



HAL
open science

Management of Big data: An empirical investigation of the Too-Much-of-a-Good-Thing effect in medium and large firms

Claudio Vitari, Elisabetta Raguseo, Federico Pigni

► **To cite this version:**

Claudio Vitari, Elisabetta Raguseo, Federico Pigni. Management of Big data: An empirical investigation of the Too-Much-of-a-Good-Thing effect in medium and large firms. *Systèmes d'Information et Management*, In press. hal-03876785

HAL Id: hal-03876785

<https://amu.hal.science/hal-03876785>

Submitted on 28 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Management of Big data: An empirical investigation of the Too-Much-of-a-Good-Thing effect in medium and large firms

*Claudio VITARI**, *Elisabetta RAGUSEO*** & *Federico PIGNI****

* Aix Marseille Univ, Université de Toulon, CERGAM, FEG, Aix-en-Provence, France

** Department of Management and Production Engineering, Politecnico di Torino, Torino, Italy

*** Department of Management, Technology and Strategy, Grenoble Ecole de Management, Grenoble, France

Abstract

Firms adopt Big data solutions, but a body of evidence suggests that Big data in some cases may create more problems than benefits. We hypothesize that the problem may not be Big data in itself but rather too much of it. These kinds of effects echo the Too-Much-of-a-Good-Thing (TMGT) effect in the field of management. This theory also seems meaningful and applicable in management information systems. We contribute to assessments of the TMGT effect related to Big data by providing an answer to the following question: When does the extension of Big data lead to value erosion? We collected data from a sample of medium and large firms and established a set of regression models to test the relationship between Big data and value creation, considering firm size as a moderator. The data confirm the existence of both an inverted U-shaped curve and firm size moderation. These results extend the applicability of the TMGT effect theory and are useful for firms exploring investments in Big data.

Keywords

Too-Much-of-a-Good-Thing effect, inverted U-shaped curve, Big data, business value, medium and large firms

1. Introduction

A growing body of evidence in both academic literature and the business field suggests that Big data can engender more problems than benefits in adopting organizations.

The Australian Broadcasting Corporation revealed that the introduction of a new procedure leveraging Big data analytics and automation in the public entity responsible for distributing subsidies and recovering debts contributed to the death of more than 2,000 people (Medhora, 2019). The multiplication of data sources, the acceleration of data processing, the uncertainty in the veracity of the available data regarding the subsidy rights of each beneficiary, and the automation of customer relationship management ended up being deleterious for the citizens. Complementarily, the Facebook–Cambridge Analytica data scandal highlighted other aspects of the potential harms of Big data initiatives: the ease of sharing large volumes of data, the speed of sharing these data, and the lack of consideration of the veracity of user consent in sharing their data contributed to the abuse of these data, privacy loss and a sharp slide of Facebook’s share prices from USD 185 to USD 152 in the weeks following the disclosure (Cadwalladr & Graham-Harrison, 2018).

These two examples are not just anecdotal. Statistically speaking, despite increasing investments in Big data, firms are failing in their efforts to become data driven (NewVantage Partners, 2019). The failure rate of Big data projects increased from 60% to 85% between 2016 and 2017, and Gartner reports that this sobering and declining trend in Big data success could have been in place for a few years prior to 2016 (Asay, 2017). While several explanations are advanced, organizations seem unable to effectively learn from the past and overcome the different obstacles in Big data projects (NewVantage Partners, 2019).

Given these uncertainties, we aimed to deepen our understanding of the negative contribution of Big data to value creation by exploring the following research question: When does the extension of Big data lead to value erosion? We originally hypothesize that the problem does not lie in the characteristics of Big data projects or implementations but rather that “too much of it” produces value erosion. Therefore, we will test for the emergence of a significant curvilinear pattern, questioning the fundamental assumption of a positive linear relationship between antecedents and outcomes. While very few studies have adopted a similar approach in the information systems field, we believe in the necessity to shed light on possible negative outcomes of Big data investments, for which the “more is better” assumption is semantically reinforced. The “Big” in Big data implicitly assumes that “Big” data are better than “small” data and that more data means better value for the organization.

Our contributions have relevant managerial implications. In fact, few organizations consider the possible danger of overinvesting in Big data initiatives. With estimates forecasting that the Big data market will be valued at USD 135.22 billion by 2025, with a growth rate of 19% from 2017 to 2025 (Navanath, 2020), determining when Big data returns become negative is crucial.

The rest of the paper is structured as follows. First, we present the theoretical background and formulate our hypotheses. We then detail the methodology we followed and present our results. We continue with a discussion of the findings, guidelines for future studies and our conclusions.

2. Theory

Our theoretical background is built on the Too-Much-of-a-Good-Thing (TMGT) effect theory. We develop our theoretical framework in a funnel approach, starting from general considerations on the meaning of this theory in the management discipline. Subsequently, we

portray the application of TMGT in Management of Information Systems (MIS). Finally, we analyze the TMGT effect in the specific domain of Big data.

2.1. The Too-Much-of-a-Good-Thing effect in management

The Too-Much-of-a-Good-Thing (TMGT) effect accounts for the paradox that typically beneficial antecedents can cause harm when taken too far (Pierce & Aguinis, 2013). This harm could be a waste of resources, in that the additional antecedent does not produce any further benefit. The harm could also be severe, in cases where the additional antecedent could produce undesirable outcomes.

Otherwise, the TMGT theory suggests that antecedent variables, generally those that lead to desirable outcomes, may generate negative effects on the dependent variable. This relation implies that the widely accepted monotonic positive relationship between antecedents and outcomes are valid only to a certain extent, exhibiting an overall curvilinear pattern that forms an inverted U-shaped curve. Therefore, the relationship between antecedents and outcomes becomes positive, in which case the relationship will start to display an asymptotic pattern or trend in the opposite direction, becoming negative. The inflection point at which the relationship becomes negative and the steepness of the curve depend on the specific context of the relationship.

While seeking the right balance between too few and too much is a long-standing philosophical debate with contemporary defenders (Illich, 1973), this principle is largely disregarded in the management literature. Empirical managerial evidence and the assumptions of management scholars remain substantially grounded on the adage that “more is better” (Gomes et al., 2003; Pierce & Aguinis, 2013). This assumption could be considered a founding principle for the establishment of the management discipline (Taylor, 2012). This assumption can be sourced to economics (Cass, 1965) and the idea of progress (Mackenzie, 1899; Nisbet, 1994), which largely permeates management and other fields.

