	0.57	1	3	2.33	0.56	1	3	2.18	0.58	1	3	0.00
Sales variance	220.78	831.63	0	31,600.00	193.77	715.44	0	31,600.00	379.47	1,311.11	0	21,000.00	0.00
Profit variance	116.55	755.24	0	30,000.00	107.25	780.35	0	30,000.00	170.06	587.57	0	8,825.00	0.00
Survival	0.91	0.29	0	1	0.92	0.28	0	1	0.88	0.33	0	1	0.00
Employee growth	0.10	0.19	-1.00	2.32	0.10	0.18	-1.00	2.32	0.14	0.22	-1.00	1.69	0.00
VC	0.13	0.34	0	1	0.11	0.32	0	1	0.22	0.42	0	1	0.00
Bank	0.54	0.50	0	1	0.55	0.50	0	1	0.49	0.50	0	1	0.00
Number of employees (t)	4.44	7.07	0	204.5	4.13	6.07	0	177.5	6.24	11.03	0	204.5	0.00
Number of employees when founding	2.80	3.42	1	93	2.73	3.33	1	93	3.23	3.87	1	53.25	0.00
R&D$_d$ (t)	0.28	0.45	0	1	0.25	0.43	0	1	0.48	0.50	0	1	0.00
Patent	0.04	0.20	0	1	0.04	0.18	0	1	0.07	0.25	0	1	0.00
Market novelty	1.36	0.88	1	4	1.31	0.82	1	4	1.65	1.11	1	4	0.00

Costs, sales, and profits are given in thousand Euros. All variables are continuous variables, except for *survival*, *VC*, *bank*, *R&D$_d$*, *patent*, and *market novelty* which are binary variables, and *profit indicator*, which is an ordinal variable. *Sales variance* and *profit variance* are normalized intra-firm variances.

firms were founded later than 2013, with, on average, only four employees. Table 1 shows the summary statistics for the most important variables used in the analysis. In terms of performance measures, we have detailed information on the start-ups' cost structures. The costs are reported in terms of investment costs, R&D costs, and operating costs, with operating costs being broken down further into personnel, material, and other operating costs. We further observe values for sales and profits, as well as the development of these measures over time.

Information on annual employment figures and bank financing is also obtained from the IAB/ ZEW Start-up Panel. The binary indicator *vc*, which equals unity if the start-up has received VC at any given point in time, is determined based on a specific question in the IAB/ ZEW Start-up survey and the information in the ZEW Transaction Data Base. Combining these two sources allows us to also capture smaller deals that may not be recorded in any of the data sources for VC deals. Table A.1 in the appendix presents a detailed description of all measures used in our analyses.

3.2. Methodology

As outlined above, adopting BDA comes with significant risks and costs. It has been shown that in the context of start-ups, comparable investment decisions are highly dependent on founder and team characteristics (Hottenrott & Chapman 2022). In line with the RBV, we expect founders to be more likely to start BDA-based businesses if they possess the necessary direct and complementary resources. Thus, we must assume that founders adopting BDA significantly differ from non-adopters in certain characteristics. Using a logistic regression on a set of founder and firm characteristics on the probability of the firm adopting BDA (compare Table A.4), we show that founders that use BDA methods are younger, found more often in teams than alone, have more often a higher education degree, but less experience in the specific sector. Yet, they are more often serial entrepreneurs, i.e., have more often founded a start-up before. Also, more recently founded ventures are more likely to adopt BDA, pointing out strong cohort effects. Even though we do not have data on the point in time when individual start-ups introduced BDA, this suggests that the diffusion of BDA is still in progress within the (German) economy.

To account for the non-randomness of BDA adoption, we introduce entropy balancing as a pre-processing method (Hainmueller 2012) addressing the problem of selection bias.[10] This means that we re-weight the sample of those start-ups that do not use BDA to match the sample of start-ups that use BDA, using a rich set of variables that account for BDA

[10] According to Hainmueller (2012) entropy balancing has several advantages as opposed to other balancing methods: It allows for a more extensive set of constraints, for weights to vary across units, and the computational complexity is relatively low. The method ensures closeness to the base weight and prevents unnecessary information loss. The resulting weights can then be used in almost any subsequent estimation.

adoption and likely also drive start-up performance more generally, such as educational background. This especially enables us to account for differences in the founders' cognitive and non-cognitive skills, but also for differences in the firms' environment. A complete list of the balancing variables, including their summary statistics, is presented in the appendix (see Table A.3). The weighting as part of the entropy balancing is done based on the first moment of these variables, and the resulting balanced sample serves as the basis for all following regression analyses.[11]

To test *Hypothesis I*, we conduct an analysis of the start-ups' sales and profits by modeling their volumes conditional on whether a venture has positive sales and profits. This model distinguishes between profits that are either greater (positive profit) or equal to zero (no profit). Losses (negative amounts) are reduced to not being a profit, i.e., considered as zero profits. Conditional on realizing profits, we analyze the magnitude of profits in the second stage. We assume that start-ups that already make positive sales or profits, i.e., successfully launched a product or service to the market, differ systematically from those that did not (yet). To account for this, we estimate a two-stage Heckman Model (Heckman 1976, 1979) using a Maximum Likelihood Estimator (MLE).[12] The first stage of the model focuses on the factors that contribute to a start-up having positive sales (respectively, profits). The second stage, accounting for the results of the first stage, estimates a model for the amount of these sales (profits). Since sales and profits are, as usually, highly skewed, we use logarithmic transformations of these measures in all following analyses.[13]

[11] The results stay qualitatively the same when balancing on second and third moments.

[12] We use an MLE which estimates both stages of the model within one single equation, and thus, allows the use of clustered standard errors accounting for repeat observations of the same company.

[13] In the case of the Heckman models, we follow a transformation of the form *ln(x)*. For all other analyses, for the sake of retaining observations with zeros, we use *ln(x+1)*, with x being the dependent variable. In a Heckman model, this would distort the first-stage estimation.

Exclusion restrictions, i.e., variables driving the first-stage but not the second-stage outcome, are usually derived from a priori theory. In this analysis, we adopt an alternative approach proposed by Farbmacher (2021) which, additionally to a priori theoretical considerations, allows for a data-driven identification of exclusion restrictions using the Least Absolute Shrinkage and Selection Operator (LASSO). Originally, the regularization technique LASSO is used to avoid overfitting of regression models by penalizing coefficients such that some of them are efficiently shrunken to zero. Farbmacher (2021) suggests using the LASSO technique on the second-stage Heckman equation and considering those variables dropping out of the set of explanatory variables as exclusion restrictions. In our case, we define a set of firm and founder characteristics as potential exclusion restrictions. This set includes *founder age*, *female*, *foreign*, *team*, *education*, *industry experience*, *founding experience*, *opportunity-driven*, *city*, *East Germany*, and *founding year*. For the remainder of this paper, these variables also serve as the set of firm controls.[14] In addition, two indicators for innovation activity, namely, one indicating whether the start-up is conducting any R&D in the respective year and one indicating whether a market novelty has been introduced, also serve as potential candidates (for summary statistics on these variables, again compare Table 1). We control for industry and year effects but do not include them in the set of potential exclusion restrictions. We rely solely on the exclusion restrictions identified by LASSO and, in the following, use these in a Heckman model applied to the balanced sample. To further understand what is driving the development of sales and profits, we estimate a multiple linear regression model on the effect of BDA on the start-ups' operating costs and the different components thereof. We focus on operating costs specifically, as

[14] As these variables correlate with the firm using BDA in the first place (compare Table A.4 in the Appendix), we additionally check whether this introduces a multicollinearity problem. Since the variance inflation factor for BDA is below 5 (for all defined models of this article), multicollinearity between those explanatory variables should not affect the estimated coefficients.

they account for more than 85% of the start-ups' total costs in our data (compare Table 1). We control for the size of the start-ups (i.e. using the number of employees in a given year), the start-ups' industry affiliation, the observation year, and the above-described firm characteristics. We expand the analysis on operating costs by running analogous regressions on the two components of operating costs, i.e., personnel and material costs.

