<s:chapter>

<p:ct>3. Big Data and SMEs

<fs:lrh>Big data in small business

<fs:rrh>Big Data and SMEs

<p:au>Vania Sena and Sena Ozdemir

<p:a>3.1 INTRODUCTION

<p:text>The concept of Big Data emerged around 20 years ago (Laney, 2001; McAfee et al., 2012) and although there was some confusion on what it meant and what was really behind the label (Laney, 2001; Gandomi and Haider, 2015), Big Data quickly became shorthand for large volumes of data – both structured and unstructured – generated by the routine activities of individuals and organisations (Manyika et al., 2011; Davenport et al., 2012). The words “Big Data” by themselves have become more and more popular (Davenport, 2014; Gandomi and Haider, 2015). The reasons why the concept of Big Data became quickly popular are well known: the cost of storing data fell substantially over time while low-cost data capture technologies (i.e. mobile phones, social media, apps) became very popular among consumers and small producers (McAfee et al., 2012; Malomo and Sena, 2017).

Twenty years afterwards and now with the words “Big Data”, we tend to refer not only to large volumes of data but also to the set of methodologies that allow us to exploit the data themselves (Chen et al., 2012; Kiron et al., 2014; Stubbs, 2014). While Big Data refers to the characteristics of the data generated through different mechanisms (like sensors, websites), data analytics is commonly used to label methodologies that allow us to make sense of Big Data (Kiron et al., 2012; Stubbs, 2014). Analytics has been defined as “the discovery of meaningful patterns – new and novel information and knowledge – in data” (Delen and Ram, 2018). While for a long time they were the preserve of computer and data scientists, these have become part of the tools organisations used to analyse the large volumes of data they produce internally (Chen et al., 2012; McAfee and Brynjolfsson, 2012; Kubina et al., 2015; Mikalef et al., 2018).

As the techniques to exploit Big Data have become common, researchers have started to analyse where and how the exploitation of Big Data can help businesses to improve their performance (La Valle et al., 2011; Schroeck et al., 2012; Davenport, 2014). Big Data have been studied by several sub-fields of management: marketing, operations, human resources management and finance (Cao et al., 2015; Schoenherr and Speier-Pero, 2015; Dubey et al., 2016; Wedel and Kannan, 2016; Wang et al., 2018). Most studies conclude that exploitation of large volumes of data stored by organisations can improve their performance suggesting a number of channels through which this may happen. Interestingly, they all share a common denominator, that is, the belief that Big Data can affect performance positively because of the changes in the way decisions are made in each functional area. In other words, thanks to the availability of data, managers can make evidence-based decisions across a number of functions and reduce coordination and transaction costs in different functional processes (Akter et al., 2016; Wamba et al., 2017; Mikalef et al., 2018). Importantly, the availability of Big Data can affect not only the operational or routine decisions but can also influence the quality of strategic decision making (MacAfee and Brynjolfsson, 2012; Provost and Fawcett, 2013; Wills, 2014).

Most of the research on Big Data and performance has focused on large firms by default as they have the financial capability to invest in Big Data technologies and infrastructures (La Valle et al., 2011; Schroeck et al., 2012; Davenport, 2014) to analyse them. How do Big Data enter in a discussion of small and medium enterprises (SMEs)? Research on Big Data and SMEs is not very large, and it lacks a framework providing insights into the implications of Big Data on SMEs' performance. Traditionally, the relationship between the different activities of a firm and performance has been analysed using the concept of Value Chain whose underpinning assumption is that all the processes within an organisation need to be coordinated in such a way that value can be created. The implication is that the drivers of business performance can be identified by breaking down the business into strategically important activities that contribute to the creation of value. The concept of Value Chain can still be useful when trying to understand how Big Data can help businesses to generate value; therefore, several authors have developed the concept of Data Value Chain that describes how data within a business is exploited to improve performance. Data Value Chains are identified by a set of sequential steps (data generation, data collection, data analysis and exploitation) that simplify the mechanisms through which businesses can extract value from data and improve business performance. The key question in this field is the following: Is the impact of Big Data on SMEs different from what we observe among large firms? No study has proposed a Data Value Chain model for SMEs. Most of the existing models propose a basic model which does not consider important phases such as value creation and impact on performance. Still, this can be an important research area built upon the assumption that SMEs are different types of firms altogether and not smaller version of large firms.

Interestingly, research on SMEs has identified a number of key differences among large firms and SMEs. First, SMEs (in particular the very small ones) may have a short lifespan; we refer here to the "liability of newness" which can be rephrased as the "liability of smallness" and refers to the poor survival rate among very small and young firms. Second, empirical research suggests that among those that survive, growth may be slow or close to zero. In other words, SMEs may experience limited growth over their lifespan and when this happens, it is similar to "growth spurts" which push SMEs towards a new steady state of close-to-zero growth. Third, the size of SMEs may imply that structures and routines within teams are not fully shaped: this may be particularly true for very small SMEs where a few employees may cover several roles and knowledge is mostly tacit and hardly codified. The result is that business performance may suffer because of the knowledge silos that this type of internal structure creates.

So far, no study has looked into how the exploitation of Big Data may address these issues. Against this background, the purpose of this chapter is multi-fold. First, we offer a short review of the main methodologies to capture, store and exploit Big Data and discuss the opportunities they offer to SMEs. Second, we discuss how Big Data can limit the liability of smallness: while there is not much theory around the topic, we try to identify the key channels through which Big Data can help reduce the liability of smallness. We start from an organisational perspective of the liability of smallness and focus on the fact that in small organisations teams cannot work together properly as routines and roles are not fixed in stone. The resulting coordination costs may make small firms difficult to manage but importantly may hinder the development of new products and eventually the survival chances. We therefore explore how the exploitation of Big Data can reduce the coordination costs and facilitate collaboration among teams which in turn may contribute to the development of new products. Finally, we discuss the relationship between business performance and investment in Big Data, both theoretically and empirically. Theoretically, we explore how Big Data help create a competitive advantage among SMEs; in particular we discuss how Big Data may help SMEs acquire specialised knowledge or intelligence which may translate into increases in performance. We also explore the type of capabilities SMEs need to exploit Big Data and how they can acquire them. Empirically, we try to quantify the impact that investments in Big Data infrastructure have on business performance. We are not planning to test specific channels of transmission but rather search for broad correlations which can give an idea of the magnitude of the impact of the investment on business performance. Eventually we hope the chapter may offer SMEs useful insights on how to exploit the Big Data they produce.

The structure of the chapter is as follows. Section 3.2 introduces the concept of Big Data. Section 3.3 presents a broad overview on the Data Value Chain and its different stages. Section 3.4 analyses how Big Data can help SMEs to overcome the “liability of smallness” starting from the literature on capabilities and Big Data. Section 3.5 presents the results of an empirical study on firms’ growth. Section 3.6 focuses on the managerial implications of the research discussed in the previous sections while Section 3.7 offers some concluding remarks.

<p:a>3.2 DEFINING BIG DATA

<p:text>As mentioned in the introduction, the volume of data available to businesses has increased exponentially (Laney, 2011; Davenport, 2014). We refer here not only to data generated by businesses themselves but also to data from other sources like government, charities and individuals (Dobre and Xhafa, 2014). What drives this growth? Several researchers and consultancy firms have suggested that this growth rate has mirrored the growth in the number of smartphones available to consumers, the development of the internet of things (IOT) and the development of social media apps (Manyika et al., 2011; Schroek et al., 2012; Hashem et al., 2015). By themselves, social media generate large volumes of data as users act as “Datastreams” contributing to the creation of new data. In a sense, Big Data are simply the by-product of the data that digital lives have become quite important. At the core, Big Data is shorthand for data that contain semi-structured, structured or unstructured data although unstructured data are the most common type of Big Data (Davenport et al., 2012). The main feature of unstructured data is that they lack a data scheme and therefore extracting meaning from them can be more complicated (although not impossible) (Davenport et al., 2012). The relative abundance of unstructured data is mostly linked to the fact that apps, cookies, social

activity, NoSQL databases and sensors tend to be the most common technologies for data capture (Hashem et al., 2015).