Even when the “more is better” approach is abandoned, it is often replaced with a categorical imperative to do more. What was originally an invitation to increase beneficial antecedents and generate higher benefits translates into organizational, economic and management actions (Goold, 1999; Rich, 2013). The implications are vast and based on the shared and accepted vision that increasing beneficial antecedents is a necessary and natural condition to avoid economic collapse or organizational failure. By taking this stance, management scholars remained blind in researching the alternative scenarios of finding the right balance between too little and too much. Nonetheless, multiple pieces of evidence of the TMGT effect have already emerged in several disciplines, including organizational behavior, human resource management, entrepreneurship, and strategic management, where 11% of the articles investigated U-shaped relationships (Pierce & Aguinis, 2013).

2.2. The Too-Much-of-a-Good-Thing effect in management information systems

We believe TMGT is meaningful and applicable in MISs, more than what is actually practiced. Notably, among the 388 articles published in the basket of eight journals in 2019, 121 (31%) explicitly state at least one hypothesis. A total of 769 hypotheses are tested, and 743 of these hypotheses (97%) propose linear relationships among variables. Only 7 articles (1.8%) and a total of 13 hypotheses (1.7%) explored inverted U-shaped relationships.

The adage “more is better” is rooted in many largely applied MIS models, including the Technology Acceptance Model (Davis et al., 1989; Venkatesh et al., 2003) and its derivatives. Notably, growth is embedded in MIS theory, not just as an opportunity but as an organizational imperative that IS can satisfy (Melville et al., 2004; Street & Meister, 2004).

Nonetheless, as in other disciplines (Fleishman, 1998), some MIS scholars have questioned the legitimacy of these assumptions, challenged the monotonic linear relationship between antecedents and outcomes and demonstrated that more is not always better (Wolff, 2016).

The concept of information overload is the most largely accepted declination of the TMGT effect in the MIS discipline. Information availability is a widely considered antecedent of effective decision-making, IS satisfaction, and intention to use, among other positive outcomes (DeLone & McLean, 2003). Instead, information overload can negatively affect these outcomes when information becomes excessive beyond a certain threshold and can start to generate negative effects that reverse the relation between antecedents and outcomes from positive to negative. Information overload may then lead to poor decisions (Schroeder & Benbasat, 1975), unsatisfied IS users, and other undesirable effects. Notably, while the inflection point and the steepness of the curve related to negative effects are context specific, the relationship between the antecedents of information overload and the corresponding effects is curvilinear rather than monotonic.

2.3. The Too-Much-of-a-Good-Thing effect in Big data

With our study, we question this fundamental assumption of a linear relationship between antecedents and outcomes, and we join the active minority of MIS scholars advancing the study of inverted U-shaped relationships. If proved, the implication is correctly that investments in Big data may not always translate into more business value.

2.3.1. Extension of Big data

The Big data phenomenon emerged from the increasing amount of digital data that organizations were generating, collecting and gathering (Cox & Ellsworth, 1997). Indeed, Big data refers to datasets so large that they fall “beyond the ability of typical database software tools to capture, store, manage, and analyze” (Chang et al., 2014; Richardson et al., 2021).

The challenges organizations faced in the early days of Big data related to different organizational and technological factors. Data were produced in different places, structured with different standards, recorded in heterogeneous formats, processed with disparate intents, and stored in multiple databases. Software applications were not able to analyze all the available data, extract information from the data or interact with the data. New architectures and applications have been developed to overcome these limits and harness these masses of data (Lynch, 2008). Hence, Big data technologies allow the extraction, processing and analysis of large amounts of data to extract value from them (Ren et al., 2016).

2.3.2. Business value of Big data

The business value of Big data can spread in different directions. Big data can provide clues about customers’ preferences, feedback about product performance and strategic insights (Cappa et al., 2021). Alibaba and Tencent were able to outperform traditional banks in loan delivery due to their ability to process Big data and compute more precise credit scores (Nonninger, 2018). These outcomes can be organized around four dimensions of business value: transactional value, strategic value, transformational value and informational value (Gregor et al., 2006; Ren et al., 2016). Transactional value refers to the degree to which the user perceives Big data solutions providing operational benefits by reducing operating costs or communication costs. Strategic value refers to the degree of perceived benefit to the organization at the strategic level by enabling a rapid response to change or improving customer relations. Transformational value refers to the degree of perceived change in the structure and capacity of a firm as a result of Big data solutions, which serve as a catalyst for future benefits and expanding company capabilities. Informational value refers to the degree

to which the user of Big data solutions benefits from information, such as improving the management of data or enabling faster access to data than previously achieved.

Nevertheless, size alone does not explain the novelty of Big data processes and impacts (Constantiou & Kallinikos, 2015). As such, three Vs have been advanced to characterize Big data: Volume, Velocity, and Variety (Lee, 2017). Volume refers to the amount of data an organization collects and/or generates, velocity refers to the speed at which data are generated and processed, and variety refers to the number of data types, both structured and unstructured.

Similarly, the promise of the value creation opportunities associated with data led firms to look at novel ways to generate, collect and gather even more data with the newly available technological capabilities. The adage of “more is better” was then applied to data, stimulating organizations to gather even more data, often without a specific purpose, to create more value and ‘bigger’ value (Manyika et al., 2011; McAfee & Brynjolfsson, 2012). The amount of digital data in existence has grown exponentially, and firms have proliferated “digital” initiatives in the constant search for new value creation opportunities based on the endless call to explore and leverage Big data (Chen et al., 2012; Grover et al., 2018). This approach is indirectly supported by the current research on Big data analytics. Notably, despite using different theoretical perspectives, tests have confirmed the hypothesized positive linear relationship between Big data and business value.