To address *Hypothesis II* and the question of differences in venture risk, we analyze the start-ups' survival probabilities and their volatility in sales and profits. We use a logistic regression analysis to model firm survival and two multiple linear regression models for the logarithmized start-ups' normalized intra-firm variance of sales and profits, respectively. Note that the binary variable *survival* is time-independent, simply reflecting whether the start-up still exists, i.e., survived until 2021. Analogously to the models on the start-ups' costs, we control for firm characteristics, industry, and size using the number of employees when modeling survival and employee growth for the intra-firm variance.To further disentangle the development of our two performance measures over the ventures' age, we additionally run a quantile regression on the ventures' sales and profit, specifically focussing on the performance development of the highest and lowest deciles of start-ups in our sample. This also gives us a preliminary impression of the start-ups' longer-term development. Again, we control for various variables, including firm characteristics, industry affiliation, observation year, and firm size. In this model, we approximate firm size by using the number of employees at the start-up foundation instead of the yearly number of employees, as it is likely that the number of employees mechanically develops with the dependent variables sales and profit.

Finally, we test the hypothesis that start-ups using BDA outperform their competitors in the longer run (*Hypothesis III*) by examining their employee growth and their probability of securing VC. To calculate employee growth, we use the Compound Annual Growth Rate (CAGR) of the number of full-time equivalents, which is calculated using

$$\mathrm{CAGR} = \left(\frac{x_n}{x_0}\right)^{\frac{1}{n}} - 1, \tag{1}$$

with x_0 (x_n) being the first (last) observation (here: number of full-time equivalents) and n being the number of years between these observations. Notice that despite being only based on the respective first and last observation, the CAGR is, at least mathematically, not biased by the maximum age of the firms in the sample.[15] We again control for firm characteristics and the start-ups' industry affiliation in the corresponding regression analyses. Similarly to the case of the quantile regressions on sales and profits, it is not particularly reasonable to control for the yearly number of employees. Instead, we again use the number of employees at the foundation date to approximate firm size.

For the second longer-term performance measure, i.e. the probability of securing VC funding, we slightly adapt the pre-processing approach for the regression analyses. As previous literature has shown (compare, e.g., Gompers et al. 2020), founder characteristics heavily influence the selection process of investors compelling us to rely once more on entropy balancing. Unlike above, however, we employ two additional variables in the balancing procedure. Firstly, to proxy innovation potential, we use information on whether the start-up had already held a valid patent at the time of foundation.[16] As shown by Haeussler et al. (2009), Hsu & Ziedonis (2013), or Hottenrott et al. (2016) start-ups holding one or more patents are indeed more likely to secure VC financing.[17] Secondly, we

[15] As the IAB/ ZEW Start-up Panel is unbalanced, i.e., the number of observations differs among the single start-ups, it is challenging to calculate growth rates based on all available observations without impairing comparability. By definition, the CAGR circumvents this issue and offers an appropriate alternative growth measure.

[16] We do not use any form of R&D intensity indicator here as this might introduce a problem of reversed causality or simultaneity. One cannot distinguish whether more intensive R&D activities lead to the start-up acquiring VC - as it seems more innovative and thus more attractive to the investor - or whether the additional funding via VC leads to the start-up being able to spend more on R&D expenditures.

[17] See also Hoenig & Henkel (2014) for a detailed overview of the literature on patents and VC financing.

balance on a discrete profit indicator capturing whether the start-up has never, sometimes, or always made positive profits. Weighting is again done based on the first moment of these variables, resulting in a matched sample of comparable start-ups in terms of their essential characteristics, their degree of innovativeness, and their profitability. Using this balanced sample, we conduct a logistic regression on the start-ups' probability of receiving VC. As a benchmark, we are also interested in the start-ups' likelihood of obtaining bank financing. Since the provision of debt, in contrast to equity, does not entitle to any upside potential beyond the upfront agreed amount, we would expect that BDA matters differently to banks. Provided that BDA activity may be risky and associated with high costs, start-ups that adopt this technology should have a higher probability of securing VC funds both of higher financing demands and higher expected future profitability. Regarding bank financing, we do not expect this positive link. In the logistic regressions, we control for firm size, industry effects, and firm characteristics.

4. Results

4.1. Short-Term Performance Measures

This section presents the results of the empirical analyses of the impact of BDA adoption on the classical performance measures. Table 2 shows the results of the first and second stages of the corresponding Heckman models, as well as the multiple linear regression on the start-up's costs. Here, the variables *founding year* and *no degree* serve as exclusion restrictions as identified via LASSO (compare also Table A.5). These seem to be plausible exclusion restrictions from a theoretical point of view, as they suggest that start-ups need some time to be able to generate sales and that the probability of bringing products or services to the market is lower for start-ups whose founders' education levels are lower. Once they succeed in entering the market, however, the start-ups' age and the entrepreneurs' lack of educational degrees have no significant effect on the sales volume, all else equal.

The findings for the start-ups' sales are twofold (second and third columns in Table 2). The adoption of BDA reduces the probability of obtaining positive sales but has no effect on the sales volume conditional on realizing any sales. Columns 4 and 5 of Table 2 show the corresponding results for profits with *industry experience*, *apprenticeship*, and *vocational school* serving as exclusion restrictions. Again, the influence of BDA is significantly negative in the first stage and positive but insignificant in the second stage. Analogous to the sales analysis, this shows that BDA reduces the probability of start-ups making profits but, conditional on at least breaking even, has no effect on the magnitude of these profits. The results in columns 6 and 7 of Table 2 confirm the hypothesis that using BDA is associated with significantly higher operating costs. This effect is mainly driven by higher personnel costs. Firms adopting BDA have approximately 50% higher personnel costs than non-adopters. Thus, we can confirm *Hypothesis I* so far, as we do not find any positive performance effects for the classical short-term performance measures regarding sales and profits, but rather find a strong effect concerning the start-ups' cost structure towards higher personnel costs. No increase in sales and profits but higher costs already hint at the higher risks that we predict BDA adopting start-ups to encounter (compare *Hypothesis II*).[18]

[18] For the sake of completeness, we also run a pooled linear regression on the logarithmized sales and profits, compare Table A.6 in the appendix. The results, however, as expected, are inconclusive.