A number of authors have tried to define Big Data. Chen et al. (2014) define Big Data in terms of volume and the velocity with which it is generated from various sources. Laney (2001) builds upon this definition and characterises Big Data using 3Vs, that is, volume, velocity and variety. Here volume refers to the size of data; velocity refers to the speed by which data are produced by the different sources while variety refers to the different formats of data which are not only numeric but also audio, picture files, text and so on. Big Data have also been defined by seven criteria, commonly referred to as 7Vs: Volume, Velocity, Variety, Veracity, Value, Variability and Visualisation. Variability refers to the fact that the time dimension of Big Data may vary. Some data can be quarterly or daily or even hourly. Such variability can create data management challenges which are far more pronounced if trying to merge structured and unstructured data. Veracity refers to the quality of the data and the extent to which they are accurate description of the underlying phenomena. Sivarajah et al. (2017) also highlight that there are three challenges created by Big Data (implying that Big Data can be rather defined by the challenges they pose). The first challenge is related to the fact that some of it lacks a scheme. The second challenge refers to the processing power as volume and complexity create problems when transforming and analysing the data. The third challenge is the governance of data which include issues around privacy, security, governance and ethics.

<p:a>3.3 DATA VALUE CHAINS

<p:text>The nature of Big Data implies that businesses have to change the way they use and exploit their data holdings as traditional tools and conventional techniques that are suitable for structured data could not be used any longer. Business analytics has been developed to examine large volumes of raw data to extract information that can be used by businesses to improve performance. The process that allows us to capture, store and analyse data is labelled as the Data Value Chain. It can be divided into several steps such as Data Generation, Data Acquisition, Data Storage, Data Analysis and finally Data Exposition (which allows us to translate data insights into improvements in performance).

<p:b>3.3.1 Data Generation and Acquisition

<p:text>Data can be generated by internal processes (such as data generated by Human Resources, for instance) or acquired by external devices such as sensors or websites. As mentioned above, data can be structured or unstructured and need to be cleaned and validated. Data can be acquired in a batch mode or in stream mode and transferred to a storage infrastructure (i.e. a Data lake or data centre) where it is pre-processed to ensure they are not noisy and redundant. Techniques for data pre-processing vary from standard cleaning to transformation and integration (where different data are merged).

<p:b>3.3.2 Data Storage and Analysis

<p:text>Systems that provide storage for Big Data have four components: a storage model (which can be either file-based, object-based or block-based), a data model (using distributed storage and NSQL databases), a storage infrastructure and a distributed processing infrastructure that allow us to share data and

tasks over several interconnected nodes. Data analysis is about analysing pre-processed stored data in order to find correlations, identify patterns and create actionable insights. Analytics research spans from developing algorithms to designing methodologies to analyse different data from a number of sources. Descriptive analytics mostly explores patterns and ultimately attempts to understand what has happened or what is happening right now. At the core, it is a cluster of techniques that can visualise data and is able to summarise data. Visualisation is facilitated by maps, graphs and 3D models which are laid over geographical open data. Descriptive analytics relies on historical data and is traditionally used by business intelligence teams (Kubina et al., 2015). There are various forms of descriptive analytics: these include the use of dashboard applications as another form of descriptive analytics that helps the firm to monitor multiple processes in its division at the same time. Diagnostic analytics explores why some phenomena happened and employs exploratory data analysis in order to identify patterns in the data that may give clues as to why some variables move in a certain direction (Delen and Ram, 2018).

Predictive analytics essentially is about estimating the future value of a variable (like efficiency or productivity) based on current and historical data (Liu, 2014). If the outcome variable is a categorical variable, we use classification models (like random forests etc.) to predict the future values of our dependent variable; if the variable is continuous, regression analysis is a possible model for prediction. If the predicted variable is time-dependent, then we use time-series forecasting. Predictive analytics uses supervised, unsupervised and semi-supervised machine learning techniques, and uses forecasting and statistical modelling techniques to determine future possibilities

Prescriptive analytics uses optimisation, simulation and heuristics-based techniques to identify potential courses of action (Delen and Ram, 2018). Many areas of business-like operations, finance and marketing heavily rely on the use of prescriptive analytics to determine their best possible strategy which would ultimately maximise revenues. Based on the feedback firms get from models of predictive analytics, firms then make use of prescriptive analytics to optimise their business models. Research that uses analytics as the core methodology typically uses methods such as statistics, econometrics, machine learning and network science. However, management research has not fully recognised its potential (Shmueli and Koppius, 2011) as researchers have placed greater focus on causality and on using the data to confirm theoretical models rather than exploring the data and letting them drive the formulation of theories.

<p:b>3.3.3 Data Exposition

<p:text>This last step allows businesses to create value and improve performance on the basis of the data that have been collected and analysed. How can Big Data help businesses to improve their performance? Most of the spending on Big Data technologies is taking place in sectors such as banking and finance, oil and gas, healthcare, mobile telecommunications, insurance, e-commerce, media and investment services. Big Data are particularly helpful to the financial sector that uses Big Data for risk management and to improve their customer services. Mobile telecommunications have been using Big Data to acquire a large customer base, study consumer behaviour and foster innovation. Within the media and e-commerce sectors,

Big Data have helped firms to understand consumers' spending habits and to tailor their products to their preferences.

Overall, Big Data may offer firms two opportunities: (a) the ability to derive insights on its operations in a very granular way and (b) the possibility of changing strategic decision-making real-time in order to respond quickly to changes of the environment. The implication is that there are two mechanisms through which data allow us to create value: (a) internal route, that is, data are used to improve internal processes and therefore value is created by reducing inefficiencies; (b) external route, that is, data are used to gather intelligence on customers and therefore value is created by increasing sales.

<p:a>3.4 BIG DATA, CAPABILITIES AND SMES

<p:text>In this section we will start exploring how the relationship between performance and Big Data and its contribution to value creation have been conceptualised by management research. Management research describes Big Data as a critical asset in line with the resource-based view (RBV) which has been the predominant theoretical framework to analyse the relationship between Big Data and performance. While RBV has been the main theoretical framework behind a number of well-known contributions, the RBV has been criticised for two main reasons. First, it describes Big Data as the amount of data stored by businesses, but it pays limited attention to the skills and additional resources that are useful to create value from the data. Importantly, access to Big Data skills is more challenging for SMEs than for larger sized firms, which have greater intra-firm resources as well as financial resources to outsource such skills. Similarly, while RBV suggests that Big Data can create value, it does not make explicit suggestions on the organisational mechanisms that transform Big Data into value. In this sense, it can be argued that the effective utilisation of a resource such as Big Data requires the availability of organisational mechanisms that would enable its coordinated usage within diverse intra-firm cross-functional units. Second, RBV has limited value in explaining how the relationship between data resources and value creation may vary over time. This is quite important as RBV may seem to imply that increasing data resources may have a growing impact on value creation while in reality there may be a case of decreasing returns to scale in the relationship between the two variables. Finally, the RBV does not pay attention to the actual value of the knowledge businesses can extract from Big Data; while it is usually assumed that all knowledge from Big Data can be valuable, in reality this is not the case as its value varies with the level of "specificity" of the knowledge itself. For instance, insights from Big Data that are common knowledge may not be very useful as they may not confer a business a competitive advantage.

Some researchers have suggested that the dynamic capabilities approach may be more useful to understand the relationship between Big Data and performance. In these studies, the emphasis is on processes and strategies that firms put in place to exploit its Big Data holdings. So Big Data can create value as long as there are some capabilities (i.e. skills and strategies) to exploit the data. A number of capabilities have been identified as being relevant to explain how value can be created from Big Data. Mikalef et al. (2016) suggest that Big Data capabilities are learning, coordinating (as Big Data allow us to coordinate activities from different business functions) and reconfiguring capabilities, among others. Other Big Data capabilities are linked to the infrastructure (i.e. the capability of storing the data in such a way they can be exploited for the creation of value), the technical expertise (so that insights can be extracted from Big Data) and organisational learning. The real advantage of this shift of focus from RBV to dynamic capabilities is that we have a better understanding of what drives performance at firm level. In other words, storing data per se is not sufficient to generate value but additional resources need to be put in place to unlock its value. Of course, Big Data capabilities and skills at the individual level are also essential for effective implementation of Big Data.

While this is all well, one interesting question is whether Big Data can help SMEs to catch up with large firms. To be able to do, Big Data should be able to alleviate some of the issues around the "liability of smallness". Like any other company, SMEs need to invest in specific internal capabilities (such as data infrastructure, skills, absorptive capacity) before they can use data to create value. However, to what extent does investing in the development of new capabilities help SMEs reduce the "liability of smallness"? This is a question that has no answer at the moment as there is currently no research on Big Data and "liability of smallness". In the remainder of the section, we will try to address this issue in the hope that a potential research agenda in this area can be formulated.