2.3.3. Limits to the business value of Big data

Nonetheless, previous studies have identified numerous challenges that firms must overcome when pursuing Big data initiatives and advancing solutions. While calls to explore the “negative” side of technology adoption have a long tradition (Illich, 1973; Loebbecke & Picot, 2015), only a few studies have discussed the possible negative effects of Big data solutions and have only focused on the increased privacy and security risks. Interestingly, studies in marketing observed an inverted U-shaped relationship between knowledge integration mechanisms, which could play a role in Big data exploitation, and product innovativeness (Tsai et al., 2015). In the Big data literature, the TMGT effect remains uncharted territory, acting as a pink elephant in the room.

We wonder, then, where the limit of positive outcomes sits and if it exists for all Big data initiatives. The Australian public entity responsible for distributing subsidies and recovering debts may have invested too much in Big data by collecting data from too voluminous and too disparate data sources to effectively handle. Additionally, the data collected may have been too uncertain in quality and veracity. The interactions with the beneficiaries may be too frequent to be manageable by beneficiaries. In particular, citizens indicated velocity as a major source of annoyance, with the targeted public contesting the reception of too many messages. This high frequency of messages to beneficiaries seems strictly related to the weak veracity of the data collected and the consequent need of the system to ask for additional evidence from the beneficiaries to reconcile contrasting data (Medhora, 2019). Similar kinds of doubts could be raised in the Facebook–Cambridge Analytica data scandal. Perhaps the collection of masses of user data has been too easy, the sharing of user data too fast, and the control of the veracity of the consent in data sharing too neglected (Cadwalladr & Graham-Harrison, 2018).

Different elements may have blinded researchers and diverted their attention from the problem. The term Big data itself may have been a factor, implicitly suggesting a novel and limitless space of opportunities and possibilities for organizations to explore. In this new Big data paradigm, there would be no space for inflection points and curvilinear patterns; instead,

there would be only room for infinite growth sustained by datasets that might appear massive today but probably small in the near future.

Another important factor concerns selection bias resulting from the restriction of predictor ranges that may afflict current studies. As previously demonstrated (Pierce & Aguinis, 2013), failing to include the entire range of predictor scores (e.g., only successful firms) may limit the possibility of observing TMGT effects. The available Big data studies, for example, tend to focus on limited or specific time frames for data collection, hindering the observability of effects that manifest over the long term. The scarcity of longitudinal studies in Big data research (e.g., Raguseo et al., 2020) is suggestive of this trend.

3. Hypothesis development

Accordingly, we hypothesize that Big data investments follow TMGT theory. Specifically, we develop two hypotheses: the first hypothesis concerns the direct relationship between the extension of Big data and the business value of Big data, and the second hypothesis considers firm size as a moderator of this direct relationship.

3.1. The effect of the extension of Big data on the business value of Big data

We contend that assuming a monotonic linear relationship between Big data and business value could lead to the development of deficient theories and to the failure of many organizational initiatives. We hypothesize that the TMGT effect characterizes Big data investments and that Big data initiatives may harm competitiveness when outgrowing their beneficial potential. Therefore, following an initial phase in which increases in Big data investments positively grow infrastructures and (Grover et al., 2018) benefit the organization, additional increases would first result in a diminishing return for various outcomes. Moreover, after a certain and context-specific threshold is reached (the inflection point), outcomes may become negative. Past this phase, the negative outcomes will indirectly harm a firm's business value and negatively affect transactional, strategic, transformational, or informational business value (Gregor et al., 2006; Raguseo, 2018; Ren et al., 2016). The first hypothesis has been empirically confirmed in the public press (Cadwalladr & Graham-Harrison, 2018; Medhora, 2019). Nonetheless, little quantitative research has been performed by scholars.

3.1.1. Plateauing revenues from investments in Big data

Two countervailing forces would explain this inverted U-shaped curve. On the one hand, Big data investments follow a pattern similar to that of traditional IT investments and are positively associated with business value (Mithas et al., 2012; Mitra, 2005). Then, Big data investments will start to manifest diminishing returns similar to such as other production factors, IT investments, economic growth, and IT productivity (Lin & Shao, 2006). Hence, we suggest that Big data extension could be positively associated with increases in output variables and indirectly associated with a firm's competitive performance. However, we hypothesize a curvilinear relationship sloping upward with diminishing returns and evolving in a logarithmic-concave shape, as progress related to technical effectiveness tends to decrease as IT investments increase (Lin & Shao, 2006). In practice, we assume that the first Big data initiatives will focus on the greatest potential for increased revenue. Hence, the generated revenues could be very large for a single extension of Big data. The following initiatives, implying an additional extension of Big data, would increase revenues but to a lesser extent than that traditionally hypothesized (Figure 1.a on revenues of Big data).

3.1.2. Increasing costs of investments in Big data

On the other hand, the second force that explains the inverted U-shaped curve between the extension of Big data and their business value should manifest as a positive association between the costs and the extension of Big data. Maintaining and refining the current IT for Big data, without the additional extension of Big data, would cost less than adopting new IT and, hence, extending Big data (Steelman et al., 2019). Therefore, we posit that each additional extension of Big data costs more than the previous extension. Notably, each extension would demand increasingly sophisticated technologies and new and rarer skills, increasing the costs (Figure 1.b on costs of Big data).

3.1.3. Combination of plateauing revenues and increasing costs of investments in Big data

The additive combination of these two kinds of latent mechanisms of revenues and costs results in an inverted U-shaped relationship (Haans et al., 2016). The two mechanisms combined negatively affect the net benefits and progressively reduce business value. Hence, the overall value created will diminish to a point at which an additional extension of Big data would no longer create business value. Further extensions of Big data would then destroy even more business value because the marginal costs would overrun the small marginal revenues of the previous Big data initiative (Figure 1.c on value of Big data equals revenues minus costs).

For example, let us consider a Big data initiative focused on customer relationship management that is capable of providing, in near real time, a whole and multifaceted representation of a firm's customers. In this case, additional speed in data frequency, new sources of data, and the additional capacity to store information about customers would unlikely result in better products or services. Conversely, while larger investments in Big data may provide additional frequency or increases in the precision of the models, the additional effort required to achieve these objectives will be affected by diminishing returns, yielding a plateau relationship. Ultimately, the net business value may be negative.

This situation is common in practice, as firms often operate under uncertainty and lack precise information for a project (Tiwana et al., 2007). Therefore, a portion of companies do not realize that an additional extension of their Big data could be profitable to pursue; hence, this portion of companies does not pursue this beneficial extension of Big data. Similarly, another portion of companies does not recognize that additional extensions of their Big data would destroy their business value, and hence, this portion launches such extensions.