Table 2 Pooled Regression Results on Start-ups' Sales, Profits, and Costs

	Sales		Profits		Costs		
	Sales [0/1]	Ln(sales)	Profits [0/1]	Ln(profits)	Ln(operating costs)	Ln(personnel costs)	Ln(material costs)
BDA	-0.190***	0.078	-0.130***	0.082	0.188**	0.515***	0.116
	[-0.323, -0.057]	[-0.034, 0.189]	[-0.221, -0.040]	[-0.058, 0.222]	[0.033, 0.343]	[0.193, 0.836]	[-0.096, 0.328]
Number of employees	0.132***	0.065***	0.006	0.052***	0.091***	0.153***	0.079***
	[0.094, 0.170]	[0.049, 0.082]	[-0.004, 0.017]	[0.036, 0.067]	[0.061, 0.121]	[0.095, 0.211]	[0.056, 0.103]
Constant	574.510***	12.239***	356.765***	-16.771	10.013***	6.027***	8.601***
	[429.527, 719.493]	[11.827, 12.651]	[286.951, 426.579]	[-121.547, 88.005]	[9.374, 10.652]	[4.452, 7.601]	[7.588, 9.614]
Wald-test firm controls	392.89***		380.18***		8.96***	10.45***	2.73***
Wald-test sector controls	141.64***		99.40***		10.69***	2.63***	43.12***
Wald-test year controls	457.59***		270.68***		9.25***	12.25***	5.18***
Observations	14,228	14,228	13,289	13,289	13,289	13,932	13,449
R^2					0.243	0.195	0.166
Mean of dependent variable	0.94	11.26	0.69	10.34	11.25	8.40	9.72

The table shows the regression coefficients of the Heckman Models on the start-ups' sales and profits and the results of the linear regression model on the start-ups' costs with 95% confidence intervals in brackets. The sample is entropy balanced using weights for the founders' age, gender, nationality, team composition, educational background, industry experience, founding experience, motive, geographic location, sector, and founding year. The Heckman Models use the exclusion restrictions proposed by the Farbmacher extension, resulting in the following exclusion restrictions: *founding year*, and no degree for the model on sales and *apprenticeship*, *vocational school*, and *industry experience* for the model on profits.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

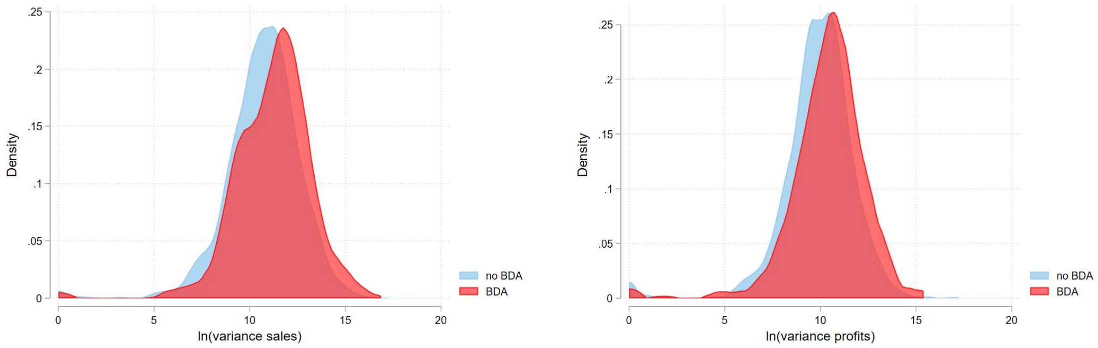

Figure 2 **Kernel Density Estimates of Sales and Profit Variances (normalized and log-transformed).**

4.2. Venture Risk

As suggested by the first set of results on short-term performance, Table 3 shows that BDA is indeed associated with higher risk as manifested in higher volatility in sales and profits and a lower survival probability. This confirms *Hypothesis II*: First, we find that start-ups adopting BDA have a higher chance of failure, i.e., compared to non-adopters, their probability of survival decreases on average by approximately 4%. This is consistent with the results from above that firms using BDA have higher costs, are less likely to generate sales, and, as a result, also less likely to generate positive earnings. At the same time, start-ups adopting BDA are exposed to significantly higher variances in sales and profits. Variance homogeneity tests, i.e., t-tests on the differences in variances and Levene-tests, report significantly higher variances in overall sales and profits for start-ups using BDA.[19] These results also hold for the logarithmized normalized intra-firm variance, capturing the volatility of sales and profits on the level of the individual firm. For both sales and profits, Figure 2 shows the shifts of the logarithmized intra-firm variance density distributions of start-ups using BDA to the right. The regression results in columns 3 and 4 of Table 3 confirm the statistical significance of these shifts.

Table 3 Pooled Logistic and Linear Regression Results on Start-up Survival and Dispersion of Sales and Profits

	Survival [0/1]	Dispersion	
		Ln(sales variance)	Ln(profit variance)
BDA	0.672***	0.318***	0.364***
	[0.502, 0.899]	[0.138, 0.498]	[0.133, 0.596]
Number of employees	1.027		
	[0.990, 1.065]		
Employee growth		2.841***	1.486***
		[2.354, 3.326]	[0.977, 1.996]
Constant		10.089***	8.537***
		[9.428, 10.749]	[7.615, 9.459]
Wald-test firm controls	11.92	6.87***	3.52***
Wald-test sector controls	17.52*	6.15***	5.60***
Observations	3,659	3,484	3,391
(pseudo) R^2	0.031	0.194	0.066
Mean of dependent variable	0.899	10.771	9.951

The table shows the odds ratios of the logistic regression on survival and the regression coefficients of the linear regressions on sales and profits variances with 95% confidence intervals in brackets. The sample is collapsed, i.e. one observation per start-up is used, and entropy is balanced using weights for the founders' age, gender, nationality, team composition, educational background, industry experience, founding experience, motive, geographic location, sector, and founding year.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To better understand the dynamics of the variations of the start-ups' sales and profits, we map their development over the ventures' age in Figure 3. The upper half of the figure shows that among the start-ups using BDA, the highest 90th quantile generates higher sales than their counterparts not using BDA. It also illustrates that the growth trajectory is much steeper. For profits, as depicted in the lower half of Figure 3, this effect is also apparent, yet less pronounced. For the lowest decile, we see, however, a clear difference between adopters and non-adopters, indicating that for the less performant ventures the usage of BDA only worsens their profit outcomes. Table 4 shows the results of the corresponding

[19] Levene-tests are conducted using centering on the median to account for the asymmetry of the data (Conover et al. 1981). The results are statistically significant at the 1%-level.