<p:b>3.4.1 Liability of Smallness

<p:text>The concept of "liability of newness" has its origins among organisational theorists who noticed that young firms are more likely to fail than established companies (Aldrich & Auster, 1986; Stinchcombe, 1965). The researcher who first introduced the concept was Stinchcombe (1965) who noticed that mortality rates among young firms are larger than among older firms. We can adapt this concept to the case of SMEs which may suffer from poor survival rates or limited growth even when they are not young. In this case, we prefer to talk about the "liability of smallness" rather than the "liability of newness". There are five reasons why we observe the liability of smallness. First, new ventures are initially characterised by low levels of role formalisation and typically lack functional completeness at inception (Stinchcombe, 1965). Small firms are characterised by a fluid organisational structure as roles have to be defined while on the job and as a result, tasks are not carried out efficiently. Importantly, expectations about the roles may be different from their actual content and it can be difficult to ensure they are consistent among each other as there are no former role-holders. The result is that small firms may have to use resources to specify the roles and relationships of individuals in the organisation. In addition, coordination costs can be substantial and can only be reduced over time when roles are crystallised and formalised.

Second, adjustment of roles to personalities may be the norm among small organisations but these can be costly and may not be easy to manage. In a new company, processes are not very bureaucratic and characterised by hierarchical thinking. In other words, they have to reorganise continuously to be able to survive. Indeed, fluid structures and participative coordination should create an environment where information can be shared quickly but at the same time, can create uncertainty about direction of travel and roles (Autio et al., 2000; Choi & Shepherd, 2005). Third, some organisations may not be part of social networks and therefore have limited legitimacy in the eyes of potential workers. In fact, it is rare that small firms hire established teams and therefore more often than not, they rely on hiring a workforce who may not be interested in working in new or small firms. Fourth, small organisations face challenges in finding customers as these firms often lack reputation in their markets. The importance of this issue for survival has been highlighted by Aldrich and Fiol (1994). They found that new firms have to build their legitimacy on two fronts: on the one hand, they have to build their reputation with external stakeholders and on the other hand, they have to work to establish legitimacy internally with their own teams. In practice, this implies that senior management firms in these businesses have to justify themselves continuously to build trust among employees and customers (Aldrich & Fiol, 1994). Finally, in small organisations where roles and positions are fluid, individuals may try to increase their sphere of influence by hoarding knowledge. In turn, this may create knowledge silos where teams or individuals are a repository of knowledge which they may not be willing to share with other teams even if this may have a negative impact on business performance.

<p:b>3.4.2 Big Data and Liability of Smallness

<p:text>How do Big Data enter into a discussion on the sources of the "liability of smallness"? Generally speaking, the liability of smallness is considered to be the result of the internal workings of the SMEs and therefore it is useful to start from how Big Data can influence the behaviour of teams within an SME. At the moment, there is no framework that allows us to explain how Big Data can have an impact on the way different teams interact among each other in the context of SMEs; still, it is important to develop a framework that allows us to interpret how the availability of Big Data can change the relationship among teams in SMEs. A useful starting point is a discussion of how teams interact in small companies. This may lead to an understanding of the conditions that lead to the emergence of the liability of smallness usually observed empirically. Sometimes, teams do not work together well because their activities are not well coordinated and integrated into each other. Additionally, there may be no clarity about the behaviour of other teams, and this may lead to asymmetric information, agency issues and eventually misalignment of objectives among teams.

As a result, a number of mechanisms have been suggested to address such a misalignment ranging from the introduction of performance-contingent incentive contracts to the development of direct monitoring mechanisms. Still, while these solutions would work very well in theory, the reality of the SMEs makes their implementation complicated. First of all, the modern structure of SMEs (i.e. a flat structure) implies there may be too many principals whose activities need to be coordinated. As a result, incentive contracts for the agents can be difficult to design given the number of objectives the different teams may have. As for monitoring, with such a structure, one principal may have too many teams to monitor. In these cases, agency problems will be exacerbated and the expectation is that these will be particularly pronounced in the case of young SMEs.

This situation implies that coordinating devices are needed so that the actions of the different teams are aligned to overarching survival goals. More importantly, their performance has to be tied to the capability of the managing team to develop credible strategy goals and to monitor the outcome of the teams' actions. Still, SMEs vary in their ability to set up targets that have clear and achievable goals that can be measured. In a sense, the belief that small companies have to be free of routines that are typically associated with large firms is so strong that no manager tends to believe that the relationship among teams has to be managed in ways that are technocratic but still compatible with entrepreneurial objectives. In other words, SMEs could invest in the development of a number of "technocratic" competencies that would allow the management of the teams as well as to insulate them from the risks of the liability of smallness. This way, the senior management team can alleviate simultaneously the agency issues discussed above and the problems created by the liability of smallness. Importantly, the development of technocratic competencies would be a building block of a new internal setting with checks and balances which would allow the internal processes and their outcomes to be monitored. The expectation would be then that the SMEs will become more efficient and grow faster so to reduce the gap with large companies.

This is the area where Big Data can trigger a major step-change. Generally, Big Data and associated methodologies tend to be complementary to existing practices but in reality, they can trigger major changes in the way companies are run. At the moment, there is agreement they can improve procedures and improve services and procedures, but they can change the way decisions are made and implemented within a company. In some sense, these arguments are not very different from pointing out that new technologies can improve the design and implementation of new products, change the internal organisation of a company and enhance internal accountability. Big Data can be considered as a coordination device that enhances the collection and the exploitation of information that can support coordinated decision making.

In the previous section, we have introduced the concept of capabilities and their relationship with the performance of SMEs. Like large companies that need to develop specific capabilities to exploit their Big Data, SMEs – often constrained by resources and capabilities at the organisational (e.g. Big Data infrastructure) and individual (e.g. Big Data skills) levels – need to develop some specific capabilities to be able to exploit data to support their decision making. At the individual level, the first capability refers to the ability to collect new forms of data (such as unstructured data) and link them to the administrative data (e.g. reports on sales) that are typically produced by businesses. The second capability refers to the ability of analysing both structured and unstructured data and to extract insights in real time. At the organisational level, the first capability refers to the ability to develop and maintain data infrastructures and governance policies for Big Data. The second important organisational capability refers to the company's ability to redesign decision making in such a way that insights from data can be fruitfully shared among teams and used to improve the quality of the decisions. Facilitating data sharing is a prerequisite for a redesign of the decision-making processes. As mentioned above, teams tend to be repositories of knowledge which they may hoard to retain their influence and may decide to share it with other teams only if it can perceive it as beneficial (Huber, 1991; Kogut & Zander, 1996; Tsai, 2001).

To understand how Big Data can support the decision-making process at SME level, we try to identify the different stages of the process. These include a first stage where priorities are decided and where coordination among priorities is needed. After that, priorities need to be translated into actions and targets. Finally, interventions are followed by the evaluation stage where outcomes are evaluated against the initial targets. Big Data can help in every stage of the process. Generally speaking, Big Data generates insights that can underpin the first stage. Indeed, granular information on markets and customers may suggest the need for a specific set of products and help to activate the supply chain. Literature has highlighted the importance of social media in this area as social networks (created by social media) can help identify emerging needs. Analytics can be useful as well: use of Natural Language Processing (NLP) can help SMEs to make use of unstructured data which can enhance the informational content of the administrative data. Budgeting is typically considered a sensitive area in young firms. Evidence in this area suggests that Big Data can increase the efficiency of budgeting as they allow us to develop an outcome-based budget that would allow resources to be redeployed where they are needed in the future and not where they have been needed in the past. Importantly, the budgetary process can by itself produce data that can help to detect patterns and used to develop better interventions at a later stage. Big Data can help streamline the number of targets associated with the design of the teams' activities; they can support the development of early-warning systems that can inform the decision-making process while sentiment analysis or real-time decision support systems can influence implementation. Some of the information sources can be supplemented by Big Data stored by the SMEs which would allow a more granular view of what happens in each team and among teams. This issue touches upon knowledge sharing and the mechanisms that facilitate such processes. Importantly, simple exposure to knowledge is not sufficient to generate knowledge acquisition (Van Wijk et al., 2008). Knowledge acquisition is supported by "trust" and Granovetter (1985) suggested that teams value a trusted source over a reliable one. Indeed, a substantial body of research has shown that with high levels of trust, parties are more willing to engage in knowledge exchange (Coleman, 1990; Nahapiet & Ghoshal, 1998; Tsai & Ghoshal, 1998; Levin & Cross, 2004, Moran, 2005; Peters & Karren, 2009), irrespective of the type and content of knowledge exchanged (García et al., 2008).