The above considerations, including the net effect of Big data revenues based on a positively sloping curve with decreasing returns and the costs of Big data based on a linear relationship, would produce an inverted U-shaped relationship between the extension of Big data and value creation. Hence, we posit the following hypothesis.

Hypothesis 1: A curvilinear, inverted U-shaped relationship exists between the extension of Big data and business value. Specifically, as the volume of Big data increases, business value increases until a maximum point is reached; as the Big data extent continues to increase, business value decreases.

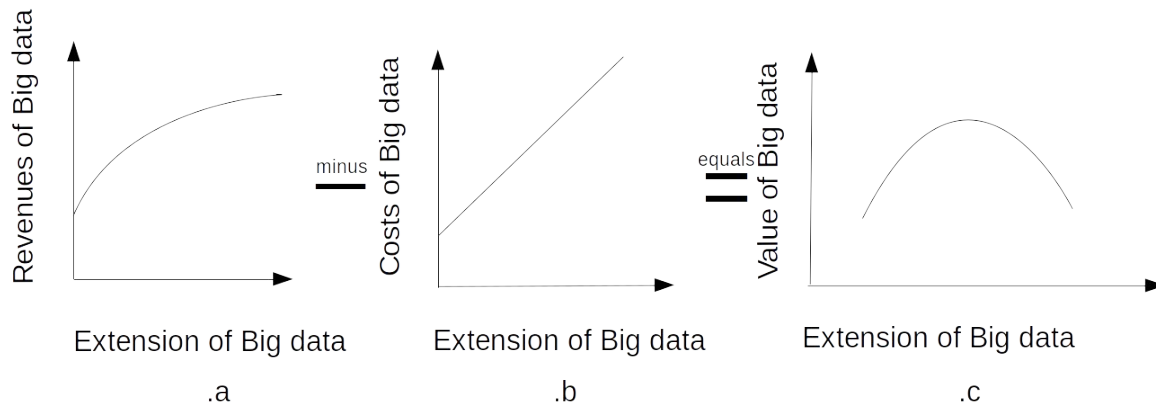


Figure 1: Graphical representation of the theoretical development of the curvilinear, inverted U-shaped relation.

3.2. The moderating effect of firm size on the extension of Big data and the business value of Big data

Although theory seems to support the possible inverted U-shaped relation between the extension of Big data and business value, the determination of the inflection point and the slope of the curve remain uncertain.

3.2.1. Inflection point

Evidence suggests that the localization of the inflection point and the slope of the curve is context specific (Pierce & Aguinis, 2013). What could be excessive in one context might be insufficient in another, and the amount of value created or destroyed for a tiny extension of Big data could be large or small depending on the particular context. Stated otherwise, each organization could reach its inflection point at a different level of Big data extension; notably, when one organization might start suffering from “too much” Big data, a second organization might not experience any negative effect well beyond the inflection point of the first organization. Additionally, each organization could create or destroy more or less value for the same extension of Big data: while one organization might largely increase the business value of its Big data with a tiny increase in Big data extension, a second organization might generate only a small increase in business value for the same extension of Big data. After the inflection point is reached, the consequences would be reversed; the organization with large increases in business value will see a larger reduction in business value with a tiny increase in Big data extension, while the second organization will see only a marginal drop in business value.

This important context-specific factor related to the inflection point and the slope of the curve takes the role of a moderating variable between the extension of Big data and the business value of Big data (Haas et al., 2016). Indeed, the moderators of U-shaped relationships can shift the turning point of the curve left or right, and they can flatten or steepen the curve.

Empirical evidence suggests that environmental dynamism could play a role in the determination of the inflection point (Haas et al., 2016). Several variables could moderate the relationship between the extension and value of Big data, such as the degree of digitalization, the extent of dynamic capabilities, and the industry. Nonetheless, for the sake of empirical feasibility and theoretical parsimony as well as to reduce the complexity and the number of

statistical tests and models, we limited the exploration of the moderating effects to one single variable.

3.2.2. The firm size contribution to the inflection point

We suggest that the size of the organization could be the most important determinant of this inflection point. Size is a structural characteristic of organizations; what is too large for one company may not be large enough for another. Firm size is largely recognized as a significant moderator of firm performance; it can change the degree to which certain postures, structures, and tactics boost firm performance in consideration of different strategic goals. Specifically, firm size is a variable that can affect how firms invest in and profit from IT (Oliveira & Martins, 2011). Several explanations are proposed for the moderating effect of firm size on the relationship between IT investments and firm performance, including resource bases, economies of scale and scope, pre-emptive move capabilities, formalization levels, decentralization patterns, specialization trends, management schemes, and innovativeness levels. Larger organizations exhibit higher degrees of differentiation and formalization, more decentralized managerial decision-making authority systems, higher levels of task specialization, more complex forms of communication and greater organizational inertia than smaller organizations; thus, IT investments by large organizations are more expensive than the same investment in smaller organizations. Furthermore, a high level of bureaucracy in larger firms makes them slower in making strategic moves and can lead to a struggle to respond to changing conditions (Eisenhardt & Martin, 2000), making them less innovative than smaller firms.

A similar debate exists concerning Big data solutions, where some aspects would favor smaller companies and others benefit larger companies (Dong & Yang, 2020; Mikalef et al., 2019). Earlier studies found few differences based on firm size (Bughin, 2016). For example, Big data as a service is considered particularly beneficial for smaller companies, as it grants radical innovation possibilities, flexibility, and efficiency, as well as the possibility of overcoming the complementary resource gap with larger companies. Conversely, stronger IT capabilities, advanced IT, better training and more secured data management seem to explain why larger companies could be better off taking advantage of Big data than are small companies (Kamioka & Tapanainen, 2014).

We think that all these reasons contribute to the quadratic effect. The two mechanisms, i.e., revenues and the cost of Big data, that define the inverted U-shape between the extension of Big data and the business value of Big data are related to the moderating role of firm size in determining the business value of Big data for any extension of Big data.