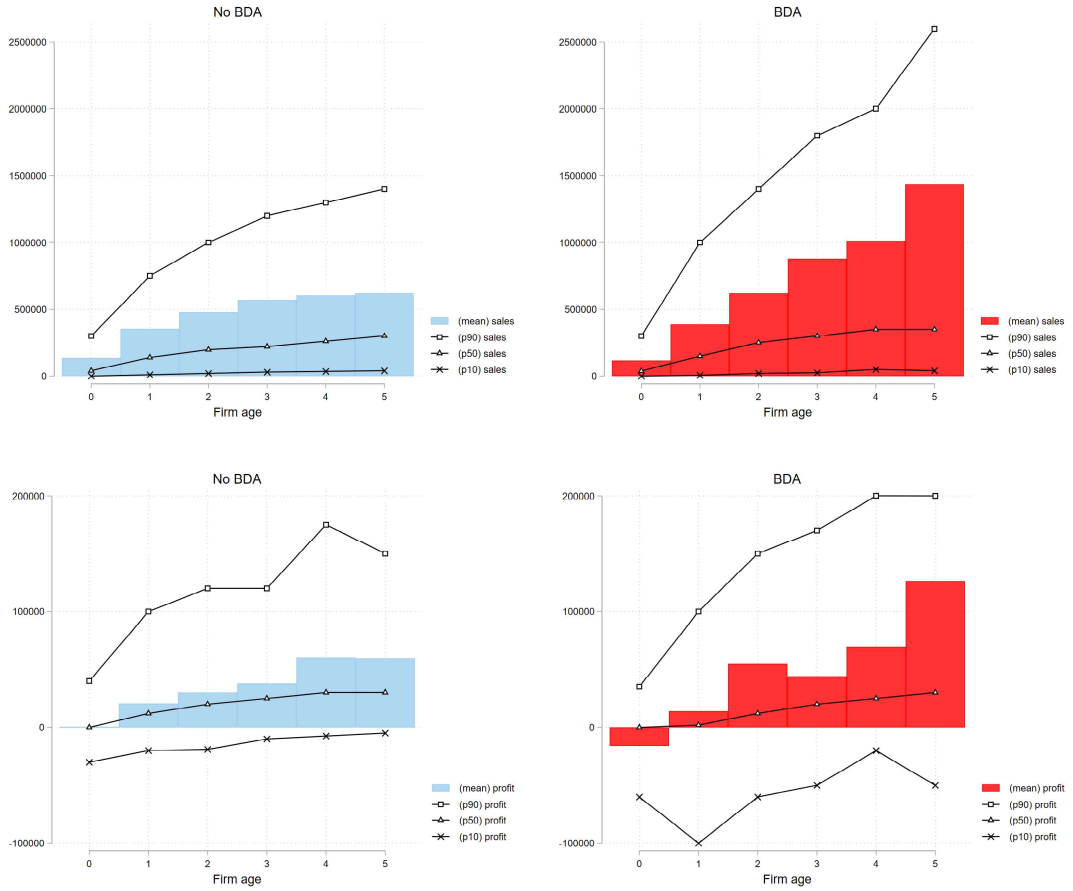

Figure 3 Sales and Profits Quantiles over Firm Age

Note: Mean, median, as well as the 10%- and 90%-quantiles of both, sales and profits, are calculated separately for *No BDA* (left) and *BDA* (right) and for each year of the firm age. Due to the unbalanced nature of the panel data set, the number of observations varies across firm age and are as follows: *No BDA*: 3,144 / *BDA*: 565 (Firm age = 0); 3,144 / 565 (1); 2,961 / 528 (2); 2,624 / 446 (3); 1,952 / 331 (4); 1,377 / 214 (5). Due to the small number of observations for BDA adopters with firm age = 6 (number of observations = 109), observations with firm age = 6 are not included in the graphics.

quantile regressions confirming these descriptive results. For the highest 90th quantile of the logarithmized sales, both the variable *BDA* as well as the interaction term of *BDA* and *firm age* are positive and strongly statistically significant. The respective coefficients on the 90th quantile of the logarithmized profits are smaller and only statistically significant for the interaction term. However, regressing on the 90th quantile of losses (i.e., using the logarithm of the absolute value of negative profits) gives us both a statistically significant coefficient on *BDA*, as well as on the interaction term of *BDA* and *firm age* illustrating the higher risk that these start-ups encounter. This leads to the question of whether this risk will pay off in the longer term.

Table 4 Pooled Linear Quantile Regression Results on Start-ups' Sales and Profits

	0.9 ln(sales)	0.1 ln(sales)	0.9 ln(profits)	0.9 ln(losses)
BDA	0.282***	-1.713	0.004	1.056***
	[0.117, 0.448]	[-5.878, 2.452]	[-0.160, 0.168]	[0.362, 1.749]
Firm age	0.272***	0.620***	0.202***	-1.228***
	[0.228, 0.317]	[0.505, 0.735]	[0.166, 0.238]	[-1.421, -1.034]
BDA × firm age	0.131***	0.714	0.081**	0.475***
	[0.056, 0.207]	[-0.547, 1.976]	[0.008, 0.154]	[0.172, 0.779]
Number of employees at foundation	0.182***	0.187**	0.088***	0.154***
	[0.155, 0.209]	[0.036, 0.339]	[0.061, 0.116]	[0.117, 0.191]
Constant	11.922***	3.947	10.901***	4.872***
	[11.503, 12.341]	[-14.103, 21.997]	[10.456, 11.345]	[3.347, 6.397]
Wald-test firm controls	9.95***	2.61***	10.11***	19.65***
Wald-test sector controls	11.95***	4.78***	5.83***	12.67***
Wald-test year controls	5.76***	41.65***	2.58***	6.96***
Observations	16,410	16,410	14,451	18,520
R^2	0.123	0.103	0.072	0.078
Mean of dependent variable	11.26	11.26	7.03	1.77

The table shows the regression coefficients of quantile regression on the 0.1 and 0.9 quantiles of the logarithmized sales, profits, and losses with 95% confidence intervals in brackets. The variable *agefirm* is demeaned to ensure a meaningful coefficient for the variable *BDA*. Note, the sample is not balanced.[20]
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.3. Start-up Growth and Access to Financing

The regression on employee growth yields a significant positive effect for new ventures that adopt BDA (Table 5). They have a 4.2 percentage points higher compound annual employee growth rate than non-adopters (compare column 2 in Table 5). Next, we analyze how the use of BDA affects the start-ups' financing sources, focusing particularly on venture capital, using it as a proxy for expected future firm performance (compare *Hypothesis III*). We anticipate that start-ups using BDA are more likely than others to raise VC funding as VC investors are willing to bear higher risks if they predict to profit from the upside

[20] A robustness check using an entropy balanced panel can be found in Table A.7 in the Appendix. However, due to limitations in the statistical software, standard errors cannot be clustered when using weights.

Table 5 Regression Results on Firm Growth and Financing Source

	Employee Growth	Financing	
		VC [0/1]	Bank [0/1]
BDA	0.042***	1.467***	1.046
	[0.020, 0.063]	[1.105, 1.948]	[0.825, 1.327]
Number of employees		1.051***	1.069***
		[1.020, 1.079]	[1.040, 1.122]
Number of employees when founding	-0.008***		
	[-0.013, -0.003]		
Constant	0.153***		
	[0.074, .231]		
Wald-test firm controls	4.41***	44.53***	100.80***
Wald-test sector controls	2.48**	60.70***	28.79***
Observations	3,651	3,643	2,739
(pseudo)R^2	0.065	0.221	0.106
Mean of dependent variable	0.109	0.126	0.500

The table shows the regression coefficients of the linear regression on firm growth and the odds ratios for the logistic regressions on venture capital and bank financing with 95% confidence intervals in brackets. The sample is entropy balanced using weights for the founders' age, gender, nationality, team composition, educational background, industry experience, founding experience, motive, geographic location, sector, and founding year. For the linear regressions on financing, we additionally use weights for a patent and a profit indicator.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

potential, i.e. if they expect the start-ups to perform well in the future. Table 5 shows the results of the logistic regression on VC funding as well as on bank funding, which serves as a counterfactual. Again, we use a weights-balanced sample to account for the fact that venture capitalists include similar firm and founder characteristics in their decision-making process that are also endogenous to the decision of whether or not start-ups use BDA. Accounting for firm and entrepreneurs' characteristics, firm size, industry affiliation, profit indication, and degree of innovativeness (as indicated by whether the start-up already held a patent before founding), the odds of a start-up receiving VC funding are by a factor of 1.376 higher (equivalent to a 4 percentage points higher probability) if it uses BDA compared to start-ups not using BDA (compare column 2 in Table 5). The results for the effect of adopting BDA on the start-ups' likelihood to receive financing from banks are statistically insignificant, indicating that whether firms use BDA or not seems not to influence the banks' lending decisions.

