Big Data Sources and Digitalization Awareness: a Dynamic Capabilities Perspective

Mattia Pedota
Department of Management, Economics and Industrial Engineering
mattia.pedota@polimi.it

Abstract

Recent research frames big data as a resource enhancing dynamic capabilities through improved prediction, decision-making and data-driven innovation. In contrast, this study frames big data as an evolutionary driver that channels firms' knowledge and attention in specific directions, implying that firms need multiple big data sources to be receptive and dynamically capable. I apply this framework to the context of the digital revolution and focus on the impact of big data on firms' digitalization priorities. By leveraging a large-scale survey of more than twenty thousand Italian firms of all sizes, I find that big data improves the digitalization awareness of firms only if they gather big data from more than one source (otherwise, counterintuitively, it may even decrease it). I also find a positive effect of source variety both on the likelihood of prioritizing individual digitalization factors and on the variety of digitalization factors prioritized. Such effects appear to be stronger for small firms relative to their larger counterparts. Given the path dependence of digitalization trajectories, these findings have relevant policy implications in the context of initiatives like the European strategy for data and the SME strategy for a sustainable and digital Europe.

Big Data Sources and Digitalization Awareness: a Dynamic Capabilities Perspective

Abstract

Recent research frames big data as a resource enhancing dynamic capabilities through improved prediction, decision-making and data-driven innovation. Instead, this study frames big data as an evolutionary driver that channels firms' knowledge and attention in specific directions, implying that firms need multiple big data sources to be receptive and dynamically capable. I apply this framework to the context of the digital revolution and focus on the impact of big data on firms' digitalization priorities. By leveraging a large-scale survey of more than twenty thousand Italian firms of all sizes, I find that big data improves the digitalization awareness of firms only if they gather big data from more than one source (otherwise, counterintuitively, it may even decrease it). I also find a positive effect of source variety both on the likelihood of prioritizing individual digitalization factors and on the variety of digitalization factors prioritized. Such effects appear to be stronger for small firms relative to their larger counterparts. Given the path dependence of digitalization trajectories, these findings have relevant policy implications in the context of initiatives like the European strategy for data and the SME strategy for a sustainable and digital Europe.

Keywords: attention; artificial intelligence; big data; digitalization; digital revolution; dynamic capabilities; knowledge

1. Introduction

Recent developments in artificial intelligence enable multipurpose exploitations of data, spurring new business models and improving the efficiency of existing ones (Garbuio & Lin, 2019; Liang et al., 2018; Reim et al., 2020). As algorithms and techniques for statistical inference improve, the world becomes more interconnected: besides people producing textual, numerical and audiovisual data through their online activities, increasingly many devices act as vehicles for data sharing and transmission (Guo et al., 2013). The surge in data availability coupled with the development of methods for data analysis easily explains the ubiquity of big data.

The baseline definition of big data refers to structured or unstructured data that is too large for traditional data-processing software (Lansley & Longley, 2016; Sestino et al., 2020). However, the definition can be broadened to include the related data analytics, storage and management (Boyd & Crawford, 2012; Wamba et al., 2015). In line with this holistic approach, the present work assumes that the act of sourcing big data is almost always coupled with some analysis. Accordingly, I hereafter refer to big data as a singular expression denoting the whole phenomenon rather than the data themselves.

Big data is acknowledged to improve various aspects of firm performance (Brynjolfsson & McElheran, 2016; McAfee et al., 2012). In the realm of innovation, recent research suggests that it can enhance a firm's dynamic capabilities (Teece et al., 1997) through improved prediction, data-driven innovation, and better sensing of opportunities (Conboy et al., 2020; Côrte-real et al., 2017; Rialti et al., 2019). However, while this covers the resource-based side of dynamic capabilities, it overlooks its evolutionary-based nuances, such as bounded rationality, learning and path dependence (Barreto, 2010; Nelson et al., 2018). Big data underlies information, which channels organizational knowledge and attention and contributes to shaping a firm's perception of the current and future states of the environment (Ocasio, 1997; Zollo & Winter, 2003). Thus, I suggest that the variety and typology of big data sources are likely to affect a firm's ability to sense opportunities and reconfigure resources in specific directions. Furthermore, while expecting these effects to hold for all firms, I suggest they may be even stronger for small firms: being characterized by lower bureaucratization and less coexisting perspectives (i.e. less employees), small firms may be more susceptible to both informational gains and potential biases from big data.

The digital revolution constitutes the ideal context to study these dynamics, as it requires multifaceted adaptation. Several studies suggest that not only technological innovation, but also strategic planning, collaboration, skill sourcing and skill consolidation are essential for firms to thrive in the digital era (Ciarli et al., 2021; Pedota et al., 2023). Lacking big data may weaken firms' (dynamic) capabilities of sensing the potential of advanced digital and automation technologies and reconfiguring their strategy and resources accordingly. Also, extracting big data from only few sources may bias firms toward few digitalization factors at the expense of others. Due to the path dependence of knowledge accumulation (Cohen & Levinthal, 1990; Nelson & Winter, 1982) and the hierarchical nature of digital technologies (European Patent Office, 2017; Zolas et al., 2021), both effects may engender self-reinforcing dynamics, ultimately shaping the dynamic efficiency of firms (and in turn productive systems). Thus, investigating the interaction between big data and dynamic capabilities in the context of the digital revolution carries both theoretical and practical relevance.

This study leverages a large cross-sectional survey developed by the Italian National Institute of Statistics (ISTAT) in 2018, covering 21,934 Italian firms of all sizes. Among other topics, the survey investigates whether firms ascribe competitive relevance to digitalization factors from a given list, namely infrastructure, fiscal incentives supporting digitalization, digital initiatives of the government, capability of networking with other firms and research centers, skill sourcing, skill consolidation, development of a digitalization strategy. Dependent variables capture whether respondents flagged a given digitalization factor as important for the competitiveness and development of the firm in the following two years. Regressors of interest capture whether respondents sourced big data from social media, sensors, portable devices and/or other sources in the year before (thus, although the survey is cross-sectional, the temporal antecedence of regressors is ensured). Control variables include firm size, industry, geographic location and degree of ICT intensity, at the highest level of detail provided by the database. With this setup, I perform a series of multiple logistic and ordered logistic regressions aimed at determining: 1) whether big data enhances firms' awareness of digitalization factors; 2) if and to what extent the kind and/or variety of sources of big data make a difference in the kind and/or variety of digitalization factors prioritized; 3) whether small firms exhibit any relevant difference in such dynamics relative to their larger counterparts.

I find that firms using big data (regardless of the source) have a significantly lower probability of not regarding any listed digitalization factor as important, as well as a significantly lower probability of being unable to identify priorities among the listed factors. This is coherent with the enhancing effect of big data on dynamic capabilities. However, most interestingly, firms that extract big data from only one source have a level of digitalization awareness comparable to that of firms that do not use big data at all (and, in some cases, even lower). Furthermore, not only does source variety increase the probability of regarding any given digitalization factor as important, but it also increases the variety of digitalization factors considered important. As hypothesized, most of these effects are stronger for small firms rather than medium and large ones.

From a theoretical viewpoint, this study contributes to the literature on big data and dynamic capabilities by framing big data as an evolutionary driver rather than a mere resource, thus reconciling the resource-based and the evolutionary traditions in dynamic capabilities research. From this vantage point, the study provides evidence that the typology of big data sources that firms rely on shapes the path ahead. Consequently, firms need big data source variety to be receptive to multiple facets of the environment and thereby dynamically capable. From a practical viewpoint, results highlight an additional reason why big data is a crucial enabler of the digital revolution: not only is big data complementary to other digital technologies functionally, but it also guides firms' digitalization trajectories by shaping firms' digitalization awareness. Furthermore, results draw managerial, entrepreneurial, and institutional attention not only to the opportunities stemming from big data utilization, but also to the traps inherent in relying on too few big data sources. Given the path dependence of digitalization trajectories, this realization matters for both firms and economic systems, thus bearing relevance to initiatives like the “European strategy for data” (European Commission, 2020a).

2. Theoretical Background

2.1. Dynamic Capabilities and Big Data

Dynamic capabilities are defined as the firm’s ability to integrate, build and reconfigure internal and external competences to address rapidly changing environments (Teece et al., 1997: p. 516). The concept originally stressed the interdependence between the position of the firm (e.g. its asset endowment), the processes for coordination, learning and reconfiguration, and the paths lying ahead (Teece et al., 1997). Subsequent elaborations better specified the business dimensions involved (Eisenhardt & Martin, 2000) and the different functions of dynamic capabilities: sensing, seizing and transforming. Sensing refers to the ability by the firm to scan the environment for opportunities, seizing indicates the ability to exploit them, while transforming encompasses the learning and resource reconfiguration mechanisms underlying competitive advantage renewal (Teece, 2007).