3.2.2. The firm size moderation of the revenues from investments in Big data

Concerning the revenues generated by a Big data initiative, smaller companies would gain greater advantages from the extension of their Big data compared with the income increase achieved by larger companies. With their flexibility, smaller companies have an advantage over larger companies that suffer from increased communication complexity and organizational inertia (Eisenhardt & Martin, 2000). Moreover, the time required for larger companies to overcome organizational inertia would allow the potential value creation of a Big data initiative to be captured, at least partially, by some competitors, which may indeed be the flexible, smaller companies cited above. Hence, since the relationship between Big data revenues and Big data extension follows a positive curvilinear function with decreasing returns, an increase in the size of the organization would flatten the slope of the Big data revenue curve (Figure 2.a on revenues of Big data). The effect of this flattening of the slope of the Big data revenue curve is also a flattening of the inverted U-shaped curve between the extension of Big data and the business value they generate (Figure 2.c on value of Big data).

3.2.3. The firm size moderation of the costs of investments in Big data

Similarly, the costs of the same extension of Big data would be less important for smaller companies in comparison to the costs to the larger companies have to support the same extension of Big data. An extension of Big data could cost more in large companies because this extension must be integrated into a complex environment, technically as well as organizationally, and requires considerable management involvement (Love et al., 2005). In contrast, a small company would have simpler infrastructures, thus requiring less integration efforts and management time when an additional IT is added, which would translate into lower costs. Finally, larger companies would support higher costs for the same increase in Big data. Graphically speaking, the increase in the size of the organization would accentuate the steepness of the Big data linear cost function (Figure 2.b on costs of Big data).

3.3.4. Combination of the firm size moderation, revenues and costs of investments in Big data

The steeper Big data cost line reflects a shift of the inflection point to the left in the inverted U-shaped curve linking the extension of Big data with their business value Big data (Figure 2.c on the value of Big data). This additive combination of the two mechanisms of Big data, the revenues minus the costs of Big Data, would produce a flattening of the inverted U-shaped curve and a shift to the left of the turning point. This shift would increase with increasing company size. Hence, we state the following hypothesis (Figure 3).

Hypothesis 2: *The curvilinear, inverted U-shaped relation between the extension of Big data and business value is stronger for smaller companies than for larger ones.*

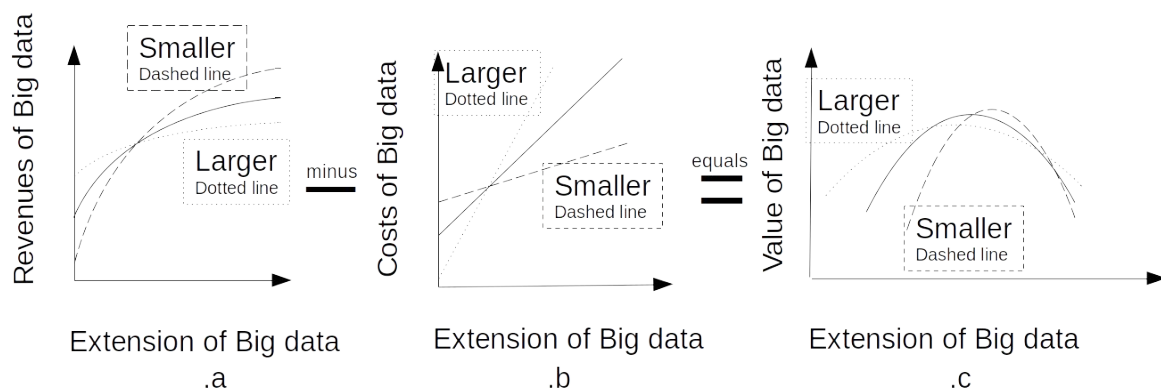


Figure 2 Graphical representation of the theoretical development of the moderation of firm size on the curvilinear, inverted U-shaped relation.

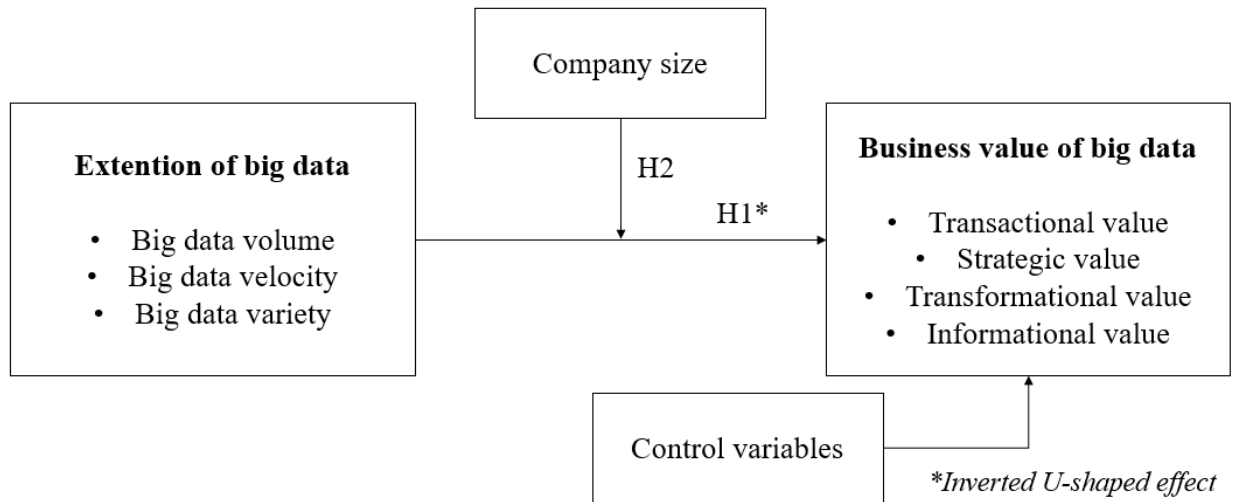


Figure 3 Research model

4. Materials and methods

This section shows the following methodology based on four steps. First, we present the data collection process of our study. Second, we show the operationalization of the variables included in our model for hypothesis testing. Third, we included details about the empirical models. Finally, we present the results of our study.