5. Conclusion and Outlook

This study focused on the relationship between BDA adoption and economic performance. Our analysis contributes to earlier research that studied performance effects of systematic data analysis in companies and how it affects their operations and productivity. While most research considered BDA in larger established companies (e.g., Brynjolfsson et al. 2011, Müller et al., 2018, Brynjolfsson et al. 2021), the presented results extend this work to the context of young and small companies. Newly founded businesses face different opportunities and challenges. Drawing from resource theory, we hypothesized that the cost-benefit relation of BDA adoption is different in start-ups due to their challenge of building up complementary resources in stages of financial and other constraints. We, therefore, differentiated between short-term performance measures, risk evaluation, and forward-looking measures. Our findings indeed show that for start-ups, the adoption of BDA is, on average, not associated with higher firm performance in the short run. Instead, we find that ventures adopting BDA have a lower probability of making profits as their operating costs are significantly higher, mainly driven by personnel costs. Further, we show that BDA-adopting start-ups are riskier. In addition to being less likely to generate sales and profits, they also have a higher volatility in these performance measures and are more likely to fail. The observed high intra-firm variances of sales and profits suggest that there is significant uncertainty of profitability.

When taking the development of sales and profits over time into account, a different picture emerges. For start-ups in the upper deciles of the sales and profit distribution, adopting BDA has a performance-enhancing effect that increases over the firm's life cycle. This effect is reversed for firms in the lowest decile. This underpins the higher risk and hints at potentially positive longer-term effects, conditional on firm survival. Despite the additional risk we show, adopting BDA can be rewarded in the longer run. New ventures that employ BDA show higher employment growth and are more likely to receive VC financing.

Our findings complement the empirical literature on the impact of BDA on firm performance by showing that a profitable use for newly found ventures is not self-evident, and that adoption can come with significant costs and risks. The results are of particular interest as, to our knowledge, empirical papers so far focus on large, established firms for which BDA seems to more generally result in significant performance improvements - even in the short run. In the context of young companies, these results emphasize the importance of differentiating between short-term and longer-term performance indicators when measuring the impact of the adoption of new technologies by younger firms. Further, our results contribute theoretically to our understanding of data as a critical resource in the RBV. Even though start-ups benefit from low adjustment costs when adopting BDA in their existing resource structure, i.e., it might be easier for them to shape their human capital according to the requirements of BDA, build up data-driven business processes, or a data-oriented organizational culture, we show that it takes time until they eventually are able to profit from the resulting performance advantages. Hence, more generally, investments in BDA may not pay off in the first phase of the firm's life cycle but rather require time and additional (financial) resources to accumulate the required complementary resources.

This supports the notion of the Schumpeterian Entrepreneur founding new firms at the frontier of innovation with a considerably higher risk that is not necessarily immediately rewarded (Aghion et al. 2021, Botelho et al. 2021).

Our study also sheds some light on which founders adopt BDA. We observe the use of BDA in all sectors that are covered in our study, with much variation in adoption being explained by founder and team characteristics. Our findings confirm that often a diverse set of skills is present in BDA-adopting startups and that the advantage of teams might stem from the interdisciplinary mix of knowledge they draw upon, for example, knowledge in computing, statistics, and machine learning, but also business domain knowledge (Akhtar et al. 2019). Apart from their education, the founders' experience in previous business foundations indicates a set of soft skills, including management expertise. The RBV literature has repeatedly pointed out the importance of such managerial skills when deploying BDA (Gupta & George 2016, Wang et al. 2019). Finally, we contribute to the literature on VC funding by showing that it plays a crucial role in financing BDA-using startups. These findings are relevant for entrepreneurs and innovation policy. While we find a lower survival probability for BDA-adopting start-ups, this could also be the result of better BDA-based managerial decisions, such as closing the business after learning about the low expected returns. From a policy perspective, BDA adoption can, despite its cost- and risk-driving potential, still be beneficial from an ecosystem perspective. Knowledge spillovers through employee mobility, cooperation, or competition can benefit the innovation system more generally, even if the returns to the individual entrepreneur are uncertain or low.

Our study is, however, not without its limitations. Even though the analysis is based on survey data that allows us to capture BDA usage across sectors, the definition of BDA is rather broad and includes concepts, processes, technologies, and software applications in one single binary indicator. A route for future research could be a more detailed differentiation of the use of BDA. This is also highlighted by Ghasemaghaei (2021), who indicates the necessity of conceptually differentiating among BD characteristics rather than treating big data as a holistic variable. Drawing on the RBV, one could, for example, distinguish between tangible, human, and intangible resources related to BDA. Whereas the currently used definition of BDA in the IBA/ ZEW Start-up Panel already includes tangible and intangible resources, it misses the component of human skills related to BDA, for example, concerning general expertise or specialized training. Another dimension worthwhile investigating would be the intensity of BDA usage or its relevance for the start-ups' business models. From a methodological point of view, we address the problem of endogeneity using entropy balancing. This is especially important since it has been shown that in a start-up, strategic decisions, like adopting BDA, are highly influenced by their founders (Hottenrott & Chapman 2022). However, entropy balancing only corrects for selection bias in the observable variables, not for endogeneity in general. Although our rich data set allows us to capture the most important dimensions of selection bias, we may still be concerned with factors that drive both BDA use and performance but were not accounted for.

Finally, the conclusions derived from our results are partly based on the assumption that VC funding is an indicator of future firm performance. To maximize their pay-off in an exit event, venture capitalists are interested in investing in start-ups they attribute a high future market value. Since it has been shown that venture capitalists conduct sophisticated and complex analyses when selecting their investments (Gompers et al. 2020, Hellmann & Puri 2000, Kaplan & Strömberg 2004), we argued that the reception of VC funding is a reliable forward-looking measure. However, one could question to what degree they can reliably identify such start-ups. It has been shown that venture capitalists are biased in their investment decisions, among others, concerning location (Cumming & Dai 2010), gender (Eddleston et al. 2016), or similarity bias regarding education and professional experience (Franke et al. 2006). Although we try to control for most of these attributes, we cannot rule out that BDA use is subject to some form of 'buzzword bias,' hype, or herding that explains the observed investments.

We encourage more research on the performance effects of BDA use. As BDA is becoming more intensively used by private and public actors, we expect learning effects at both individual and collective levels. It seems, therefore, crucial to study differences in the performance effects between early and late adopters and whether there are returns to specialization in BDA that lead to companies in-sourcing BDA capacities.

References

Aghion, P., Antonin C., and Bunel S. 2021. The Power of Creative Destruction: Economic Upheaval and the Wealth of Nations. *Harvard University Press.*

Akhtar, P., Frynas, J.G., Mellahi, K., and Ullah., S. 2019. Big Data-savvy Teams' Skills, Big Data-Driven Actions and Business Performance. *British Journal of Management* **30** 252–271.

Akter, S., Wamba, S., Gunasekaran, A., Dubey, R., and Childe, S. 2016. How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics* **182** 113–131.

Alvarez, S. A., and Barney, J. B. 2005. How Do Entrepreneurs Organize Firms Under Conditions of Uncertainty? *Journal of Management* **31** 776–739.

Aydiner, A. S., Tatoglu, E., Bayraktar, E., Zaim, S., and Delen, D. 2019. Business Analytics and Firm Performance: The Mediating Role of Business Process Performance. *Journal of Business Research* **96** 229–237.

Bajari, P., Chernozhukov, V., Hortaçsu, A., and Suzuki, J. 2019. The Impact of Big Data on Firm Performance: An Empirical Investigation. *AEA Papers and Proceedings* **109** 33–37.

Barney, J. 1991. Firm Resources and Sustained Competitive Advantage. *Journal of Management* **17** 99–120.

Berger, M., and Hottenrott, H. 2021. Start-up Subsidies and the Sources of Venture Capital. *Journal of Business Venturing Insights* **16**.

Bertoni, F., Colombo, M. G., and Quas, A. 2015. The Patterns of Venture Capital Investment in Europe. *Small Business Economics* **45** 543–560.