We argue that the use of Big Data can help create this trust. Indeed, Big Data can help develop a common language around processes and targets and facilitate the communication among roles and functions. This can be particularly relevant in the case of complex knowledge which is difficult to share and hence to acquire (Kogut & Zander, 1992, Szulanski, 1996). Use of Big Data technologies can reduce the complexity of knowledge and can facilitate its sharing among teams even in situations where face-to-face interactions are not allowed. Communicating in a shared language may enhance knowledge acquisition (Tsai, 2001) and reduces misunderstandings (Szulanski, 1996; Nahapiet & Ghoshal, 1998).

Finally, Big Data can play a major role in the evaluation phase. Evaluation plays a key role when planning marketing campaigns whose effectiveness needs to be assessed in a continuous way. Real-time data can be used to assess quickly the impact of the interventions and whether corrective actions are needed. Big Data allow real-time decisions while being transparent. All this opens up the possibility of an evaluation cycle where evaluation is carried out continuously.

<p:b>3.4.3 Liability of Smallness and Absorptive Capacity: Big Data and Specialised Knowledge

<p:text>How can SME build up their capabilities to store and analyse Big Data? As SMEs may be financially constrained, they may rely on the external providers. Empirical studies suggest that as businesses grow, they start to engage in more in-house exploration activities (Rothaermel & Deeds, 2004). Similarly, SMEs striving to grow are more likely to outsource these services than large firms (Barge-Gil, 2010; Zeng et al., 2010). However, from a transaction cost economics (TCE) perspective, this can be a problematic option particularly in uncertain environments the quality of the transaction cannot be ascertained (Rindfleisch & Heide, 1997). In such a scenario, small firms can be easily exposed to the opportunistic behaviour of the external providers as SMEs may not have the knowledge and skills to assess the quality of the services received (Anderson, 1985; Nooteboom, 1993; Rindfleisch and Heide 1997; Heide 2003). As a result, it is usually argued that SMEs should try to internalise their data infrastructure where possible (Williamson 1985, 1989).

In this section, we argue that this is not always the best option. In this context, it is essential to think about the value of storing and analysing Big Data internally; in turn, this depends on a variety of factors such as the *value* of the knowledge extracted from Big Data as well as the availability of skills that enable effective exploitation of Big Data (Sena and Ozdemir, 2020). The value of Big Data needs to be considered at the broader level by considering how much value a firm may generate by exploiting Big Data as well as the market orientation[1] of the firm. We refer here to the proactive and responsive market orientation concepts (Narver et al., 2004) which consider the different types of market intelligence that are needed by market-oriented firms. While responsive market orientation is about new knowledge that is related to previous experiences of the firm and customers (e.g. identification of existing needs), proactive market orientation is about the search for radically new information and knowledge (e.g. exploration of latent needs) (Narver et al., 2004; Tsai et al. 2008). This implies that while responsive market-oriented firms engage in adaptive (i.e. exploitative) learning, proactive market-oriented firms engage in explorative learning (Jaworski et al. 2000; Narver et al. 2004; Tsai et al. 2008; Ozdemir et al., 2017).

Proactive market orientation requires exploratory search for radically new information that is highly specific within an industry. We expect that the acquisition of such information will increase the opportunity costs for outsourcing SMEs since the process of knowledge acquisition will require specific knowledge (i.e. less likely to be held by the outsourcing firm) to effectively filter (or detect) the useful information out of a large amount of outsourced data. In addition, since outsourcing firms will be less likely to own the knowledge base to assess the quality of the new information, the cognitive costs of verifying the performance of the information provider are likely to be high. These potential risks are vital when acquiring new information since they are costly to recover in the latter stages of the information processing. Similarly, there will potentially be a limited number of SMEs with dedicated marketing analytics functions and capabilities to unlearn existing knowledge to think “out of the box” to be able to provide valuable insights on radically new information attained through Big Data.

Responsive market orientation, on the other hand, necessitates exploitative search that enables the acquisition of knowledge which may not be too specific. Outsourcing analytical services with lower degrees of novelty will be more cost efficient than trying to acquire it internally. Since small firms will be more familiar with the new information, the transaction costs will also be lower. This view is further supported by the RBV perspective which states that previous knowledge can help businesses to acquire, assimilate and transform externally generated knowledge which is critical for their operations (Cohen & Levinthal, 1990; Zahra & George, 2002). Since non-specific knowledge is commonly available and easier to imitate and substitute, firms should spend less time and effort to acquire such information and rather focus on acquiring inimitable and radically new information and knowledge. Similarly, Choudhury and Sampler (1997) link one of the key concepts of TCE, that is, asset specificity, with the concept of organisational knowledge specificity: knowledge is specific if it is owned by a single firm or by a limited number of firms. For instance, knowledge on radically new technologies would only be owned by a small number of firms within an industry. Choudhury and Sampler (1997: 37) extended the TCE theory by asserting that “in deciding between outsourcing the task of monitoring an environmental information source and retaining the responsibility internally, an organization will choose the option that minimizes the sum of the surveillance costs, the coordination costs, the behavioural contractual costs, and the cognitive transaction costs”. Following this line of thought. SMEs with scarce resources and limited growth opportunities will attempt to outsource the analytical function (in charge of the exploitation and analysis of Big Data) even if it is costly to do so and makes them vulnerable to opportunistic behaviour. This effect will be more pronounced if the knowledge is time specific (i.e. the extent to which knowledge loses value unless used immediately after it becomes available) and the environment is turbulent (i.e. there is high environmental uncertainty and rapid market changes) (Boyd & Fulk, 1996).

<p:a>3.5 MODELLING GROWTH

<p:text>In this section, we try to model SMEs’ growth and show how access to resources which can help Big Data exploitation can accelerate growth among SMEs. We have previously discussed the concept of Big Data capabilities and highlighted the fact that these can be of several types. However, most capabilities that are useful to exploit Big Data tend to be embodied in human capital and therefore having to access suitably skilled human capital is a precondition for the development of the internal capabilities necessary to exploit Big Data. For these reasons, in this section, we will focus on access to skilled human capital as a proxy for the access to Big Data capabilities.

Typically, business growth is modelled through the Gibrat's law, a favourite topic among applied industrial economists for a long time. First proposed by Gibrat in 1931, it states that a firm's growth rate is independent of its size and is based on the idea that firms (within an industry) draw growth rates from a distribution that is the same for all firms regardless of their previous size. Several studies have then tested the empirical validity of the Gibrat's law (see Sutton, 1997 for a survey) in an attempt to understand when the Gibrat's law applies and what drives the firms' growth in an industry. The empirical literature on the determinants of firms' growth has a long and illustrious history. Gibrat (1931) was the first to present an empirical model of the dynamics of the firms' size and its growth, which has then become known as the Gibrat's law. According to the Gibrat's law, firms face the same probability distribution of growth rates, with each firm's observed growth determined by a random sampling from that distribution. The main implication from the Gibrat's law is that a firm's growth rate is independent of its size.

A rich body of empirical evidence has been produced that has tested the empirical validity of the Gibrat's law, spanning numerous countries and time periods although mostly focusing on manufacturing. The results of tests based on this kind of model have been mixed. Earlier studies (Hart, 1962) which typically included large manufacturing firms provided compelling evidence supporting the Gibrat's law. Some studies have included small firms in the sample and found a negative relationship between firm growth and firm size, so rejecting the Gibrat's law (Evans, 1987; Hall, 1987; Dunne et al., 1989; Dunne & Hughes, 1994; Mata, 1994; Audretsch, 1995; Hart & Oulton, 1996; Audretsch et al., 1999; Almus & Nerlinger, 2000; Becchetti & Trovato, 2002; Goddard et al., 2002). In an attempt to reconcile the contrasting evidence, some studies have changed the approach to the estimation of the Gibrat's law and therefore have started to test whether the firms' growth follows a random walk. Goddard et al. (2002), Del Monte and Papagni (2003), Oliveira and Fortunato (2003) and Chen and Lu (2003) carried out panel unit root tests with contradictory results.[2] As Sutton (1997) pointed out, the reason for these contradictory results lies in systematic differences in the samples selected. The Gibrat's law holds when only large firms or firms that have exhausted scale economies are included in the sample. As explained by Audretsch et al. (2004), the firms' growth rates will be independent of their size as long as their likelihood of survival is independent of their size. However, when the likelihood of survival is positively related to firm size, the observed growth rates are no longer normally distributed for each firm size or firm-size class. If size is a requirement for survival, or at least positively influences the likelihood of survival, the consequences of not growing or even experiencing negative growth has a different impact across size classes. The propensity for small firms experiencing low (or negative) growth to exit compared to low-growth large firms biases the samples of surviving small firms towards higher growth enterprises. By contrast, a sample of surviving large firms consists of both low- and high-growth enterprises; thus, when the consequences of not obtaining a high growth opportunity differ systematically between large and small firms in terms of the likelihood of survival, the resulting distributions of actual observed growth patterns across different firm size classes will also vary systematically between large and small firms. Therefore, the Gibrat's law will tend to hold for larger firms but not for smaller enterprises and therefore growth rates will be negatively related to firm size for samples including a full spectrum of large and small firms.