4.1. Data collection

We administered a questionnaire to medium- and large-sized French firms considering the number of employees since in small companies, the rate of Big data adoption is very low. As our study considers the effects at the firm level, we followed previous studies that targeted the chief information officer (CIO) as the main informant. We implemented a random sampling method to select 1,962 medium and large French companies to interview from a population of 19,875 medium and large companies belonging to the sectors shown in Table 1 and registered in Bureau Van Dijk's DIANE database, which is one of the main sources of financial information on firms in France. We aimed to gather 200 questionnaires, assuring a 95% confidence level and a 6.9% confidence interval. We conducted a pilot study on a randomly selected subsample of the population to gather the first 30 answers from companies. The gathering of these first 30 questionnaires allows us to test the comprehensibility of the questions, to identify possible response issues and to establish the expected response rate. To collect 30 questionnaires, we needed 142 randomly selected companies, which made the response rate of the pilot test 21.13%. All the questions led to appropriate answers and, thus, did not require further changes. Therefore, the final questionnaire remained the same as the pilot.

The data gathering process involved three steps. In the first step, we contacted the company to inform them of the aim of the study and to request permission to contact the CIO. In the second step, the CIO was contacted and asked about his or her willingness to participate in the survey. When the CIO was not available at the time agreed upon in the first call, we made a second appointment. Therefore, the questionnaire was completed, either in the second or third

step, according to the availability of the CIO. When the CIO was unable to answer the questionnaire, we identified another qualified respondent knowledgeable about the firm's investments and the adoption of Big data solutions. Overall, 200 questionnaires, 30 from the pilot test and 170 from the survey delivery phase, were completed (Table 1).

Dimension	Category	Percentage (%)
Role	General Director	26.54%
	Chief Executive Officer	4.08%
	Chief Information Officer	63.26%
	Other person qualified to make Big data investments	6.12%
Size of the company	Medium-sized firms (between 50 and 249 employees)	84.21%
	Large companies (more than 250 employees)	15.79%
Industry	Manufacturing	33.50%
	Wholesale and retail trade	19.50%
	Professional, scientific and technical activities	11.50%
	Construction	6.50%
	Transportation and storage	6.50%
	Administrative and support service activities	6.00%
	Accommodation and food service activities	4.50%
	Human health and social work activities	4.00%
	Other sectors	8.00%

Table 1 Sample characteristics

4.2. Measures

The questionnaire consisted of two sections. The first section, which all companies answered, determined the presence and extension of Big data. The second section was made available to those companies that had been found to have adopted Big data solutions.

4.2.1. Extension of Big data

To define the extension of Big data for every company, we followed a two-step method. First, to determine the presence or lack of Big data, we set inclusion criteria for each question. Regarding velocity, we looked for real-time or near real-time latency. The presence of more than one heterogeneous data source was assessed to establish variety. For volume, we included firms with stored data exceeding a petabyte. If the respondents satisfied these inclusion criteria, they were asked to confirm our conjecture that they had Big data solutions via a dichotomous (yes/no) question. If their answer was affirmative, we asked which Big data solutions the companies adopted among the following options:

- *Visual analytics software or other software used to display analytical results in visual formats,*
- *Scripting languages or other programming languages that work well with Big data (e.g., Python, Pig, and Hive),*
- *In-memory analytics software or other Big data processing software used in computers to increase speed,*
- *MapReduce and Hadoop software or other software used to process Big data across multiple parallel servers,*
- *Machine learning software or other software is used to rapidly find the model that best fits a dataset,*
- *Natural language processing or other software used for texts, including for information extraction, text summarization, question answering, or sentiment analysis,*
- *Social media analytics software (content-based analytics and structure-based analytics), and*

- *Predictive analytics software was used to extract information from the data and predict trends and behavior patterns.*

Second, to obtain a synthetic value of the extension of the Big data for each company, we computed an index of the three V dimensions that characterize Big data (I. Lee, 2017): velocity, variety and volume. This index is the mean value of three standardized V dimensions: extension of Big data = (standardized velocity + standardized variety + standardized volume)/3. Details about the operationalization steps and the questions included in the questionnaire for every dimension are provided in Table 2.

Dimension of the extension of Big data	Description	Operationalization (choices of the respondent)
Velocity	We designed a three-level scale variable that considers the shortest latency of company data. The following question was included: To date, what is the shortest latency of your data?	<ol style="list-style-type: none"> 1. After a long time (data are updated in the database only once or irregularly), 2. Near time (data are updated in the database at set and regular time intervals), 3. Real time (data are updated in the database as they are obtained, with little or no latency).
Variety	We counted the number of sources of data for each company. The following question was included: To date, what are the sources of data of your company beyond traditional databases? (multiple choice)	<ol style="list-style-type: none"> 1. Radio frequency identification system data, 2. Clickstream data, 3. Smart/intelligent/connected meter data or other smart/intelligent/connected object data, 4. Global positioning system data, 5. Point of sales data or other transactional data sources, 6. Social media posts, 7. Weblog posts, 8. Microblog (e.g., Tweets) posts, 9. Online portal content, 10. Email message content, 11. Other natural language text sources, 12. Audio sources, 13. Image sources, 14. Video sources, 15. Other sources, please specify: _____.
Volume	We built a five-level variable that operationalizes the total amount of data stored in all the databases of the company. The following question was included: To date, what is the total amount of data stored in all the databases of your company?	<ol style="list-style-type: none"> 1. Less than 1 terabyte, 2. Between 1 terabyte and 1 petabyte, 3. Between 1 petabyte and 1 exabyte, 4. Between 1 exabyte and 1 zettabyte, 5. More than 1 zettabyte.

Table 2. Dimensions of the extension of Big data

4.2.2. Business value of Big data

The business value of Big data is operationalized as a second-order construct (Figure 4) by combining the four dimensions of business value already validated in the literature: transactional, strategic, transformational and informational value (Gregor et al., 2006; Ren et al., 2016). The four dimensions were assessed using seven-point Likert scales, with responses ranging from “completely disagree” (-3) to “completely agree” (+3) (Gregor et al., 2006; Ren et al., 2016) (Table 3).