Botelho, T. L. and Fehder, D. and Hochberg, Y. 2021. Innovation-driven Entrepreneurship. *National Bureau of Economic Research* **No. w28990**.

Brynjolfsson, E., Hitt, L. M., and Kim, H. H. 2011. Strength in Numbers: How Does Data-Driven Decision-making Affect Firm Performance? *Available at SSRN 1819486*.

Brynjolfsson, E. and McElheran, K. 2016. The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review* **106** 133–39.

Brynjolfsson, E., Jin, W., and McElheran, K. 2021. The Power of Prediction: Predictive Analytics, Workplace Complements, and Business Performance. *Business Economics* **56** 217–239.

Chandler, G. and Hanks, S. 1993. Measuring the performance of emerging businesses: A validation study. *Journal of Business Venturing* **8** 391–408.

Chatterji, A., Delecourt, S., Hasan, S., Koning, R. 2019. When does advice impact startup performance? *Strategic Management Journal* **40** 331–356.

Ciampi, F., Demi, S., Magrini, A., Marzi, G., and Papa, A. 2021. Exploring the impact of big data analytics capabilities on business model innovation: The mediating role of entrepreneurial orientation. *Journal of Business Research* **123**, 1–13.

Conover, W.J., Johnson, M. E.. and Johnson, M. M. 1981. A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics* **23** 351–361.

Côrte-Real, N., Oliveira, T., and Ruivo, P. 2017. Assessing Business Value of Big Data Analytics in European Firms. *Journal of Business Research* **70** 379–390.

Cumming, D. and Dai, N. 2010. Local Bias in Venture Capital Investments. *Journal of Empirical Finance* **3** 362–380.

Eddleston, K. A., Ladge, J. J., Mitteness, C., and Balachandra, L. 2016. Do You See What I See? Signaling Effects of Gender and Firm Characteristics on Financing Entrepreneurial Ventures. *Entrepreneurship Theory and Practice* **40** 489–514.

Farbmacher, H. 2021. Selection models with data-driven exclusion restrictions. https://github.com/farbmacher/heckman_lasso. Retrieved September 2022.

Farboodi, M., Mihet, R., Philippon, T., and Veldkamp, L. 2019. Big Data and Firm Dynamics. *AEA Papers and Proceedings* **109** 38–42.

Feeser, H., and Willard, G. 1990. Founding strategy and performance: A comparison of high and low growth high tech firms. *Strategic Management Journal* **11** 87–98.

Franke, N., Gruber, M., Harhoff, D., and Henkel, J. 2006. What You Are Is What You Like — Similarity Biases in Venture Capitalists' Evaluations of Start-up Teams. *Journal of Business Venturing* **21** 802–826.

Fryges, H., Gottschalk, S., and Kohn, K. 2009. The KfW/ZEW Start-up Panel: Design and Research Potential. *ZEW-Centre for European Economic Research Discussion Paper* **4441** 488–500.

Gandomi, A., and Haider, M. 2015. Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management* **35** 137–144.

Ghasemaghaei, M. 2021. Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics. *International Journal of Information Management* **57**.

Gilbert, B., McDougall, P., and Audretsch, D. 2006. New Venture Growth: A Review and Extension. *Journal of Management* **32** 926–950.

Gobble, M. M. 2015. Big Data: The Next Big Thing in Innovation. *Research-technology Management* **56** 64–67.

Gompers, P. A., Gornall, W., Kaplan, S. N., and Strebulaev, I. A. 2020. How do Venture Capitalists Make Decisions? *Journal of Financial Economics* **135** 169–190.

Grant, R. M. 1996. Prospering in Dynamically-competitive Environments: Organizational Capability as Knowledge Integration. *Organization Science* **4** 375–387.

Gupta, M., and George, J. F. 2016. Toward the Development of a Big Data Analytics Capability. *Information & Management* **53** 1049–1064.

Hainmueller, J. 2012. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis* **20** 25–46.

Haeussler, C., Harhoff, D., and Mueller E. 2009. To Be Financed or Not – The Role of Patents for Venture Capital Financing, Working Paper.

Heckman, J. 1976. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models. *Annals of Economic and Social Measurement* **5** 475—492.

Heckman, J. 1979. Sample Selection as a Specification Error. *Econometrica* **47** 153—161.

Helfat, C., Kaul, A., Ketchen Jr., D., Barney, J., Chatain, O., and Singh, H. 2023. Renewing the resource-based view: New contexts, new concepts, and new methods. *Strategic Management Journal* **44** 1357–1390.

Hellmann, T., and Puri, M. 2000. The Interaction Between Product Market and Financing Strategy: The Role of Venture Capital. *Review of Financial Studies* **13** 959–984.

Henderson, R. 2006. The Innovator's Dilemma as a Problem of Orgaizational Competence. *Journal of Product Innovation Management* **47** 5–11.

Hoenig, D., and Henkel J. 2014. Quality Signals? The Role of Patents, Alliances, and Team Experience in Venture Capital Financing. *R*esearch Policy **44** 1049—1064.

Hitt, L. M., Wu, D. J., and Zhou, X. 2002. Investment in Enterprise Resource Planning: Business Impact and Productivity Measures. *Journal of Product Innovation Management* **19** 71–98.

Hottenrott, H., and Chapman, G. 2022. Green Start-ups and the Role of Founder Personality. *Journal of Business Venturing Insights* **197**.

Hottenrott, H., Hall, B. H., and Czarnitzki D. 2016. Patents as quality signals? The implications for financing constraints on R&D. *Economics of Innovation and New Technology* **25** 197–217.

Hsu, D., and Ziedonis. R. 2013. Resources as Dual Sources of Advantage: Implications for Valuing Entrepreneurial-Firm Patents. *Strategic Management Journal* **34** 761—781.

Janney, J. J., & Dess, G. G. 2006. The Risk Concept for Entrepreneurs Reconsidered: New Challenges to the Conventional Wisdom. *Journal of Business Venturing* **21** 385–400.

Kaplan, S. N., and Strömberg, P. E. 2004. Characteristics, Contracts, and Actions: Evidence from Venture Capitalist Analyses. *Journal of Finance* **59** 2177–2210.

Lange, J., Mollov, A., Pearlmutter, M., Singh, S., and Bygrave, W.D. 2007. Pre-start-up formal business plans and post-start-up performance: A study of 116 new ventures. *Venture Capital.* **9** 237–256

Lerner, J. and Nanda, R. 2020. Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn. *Journal of Economic Perspectives* **34** 237–261.

Lozada, N., Arias-Pérez, J., and Perdomo-Charry, G. 2019. Big data analytics capability and co-innovation: An empirical study. *Heliyon* **5**.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*

McAfee, A., and Brynjolfsson, E. 2012. Big Data: The Management Revolution. *Harward Business Review* **90** 60–68.

Mikalef, P., Boura, M., Lekakos, G., and Krogstie, J. 2019. Big Data Analytics Capabilities and Innovation: the Mediating Role of Dynamic Capabilities and Moderating Effect of the Environment. *British Journal of Management* **30** 272–298.

Miller, K. D., and Bromiley, P. 1990. Strategic Risk and Corporate Performance: An Analysis of Alternative Risk Measures *Academy of Management Journal* **33** 756–779.

Müller, O., Fay, M., and Vom Brocke, J. 2018. The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems* **35** 488–509.

Nielsen, K. 2015. Human capital and new venture performance: the industry choice and performance of academic entrepreneurs. *The Journal of Technology Transfer* **40** 453–474.