One remarkable fact about the Gibrat's law is its lack of micro-economic foundations. Some authors have tried to add economics to this model, particularly by exploring why the Gibrat's law does not hold. These models point to a number of socio-economic variables influencing firm performance and, moreover, they provide a theoretical explanation for the relationship between size and growth. Jovanovic (1982) developed a theoretical model that could account for possible departures from the Gibrat's law. The model assumes that firms are heterogeneous and that they learn about their true efficiency as they operate in an industry. Failure and growth rates decrease with size and age. Cabral (1995) suggests that the negative relationship between the growth and size can be explained by the fact that entering in a new market requires a sunk investment in capacity. Since small entrants are more likely to exit than large entrants, it is optimal for small entrants to invest gradually which can explain why they tend to grow faster than large entrants. Cooley and Quadrini (2001) develop a theoretical model that introduces financial market frictions and persistent shocks into a learning framework of firm dynamics and produces results consistent with the empirical regularities of the negative effects of initial firm size and of firm age on firm growth. Other empirical evidence also includes the roles of share of foreign participation (Fotopoulos & Louri, 2002) and financial structure (Becchetti & Trovato, 2002; Fotopoulos & Louri, 2002).

Virtually, no empirical paper has analysed how use of Big Data can affect the relationship between a firm's growth and its size. This is not for lack of theories: there are many channels through which Big Data exploitation can condition the growth-size relationship. Exploitation of Big Data may matter to small firms because it facilitates fast growth and allows them to compete with the large firms, so creating the conditions for the rejection of the Gibrat's law. However, the exploitation of Big Data assumes the firms may have access to a set of capabilities which can help them exploit Big Data. Empirically, proxies for capabilities cannot be identified easily and this explains why empirical work in this field is virtually non-existent.

Another issue to consider is the type of industries that may benefit from the exploitation of Big Data. Most empirical analysis on the Gibrat's law is carried out on manufacturing[3] and very little attention is given to services. The argument underlying the preference for manufacturing is that services are serving only localised markets; therefore, service firms tend to operate in markets where economies of scale can be exploited at relatively small levels of output and therefore these firms do not have to grow to overcome any problem related to surviving. On the contrary, services are a diverse group of industries. Indeed, some service sectors are dominated by large firms organised in networks (e.g. retail, banking, hospitality) with small firms competing with large firms by specialising in niche markets. Investments to store and eventually exploit Big Data are quite common in services so creating the conditions for the rejection of the Gibrat's law. Equally, single-unit firms (i.e. firms that are not organised in networks), which may operate niche markets, may still experience episodes of high growth if they invest in technologies for the exploitation of Big Data. However, as we have pointed out above, investment in Big Data technologies has to be accompanied by the development of internal capabilities for the exploitation of Big Data and ultimately this requires access to a skilled workforce. Therefore, one implication is that the relationship between growth and size in services may be mediated by the access to a skilled workforce and this needs to be taken into account when testing for the Gibrat's law.

The empirical framework we use for our analysis is rather straightforward. As mentioned above, according to the Gibrat's law, firms face the same probability distribution of growth rates, with each firm's observed growth determined by a random sampling from that distribution. If the law holds, we would expect no differences in the mean and variance of growth rates across size classes of firm. If this is not the case, firm sizes regress towards or away from their mean in the Galtonian sense. The company growth path can be explosive, that is, firms tend to grow faster as they get larger (large firms grow faster than small ones). Alternatively, small firms tend to grow faster than larger firms (mean-reverting argument), which corresponds to the tendency for a variable to return to the mean size. It is often the case that the variance of the growth rates decreases as the size of firm increases. In practice, we should expect the size distribution of firms to be approximately lognormal (Hart & Oulton, 1996). The fact that the firm size distribution is approximately lognormal is consistent with the hypothesis that a firm's size is heavily influenced by multiplicative stochastic shocks. Chesher (1979) has shown that the law will not hold if the error terms are serially correlated. Serial correlation in proportionate growth rates can be ascribed to persistence of chance factors which make a company grow abnormally fast or abnormally slowly. In this case, size encourages (or discourages) growth, and when there is serial correlation in growth rates, that growth encourages (or discourages) growth. Thus, departures from the Gibrat's law arise: if sizes regress towards or away from the mean size, if above average growth in one period persists into the next, or if a period of above average growth is followed by one of below average growth.

To test whether the law continues to hold for firms that have access to the skilled workforce which may help build up Big Data capabilities, we interact the size of the firm in the previous period with the previous period's proxy of such access. In this case, small firms grow faster than large ones and this effect is particularly true for small firms that have access to the skilled workforce. However, as the coefficient associated with an interaction term varies, whether the null hypothesis is rejected or not depends on the firm-specific size of the investment. In terms of estimation, the presence of firm-specific effects in the growth model leads to a correlation between a regressor and the error term, hence OLS is a biased and inconsistent estimator. Therefore, we use an estimator (such as the sys-GMM estimator) that allows us to control simultaneously for the presence of firm-level heterogeneity and autocorrelation in the residuals.

Our empirical analysis is conducted on a sample of British plants, sourced from the British Annual Business Inquiry (ABI). The dataset covers both the production (including manufacturing) and the non-production sector (services). However, the time-series dimension varies across the twos: while for the production sector it is possible to have information available up to 1980 (and the early !970s for some industries), the data for the services sector are available only after 1997. The size of the plant is measured by its total number of employees while its growth is computed as the annual growth of the number of employees. Information on the plants' age (age) has been sourced from the business registry (BSD). The information on the number of local units pertaining to the same firm allows us to compute the number of plants (or local units) owned by each firm; equally firms that have only one single unit are considered to be single-unit firms. As for the proxies of Big Data capabilities, we decide to focus on the access to such capabilities and more precisely the density of graduate workforce in the local authorities.

Table 3.1 reports the estimates of the Gibrat's law for all service firms in our sample. Standard errors are clustered around industry and local authority. In terms of our parameters of interest, the results suggest that the key parameter is always smaller than one and significant for the sample that includes all the firms and for any value of the density of human capital. Indeed, we test whether it is different from one for different levels of density of human capital (namely, at the minimum, the mean and its maximum value) and we find that for each value, the coefficient is significantly different from one. In other words, in our sample small firms grow faster than large firms and this explains why the value of the coefficient attached to the lagged size of the firm is negative and significant. The classical explanation for the fact that small firms seem to grow faster is that there is a minimum efficient scale of firm and, until this size is reached, the firm experiences decreasing average costs and can therefore enjoy rapid growth. After this point, its average cost curve flattens out and therefore firms experience constant average and marginal costs. However, this also suggests that services in our sample behave like manufacturing firms in that small services firms need to grow faster to be able to compete effectively with the large firms (small firms' selection bias). This is different to what Audretsch et al. (2004) find for the Dutch hospitality sector (sector dominated by small firms) but it is consistent with the findings of Hart and Oulton (1999) who found a negative relationship between size and growth for the British "distribution and hotel" sector and with what Petrunia (2008) finds for the Canadian retail trade. Our results show first that the relationship between growth and size is negative in British services and second that this negative relationship is driven by the fact that smaller firms are growing faster as long as they have access to a skilled workforce. Also, the Gibrat's law holds for firms that are organised. One possible explanation for the negative relationship between size and growth found is related to the age of firms. Evans (1987) notes that theories of firm dynamics generate growth patterns that vary with the age of the firm. The growth process of small firms may appear different from the growth process of large firms because of age effects. Younger firms, typically with substantial growth rates and highly volatile growth, tend to make up the majority of small firms and the minority of large firms. The negative relationship between firm age and growth has been discussed by a number of empirical studies and in different countries: Evans (1987) and Dunne et al. (1989) for the US, and Dunne and Hughes (1994) for the UK). We therefore consider the impact of age on growth and unsurprisingly, we find that age has a negative and highly significant influence on business growth, as younger firms tend to grow faster in our sample.