Dimensions	ID	Items
<i>Transactional value</i> (Gregor et al., 2006; Ren et al., 2016)	TAB1	Savings in supply chain management
	TAB2	Reducing operating costs
	TAB3	Reducing communication costs
	TAB4	Enhancing employee productivity
<i>Strategic value</i> (Gregor et al., 2006; Ren et al., 2016)	SB1	Creating competitive advantage
	SB2	Enabling quicker response to change
	SB3	Improving customer relations
<i>Transformational value</i> (Gregor et al.,	TFB1	An improved skill level for employees

2006; Ren et al., 2016)	TFB2	Developing new business opportunities
	TFB3	Expanding capabilities
	TFB4	Improving organization
<i>Informational value</i> (Gregor et al., 2006)	IB1	Enabling faster access to data
	IB2	Improving management data
	IB3	Improving data accuracy

Note: "ID" stands for identity number of the item

Table 3 Operationalization of the business value of the Big data construct based on Likert scales (Gregor et al., 2006; Ren et al., 2016)

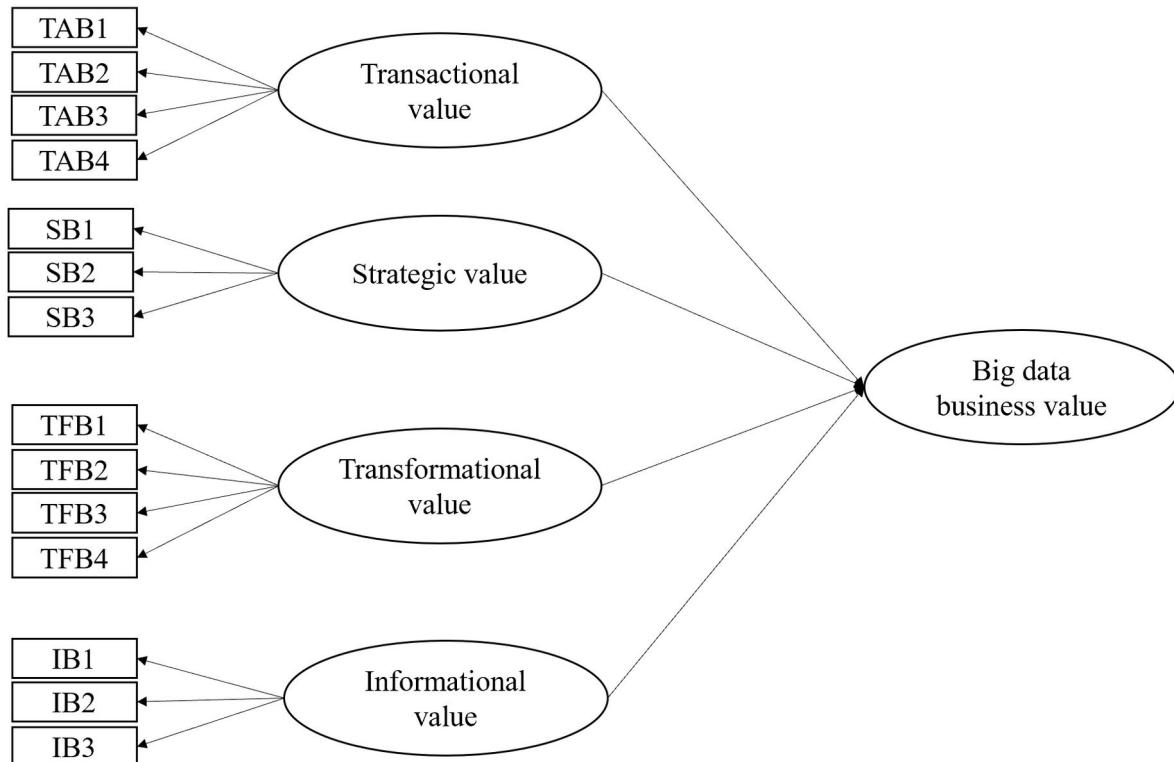


Figure 4 Business value of Big data as a second-order construct

4.2.3. Firm size

Firm size is operationalized as the logarithmic form of the number of employees of every company (Qian & Li, 2003). The data on the number of employees were extracted from the Bureau Van Dijk DIANE database.

4.2.4. Control variables

We included the firm's industry as a control variable because, on the one hand, we acknowledge that the inflection point could be context specific, and on the other hand, Big data solutions seem applicable regardless of the industry (Lehrer et al., 2018), even if returns may be different (Lehrer et al., 2018; Müller et al., 2018). The industry code was based on the 1-digit Standard Industry Classification (SIC) level extracted from the Bureau Van Dijk DIANE database. We operationalized the firm's industry by generating a set of dummy variables, one for each sector (Qian & Li, 2003). To simplify the readability of the models, we omitted the coefficients of these variables in the table showing the regression results. As far as

we consider the industry a control variable, we do not design the study to specifically explore or understand the eventual differences between industries.

4.3. The empirical models

Because not all the measures have multiple items and we used an index to operationalize the extent of Big data, there is no need to develop a separate measurement model for structural equation modeling (SEM) (Dong & Yang, 2020). Hence, we tested our hypotheses via an ordinary least squares (OLS) regression model, which is essentially equivalent to estimating a structural model with a single measurement in SEM. The equation of the model for the first hypothesis regarding the impact of the extension of Big data on business value takes the following form:

Business value of Big data = $\beta_0 + \beta_1$ Extension of Big data + β_2 Extension of Big data ²+ β_3 Firm size + β_4 CV+ ϵ (Equation 1), where CV is the set of control variables.

We included the first-order independent variable, namely, the extension of Big data, in the equation because leaving it out would be tantamount to assuming that the turning point is at $X=0$, which is a very strong assumption to make a priori (Aiken et al., 1991). Overall, a significant and negative β_2 indicates an inverted U-shaped relationship. However, even though the analysis shows a significant and negative β_2 , it is not sufficient to establish a quadratic effect. According to this result, we proposed a three-step procedure to establish the U-shaped relation, as discussed by Haans et al. (2016) in their literature review on theorizing and testing U-shaped relationships. They wrote, “First, β_2 has to be significant and of the expected sign. Second, the slope must be sufficiently steep at both ends of the data range [...] Third, the turning point needs to be located well within the data range” (Haans et al., 2016, p. 1182). In the regression results section, the satisfaction of these conditions will be discussed for the first hypothesis.