Paravisini, D. and Schoar, A. 2013. The Incentive Effect of Scores: Randomized Evidence from Credit Committees. *NBER working paper*

Raguseo, E., Vitari, C., and Pigni, F. 2020. Profiting from big data analytics: The moderating roles of industry concentration and firm size. *International Journal of Production Economics* **229**.

Rammer, C., and Es-Sadki, N. 2022. Using Big Data for Generating Firm-Level Innovation Indicators–A Literature Review. *ZEW-Centre for European Economic Research Discussion Paper* **22-007**.

Reid, G. and Smith, J. 2000. What Makes a New Business Start-Up Successful? *Small Business Economics* **14** 165–182.

Sena, V., Bhaumik, S., Sengupta, A., and Demirbag, M. 2019. Big Data and Performance: What Can Management Research Tell Us? *British Journal of Management* **30** 219–228.

Suoniemi, S., Meyer-Waarden, L., Munzel, A., Zablah, A., and Straub, D. 2020. Big data and firm performance: The roles of market-directed capabilities and business strategy. *Information & Management* **57**.

Tambe, P. (2014). Big Data Investment, Skills, and Firm Value. *Management Science* **60**, 1452–1469.

Wallace, N. 2020. European Union Gets in the Venture Capital Game. *Science* **368** 120–121.

Wang, Y., Kung, L., Gupta, S., and Ozdemir, S. 2019. Leveraging Big Data Analytics to Improve Quality of Care in Healthcare Organizations: A Configurational Perspective. *British Journal of Management* **30** 362–388.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., and Childe, S. J. 2017. Big Data Analytics and Firm Performance: Effects of Dynamic Capabilities. *Journal of Business Research* **30** 356–365.

Winter, S. G. 2003. Understanding Dynamic Capabilities. *Strategic Management Journal* **24** 991–995.

Yu, W., Jacobs, M. A., Chavez, R., and Feng, M. 2019. Data-driven Supply Chain Orientation and Financial Performance: the Moderating Effect of Innovation-focused Complementary Assets. *British Journal of Management* **30** 299–314.

Appendix

Table A.1 Description of Variables

Variable Name	Variable Description
BDA	The start-up is using BDA methods.
Founder age	Age of the founder when founding the start-up. For teams, it is the average age.
Female	At least one founder is female.
Foreign	At least one founder is not German.
Team	The start-up was founded by more than one person.
Education	Highest educational level of the founders, ranging from 1 to 9 corresponding to whether the founders have (1) no degree, a degree from (2) an apprenticeship, (3) a vocational school, (4) a master school, (5) an academy for civil servants, (6) a vocational academy, (7) a technical college, (8) a university, or (9) a PhD/ habilitation.
Industry experience (t)	Years of experience in the industry the start-up operates in.
Founding experience	At least one founder has founded a start-up before.
Opportunity-driven	The start-up was founded to implement a business idea, or to increase earnings, or to allow more self-determined work.
Urban	The start-up was founded in a city.
East Germany	The start-up was founded in East Germany (Brandenburg, Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, Thuringia).
Firm age (t)	Age of the firm in the respective year.
Founding year	Year the start-up was founded in.
Industry	The main industry the start-up operates in. Industry classification is defined according to WZ2008 with aggregation to 11 sectors (cutting-edge tech manufacturing, high-tech manufacturing, technical services, software, low-tech manufacturing, knowledge-intensive services, other company services, creative services, other services, construction, and trade).
Total costs (t)	Total start-up costs in the respective year. They are composed of investment costs, R&D costs, and operating costs.
Investment costs (t)	Investment costs of the start-up in the respective year.
R&D costs (t)	Research and Development expenditures of the start-up in the respective year.
$R\&D_d$ (t)	The start-up has conducted R&D in the respective year.
Operating costs (t)	Operating costs of the start-up in the respective year.
Material costs (t)	Material costs of the start-up in the respective year.
Personnel costs (t)	Personnel costs of the start-up in the respective year.
Other operating costs (t)	Operating costs that do not fall in the category of material costs or personnel costs of the start-up in the respective year.
Sales (t)	Sales of the start-up in the respective year.
Profits (t)	Profits of the start-up in the respective year.
Profit indicator	Indicating whether the start-up has never (1), sometimes (2), or always made positive profits (3).
Sales variance	Normalized intra-firm variance of sales.
Profit variance	Normalized intra-firm variance of profits.
Survival	The start-up is still operating in 2021.
Employee growth	CAGR of employee growth.
Number of employees when founding	Number of employees at the start-up's foundation date.
Number of employees (t)	Number of employees of the start-up in the respective year.
Patent	The start-up has already held a patent before the foundation.
Market novelty (t)	The start-up has introduced a market novelty in the respective year.
VC	The start-up has received at least once venture capital funding.
Bank	The start-up has received at least once funding from a bank.

Table A.2 Distribution of Start-ups across Industries

Industry	Frequency	Percent
Cutting-edge tech manufacturing	285	7.68
High-tech manufacturing	236	6.36
Technical services	765	20.63
Software	332	8.95
Low-tech manufacturing	365	9.84
Knowledge-intensive services	354	9.54
Other company services	265	7.14
Creative services	231	6.23
Other services	207	5.58
Construction	323	8.71
Trade	346	9.33

Industries are aggregated according to the WZ2008 definition.

Table A.3 Summary Statistics: Firm and Founder Characteristics

Variable	Mean	SD	Min.	Max.
BDA	0.15	0.36	0	1
Founder age	41.79	9.99	18	95
Female	0.18	0.39	0	1
Foreign	0.11	0.31	0	1
Team	0.34	0.47	0	1
Education	5.58	2.62	1	9
Industry experience	15.19	10.06	1	60
Founding experience	0.44	0.50	0	1
Opportunity-driven	0.83	0.37	0	1
Urban	0.39	0.49	0	1
East Germany	0.13	0.34	0	1
Founding year	2.27	1.77	0	6
Observations				3,709

All variables are binary variables except for *founder age*, *industry experience*, and *firm age*, which are measured in years, *education*, which is an ordinal variable, and *number of employees*. Summary statistics are calculated based on the collapsed sample, i.e., one observation per start-up is used, except for *number of employees* and *market novelty*, as they vary over time.

Table A.4 Logistic Regression Results on Firm Characteristics

	BDA [0/1]
Founder age	0.987**
	[0.975, 0.999]
Female founder	0.806
	[0.619, 1.049]
Foreign founder	1.099
	[0.815, 1.482]
Team	1.390***
	[1.128, 1.712]
Academic founder	1.475***
	[1.170, 1.860]
Industry experience	0.989*
	[0.977, 1.001]
Serial entrepreneur	1.494***
	[1.216, 1.837]
Opportunity Driven	1.197
	[0.893, 1.604]
City	1.128
	[0.928, 1.371]
East Germany	0.830
	[0.618, 1.116]
Founding year	1.059*
	[0.999, 1.123]
Constant	-117.948*
	[-236.298, 0.403]
Wald-test sector controls	74.38***
Observations	3,670
R^2	0.077
Mean of dependent variable	0.152

The table shows the odds ratios of the logistic regression on firm characteristics with 95% confidence intervals in brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5 Heckman MLE Regression Results on Start-up Sales and Profits using LASSO