Dynamic capabilities are deeply rooted in the evolutionary view (Nelson & Winter, 1982), stressing path dependence, learning and innovation. They can be regarded as high-level collections of routines through which a firm dynamically adjusts its lower-level routines and resources as a function of its perceived position in the environment (Winter, 2003). In this respect knowledge is key, as it shapes the firm’s awareness of the environment and its expectations about it, including the perceived strategic importance of change (Zollo & Winter, 2002). Furthermore, knowledge underlies the quintessential dynamic capabilities: those related with learning and innovation (Nelson et al., 2018). Absorptive capacity, the ability of the firm to recognize the value of new knowledge, assimilate it and apply it to commercial ends (Cohen & Levinthal, 1990), is often regarded as a dynamic capability itself (Volberda et al., 2010; Zahra & George, 2002). Knowledge articulation and, to a greater extent, knowledge codification enhance organizational learning, by fostering knowledge sharing and facilitating the identification of causal mechanisms (Nonaka, 1994; Zollo & Winter, 2002). Various studies provide evidence that dynamic capabilities for knowledge absorption, integration and reconfiguration are drivers of success in innovation-based Schumpeterian competition (Danneels, 2008, 2012; Nelson et al., 2018; Verona & Ravasi, 2003).

The interplay between dynamic capabilities, knowledge and innovation has recently drawn attention due to globalization and digitalization. Key knowledge is now dispersed across geographies, sources and media. This reinforces the open innovation paradigm (Bogers et al., 2018; Chesbrough, 2003), by making it easier and more beneficial to tap into external sources of knowledge. As a result, dynamic capabilities become even more important: sensing capabilities become critical to identify relevant external knowledge, as well as licensing out opportunities; seizing capabilities are needed to set incentives, processes and governance mechanisms to leverage collaboration; transformation capabilities become key factors underlying the dynamic integration of internal and external knowledge (Bogers et al., 2019).

In this highly globalized and digitalized context, big data bears relevance to dynamic capabilities due to its knowledge-related properties (Ferraris et al., 2018). When complemented by machine learning techniques (Zhou et al., 2014), analytical skills (LaValle et al., 2011) and a supportive culture (Frisk & Bannister, 2017), big data brings significant value to firms (Ciampi et al., 2021; McAfee et al., 2012; Wamba & Mishra, 2017). Depending on the effectiveness of analytical capabilities and diffusion mechanisms, big data may translate into knowledge that significantly affects the kind and quality of decisions. When managed through appropriate routines, such knowledge can be leveraged in different instances (Ervelles et al., 2016). Thus, big data has the potential to improve several functional areas, ranging from marketing to supply chain management (Chehbi Gamoura et al., 2020; Choi & Chen, 2021; Lo & Campos, 2018).

Given these features, recent research on big data often employs a dynamic capabilities perspective. Empirical evidence has been offered that process-oriented dynamic capabilities mediate the relationship between big data analytics capabilities and firm performance (Wamba et al., 2017). Furthermore, several studies underscore the importance of big data for innovation, both directly and indirectly. By improving knowledge building and prediction, big data makes it easier to conduct R&D (a first-order dynamic capability) and engage in innovative projects, which often entail risk and uncertainty (Niebel et al., 2019). Big data also enables data-driven innovation, by providing direct inputs (e.g. customer data) for the innovation process (Bresciani et al. 2021; Sultana et al., 2021). More indirectly, big data enhances the ability by firms to sense the opportunities that come up to them and reconfigure resources accordingly (Conboy et al., 2020; Mikalef et al., 2021). Improved predictive capabilities also heighten the promptness and effectiveness of firms' reactions to sudden change, leading to information-driven competitive advantage (Côrte-real et al., 2017; Rialti et al., 2019).

These studies have marked important steps toward understanding the relationship between big data and dynamic capabilities. However, while dynamic capabilities bring together resource-based and evolutionary views (Barreto, 2010), current treatments of big data in relation to dynamic capabilities lean heavily toward the former. Big data tends to be regarded as a resource to be deployed through an ad hoc bundle of IT and managerial capabilities, often referred to as big data analytics capabilities (Mikalef et al., 2020). When adequately complemented, big data is acknowledged to improve dynamic capabilities through better decision-making, prediction, and responsiveness, enabling firms to navigate through change. However, this widespread framing has so far overlooked the more evolutionary-based aspects of dynamic capabilities: those related with bounded rationality, learning and path dependence. Even assuming adequate complements, I argue that the postulation that big data improves dynamic capabilities generically may be too coarse-grained. Given bounded rationality, big data is also a powerful channeler of organizational attention and knowledge search. I advance that the number and typology of big data sources a firm relies on is likely to affect its perception of the environment (including its future states), and in turn its ability to sense opportunities and reconfigure its resources.

Thus, on the one hand, I aim to provide further empirical evidence on the fact that big data enhances firms' dynamic capabilities. Although the mechanism is theoretically clear, still relatively few studies address it empirically. On the other hand, more importantly, I aim to take a step more, by investigating the role of big data source typology and big data source variety in shaping the dynamic capability of sensing opportunities and adapting to technological change. To this end, the ongoing digital revolution constitutes the ideal context, as it marks a fundamental shift in the way firms operate, compete and innovate (Osterrieder et al., 2020; Stornelli et al., 2021). Adaptation to the digital revolution requires firms to innovate technologically, while also adapting operationally and strategically (Ciarli et al., 2021; Pedota et al., 2023). In other words, firms need to simultaneously advance along a series of digitalization factors. I propose that not only the mere adoption of big data, but also the typology and variety of big data sources play a role in this transition, dynamically shaping the digitalization path of firms (and in turn economic systems).

2.2. Hypotheses Formulation

Knowledge has a well-established role in shaping firms' evolutionary trajectories (Greiner, 1998; Scott & Bruce, 1987). The bulk of current knowledge is a crucial driver of the intensity and direction of the process of knowledge search. Typically referred to as the absorptive capacity of a firm (Cohen & Levinthal, 1990; Zahra & George, 2002), it is a fundamental enabler of innovation. As stressed by more recent conceptual refinements, the first component of absorptive capacity is the ability to recognize the value of new knowledge (Todorova & Durisin, 2007). Without prior related knowledge, a firm lacks the very cognitive and informational prerequisites to grasp the importance of further knowledge. This is not only due to the difficulty of integrating such new knowledge into extant cognitive structures, but also to the incapability of estimating the prospective implications of that knowledge. Firms enact a series of external routines aimed at enhancing their value recognition capabilities (Lewin et al., 2011). Many of such routines are data-driven, including the mining of patent literature (Cohen et al., 2002) and the administration of end user surveys (Kohli et al., 1993).

Hence, I argue that big data is likely to play a significant role in the ability of firms to recognize the value of digitalization factors. First, given that big data comprises both data sourcing and analytics (Boyd & Crawford, 2012; Wamba et al., 2015), it constitutes a form of embedded knowledge. Firms sourcing large quantity of data also tend to adopt advanced analytical techniques to exploit them (with varying extents of success). Thus, they possess knowledge on digitalization and data analytics. Coherently with the theory of absorptive capacity, this is likely to facilitate the recognition of the value of digital knowledge and thereby the importance of digitalization factors. Second, big data translates into information. By processing large quantities of data, firms can better intercept technoeconomic trends (Perez, 2010), as well as technological trajectories and macrotrajectories (Dosi, 1982; Pedota et al., 2021). At the same time, they are better positioned to assess their strengths and weaknesses relative to their competitive environment. Therefore, big data adopters are also likelier to know which digitalization factors to prioritize based on their needs, which leads to targeted adaptation and resource reconfiguration efforts. Possible examples are the preemptive identification of a competence-destroying technology (Abernathy & Clark, 1985) triggering the prioritization of skill sourcing, or the discovery of a market opportunity foregrounding product-oriented collaboration.

This leads to the formulation of the following two hypotheses:

HP1a: Big data allows firms to recognize the competitive relevance of digitalization factors.

HP1b: Big data allows firms to identify which digitalization factors to prioritize.

Extant research has suggested that big data may improve dynamic capabilities through information-related advantages (Conboy et al., 2020; Côrte-real et al., 2017; Rialti et al., 2019), but it is largely silent on the role of big data source typology and variety in this respect. I maintain that different big data sources lead to different kinds of information, which may affect firms' perception in different ways. Sources of information have been shown to play a pivotal role in a number of different areas, including innovation development (Medase & Abdul-Basit, 2020). Firms relying on a larger variety of sources of information are likelier to develop innovations considered as national or world premieres (Amara & Landry, 2005), with probable underlying reasons being the non-redundancy of the information obtained (Burt, 1992) and the creativity-enhancing properties of knowledge diversity (Taylor & Greve, 2006).

I propose that a higher variety of sources of information provides managers in the firm with heterogeneous knowledge, which stimulates their creativity by expanding their domain-relevant knowledge (Amabile, 1983; Amabile & Pratt, 2016). With more boundary-spanning domain-relevant knowledge, managers are likely to give weight to factors that less knowledgeable individuals would disregard. This is primarily a consequence of the notion that knowledge attracts similar knowledge (Cohen & Levinthal, 1990; Zahra & George, 2002): with more heterogeneous knowledge, managers have a larger number of hooks to recognize the value of further knowledge. However, it is also a consequence of the higher number of connections that they can make and possibly their higher motivation: with boundary-spanning knowledge at their disposal, managers can recognize the value of further knowledge not only for its immediate relevance to the current bulk of knowledge, but also for its prospective recombinatory potential (Fleming, 2001; Pedota & Piscitello, 2022). Furthermore, a higher level of creativity typically entails a surge in intrinsic motivation and positive affect (Amabile & Pratt, 2016). This may further enhance the ability by firms to recognize the value of new knowledge, as eagerness to learn is a crucial component of absorptive effort (Song et al., 2018; Srivastava et al., 2015).