When estimating separately the same model for single-unit firms and multi-unit firms (Table 3.2), we notice that the growth patterns of multi-unit firms differ. Their growth is not dependent on their size. These results are consistent with the previous evidence that the Gibrat's law holds for larger firms. Finally, we estimate our model only on surviving firms (Table 3.3). The results do not change and this suggests that the negative relationship between size and growth is not driven by small firms exiting quickly but rather by the fact that surviving small firms tend to grow faster than larger firms.

<c:table>

<p:table_title>Table 3.1 Sys-GMM estimates for the whole sample

<t:table_style_2>

<table_header=1>

INDEPENDENT VARIABLES		
(log)employment (lagged one period)	-0.102 (-4.29)	-0.208 (-3.78)
(log) access to Big Data capabilities (lagged one period)	0.1065 (3.11)	0.098 (2.17)
(log)employment (lagged one period)* (log) access to Big data capabilities (lagged one period)	-0.011 (-2.59)	-0.012 (-3.01)
Age	-	-0.291 (-3.81)
# local units	0.0024 (2.84)	0.0020 (1.98)
Hansen test (p-value)	0.192	0.238
Ar(1) (p-value)	0.000	0.000
Ar(2) (p-value)	0.411	0.561

Note: Time, sectoral and local authorities' dummies are included in all specifications. T-ratios computed using standard errors clustered around industry and local authority.

Source: ONS.

</t:table_style_2>

</c:table>

<c:table>

<p:table_title>Table 3.2 Sys-GMM estimates for the single and multi-unit firms

<t:table_style_2>

<table_header=1>

INDEPENDENT VARIABLES	SINGLE-UNIT FIRMS	SINGLE-UNIT FIRMS	MULTI-UNIT FIRMS	MULTI-UNIT FIRMS
(log)employment	-0.42	-0.49	-0.25	-0.29
(lagged one period)	(-5.06)	(-5.12)	(-1.19)	(-0.16)
(log) Access to Big Data capabilities (lagged one period)	0.07 (1.77)	0.09 (1.87)	0.08 (0.42)	0.10 (0.52)
(log)employment (lagged one period)* (log) Access to Big Data capabilities (lagged one period)	-0.02 (-2.69)	-0.02 (-1.89)	-0.07 (-0.78)	-0.05 (-0.68)
Age		-0.29 (-3.91)		-0.32 (-2.89)
# local units			0.002 (3.17)	
Hansen test (p-value)	0.136	0.167	0.234	0.240
Ar(1) (p-value)	0.000	0.000	0.000	0.000
Ar(2) (p-value)	0.569	0.567	0.446	0.509

Note: See Table 3.1.

</t:table_style_2>

</c:table>

<c:table>

<p:table_title>Table 3.3 Sys-GMM estimates for surviving firms

<t:table_style_2>

<table_header=1>

INDEPENDENT VARIABLES		
(log)employment (lagged one period)	-0.102 (-4.29)	-0.121 (-2.56)
(log) Access to Big Data capabilities (lagged one period)	0.1065 (3.11)	0.089 (2.67)
(log)employment (lagged	-0.011	-0.009

one period)* (log) Access to Big Data capabilities (lagged one period)	(-2.59)	(-2.89)
Age	-	-0.301
		(-3.41)
# local units	0.0024	0.0019
	(2.84)	(2.84)
Hansen test (p-value)	0.192	0.591
Ar(1) (p-value)	0.000	0.000
Ar(2) (p-value)	0.411	0.219

Note: See Table 3.1.

</t:table_style_2>

</c:table>

<p:a>3.6 MANAGERIAL IMPLICATIONS

<p:text>The literature on SMEs and Big Data (summarised above) identified a number of drivers of the performance gap between large firms and SMEs and suggests a potential number of mechanisms through which the exploitation of Big Data can reduce the gap. In this section, we plan to draw a number of useful lessons for managers that are planning to deploy and exploit their Big Data to improve their business' performance. We want to arrange our managerial implications using the three steps of the Data Value Chain.

<p:b>3.6.1 Data Capture

<p:text>As mentioned at the beginning of the chapter, data capture is typically associated with the investment in the technologies and systems that allow us to retain and store the data that are produced by SMEs during their routing operations. While it can be argued that this is a phase that is generally managed by Chief Technical Officers, in reality it is important to bear in mind that the nature of the investment has to be aligned to the overall needs of the business. In other words, the new systems have to be able to capture data that can be useful to other teams in the business to identify in a clear way how they can create value; therefore, managers in SMEs have to able to articulate in a clear way the actual benefits that investments in data capture system can deliver to the business and then eventually delegate the management of the investment to the Chief Technical Officer.

Another important lesson is that value from data capture systems cannot be created by one system only; for instance, being able to capture sensor data may not generate value unless there is support data capture investment that allows us to integrate them with other types of unstructured data. Indeed, one IT system alone may not be sufficient to exploit the data businesses produce and additional investments may be required because of the synergies that exist among systems. In other words, managers may find it difficult to translate the investment in data capture systems into improvements in performance if not enough attention is paid to the interdependencies among data and systems. In turn, this requires managers to plan in advance the investments they are planning to make in this area and to develop a strong business case around such investments.

<p:b>3.6.2 Data Analysis

This is an important stage of the Data Value Chain that value creation from data hinges upon. The role of the firm's internal capabilities is crucial as existing skills and knowledge can define the extent to which data are analysed correctly and in such a way the analysis is aligned and useful to a company's strategic objectives. In this context it is important to highlight that most research has emphasised the need for analytical skills that allow us to analyse and make sense of Big Data; in reality, data analysis requires businesses to acquire "translational" skills which allow us to understand the implications of the analytical findings for the business and its strategic objectives. In large firms, these skills tend to be embedded into middle management sitting between the analytical team and the executive team. In SMEs – where some roles are not well defined and because of financial constraints there may be no dedicated analytical team – the best way forward could be to equip each member of the strategic management team with enough understanding of analytics so that they can use the results of the data analysis to drive their decision making. Crucially, this does not imply that every member of the management team has to be proficient in analytics but simply that they can assess the implications of the results for the performance of the firm.

Another important lesson for SMEs is related to the type of analytical methodologies that can be used to drive business performance. A lot of emphasis is given to data visualisation as a simple suite of methodologies that allow businesses to immediately identify problems and potential patterns of interest. However, data visualisation requires a trained "eye" to make sense of the results and in the context of SMEs, these skills may be missing. Alternative methodologies such as scenario planning and predictive analytics may be more insightful when trying to identify the drivers of future business performance as they focus not so much on the evolution of performance over time but rather on the contribution of different factors to future performance. For instance, senior management teams can use predictive analytics to quantify the impact of a new marketing campaign on future performance and whether it should continue in the future.

<p:b>3.6.3 Data Exploitation

<p:text>Once the data have been analysed, they need to be exploited and used to drive performance upwardly. One key lesson from the literature reviewed in the chapter is that for this to happen the knowledge generated by the data analysis has to be sufficiently specific that it can be a source of competitive advantage for the business. But how can businesses – in particular SMEs – assess whether the knowledge the data analysis has provided is specific enough to drive performance? This is not an easy question to answer as knowledge specificity has to be assessed in the light of the industry characteristics as well as of the competitors; in reality, each business may have a different answer to this question. Importantly, though, businesses need to put in place a process to be able to assess the quality of the knowledge produced from the analysis of Big Data as well as to be able to act upon it. This process has to be managed at the senior level and allow different functions of the business to input into it. Naturally, the implication is that knowledge has

to be able to flow and be shared across different parts of a firm; this may have an implication on the structure of firms itself as it requires a concerted effort to avoid the creation of data silos that stop the circulation of knowledge in the company.