	Sales		Profits	
	Sales [0/1]	Ln(sales)	Profits [0/1]	Ln(profits)
Big data	-0.182**	0.029	-0.123***	0.084
	[-0.323, -0.040]	[-0.084, 0.141]	[-0.217, -0.030]	[-0.056, 0.224]
Founder age	-0.010***	-0.001	-0.014***	0.008***
	[-0.017, -0.003]	[-0.006, 0.005]	[-0.018, -0.010]	[0.003, 0.013]
Female founder	-0.090	-0.282***	-0.164***	-0.112*
	[-0.216, 0.036]	[-0.382, -0.182]	[-0.249, -0.078]	[-0.230, 0.006]
Foreign founder	0.012	-0.028	-0.034	0.134*
	[-0.160, 0.183]	[-0.162, 0.102]	[-0.137, 0.069]	[-0.005, 0.274]
Team	-0.175**	0.325***	0.011	0.200***
	[-0.314, -0.035]	[0.227, 0.424]	[-0.077, 0.099]	[0.090, 0.309]
Industry experience	0.005	0.016***	0.014***	
	[-0.002, 0.012]	[0.011, 0.021]	[0.010, 0.018]	
Serial entrepreneur	-0.135**	0.080*	-0.140***	0.046
	[-0.243, -0.027]	[-0.003, 0.162]	[-0.209, -0.072]	[-0.048, 0.141]
Opportunity driven	-0.110	-0.020	-0.103**	0.030
	[-0.252, 0.032]	[-0.119, 0.079]	[-0.185, -0.021]	[-0.077, 0.138]
City	-0.184***	0.097**	-0.110***	0.087*
	[-0.285, -0.082]	[0.019, 0.174]	[-0.175, -0.046]	[-0.004, 0.178]
East Germany	-0.024	-0.181***	-0.053	-0.112*
	[-0.168, 0.121]	[-0.292, -0.071]	[-0.141, 0.035]	[-0.226, 0.002]
Founding year	-0.294***		-0.171***	-0.014
	[-0.348, -0.240]		[-0.194, -0.147]	[-0.045, 0.016]
Number of employees	0.148***	0.079***	0.011***	0.050***
	[0.111, 0.184]	[0.062, 0.096]	[0.003, 0.018]	[0.039, 0.061]
R&D$_d$	-0.221***	0.163***	-0.256***	0.165***
	[-0.322, -0.120]	[0.083, 0.243]	[-0.326, -0.186]	[0.065, 0.265]
Market novelty	0.068***	0.015	-0.036**	0.085***
	[0.018, 0.117]	[-0.020, 0.050]	[-0.067, -0.005]	[0.038, 0.132]
No degree	-0.373***		-0.246***	-0.127
	[-0.590, -0.156]		[-0.416, -0.076]	[-0.358, 0.104]
Apprenticeship	-0.111	-0.062	-0.105*	
	[-0.276, 0.054]	[-0.181, 0.058]	[-0.217, 0.006]	
Vocational school	-0.090	-0.207	-0.111	
	[-0.447, 0.266]	[-0.457, 0.043]	[-0.308, 0.086]	
Master school	-0.012	-0.078	-0.045	0.064
	[-0.191, 0.167]	[-0.198, 0.043]	[-0.160, 0.070]	[-0.041, 0.170]
Academy for civil servants	0.446	-0.870*	-0.383	0.591
	[-0.218, 1.111]	[-1.889, 0.149]	[-1.181, 0.415]	[-0.506, 1.689]
Vocational academy	-0.125	0.058	-0.041	-0.134
	[-0.495, 0.245]	[-0.294, 0.411]	[-0.274, 0.193]	[-0.450, 0.183]
Technical college	-0.099	0.066	-0.068	0.047
	[-0.272, 0.075]	[-0.054, 0.186]	[-0.181, 0.045]	[-0.072, 0.166]
University	-0.114	0.055	-0.171***	0.201***
	[-0.277, 0.049]	[-0.065, 0.175]	[-0.286, -0.056]	[0.078, 0.325]
PhD/ habilitation	-0.292***	-0.038	-0.294***	0.171*
	[-0.480, -0.104]	[-0.205, 0.130]	[-0.442, -0.145]	[-0.030, 0.371]
Other education	-0.809***	0.051	-0.852***	0.908***
	[-1.283, -0.335]	[-0.450, 0.553]	[-1.352, -0.352]	[0.256, 1.560]
Constant	1.960***	12.397***	344.857***	39.327
	[1.415, 2.505]	[12.112, 12.683]	[298.079, 391.634]	[-22.534, 101.187]
Sector controls	223.09***		574.03***	
Year controls	810.38***		145.75***	
Observations	14,266	14,266	13,327	13,327
Mean of dependent variable	0.94	11.26	0.69	10.34

The table shows the coefficients of the Heckman Model using the Farbmacher extension with 95% confidence intervals in brackets. The model results in the following exclusion restrictions: *founding year* and *no degree* for the model on sales and *industry experience*, *apprenticeship*, and *vocational school* for the model on profits.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6 Pooled Linear Regression Results on Profits

	ln(sales)	Profit
BDA	-0.173	960.458
	[-0.421, 0.074]	[-17785.554, 19706.471]
R&D$_d$	-0.076	-14529.628
	[-0.298, 0.147]	[-39688.793, 10629.538]
Market novelty	0.186***	-198.932
	[0.083, 0.289]	[-18070.400, 17672.536]
Number of employees	0.082***	3649.137
	[0.058, 0.105]	[-4586.976, 11885.251]
Constant	9.165***	-12328.441
	[7.691, 10.640]	[-122758.103, 98101.221]
Wald test firm controls	13.00***	3.88***
Wald test sector controls	8.23***	1.97**
Wald test year controls	58.51***	8.71***
Observations	14,228	12,880
R^2	0.182	0.043
Mean of dependent variable	11.2617	33,048.81

95% confidence intervals in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table shows the regression coefficients of the linear regression on the start-ups' logarithmized sales and profits with 95% confidence intervals in brackets. The sample is entropy balanced using weights for the founders' age, gender, nationality, team composition, educational background, industry experience, founding experience, motive, geographic location, sector, and founding year.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7 Pooled Linear Quantile Regression Results on Start-ups' Sales and Profits using Weights

	0.9 ln(sales)	0.1 ln(sales)	0.9 ln(profits)	0.9 ln(losses)
BDA	0.206***	-0.598	-0.017	0.619***
	[0.122,0.291]	[-1.438,0.242]	[-0.102,0.067]	[0.378,0.860]
Firm age	0.326***	0.978***	0.222***	-0.666***
	[0.298,0.354]	[0.764,1.192]	[0.200,0.245]	[-0.768,-0.563]
BDA × firm age	0.088***	0.211	0.077***	0.146**
	[0.040,0.137]	[-0.118,0.541]	[0.028,0.127]	[0.001,0.292]
Number of employees at foundation	0.165***	0.189***	0.090***	0.211***
	[0.156,0.175]	[0.130,0.248]	[0.055,0.126]	[0.159,0.263]
Constant	12.274***	3.555***	11.005***	6.236***
	[11.992,12.556]	[1.374,5.737]	[10.510,11.500]	[5.233,7.238]
Wald-test firm controls	21.53***	6.53***	14.82***	29.40***
Wald-test sector controls	20.53***	18.36***	21.27***	46.01***
Wald-test year controls	2.61***	10.11***	2.29***	11.52***
Observations	16,410	16,410	14,451	18,520
Pseudo R^2	0.154	0.158	0.054	0.070
Mean of dependent variable	11.26	11.26	7.03	1.77

The table shows the regression coefficients of the quantile regression on the 0.1 and 0.9 quantiles on the logarithmized sales, profits, and losses with 95% confidence intervals in brackets. The sample is entropy balanced using weights for the founders' age, gender, nationality, team composition, educational background, industry experience, founding experience, motive, geographic location, sector, and founding year. Standard errors are not clustered. The variable *agefirm* is demeaned to ensure a meaningful coefficient of variable *BDA*.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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