Information also shapes attentional focus. Bounded rationality (Simon, 1991) implies that managers can never attain a complete and objective representation of the world (Bogner & Barr, 2000; Fiol & O'Connor, 2003). Instead, they construct a subjective framing of the external environment based on drivers like firm context (Ocasio, 1997), industry context (Sutcliffe & Huber, 1998) and past performance (March & Shapira, 1992). Depending on whether they perceive the external environment as malleable or fixed, they may adhere to proactive or deterministic causal logic, whereby they attempt to shape the environment through strategic action or merely react to environmental signals, respectively (Nadkarni & Barr, 2008). In both cases, available information forms the basis on which the subjective representation is built, thereby molding perception and in turn action. Low breadth of information may constrain the attentional focus of firms within narrow limits, with direct implications on the comprehensiveness of their environmental perception.

I argue that, by constituting information, big data shapes both knowledge and attention within firms. When relying on a variety of big data sources, firms benefit from heterogeneous knowledge flows, and they are likelier to be receptive to different facets of the environment. Thus, they are better positioned to sense and dynamically build on a wider range of opportunities. In the context of the digital revolution, this may include a reconfiguration of the skill base, the formulation of a digitalization strategy and the engagement in collaborative opportunities. Furthermore, heterogeneous knowledge boosts organizational creativity, prompting firms to consider a wider range of possibilities (e.g. visionary digitalization strategies or ambitious collaboration plans). Both these effects are likely to increase their proclivity to recognize the value and competitive relevance of digitalization factors. On the side of attention, I maintain that firms relying on various big data sources have a subjective representation of the environment that is richer and more complex than the one they would have by relying on fewer sources. This is likely to reinforce the perceived importance of digitalization factors, both due to a more comprehensive understanding of the current state of the environment and a better ability to anticipate its future states. Therefore, I suggest that relying on a higher variety of big data sources has a twofold effect: not only does it increase the likelihood of considering any given digitalization factor as important, but it also augments the variety of digitalization factors considered important. This is condensed in the following hypotheses:

HP2a: As the variety of sources of big data increases, both the awareness and the capability by firms to identify competitively relevant digitalization factors increases.

HP2b: As the variety of sources of big data increases, the variety of digitalization factors considered important for competitive advantage increases.

I also observe that small firms have a set of distinctive features relative to their larger counterparts. Younger and less resourceful, their geographical, lateral and vertical scope is typically limited. They have a lower number of employees, a shorter hierarchical chain and less structured mechanisms for carrying out core and support activities, including knowledge gathering and decision-making (Gibcus et al., 2009; Penn et al., 1998). For all these reasons, they are not (yet) stuck in path-dependent trajectories of organizational development. Instead, they tend to be relatively flexible, forward-looking and receptive to the surrounding environment (Csath, 2022), as well as proactive and entrepreneurial (Miller, 2011; Mthanti & Ojah, 2017). Furthermore, lacking dedicated staff to scan the environment for information, they often have to rely on heuristics and rules of thumb for making decisions, which makes them prone to biases (Busenitz & Barney, 1997; Gibcus et al., 2009).

This can be condensed into two relevant preculiarities of small firms. On the one hand, they have more to gain from big data. Having a lower initial bulk of information, the marginal benefit of big data is likely to be higher in their case. In other terms, ceteris paribus, I expect a higher informational difference between a small firm with big data and one without than between a large firm with big data and one without. Considering the awareness of competitively relevant digitalization factors, the effect is made even stronger by the entrepreneurial orientation and receptivity of small firms to the external environment, which increases the likelihood of them capitalizing on big data to get acquainted with different digitalization factors. On the other hand, as they are smaller, less structured and more prone to bias, I also expect small firms to benefit from big data source variety to a higher extent. This is because big data is likely to constitute a relatively large proportion of their information, and such information is going to circulate many times among a restricted number of employees. When lacking variety in data sources, this may engender a sort of echo chamber where priorities and decisions are dictated based on a very partial snapshot of the world. Unlike their larger counterparts, small firms cannot rely on a large corpus of alternative information sources, nor do they have a variety of coexisting perspectives (given the small number of employees). Therefore, I put forward the following hypotheses:

HP3a: Both the awareness-enhancing effect (HP1a) and the identification-enhancing effect (HP1b) of big data are stronger for small firms relative to medium and large ones.

HP3b: The effect of big data source variety on the awareness and capability to identify competitively relevant digitalization factors (HP2a) is stronger for small firms relative to medium and large ones.

HP3c: The effect of big data source variety on the variety of digitalization factors considered important for competitive advantage (HP2b) is stronger for small firms relative to medium and large ones.

3. Empirical Analysis

3.1. Sample Description

The sample comes from the "survey on information and communication technologies in firms", administered by the Italian National Institute of Statistics in 2018 in collaboration with the European Commission[1]. Its objective is to provide comprehensive information on the integration of ICT technologies in Italian companies that have a minimum workforce of 10 employees. Data cover a

[1] The entire questionnaire is available at the following link: https://listarilevazioni.istat.it

range of topics, including firms' training of ICT staff, utilization of e-commerce and social media platforms, implementation of electronic invoicing, approach toward the digital revolution, and commitments to emerging technologies (including big data). The survey is organized in four sections: general information (A), ICT competences (B), internet usage and connection (C), cloud computing services (D), 3D printing (E), robotics (F), big data analytics (G), invoicing (H), sales through ICT networks (I), determinants of the firm's digital transformation (J). The present study takes its key variables from sections G and J, along with various controls from section A.

The survey aims at the population of Italian firms with at least 10 employees, from any of the following sectors (letters refer to the Italian ATECO classification): manufacturing (C); supply of electricity, gas, steam and air conditioning (D); water supply, sewerage and waste management (E); construction (F); wholesale and retail trade and repair of motor vehicles and motorcycles (G); transport and storage (H); accommodation and catering services (I); information and communication services (J); real estate activities (L); professional, scientific and technical activities (M, except division 75: veterinary services); rental, travel agencies and business support services (N); repair of computers and communications equipment (group 95.1 of section S: other service activities).

The whole population of firms with at least 250 employees is included. Firms with a lower number of employees are stratified random sampled according to industry (at given levels of aggregation), number of employees (10-49, 50-99, 100-249) and geographical location (northeast, northwest, center, south, islands). The total number of firms is 21,934. To segment the sample according to size, I followed the revenue criterion of the EU recommendation 2003/361 and recognized firms with revenues lower than or equal to 10 million as small firms, firms with revenues between 10 and 50 million as medium-sized firms, and the remaining firms as large ones. This way, I identified 13,761 small firms, 4,716 medium-sized firms and 3,457 large firms.

3.2. Methodology

To test my hypotheses, I relied on selected parts of sections G and J of the survey. Subsection G1 requires firms to indicate whether they have gathered big data from sensors, portable devices, social media and/or "other sources" in year 2017. Subsection J3 requires firms to indicate which of the following digitalization factors are relevant for the competitiveness and development of the firm during years 2018 and 2019: a) infrastructure and ultra-bandwidth connection (hereafter infrastructure); b) subsidies, financing and fiscal incentives in favor of digitalization (hereafter subsidies); c) digital initiatives of the government (hereafter governmental intervention); d) networking through collaboration with other firms and research centers (hereafter collaboration); e) acquisition of new technological competences through hiring (hereafter skill sourcing); f) development/consolidation of extant technological competences through training of current personnel (hereafter skill consolidation); g) development of a digitalization strategy (hereafter digitalization strategy); h) other factors; i) no digitalization factor matters; j) I don't know. Both sections allow for multiple responses, but subsection J3 requires firms to select at most 3 digitalization factors. While constraining the variability of factors selected, this restriction has the advantage of forcing respondents to reflect more carefully about which factors to include, thereby avoiding the risk that respondents may carelessly flag all (or most) factors as important.

From these sections, I generated a series of key dummy variables. From Subsection G1, for each possible source (sensors, portable devices, social media, other), I generated a variable taking the value of 1 if the company gathered big data from it, and zero otherwise (hereafter "source dummy"). To measure the extent to which a firm relies on different big data sources, I also generated a variable that is the sum of the dummy variables associated with each source (hereafter "source variety"). Furthermore, I generated another variable taking the value of 1 if the company gathered big data *exclusively* from that source, and zero otherwise (hereafter "exclusive source dummy"). From Subsection J3, for each digitalization factor, I generated a variable taking the value of 1 if the company flagged that factor as important, and zero otherwise (hereafter "digitalization factor dummy"). Moreover, to measure the extent to which a firm considered different digitalization factors as important, I also generated a variable that is the sum of the dummy variables associated with each digitalization factor (hereafter "digitalization factor variety").