<p:a>3.7 CONCLUSIONS

<p:text>This chapter has explored the relationship between Big Data and SMEs. Our starting premise is that Big Data offer a number of opportunities to small firms. Like any other organisation, SMEs produce large volumes of data while undertaking their routine activities and as a result they end up storing data of different types and complexity that can be used by researchers to improve our understanding of the drivers of their performance and by practitioners to improve the SMEs' performance. The expectation is that ultimately the exploitation of Big Data can provide managers with a clearer understanding of the drivers of the "liability of smallness" that allow us to overcome it eventually.

<p:a_no_indent>NOTES

<p:a_no_indent>REFERENCES

<c:references>

Akter, S., Wamba, S.F., Gunasekaran, A., Dubey, R., and Childe, S.J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, 182, 113–31.

Aldrich, H., and Auster, E. (1986). Even dwarfs started small: liabilities of age and size and their strategic implications. In L. Cummings and B. Staw (eds), *Research in Organizational Behavior* (vol. 8), San Francisco, CA: Jai Press, pp. 165–98.

Aldrich, H.E., and Fiol, C.M. (1994). Fools rush in? The institutional context of industry creation. *Academy of Management Review*, 19, 645–70.

Almus, M., and Nerlinger, E. (2000) Testing Gibrat's law for young firms – empirical results for West Germany. *Small Business Economics*, 15(1), 1–12.

Anderson, E. (1985). The salesperson as outside agent or employee: a transaction cost analysis. *Marketing Science*, 4(summer), 234–54.

Audretsch, D. (1995). *Innovation and Industry Evolution*, Cambridge, MA: MIT Press.

Audretsch, D., Santarelli, E., and Vivarelli, M. (1999) Start-up size and industrial dynamics: some evidence from Italian manufacturing. *International Journal of Industrial Organization*, 17(7), 965–83.

Audretsch, D.B., Klomp, L., Santarelli, E., and Thurik, A.R. (2004). Gibrat's Law: are services different? *Review of Industrial Organisation*, 24, 301–24.

Autio, E., Sapienza, H.J., & Almeida, J.G. (2000). Effects of age at entry, knowledge intensity, and imitability on international growth. *Academy of Management Journal*, 43, 909–24.

Barge-Gil, A. (2010). Cooperation-based innovators and peripheral cooperators: an empirical analysis of their characteristics and behaviour. *Technovation*, 30(3), 195–206.

Becchetti, L., and Trovato, G. (2002). The determinants of growth for small and medium sized firms. *Small Business Economics*, 19(4), 291–306.
Boyd, B.K., and Fulk, J. (1996). Executive scanning and perceived uncertainty: a multidimensional model. *Journal of Management*, 22, 1–21.
Cabral, L. (1995). Sunk costs, firm size and firm growth. *Journal of Industrial Economics*, 43(2), 161–72.
Cao, M., Chychyla, R., and Stewart, T. (2015). Big Data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423–9.
Chen, H., Chiang, R.H., and Storey, V.C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, 36, 1165–88.
Chen, J., and Lu, W. (2003). Panel unit root tests of firm size and its growth. *Applied Economics Letters*, 10(6), 343–5.
Chesher, A. (1979). Testing the law of proportionate effect. *Journal of Industrial Economics*, 27(4), 403–11.
Choi, Y.R., and Shepherd, D.A. (2005). Stakeholder perceptions of age and other dimensions of newness. *Journal of Management*, 31, 573–96.
Choudhury, V., and Sampler, J. (1997). Information specificity and environmental scanning: an economic perspective. *MIS Quarterly*, 21(1), 25–53.
Cohen, W., and Levinthal, D. (1990). Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 128–52.
Coleman, J.S. (1990). *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.
Cooley, T., and Quadrini, V. (2001). Financial markets and firm dynamics. *American Economic Review*, 91(5), 1286–310.
Davenport, T. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*, New York: Harvard Business Review Press.
Davenport, T.H., Barth, P., and Bean, R. (2012). How big data is different. *MIT Sloan Manage Review*, 54(1), 43.
Day, G.S. (1994). The capabilities of market-driven organizations. *Journal of Marketing*, 58(4), 37–52.
Del Monte, A., and Papagni, E. (2003). R&D and the growth of firms: an empirical analysis of a panel of Italian firms. *Research Policy*, 32(6), 1003–14.
Delen, D. and Ram, S. (2018). Research challenges and opportunities in business analytics. *Journal of Business Analytics*, 1, 2–12.
Dobre, C., and Xhafa, F. (2014). Intelligent services for Big Data science. *Future Generation Computer Systems*, 37, 267–81.
Dubey, R. et al. (2016). The impact of big data on world-class sustainable manufacturing. *The International Journal of Advanced Manufacturing Technology*, 84(1–4), 631–45.
Dunne, P., and Hughes, A. (1994). Age, size, growth and survival: UK companies in the late 1980's. *Journal of Industrial Economics*, 42(2), 115–40.

Dunne, T., Roberts, M.J., amd Samuelson, L. (1989). The growth and failure of US manufacturing plants. *The Quarterly Journal of Economics*, 104(4), 671–98.
Dutta, S., Narasimhan, O., and Rajiv, S. (1999). Success in high-technology markets: is marketing capability critical? *Marketing Science*, 18(4), 547–68.
Evans, D. (1987). Tests of alternative theories of firm growth. *Journal of Political Economy*, 95(4), 657–74.Fotopoulos, G., and Louri, H. (2002). Corporate growth and FDI: are multinationals stimulating local industrial development? Discussion paper no. 3128, Centre for Economic Policy Research.
Gandomi, A., and Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–44.
García, N., Sanzo, M.J., and Trespalacios, J.A. (2008). New product internal performance and market performance: evidence from Spanish firms regarding the role of trust, interfunctional integration, and innovation type. *Technovation*, 28, 713–25.
Gibrat, R. (1931). *Les inequalites economique*, Paris: Sirey.
Goddard, J., Wilson, J., and Blandon, P. (2002). Panel tests of Gibrat's law for Japanese manufacturing. *International Journal of Industrial Organization*, 20(3), 415–33.
Granovetter, M. (1985). Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, 91(3), 481–510.
Hall, B. (1987). The relationship between firm size and firm growth in the US manufacturing sector. *Journal of Industrial Economics*, 35(4), 583–606.
Hart, P.E. (1962). Size and growth of firms. *Economica*, 29(1), 29–39.
Hart, P.E. and Oulton, N. (1999). Gibrat, Galton and Job Generation. *International Journal of the Economics of Business,* 6(2), 149-164.
Hart, P.E., and Oulton, N. (1996). The size and growth of firms. *Economic Journal*, 106(438), 1242–52.
Hashem, I.A.T., Yaqoo, I., Anuar, N.B., Mokhtar, S., Gani, A., & Khan, S.A. (2015). The rise of "big data" on cloud computing: review and open research issues. *Information Systems* 47, 98–115.
Heide, J.B. (2003). Plural governance in industrial purchasing. *Journal of Marketing*, 67(4), 18.
Huber, G.P. (1991). Organizational learning: the contributing processes and the literatures. *Organization Science*, 2, 88–115.
Hunt, S.D., and Morgan, R.M. (1995). The comparative advantage theory of competition. *Journal of Marketing*, 59(2), 1–15.
Jaworski, B.J., and Kohli, A.K. (1993). Market orientation: antecedents and consequences. *Journal of Marketing*, 57(3), 53–70.
Jaworski, B., Kohli, A.K., and Sahay, A. (2000). Market-driven versus driving markets. *Journal of the Academy of Marketing Science*, 28(1), 45–54.
Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica*, 50(3), 649–70.

Kirca, A.H., Jayachandran, S., and Bearden, W.O. (2005). Market orientation: a meta-analytic review and assessement of its antecedents and impact on performance. *Journal of Marketing*, 69(2), 24–41.

Kiron, D. Shockley, R., Kruschwitz, N., Finch, G., and Haydock, M. (2012). Analytics: the widening divide. *MIT Sloan Manage Rev*iew, 53(2), 1.

Kiron, D., Prentice, P.K., and Ferguson, R.B. (2014). The analytics mandate. *MIT Sloan Manage Review*, 55(4), 1.

Kogut, B., and Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3, 383–97.

Kogut, B. and Zander, U. (1996). What firms do? Coordination, identity, and learning. *Organization Science*, 7(5), 502-18.

Kubina, M., Koman, G., and Kubinova, I. (2015). Possibility of improving efficiency within business intelligence systems in companies. *Procedia Economics and Finance*, 26, 300–5.

Laney, D. (2001). 3D *Data Management: Controlling Data Volume, Velocity and Variety*, Stamford: META Group.