Control variables include size, industry, geographical location and ICT intensity, at the highest level of detail provided by the database. Besides performing separate analysis based on firm size (small, medium/large, whole sample), I used revenue classes to control for size at a higher level of granularity within each subsample. To control for industry, I used dummy variables capturing whether the firm belongs to any of the aforementioned sectors (letters C to S of the ATECO classification, see page 12). To account for geographical location, and in particular for the technological divide between different regions of Italy, I used dummy variables indicating the position of the firm in the northwest, northeast, center, south or islands. Finally, I proxied the degree of ICT intensity of the firm through the percentage of ICT workers employed, as ICT intensity may drive both big data adoption and digitalization awareness, thus potentially confounding the estimates.

The analysis starts with three sets of logistic regressions. Each set consists of ten logistic regressions, each of which employs a digitalization factor dummy as a dependent variable (including "I don't know" and "no digitalization factor matters"). The aim is to grasp the effect of a series of regressors of interest (outlined below) on the odds of regarding a digitalization factor as important for the competitiveness and development of the firm in the future. For simplicity, for each set, I estimated all the equations separately through logistic regressions. In the first set, for each regression, I used all source dummies as regressors of interest. In the second set, for each regression, I used all exclusive source dummies as regressors of interest. In the third set, for each regression, I used source variety as a regressor of interest.

I also considered that estimating the equations separately may entail a risk of bias, due to a possible correlation between residuals. This would happen if there were relevant omitted explanatory variables influencing some of the dependent variables jointly. Thus, as a robustness check, I grouped digitalization factors based on the similarity in their possible determinants. I grouped digitalization strategy, collaboration, skill sourcing and skill consolidation, as they may be jointly influenced by determinants related to the digital proactiveness of the firm. I grouped infrastructure, subsidies and governmental intervention, as they may be jointly influenced by determinants related to the (real or perceived) lack of enablers by the firm. I grouped "other factors", "no digitalization factor matters" and "I don't know", as they comprise the category of "alternative answers". Then, I performed three sets of multivariate probit regressions (Cappellari & Jenkins, 2003) analogous to the ones described in the previous paragraph, the only difference being the joint estimation for the three aforementioned groups of dependent variables. Coefficients and standard errors are reported in Table A1 and Table A2 in the Appendix. The signs of coefficients, the relative magnitudes and the levels of statistical significance are substantially equivalent to the ones obtained through the separate estimates, leading to the conclusion that the latter are reliable.

Subsequently, I performed three ordered logistic regressions adopting digitalization factor variety as a dependent variable. In the first, I used all source dummies as regressors of interest. In the second, I used all exclusive source dummies as regressors of interest. In the third, I used source variety as a regressor of interest. All controls mentioned previously have been included in every logistic and ordered logistic regression. Finally, I isolated small firms from the rest of the sample and replicated all the analyses separately on each of the two resulting subsamples (small firms vs all the other firms). Results are reported in the next subsection.

3.3. Results

For simplicity, I start by discussing results on the whole sample (the first third of all Tables) and proceed to focus on the peculiarities of small firms in the last part of the subsection. As Table I shows, big data coming from any source reduces considerably the probability that a firm dismisses the listed digitalization factors as irrelevant. Gathering big data from either sensors, social media or "other sources" cuts almost in half[2] the odds of not regarding any listed digitalization factor as important. Although big data coming from portable devices shows both a weaker effect and a weaker statistical significance (still within the 10% level), it goes in the same direction. At the same time, gathering big data from sensors, social media or portable devices significantly increments the capability by firms to identify which digitalization factors to prioritize, as it reduces the odds of "not knowing" in a range from 20% (portable devices) to 31% (sensors). As for "other sources", the effect is not statistically significant.

Table I also shows that source matters. The effect of big data on the prioritization of digitalization factors strongly depends on the type of source. Big data coming from social media increases the odds of prioritizing collaboration by more than 50%, whereas it has a weak effect on digitalization strategy and no other statistically significant effect. Big data coming from sensors makes firms lean toward skill sourcing (with an odds increase of 23%), while it has a weak effect on infrastructure, subsidies and skill consolidation, and no other statistically significant effect. Big data coming from portable devices significantly affects only the importance ascribed to subsidies, with an odds increase of 20%. Big data coming from other sources increments the prioritization of collaboration and that of skill sourcing, by 35% and 43% respectively (and nothing else).

TABLE I AROUND HERE

As hypothesized, source variety plays a key role as well. Insights in this sense come from the last column of Table I and the whole Table II. Table II reports the coefficients of the logit regression of digitalization factors on exclusive source dummies. While all source dummies trigger a significant reduction in the probability of not regarding any listed factor as important and that of ignoring which factors matter most (with only one exception), exclusive source dummies do not (once again, with just one exception). In other words, firms that exclusively gather big data from a single source do not

[2] This and all the subsequent odds estimates are obtained by exponentiating the logit coefficients in the corresponding Tables.

benefit from any increase in digitalization awareness: they are not significantly more prone to recognizing the importance of digitalization factors, nor are they significantly more capable of identifying those with the highest competitive relevance. Gathering big data exclusively from portable devices appears to even reduce digitalization awareness, by doubling the odds of not regarding any listed digitalization factor as important.

The effects of exclusive source dummies on individual digitalization factors reinforce the picture, as the vast majority of them are not statistically significant. Not only are there very few positive effects (e.g. big data coming exclusively from sensors and social media weakly increasing the prioritization of skill consolidation), but some are even negative, the most relevant being big data from portable devices cutting in half the odds of prioritizing collaboration. Hence, while intuition would suggest that the mere adoption of big data is enough to increase digitalization awareness, the fact that four out of the only seven statistically significant effects here are negative highlights that this is far from being true. For example, gathering big data from sensors increases the odds of prioritizing skill sourcing (see Table I), but gathering big data exclusively from sensors does not (see Table II). Gathering big data from "other sources" increases the odds of prioritizing collaboration (see Table I), but gathering big data exclusively from "other sources" reduces such odds by roughly the same amount (see Table II). This suggests that big data does improve dynamic capabilities related to the sensing of digitalization opportunities and the identification of digitalization priorities, but only when coming from multiple sources. On the contrary, employing a single source may even decrease them.

TABLE II AROUND HERE

The last column of Table I complements the findings of Table II by providing a more fine-grained account of source variety. The column reports the coefficients of the logit regression of digitalization factors on source variety (S.V. in the Table), a variable counting the number of sources of big data employed. This variable appears to crucially affect digitalization awareness. Adding one source multiplies the odds of not regarding any listed digitalization factors as important and "not knowing" by 0.57 and 0.78, respectively. Thus, going from zero to four sources decimates the odds of not regarding any listed digitalization factors as important, and it reduces the odds of "not knowing" by roughly two thirds. Even considering digitalization factors individually, source variety increments the prioritization of almost all of them in a sizable and statistically significant way, the only two (trivial) exceptions being "other factors" and governmental intervention (which are not impacted by any individual source).

The picture emerging from the previous analyses holds not only for individual digitalization factors, but also for the variety of digitalization factors prioritized, as evidenced by Table III and Table IV. All big data sources, and especially source variety, increase the odds of prioritizing a higher number of digitalization factors (see Table III). However, they do so only when used in some conjoint manner: analogously to the individual cases (Table II), the exclusive use of a single big data source does not increase the odds of prioritizing a higher number of digitalization factors and may even decrease them, as in the case of portable devices (see Table IV).

TABLE III AROUND HERE

TABLE IV AROUND HERE

Shifting the focus on small firms, it is worth noting the higher prevalence of sizable and statistically significant individual logit coefficients relative to both the medium/large firms subsample and the whole sample (see Table I). Of relevance is the effect of big data coming from social media, which augments the odds of prioritizing skill sourcing and collaboration by 43% and 79% respectively, as well as infrastructure (27%), subsidies (35%) and digitalization strategy (31%). However, these effects are completely lost in the case of small firms that gathered data exclusively from social media (see Table II). More generally, not counting "nothing" and "I don't know", small firms feature 12 significant positive source dummies, 2 significant positive exclusive source dummies and 2 significant negative exclusive source dummies; the whole sample features 9 positive significant positive source dummies, 3 significant positive exclusive source dummies and 4 significant negative source dummies; medium/large firms feature 4 significant positive source dummies, 3 significant positive exclusive source dummies and 3 significant negative exclusive source dummies. Of course, statistically significant source and exclusive source dummies vary across subsamples, due to structural differences (e.g. small firms having an inherently stronger need to prioritize infrastructure and subsidies). For brevity, the present work will not delve into such differences. What is interesting to note here, however, is that small firms appear to exhibit the greatest loss in the ability to sense digitalization opportunities when relying on a single source of big data rather than two or more, considering both magnitude and statistical significance.

This is confirmed, from a different angle, also from the last column of Table I. With the exception of collaboration[3], where the coefficient is roughly the same, the effect of source variety on the prioritization of digitalization factors is always dramatically higher (often nearly double or even more) in the case of small firms relative to both medium/large firms and the whole sample. Small firms also exhibit the highest effect of each big data source, as well as source variety, on the variety of digitalization factors prioritized (see Table IV). However, neither individual big data sources nor source variety seem to reduce the likelihood of small firms not prioritizing any of the listed digitalization factors in a significantly different way than medium/large firms. The last column of Table II makes it clear that the increase in the odds of not prioritizing any factor prompted by the absence of big data is very high in both subsamples (and even higher in the case of medium/large firms). However, the increase in the odds of "not knowing" is considerably higher in the case of small firms (roughly 100% vs 50%).

Taken together, these results confirm hypotheses HP1a and HP1b, with a further specification. Big data seems to improve digitalization awareness in general, as evidenced by the drastic reduction in

[3] Governmental intervention and "other factors" also constitute (trivial) exceptions, as they are not statistically significant

the probability of not regarding any listed digitalization factor as important and that of not knowing which factors matter most, regardless of source (the only exception being "other sources" in the latter case). Still, different sources of big data affect differently the probability of regarding any individual digitalization factor as important (see Table I). Thus, each source of big data appears to systematically guide firms toward the prioritization of specific digitalization factors. This is in line with the knowledge-shaping and attention-shaping effects postulated in the theoretical background. HP2a and HP2b are also verified (see Table III and IV), with a most interesting addition: not only does source variety increase the digitalization awareness of firms, but a minimum threshold of source variety also appears to be a necessary condition for it, as gathering big data exclusively from a single source either has no effect on digitalization awareness or even decreases it.

Hypothesis HP3a is only partially verified: as for the overall digitalization awareness, big data seems to affect small firms in roughly the same way as their larger counterparts (despite obvious systematic differences in individual priorities). However, big data enhances small firms' capability to identify specific digitalization priorities to a considerably higher extent than their larger counterparts. Instead, hypotheses HP3b and HP3c are fully verified: small firms appear to be the ones that lose the most from relying on a single big data source and gain the most from big data source variety.

4. Discussion and Conclusion

Big data is famous for being the main complement for artificial intelligence: the availability of multi-source, massive amounts of data is the reason why artificial intelligence has developed exponentially, spurring the emergence of digital-intensive business models (Fanti et al., 2021; Reim et al., 2020). This core complementarity, in turn, induces the adoption of other advanced technologies, such as cloud computing (for data storage) and cyber-physical systems (for further data acquisition). However, findings reveal that the role of big data in shaping digitalization trajectories goes beyond technological complementarities. The digital revolution has induced a rapidly changing business environment requiring the dynamic capability of advancing along a series of digitalization factors

(Ciarli et al., 2021; European Patent Office, 2017; Pedota et al., 2023). Big data adopters are much less likely to disregard digitalization factors and likelier to be able to identify which digitalization factors to prioritize for achieving competitive advantage.

Most importantly, results foreground the relevance of source variety. Firms relying on a higher variety of big data sources have a stronger awareness of each individual digitalization factor, and they also tend to prioritize a wider range of digitalization factors. Beyond my expectations, empirical evidence goes so far as indicating that reliance on a single big data source may even nullify (and possibly revert) the increase in digitalization awareness prompted by big data adoption. I find this intriguing on two grounds. First, it suggests that big data fosters digitalization also through its ability to shape knowledge and attention: if it were only for technological complementarities, reliance on a single big data source would still enhance digitalization awareness (contrary to the present evidence). Second, it pinpoints the relevance of big data sources instead of merely reasserting the value of big data. Even with its core complements (e.g. artificial intelligence), big data may not be enough to enhance the ability to sense opportunities and adapt to technological change: relying on a single big data source may trap firms in an even narrower mindset than not relying on big data at all.

Thus, the present study contributes to filling an important gap in the research on big data and dynamic capabilities (Conboy et al., 2020; Côrte-real et al., 2017; Rialti et al., 2019). While results resonate with the proposition that big data enhances dynamic capabilities by helping firms sense and seize opportunities, they highlight that complementarities in big data sources must be in place for this to occur. In this sense, I bring forward the evolutionary-based side of dynamic capabilities (Barreto, 2010; Nelson et al., 2018.). Given bounded rationality and path dependence, I argue theoretically and show empirically that the typology and of big data sources significantly shapes the way firms adapt to technological change. Hence, being exposed to big data coming from various sources is essential to dynamic capabilities. The present results on the adaptation by firms to the digital revolution are coherent with this proposition.

I also contribute to the debate on the digital revolution by shedding new light on the enabling role of big data. Among the enabling technologies of the digital revolution (EPO, 2017; Martinelli et al., 2021), results suggest that big data deserves special consideration. Besides interacting with other digital technologies (thereby functionally enabling them), it also boosts firms' awareness of other digitalization factors. Such factors enable the digital transition both at the level of the firm and at the level of economic systems. Along this line, a potential implication of my findings is a form of path dependence where firms lacking big data (of adequate variety) may fail to integrate digitalization in their vision and high-level routines. Thus, they may be even less prone to and capable of recognizing the value of digitalization in subsequent periods, triggering a vicious circle. Conversely, firms relying on multiple sources of big data are likely to make progress on multiple, complementary digitalization enablers. In subsequent periods, this may increase the amount and variety of big data at the firms' disposal, as well as their proficiency in analyzing them, thus triggering a virtuous circle of technological advancement.

Significant policymaking implications also emerge from the additional findings on small firms. Relative to their larger counterparts, small firms' digitalization awareness seems to be even more sensitive to big data, and especially to big data source variety. The detrimental effect of relying on a single big data source is of particular concern to small firms, in line with the echo chamber effect postulated in the theoretical background. This potential issue is accentuated by the weaker structure and higher adaptability of small firms. On the one hand, the magnitude of the direct effects of big data and source variety is even higher for small firms. On the other hand, small firms are even more susceptible to the potential form of path dependence described above: being still in fieri, their dynamic capabilities are likely to be even more big data-dependent, triggering stronger virtuous and vicious circles.

In its "SME Strategy for a sustainable and digital Europe", the European Commission acknowledges that SMEs do not fully benefit from data (e.g. due to unequal access to big data repositories) and tend to lack familiarity with advanced digital technologies relative to their larger counterparts. This is in line with extant literature recognizing that small firms reap less benefits from digital technologies, due to reduced resources, lower absorptive capacity and lower availability of complements (Bugamelli et al., 2012; Cirillo et al., 2021; Fabiani et al., 2005). This is especially evident and influential in countries like Italy, where SMEs are prevalent. Thus, the European Commission plans to work on improving SMEs' data accessibility (European Commission, 2020b). The theoretical arguments and empirical evidence advanced here resonate with this plan. While larger firms typically have a higher rate of adoption thanks to their resource endowment, small adopting firms seem to benefit equally (and perhaps even more) from big data adoption, at least in terms of digitalization awareness. Thus, they should be enabled and incentivized to adopt big data.

However, the present results also put novel emphasis on the necessity to have a balanced and varied wealth of data. I suggest that big data may act as a sort of meta-digital enabler by shaping knowledge and attention. Hence, I draw an additional link between the lack of big data by small firms and their tendency to shy away from advanced digital technologies. As this link works through knowledge and attention, it requires further consideration. If firms (particularly small ones) were to rely on very large but narrow quantities of data, they may end up neglecting important pieces of their digital development. For instance, if they gathered big data only from portable devices, they may overemphasize efficiency at the detriment of exploration and networking. Hence, policymakers should become aware that big data, while indeed most valuable, is a double-edged sword, and reliance on multiple big data sources may make the difference between a virtuous or a vicious circle. This should be reflected in current and future initiatives aimed at fostering digitalization, including those focused on SMEs and those larger in scope, both at the level of the single country and beyond (e.g. "a European strategy for data"; European Commission, 2020a). Countries (like Italy) where SMEs are prevalent and innovativeness is structurally hampered require particular care, as big data may deeply affect innovation dynamics.

While this work foregrounds the knowledge-shaping and attention-shaping role of big data, as well as the importance of source variety, it is silent on the organizational moderators of such effects. Future qualitative research should explore the structures and mechanisms that enable an effective integration of big data coming from different sources to enhance dynamic capabilities and minimize biases. Regarding dynamic capabilities, the present analysis mostly relates to the dimensions of sensing opportunities and setting up priorities in terms of strategy and resource reconfiguration. Building on these findings, future quantitative research leveraging longitudinal datasets may inquire into the relationship between big data and other components of dynamic capabilities. As the increasing availability of big data shapes firms' adaptation and innovation tendencies, I do encourage continued research in this area.

References

Abernathy, W. J., & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. *Research Policy*, *14*(1), 3-22.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357-376.

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior*, *36*, 157-183.

Amara, N., & Landry, R. (2005). Sources of information as determinants of novelty of innovation in manufacturing firms: evidence from the 1999 statistics Canada innovation survey. *Technovation*, *25*(3), 245-259.

Barreto, I. (2010). Dynamic capabilities: A review of past research and an agenda for the future. *Journal of Management,* 36(1), 256-280.

Bogers, M., Chesbrough, H., Heaton, S., & Teece, D. J. (2019). Strategic management of open innovation: A dynamic capabilities perspective. *California Management Review, 62*(1), 77-94.

Bogers, M., Chesbrough, H., & Moedas, C. (2018). Open innovation: Research, practices, and policies. *California Management Review, 60*(2), 5-16.

Bogner, W. C., & Barr, P. S. (2000). Making sense in hypercompetitive environments: A cognitive explanation for the persistence of high velocity competition. *Organization Science*, *11*(2), 212-226.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society, 15*(5), 662-679.