LaValle, S. Lesser, E., Shockley, R., Hopkins, M.S., and Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT Sloan Manage Review, 52(2), 21-31.

Levin, D.Z., and Cross, R. (2004). The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. *Management Science*, 50, 1477–90.

Liu, Y. (2014). Big data and predictive business analytics. *Journal of Business Forecasting*, 33(4), 40.

Malomo, F. and Sena, V. (2017), Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. Policy & Internet, 9: 7-27.

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*, Washington: McKinsey Global Institute

Mata, J. (1994). Firm growth during infancy. *Small Business Economics*, 6(1), 27–39.

McAfee, A., and Brynjolfsson, E. (2012). Big Data: the management revolution: exploiting vast new flows of information can radically improve your company's performance. But first you'll have to change your decision making culture. *Harvard Business Review*, 90, 60–8.

Mcafee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., and Barton, D. (2012). Big Data: the management revolution. *Harvard Business Review*, 90(10), 60–68.

Mikalef, P., Pappas, I.O., Giannakos, M.N., Krogstie, J., and Lekakos, G. (2016). Big Data and strategy: a research framework, MCIS 2016 proceedings, 50.

Mikalef P. Pappas, I.O., Krogstie, J. & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. Inf Systems and e-Business Management, 16(3), 547–78.

Moran, P. (2005). Structural vs. relational embeddedness: social capital and managerial performance. *Strategic Management Journal*, 26, 1129–51.

Morgan, N.A., Douglas, W.V., and Mason, C.H. (2009). Market orientation, marketing capabilities, and firm performance. *Strategic Management Journal*, 30(8), 909–20.
Nahapiet, J. and Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23, 242–66.
Narver, J.C., Slater, S.F., and MacLachlan, D.L. (2004). Responsive and proactive market orientation and new product success. *Journal of Product Innovation Management*, 21(5), 334–47.
Nooteboom, B. (1993). Firm size effects on transaction costs. *Small Business Economics*, 5(4), 283–95.
Oliveira, B., and Fortunato, A. (2003). Does the firm growth follow a random walk? An application with panel data unit root test. Paper presented to the EUNIP conference, Porto, Portugal.
Ozdemir, S., Kandemir, D., and Eng, T-Y. (2017). The role of horizontal and vertical new product alliances in responsive and proactive market orientations and performance of industrial manufacturing firms. *Industrial Marketing Management*, 64, 25–35.
Peters, L., and Karren, R.J. (2009). An examination of the roles of trust and functional diversity on virtual team performance ratings. *Group & Organization Management*, 34, 479–504.
Petrunia, R. (2007). Persistence of initial debt in the long-term employment dynamics of new firms. *Canadian Journal of Economics*, 40(3), 861-880.
Petrunia, R. (2008). Does Gibrat's Law hold? Evidence from Canadian retail and manufacturing firms. *Small Business Economics*, 30, 201–14.
Provost, F., and Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), 51–9.
Rindfleisch, A., and Heide, J.B. (1997). Transaction cost analysis: past, present, and future applications. *Journal of Marketing*, 61(4), 30–54.
Rothaermel, F.T., and Deeds, D.L. (2004). Exploration and exploitation alliances in biotechnology: a system of new product development. *Strategic Management Journal*, 25(3), 201–21.
Schoenherr, T., and Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: current state and future potential. *Journal of Business Logistics*, 36(1), 120–32.
Schroeck M. Shockley, R., Smart, J., Romero-Morales, D. and Tufano, P. (2012). Analytics: The Real-World Use of Big Data (IBM Institute for Business Value - Executive Report). IBM Institute for Business Value. Sena, V., and Ozdemir, S. (2020). Spillover effects of investment in BDA in B2B relationships: what is the role of human capital? *Industrial Marketing Management*, 86(April), 77–89.
Shmueli, G., and Koppius, O.R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553–72.
Sivarajah, U., Kamal, M.M., Irani, Z., and Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–86.
Slater, S.F., and Narver, J.C. (1995). Market orientation and the learning organization. *Journal of Marketing*, 59(3), 63–74.

Stinchcombe, A.L. (1965). Social structure and organizations. In J.P. March (ed.), *Handbook of Organizations*, Chicago, IL: Rand Mcnally, pp. 142–93.

Stubbs E. (2014). *Big Data, Big Innovation: Enabling Competitive Differentiation through Business Analytics*, Hoboken, NJ: Wiley.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, 35(1), 40–59.

Szulanski, G. (1996). Exploring internal stickiness: impediments to the transfer of best practice within the firm. *Strategic Management Journal*, 17, 27–43.

Tsai, K., Chou, C., and Kuo, J. (2008). The curvilinear relationships between responsive and proactive market orientations and new product performance: a contingent link. *Industrial Marketing Management*, 37(8), 884–94.

Tsai, W. (2001). Knowledge Transfer in Intraorganizational Networks: Effects of Network Position and Absorptive Capacity on Business Unit Innovation and Performance. *Academy of Management Journal*, 44(5), 996-1004.

Tsai, W., and Ghoshal, S. (1998). Social capital and value creation: the role of intrafirm networks. *Academy of Management Journal*, 41, 464–76.

Urga, G., Geroski, P., Lazarova, S., and Walters, C. (2003). Are differences in firm size transitory or permanent? *Journal of Applied Econometrics*, 18(1), 47–59.

Van Wijk, R., Jansen, J.J.P., and Lyles, M.A. (2008). Inter- and intra-organizational knowledge transfer: a meta-analytic review and assessment of its antecedents and consequences. *Journal of Management Studies*, 45, 830–53.

Wamba, S.F Gunasekaran, A., Akter, S., Ji-fan Ren, S., Dubey, R., and Childe, S.J. (2017). Big data analytics and firm performance: effects of dynamic capabilities. *Journal of Business Research*, 20, 356–65.

Wang, Y., Kung, L., and Byrd, T.A. (2018). Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13.

Wedel, M., and Kannan, P. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121.

Williamson, O.E. (1985). *The Economic Institutions of Capitalism*, New York: The Free Press.

Williamson, O.E. (1989). Transaction cost economics. In R. Schmalensee and R. Willig (eds), *Handbook of Industrial Organization*, vol., Amsterdam: North-Holland pg. 135-182.

Wills, M.J. (2014). Decisions through data: analytics in healthcare. *Journal of Healthcare Management*, 59(4), 254–62.

Yli-Renko, H., Autio, E., and Sapienza, H.J. (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic Management Journal*, 22, 587.

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

Zeng, D.D., Lusch, R., and Chen, H-C. (2010). Social media analytics and intelligence. *Intelligent Systems, IEEE*, November/December, 12–16.

</c:references>

</s:chapter>

---

1. Traditionally, market orientation has been described in three major ways as a set of behaviours and activities (Jaworski & Kohli, 1993), part of an organisational culture (Day, 1994; Slater & Narver, 1995) and a resource (Hunt & Morgan, 1995; Dutta et al. 1999; Morgan et al. 2009). The behavioural perspective views market orientation as the organisation-wide generation and dissemination of information and responsiveness to market intelligence on customer needs, competitive actions and strategies, and the wider business environment (Jaworski & Kohli, 1993). The cultural perspective focuses on organisational norms and values that drive behaviours which are consistent with market orientation including customer orientation, competitor orientation and inter-functional coordination (Slater & Narver, 1995; Kirca et al. 2005). The resource-based perspective, on the other hand, perceives market orientation as a valuable, rare, socially complex and causally ambiguous resource which enables a firm to produce an offering that aligns with the specific tastes and preferences of his market segments (Hunt & Morgan, 1995; Kirca et al. 2005).

2. Urga et al. (2003) and Del Monte and Papagni (2003) find that firm growth follows a random walk and therefore the Gibrat's law holds. On the contrary, Goddard et al. (2002) and Oliveira and Fortunato (2003) using a panel data of Japanese and Portuguese manufacturing firms, respectively, provide some support for the firm sizes are mean-reverting.

3. There are some remarkable exceptions such as Audretsch et al. (2004) who analysed the Dutch hospitality sector (restaurants, cafes, hotels and camping sites) which is characterised by small and independent firms serving mostly local markets. Their results showed that the Gibrat's law is accepted in most cases consistently with the traditional view that services firms grow at the same rate independently of size. However, Oliveira and Fortunato (2004) find the opposite when analysing the growth patterns of firms from the Portuguese service sectors. Equally, Petrunia (2007) finds that the Gibrat's law does not hold for Canadian retail trade.