Bresciani, S., Ciampi, F., Meli, F., & Ferraris, A. (2021). Using big data for co-innovation processes: Mapping the field of data-driven innovation, proposing theoretical developments and providing a research agenda. *International Journal of Information Management*, *60*, 102347.

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133-39.

Burt, R.S., (1992). *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.

Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, *12*(1), 9-30.

Cappellari, L., & Jenkins, S. P. (2003). Multivariate probit regression using simulated maximum likelihood. *The STATA journal, 3*(3), 278-294

Chehbi-Gamoura, S., Derrouiche, R., Damand, D., & Barth, M. (2020). Insights from big Data Analytics in supply chain management: an all-inclusive literature review using the SCOR model. *Production Planning & Control*, *31*(5), 355-382.

Chesbrough, H. W. (2003). *Open innovation. The new imperative for creating and profiting from technology*. Harvard Business School Press, Boston, MA.

Choi, T. M., & Chen, Y. (2021). Circular supply chain management with large scale group decision making in the big data era: The macro-micro model. *Technological Forecasting and Social Change*, *169*, 120791.

Ciampi, F., Faraoni, M., Ballerini, J., & Meli, F. (2022). The co-evolutionary relationship between digitalization and organizational agility: Ongoing debates, theoretical developments and future research perspectives. *Technological Forecasting and Social Change*, *176*, 121383.

Ciarli, T., Kenney, M., Massini, S., & Piscitello, L. (2021). Digital technologies, innovation, and skills: Emerging trajectories and challenges. *Research Policy, 50*(7), 104289.

Cirillo V., Fanti, L., Mina A. and Ricci A. (2021). Digitizing Firms: are Skills Driving Change? Evidence from Firm-Level Italian Data. *LEM WP n.4/2021*, Laboratory of Economics and Management (LEM), Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128-152.

Cohen, W. M., Goto, A., Nagata, A., Nelson, R. R., & Walsh, J. P. (2002). R&D spillovers, patents and the incentives to innovate in Japan and the United States. *Research Policy*, *31*(8-9), 1349-1367.

Conboy, K., Mikalef, P., Dennehy, D., & Krogstie, J. (2020). Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *European Journal of Operational Research*, *281*(3), 656-672.

Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of Big Data Analytics in European firms. *Journal of Business Research*, *70*, 379-390.

Csath, M. (2022). Organizational learning as the best business practice for adaptation in times of great changes: a viewpoint. *Development and Learning in Organizations: An International Journal*, forthcoming.

Danneels, E. (2008). Organizational antecedents of second-order competences. *Strategic Management Journal, 29*(5), 519-543.

Danneels, E. (2012). Second-order competences and Schumpeterian rents. *Strategic Entrepreneurship Journal, 6*(1), 42-58.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*(3), 147-162.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic Management Journal*, *21*(10-11), 1105-1121.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904.

European Commission (2020). A European strategy for data. Retrieved September 27th, 2022 from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN.

European Commission (2020). *An SME Strategy for a sustainable and digital Europe.* Retrieved April 27th, 2022 from: https://ec.europa.eu/info/sites/default/files/communication-sme-strategy-march-2020_en.pdf.

European Patent Office (2017). Patents and the Fourth Industrial Revolution. Retrieved May 1, 2020 from: http://www.lemoci.com/wp-content/uploads/2017/12/Patents-and-the-Fourth-industrial-Revolution-2017.pdf

Fabiani, S., Schivardi, F., & Trento, S. (2005). ICT adoption in Italian manufacturing: firm-level evidence. *Industrial and Corporate Change*, *14*(2), 225-249.

Fanti, L., Guarascio, D., & Moggi, M. (2022). From Heron of Alexandria to Amazon's Alexa: a stylized history of AI and its impact on business models, organization and work. *Journal of Industrial and Business Economics, 49*(3), 409-440.

Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2018). Big data analytics capabilities and knowledge management: impact on firm performance. *Management Decision,* doi: https://doi.org/10.1108/MD-07-2018-0825.

Fiol, C. M., & O'Connor, E. J. (2003). Waking up! Mindfulness in the face of bandwagons. *Academy of Management Review*, *28*(1), 54-70.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, *47*(1), 117-132.

Frisk, J. E., & Bannister, F. (2017). Improving the use of analytics and big data by changing the decision-making culture: A design approach. *Management Decision, 55(*10), 2074-2088.

Garbuio, M., & Lin, N. (2019). Artificial intelligence as a growth engine for health care startups: Emerging business models. *California Management Review*, *61*(2), 59-83.

Gibcus, P., Vermeulen, P. A., & De Jong, J. P. (2009). Strategic decision making in small firms: a taxonomy of small business owners. *International Journal of Entrepreneurship and Small Business*, *7*(1), 74.

Greiner, L.E., 1998. Evolution and revolution as organizations grow, *Harvard Business Review, 76* (3), 55–68.

Guo, B., Zhang, D., Wang, Z., Yu, Z., & Zhou, X. (2013). Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. *Journal of Network and Computer Applications*, *36*(6), 1531-1539.

Kohli, A. K., Jaworski, B. J., & Kumar, A. (1993). MARKOR: a measure of market orientation. *Journal of Marketing Research*, *30*(4), 467-477.

Lansley, G., & Longley, P. (2016). Deriving age and gender from forenames for consumer analytics. *Journal of Retailing and Consumer Services*, *30*, 271-278.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2), 21-32.

Lewin, A. Y., Massini, S., & Peeters, C. (2011). Microfoundations of internal and external absorptive capacity routines. *Organization Science*, *22*(1), 81-98.

Liang, Y. C., Lu, X., Li, W. D., & Wang, S. (2018). Cyber Physical System and Big Data enabled energy efficient machining optimisation. *Journal of Cleaner Production*, *187*, 46-62.

Lo, F. Y., & Campos, N. (2018). Blending Internet-of-Things (IoT) solutions into relationship marketing strategies. *Technological Forecasting and Social Change, 137*, 10-18.

Martinelli, A., Mina, A., & Moggi, M. (2021). The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. *Industrial and Corporate Change, 30*(1), 161-188.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60-68.

Medase, S. K., & Abdul-Basit, S. (2020). External knowledge modes and firm-level innovation performance: Empirical evidence from sub-Saharan Africa. *Journal of Innovation & Knowledge*, *5*(2), 81-95.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management, 57*(2), 103169.

Mikalef, P., van de Wetering, R., & Krogstie, J. (2021). Building dynamic capabilities by leveraging big data analytics: The role of organizational inertia. *Information & Management*, *58*(6), 103412.

Miller, D. (2011). Miller (1983) revisited: A reflection on EO research and some suggestions for the future. *Entrepreneurship Theory and Practice*, *35*(5), 873-894.

Mishra, D., Luo, Z., Jiang, S., Papadopoulos, T., & Dubey, R. (2017). A bibliographic study on big data: concepts, trends and challenges. *Business Process Management Journal*, *23*(3), 555-573.

Mthanti, T., & Ojah, K. (2017). Entrepreneurial orientation (EO): Measurement and policy implications of entrepreneurship at the macroeconomic level. *Research Policy*, *46*(4), 724-739.

Nadkarni, S., & Barr, P. S. (2008). Environmental context, managerial cognition, and strategic action: An integrated view. *Strategic Management Journal*, *29*(13), 1395-1427.

Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, MA.

Nelson, R. R., Dosi, G., Helfat, C. E., Pyka, A., Saviotti, P. P., Lee, K., ... & Malerba, F. (2018). *Modern Evolutionary Economics: an Overview*. Cambridge University Press, Cambridge, UK.

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

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*(1), 14-37.

Osterrieder, P., Budde, L., & Friedli, T. (2020). The smart factory as a key construct of industry 4.0: A systematic literature review. *International Journal of Production Economics, 221*, 107476.

Pedota, M., & Piscitello, L. (2022). A new perspective on technology-driven creativity enhancement in the Fourth Industrial Revolution. *Creativity and Innovation Management*, *31*(1), 109-122.

Pedota, M., Grilli, L., & Piscitello, L. (2021). Technological paradigms and the power of convergence. *Industrial and Corporate Change*, *30*(6), 1633-1654.

Pedota, M., Grilli, L., & Piscitello, L. (2022). Technology adoption and upskilling in the wake of Industry 4.0. *Technological Forecasting and Social Change*, *187,* 122085.

Penn, D. W., Ang'wa, W., Forster, R., Heydon, G., & Richardson, S. J. (1998). Learning in smaller organisations. *The learning organization*.

Perez, C. (2010). Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, *34*(1), 185-202.

Reim, W., Åström, J., & Eriksson, O. (2020). Implementation of artificial intelligence (AI): a roadmap for business model innovation. *AI*, *1*(2), 11.

Rialti, R., Marzi, G., Ciappei, C., & Busso, D. (2019). Big data and dynamic capabilities: a bibliometric analysis and systematic literature review. *Management Decision*, *57*(8), 2052-2068.

Scott, M., & Bruce, R. (1987). Five stages of growth in small business. *Long Range Planning*, *20*(3), 45-52.

Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. *Technovation*, *98*, 102173.

Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science, 2*(1), 125-134.

Song, Y., Gnyawali, D. R., Srivastava, M. K., & Asgari, E. (2018). In search of precision in absorptive capacity research: A synthesis of the literature and consolidation of findings. *Journal of Management*, *44*(6), 2343-2374.

Srivastava, M. K., Gnyawali, D. R., & Hatfield, D. E. (2015). Behavioral implications of absorptive capacity: The role of technological effort and technological capability in leveraging alliance network technological resources. *Technological Forecasting and Social Change, 92*, 346-358.

Stornelli, A., Ozcan, S., & Simms, C. (2021). Advanced manufacturing technology adoption and innovation: A systematic literature review on barriers, enablers, and innovation types. *Research Policy, 50*(6), 104229.

Sultana, S., Akter, S., Kyriazis, E., & Wamba, S. F. (2021). Architecting and developing big data-driven innovation (DDI) in the digital economy. *Journal of Global Information Management*, *29*(3), 165-187

Sutcliffe, K. M., & Huber, G. P. (1998). Firm and industry as determinants of executive perceptions of the environment. *Strategic Management Journal*, *19*(8), 793-807.

Talwar, S., Kaur, P., Fosso Wamba, S., & Dhir, A. (2021). Big Data in operations and supply chain management: a systematic literature review and future research agenda. *International Journal of Production Research*, *59*(11), 3509-3534.

Taylor, A., & Greve, H. R. (2006). Superman or the fantastic four? Knowledge combination and experience in innovative teams. *Academy of Management Journal*, *49*(4), 723-740.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319-1350.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509-533.

Todorova, G., & Durisin, B. (2007). Absorptive capacity: Valuing a reconceptualization. *Academy of Management Review*, *32*(3), 774-786.

Verona, G., & Ravasi, D. (2003). Unbundling dynamic capabilities: an exploratory study of continuous product innovation. *Industrial and Corporate Change, 12*(3), 577-606.

Volberda, H. W., Foss, N. J., & Lyles, M. A. (2010). Perspective—Absorbing the concept of absorptive capacity: How to realize its potential in the organization field. *Organization Science, 21*(4), 931-951.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234-246.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356-365.

Wamba, S.F., and Mishra, D. (2017). Big data integration with business processes: a literature review. *Business Process Management Journal*, *23*(3), 477-492.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal, 24*(10), 991-995.

Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review*, *27*(2), 185-203.

Zhou, Z. H., Chawla, N. V., Jin, Y., & Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]. *IEEE Computational Intelligence Magazine*, *9*(4), 62-74.

Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science, 13*(3), 339-351.

Appendix

TABLE A1 AROUND HERE

TABLE A2 AROUND HERE

Table I. Logit coefficients of big data sources and source variety

Group	Digitalization Factor	Sensors	Portable	Social	Other	S.V.
Whole Sample	Infrastructure	0.12*	0.05	0.04	0.09	0.08***
		(0.07)	(0.08)	(0.07)	(0.07)	(0.02)
	Subsidies	0.11*	0.18**	0.09	-0.01	0.09***
		(0.07)	(0.07)	(0.07)	(0.07)	(0.02)
	Governmental Intervention	-0.01	0.05	-0.02	0.11	0.32
		(0.10)	(0.10)	(0.10)	(0.09)	(0.32)
	Collaboration	0.16	0.12	0.43***	0.30***	0.25***
		(0.10)	(0.12)	(0.11)	(0.10)	(0.03)
	Skill Sourcing	0.21***	0.04	0.11	0.36***	0.19***
		(0.08)	(0.10)	(0.09)	(0.08)	(0.03)
	Skill Consolidation	0.13*	0.11	0.06	0.10	0.10***
		(0.07)	(0.08)	(0.08)	(0.07)	(0.02)
	Digitalization Strategy	0.06	0.10	0.15*	0.11	0.11***
		(0.07)	(0.08)	(0.08)	(0.07)	(0.03)
	Other Factors	0.15	0.15	-0.20	-0.18	-0.04
		(0.17)	(0.19)	(0.21)	(0.18)	(0.06)
	I Don't Know	-0.37***	-0.22**	-0.30**	-0.13	-0.25***
		(0.11)	(0.11)	(0.12)	(0.10)	(0.04)
	Nothing	-0.60***	-0.37*	-0.58**	-0.70***	-0.55***
		(0.21)	(0.19)	(0.25)	(0.23)	(0.09)
Small Firms	Infrastructure	0.12	0.03	0.24**	0.23**	0.15***
		(0.13)	(0.12)	(0.12)	(0.11)	(0.04)
	Subsidies	0.08	0.19*	0.30***	-0.23	0.14***
		(0.13)	(0.11)	(0.12)	(0.11)	(0.04)
	Governmental Intervention	-0.24	0.11	0.11	0.07	0.03
		(0.19)	(0.15)	(0.16)	(0.14)	(0.05)
	Collaboration	0.10	0.07	0.58***	0.17	0.24***
		(0.21)	(0.19)	(0.17)	(0.17)	(0.06)
	Skill Sourcing	0.56***	-0.02	0.36***	0.35***	0.31***
		(0.16)	(0.17)	(0.14)	(0.14)	(0.05)
	Skill Consolidation	0.40***	0.19	0.13	0.21*	0.23***
		(0.13)	(0.13)	(0.12)	(0.14)	(0.04)
	Digitalization Strategy	-0.01	0.10	0.27**	0.24*	0.16***
		(0.16)	(0.14)	(0.14)	(0.13)	(0.05)
	Other Factors	0.48	-0.17	-0.16	-0.45	-0.07
		(0.31)	(0.31)	(0.32)	(0.33)	(0.11)
	I Don't Know	-0.35*	-0.28*	-0.65***	-0.04	-0.32***
		(0.19)	(0.16)	(0.18)	(0.14)	(0.06)
	Nothing	-0.65*	-0.19	-0.50*	-0.70**	-0.46***
		(0.36)	(0.23)	(0.30)	(0.32)	(0.12)
Medium and Large Firms	Infrastructure	0.11	0.07	-0.06	0.06	0.05*
		(0.08)	(0.10)	(0.10)	(0.08)	(0.03)
	Subsidies	0.12	0.17*	0.02	0.06	0.09***
		(0.08)	(0.10)	(0.10)	(0.08)	(0.03)
	Governmental Intervention	0.07	-0.02	-0.10	0.13	0.03
		(0.11)	(0.13)	(0.13)	(0.11)	(0.04)
	Collaboration	0.11	0.19	0.34**	0.35***	0.24***
		(0.12)	(0.15)	(0.14)	(0.12)	(0.04)
	Skill Sourcing	0.08	0.02	0.00	0.40***	0.13***
		(0.09)	(0.11)	(0.11)	(0.10)	(0.03)
	Skill Consolidation	0.05	0.00	0.05	0.08	0.04
		(0.08)	(0.10)	(0.10)	(0.09)	(0.03)
	Digitalization Strategy	0.05	0.11	0.08	0.05	0.07**
		(0.08)	(0.11)	(0.10)	(0.09)	(0.03)
	Other Factors	0.04	0.35	-0.23	-0.05	0.04
		(0.22)	(0.25)	(0.28)	(0.23)	(0.08)
	I Don't Know	-0.28*	-0.15	0.00	-0.24*	-0.18***
		(0.14)	(0.17)	(0.16)	(0.14)	(0.05)
	Nothing	-0.46*	-0.75**	-0.70*	-0.69**	-0.62***
		(0.27)	(0.36)	(0.43)	(0.33)	(0.14)

Table I shows the coefficient of the logit regressions of digitalization factors on source dummies and source variety for each sample group. Standard errors are reported in parentheses. Three stars indicate significance at 1%, two stars at 5%, one star at 10%. All control variables described in the Methodology subsection have been included in each regression.

Table II: Logit coefficients of exclusive big data sources

Group	Digitalization Factor	Sensors	Portable	Social	Other	Nothing
Whole Sample	Infrastructure	0.17	0.20	0.00	0.21**	-0.14**
		(0.11)	(0.13)	(0.12)	(0.10)	(0.06)
	Subsidies	0.02	0.20	0.00	-0.17*	-0.24***
		(0.10)	(0.13)	(0.12)	(0.10)	(0.06)
	Governmental Intervention	-0.05	0.24	0.10	0.13	-0.04
		0.15	(0.16)	(0.17)	(0.14)	(0.09)
	Collaboration	-0.28*	-0.89***	0.04	-0.28*	-0.62***
		(0.16)	(0.26)	(0.17)	(0.15)	(0.09)
	Skill Sourcing	-0.03	-0.25	-0.20	-0.05	-0.44***
		(0.12)	(0.18)	(0.15)	(0.14)	(0.08)
	Skill Consolidation	0.19*	0.08	0.24*	0.10	-0.23***
		(0.11)	(0.14)	(0.13)	(0.10)	(0.07)
	Digitalization Strategy	-0.08	-0.12	0.16	0.02	-0.24***
		(0.11)	(0.15)	(0.13)	(0.11)	(0.07)
	Other Factors	-0.30	-0.34	-0.25	-0.10	-0.08
		(0.31)	(0.38)	(0.35)	(0.27)	(0.17)
	I Don't Know	-0.32	0.01	-0.55**	0.10	0.55***
		(0.21)	(0.20)	(0.24)	(0.17)	(0.11)
	Nothing	0.37	0.74**	0.49	0.46	1.12***
		(0.36)	(0.34)	(0.40)	(0.36)	(0.24)
Small Firms	Infrastructure	0.15	-0.08	-0.01	0.10	-0.31***
		(0.21)	(0.19)	(0.19)	(0.17)	(0.11)
	Subsidies	-0.08	0.03	0.08	-0.23	-0.35***
		(0.22)	(0.19)	(0.19)	(0.17)	(0.11)
	Governmental Intervention	-0.43	0.26	0.35	0.04	-0.02
		(0.35)	(0.24)	(0.25)	(0.23)	(0.15)
	Collaboration	-0.35	-0.82**	-0.04	-0.35	-0.59***
		(0.36)	(0.38)	(0.28)	(0.26)	(0.15)
	Skill Sourcing	0.42*	-0.53*	-0.01	-0.31	-0.72***
		(0.25)	(0.29)	(0.23)	(0.21)	(0.13)
	Skill Consolidation	0.46**	-0.02	0.00	0.05	-0.46***
		(0.22)	(0.20)	(0.20)	(0.18)	(0.11)
	Digitalization Strategy	-0.24	-0.37	-0.05	-0.23	-0.39***
		(0.26)	0.25	(0.22)	(0.20)	(0.13)
	Other Factors	0.33	-0.38	0.24	-0.11	-0.42
		(0.55)	(0.59)	(0.49)	(0.48)	(0.76)
	I Don't Know	-0.31	0.18	-0.52	0.54**	0.71***
		(0.35)	(0.27)	(0.35)	(0.25)	(0.18)
	Nothing	0.34	0.72*	0.50	0.45	1.00***
		(0.57)	(0.43)	(0.52)	(0.48)	(0.34)
Medium and Large Firms	Infrastructure	0.13	0.36*	-0.06	0.26**	-0.09
		(0.12)	(0.19)	(0.16)	(0.13)	(0.08)
	Subsidies	-0.02	0.22	-0.11	-0.15	-0.24***
		(0.13)	(0.19)	(0.16)	(0.13)	(0.08)
	Governmental Intervention	0.06	0.24	-0.13	0.21	-0.04
		(0.18)	(0.24)	(0.24)	(0.17)	(0.11)
	Collaboration	-0.30*	-0.91**	0.13	-0.21	-0.59***
		(0.18)	(0.36)	(0.21)	(0.18)	(0.11)
	Skill Sourcing	-0.13	-0.13	-0.42**	0.08	-0.30***
		(0.14)	(0.23)	(0.20)	(0.14)	(0.09)
	Skill Consolidation	0.12	-0.02	0.36**	0.10	-0.12
		(0.13)	(0.20)	(0.17)	(0.14)	(0.08)
	Digitalization Strategy	-0.05	0.01	0.26	0.14	-0.17**
		(0.13)	(0.20)	(0.17)	(0.14)	(0.09)
	Other Factors	-0.51	-0.17	-0.70	-0.06	-0.23
		(0.39)	(0.50)	(0.54)	(0.33)	(0.21)
	I Don't Know	-0.39	-0.10	-0.50	-0.38	0.41***
		(026)	(0.33)	(0.34)	(0.25)	(0.14)
	Nothing	0.50	0.59	0.39	0.44	1.22***
		(0.48)	(0.57)	(0.68)	(0.53)	(0.34)

Table II shows the coefficient of the logit regressions of digitalization factors on exclusive source dummies for each sample group. Standard errors are reported in parentheses. Three stars indicate significance at 1%, two stars at 5%, one star at 10%. All control variables described in the Methodology subsection have been included in each regression.

Table III. Ordered logit coefficients of big data sources and source variety

Group	Sensors	Portable	Social	Other	S. V.
Whole Sample	0.31*** (0.06)	0.22*** (0.07)	0.30*** (0.07)	0.37*** (0.06)	0.30*** (0.02)
Small Firms	0.45*** (0.13)	0.21** (0.10)	0.54*** (0.11)	0.41*** (0.10)	0.39*** (0.03)
Medium and Large Firms	0.20*** (0.08)	0.22** (0.10)	0.14 (0.09)	0.38*** (0.08)	0.24*** (0.03)

Table III shows the coefficients of the ordered logit regressions of digitalization factor variety on source dummies and source variety for each sample group. Standard errors are reported in parentheses. Three stars indicate significance at 1%, two stars at 5%, one star at 10%. All control variables described in the Methodology subsection have been included in each regression.

Table IV. Ordered logit coefficients of exclusive big data sources

Group	Sensors	Portable	Social	Other	Nothing
Whole Sample	-0.09 (0.10)	-0.19* (0.12)	-0.02 (0.11)	-0.02 (0.10)	-0.65*** (0.06)
Small Firms	0.10 (0.21)	-0.47*** (0.17)	-0.06 (0.18)	-0.25 (0.17)	-0.85*** (0.11)
Medium and Large Firms	-0.16 (0.12)	0.02 (0.18)	-0.07 (0.16)	0.16 (0.13)	-0.52*** (0.08)

Table IV shows the coefficients of the ordered logit regressions of digitalization factor variety on exclusive source dummies for each sample group. Standard errors are reported in parentheses. Three stars indicate significance at 1%, two stars at 5%, one star at 10%. All control variables described in the Methodology subsection have been included in each regression.

Table A1. Multivariate probit coefficients of big data sources and source variety

Digitalization Factor	Sensors	Portable	Social	Other	S.V.
Infrastructure	0.07*	0.03	0.04	0.05	0.05***
	(0.04)	(0.05)	(0.05)	(0.04)	(0.01)
Subsidies	0.06*	0.11**	0.06	-0.01	0.06***
	(0.03)	(0.04)	(0.05)	(0.04)	(0.01)
Governmental Intervention	-0.01	0.03	-0.01	0.06	0.02
	(0.05)	(0.06)	(0.06)	(0.05)	(0.02)
Collaboration	0.10	0.06	0.23***	0.16***	0.13***
	(0.06)	(0.06)	(0.06)	(0.05)	(0.02)
Skill Sourcing	0.13***	0.03	0.08	0.21***	0.12***
	(0.05)	(0.05)	(0.05)	(0.04)	(0.02)
Skill Consolidation	0.09*	0.07	0.04	0.06	0.07***
	(0.04)	(0.05)	(0.05)	(0.04)	(0.02)
Digitalization Strategy	0.04	0.06	0.09*	0.07	0.06***
	(0.04)	(0.05)	(0.05)	(0.04)	(0.02)
Other Factors	0.06	0.06	-0.09	-0.07	-0.01
	(0.08)	(0.09)	(0.09)	(0.08)	(0.03)
I Don't Know	-0.17***	-0.10*	-0.13**	-0.07	-0.12***
	(0.06)	(0.06)	(0.06)	(0.05)	(0.02)
Nothing	-0.26***	-0.19**	-0.24**	-0.33***	-0.25***
	(0.09)	(0.09)	(0.10)	(0.09)	(0.04)

Table A1 shows the coefficient of the multivariate probit regressions of digitalization factors on source dummies and source variety. Dependent variables have been grouped as described in the Methodology subsection for joint estimation. Standard errors are reported in parentheses. Three stars indicate significance at 1%, two stars at 5%, one star at 10%. All control variables described in the Methodology subsection have been included in each regression.

Table A2. Multivariate probit coefficients of exclusive big data sources

Digitalization Factor	Sensors	Portable	Social	Other	Nothing
Infrastructure	0.10	0.12	0.01	0.13**	-0.09**
	(0.07)	(0.08)	(0.08)	(0.06)	(0.04)
Subsidies	0.00	0.13	0.01	-0.11*	-0.14***
	(0.06)	(0.08)	(0.06)	(0.06)	(0.04)
Governmental Intervention	-0.02	0.15	0.06	0.08	-0.02
	(0.08)	(0.09)	(0.09)	(0.08)	(0.05)
Collaboration	-0.15*	-0.46***	0.01	-0.15*	-0.33***
	(0.08)	(0.12)	(0.09)	(0.08)	(0.05)
Skill Sourcing	-0.02	-0.16	-0.11	-0.04	-0.27***
	(0.07)	(0.10)	(0.08)	(0.07)	(0.04)
Skill Consolidation	0.11*	0.05	0.15*	0.06	-0.14***
	(0.07)	(0.08)	(0.08)	(0.07)	(0.04)
Digitalization Strategy	-0.05	-0.08	0.09	0.01	-0.15***
	(0.07)	(0.09)	(0.08)	(0.07)	(0.04)
Other Factors	-0.13	-0.15	-0.11	-0.04	-0.03
	(0.14)	(0.16)	(0.15)	(0.12)	(0.07)
I Don’t Know	-0.15	-0.01	-0.28**	0.03	0.28***
	(0.10)	(0.11)	(0.12)	(0.09)	(0.06)
Nothing	0.18	0.33**	0.26	0.21	-0.50***
	(0.15)	(0.15)	(0.16)	(0.15)	(0.10)

Table A2 shows the coefficient of the multivariate probit regressions of digitalization factors on exclusive source dummies. Dependent variables have been grouped as described in the Methodology subsection for joint estimation. Standard errors are reported in parentheses. Three stars indicate significance at 1%, two stars at 5%, one star at 10%. All control variables described in the Methodology subsection have been included in each regression.