To test the second hypothesis on the moderating effect of size on the inverted U-shaped relationship investigated, we tested the following equation from the OLS regression model:

Business value of Big data = $\beta_5 + \beta_6$ Extension of Big data + β_7 Extension of Big data ²+ β_8 Firm size + β_9 Extension of Big data * Firm size + β_{10} Extension of Big data ² * Firm size + β_{11} CV + ϵ (Equation 2)

After the verification of the two hypotheses, we conducted a set of post hoc analyses to improve our understanding of the mechanisms by which the extent of Big data impacts the business value of Big data by investigating the existence, if any, of differences when separately considering the dimensions of the extent of Big data and of the business value of Big data. We conducted a post hoc analysis since the available knowledge is limited; therefore, this approach could be useful for adding knowledge to the literature by using exploratory methods.

4.4. Analyses and results

4.4.1. Descriptions

We mainly used STATA 14 to conduct analyses, starting with confirmatory factor analysis (FCA), to verify whether the variables we employed and already used in other studies have the appropriate psychometric properties. The loadings of the items on their respective constructs ranged from 0.672 to 0.840. We consider these loadings satisfactory (Hair et al., 2010). The t statistic of each factor was compounded to verify convergent validity. All the factor loadings were found to be statistically significant, and all the t values were higher than the cutoff point of 1.980. The overall constructs were significant, given that the Kaiser–

Meyer–Olkin measure of the sampling adequacy was equal to 0.843 and the Bartlett’s test of sphericity gave a statistically significant chi-square value of 1.156 (p value = 0.001). The recommended reliability levels and average variance extracted (AVE) were also determined. Cronbach’s alpha values ranged from 0.650 to 0.770, and the AVE values ranged from 0.516 to 0.603. These values were higher than the acceptability threshold values (Bagozzi & Yi, 1988; Churchill, 1979). These results revealed the presence of convergent validity in the measurement model. Unidimensionality was also confirmed from the AVE values (> 0.50).

The variance explained by each principal factor was also tested to identify any potential common method bias (Podsakoff & Organ, 1986). Harman’s one-factor test showed that the first factor accounted for only 23.542% of the total variance, which indicates that common method bias would not be a serious problem. Furthermore, the correlation matrix (Table 4) showed that the highest interconstruct correlation was 0.399, and common method bias usually displays an extremely high correlation ($r > 0.90$) (Bagozzi et al., 1991). Therefore, it can be stated that the common method bias in this research was not a serious issue.

Construct	ID	Mean	SD	CA	AVE	Factor loading	t-values
Transactional value	TAB1	0.487	1.501	0.688	0.516	0.734***	5.486
	TAB2	0.658	1.502			0.728***	5.860
	TAB3	0.013	1.553			0.697***	7.702
	TAB4	1.079	1.503			0.714***	7.160
Strategic value	SB1	0.368	1.468	0.670	0.603	0.840***	11.845
	SB2	0.855	1.430			0.730***	6.932
	SB3	0.461	1.562			0.756***	4.512
Transformational value	TFB1	0.513	1.527	0.770	0.594	0.815***	10.477
	TFB2	0.592	1.593			0.771***	10.294
	TFB3	0.908	1.298			0.815***	10.772
	TFB4	1.013	1.227			0.672***	6.412
Informational value	IB1	0.908	1.435	0.650	0.585	0.714***	5.499
	IB2	1.487	1.291			0.825***	9.576
	IB3	1.211	1.225			0.751***	4.858

Note: SD = standard deviation; CA = Cronbach’s alpha; AVE = average variance extracted; ***p < 0.001.

Table 4 Descriptive and psychometric table of measurements

Table 5 shows the discriminant validity of our variables measured with Likert scales. The square root of AVE was compared for each construct, with correlations between each construct and the remaining constructs (Fornell & Larcker, 1981). Each construct shared more variance with the related measurement variables than with the constructs of the other measurement variables. Therefore, discriminant validity was supported.

No.	Variable	1	2	3	4
1	Transactional value	0.718			
2	Strategic value	0.166	0.776		
3	Transformational value	0.381	0.399	0.770	
4	Informational value	0.288	0.092	0.273	0.765

Table 5 Correlation matrix of the measured scales for discriminant validity evaluation and square roots of the average variance extracted as diagonal elements

After the computation of the psychometric properties of the variables operationalized in this study with Likert scales, we computed the descriptive statistics for all variables involved in the study (Table 6). The analysis showed that companies, on average, adopt between three and four Big data solutions; additionally, a low percentage of companies store an amount of data

greater than one zettabyte, and data are rarely uploaded in real time. Most companies upload data at set and regular time intervals. Considering the business value of Big data, the majority of companies achieve Big data business value in terms of transformational, strategic, transactional, and informational value.

Variable	Construct	Min.	Max.	Mean	Standard deviation
Extension of Big data		-0.895	1.945	0.001	0.619
	Big data volume	1	5	1.776	0.704
	Big data velocity	1	3	1.197	0.432
	Big data variety	1	8	3.776	2.464
Business value of Big data		-1.396	2.792	0.770	0.848
	Transactional value	-2.25	3	0.559	1.088
	Strategic value	-3	3	0.561	1.154
	Transformational value	-2.25	3	0.766	1.088
	Informational value	-1	3	1.202	1.001
Firm size		3.912	7.147	4.780	0.711

Table 6 Descriptives

4.4.2. Regression results

To test the two hypotheses, we conducted regression analysis. Before running the OLS regressions, we wanted to ensure that multicollinearity effects were not an issue. The variance inflation factor (VIF) was computed for each of the variables by running separate analyses in which one variable was the dependent variable and all the other variables were considered independent. The VIF values ranged from 1.27 to 1.59. None of the VIF values reached the maximum acceptable level of 10. Thus, multicollinearity did not appear to be an issue.

Table 7 shows the results of the OLS regression models. Overall, four models were run. Specifically, Model 1 contained only the control variables. Model 2 included the first-order effects of the independent variables (extension of Big data and firm size). With Model 3, we tested Hypothesis 1 and added, as an independent variable, the squared value of the extension of Big data. Model 4 tested Hypothesis 2, and we added the interaction effect of the extension of Big data with firm size and the same interaction effect with the squared value of the extension of Big data.

Dependent variable Independent variables	Big data business value			
	M1	M2	M3	M4
Model			H1	H2
Hypothesis				
Direct effect				
Extension of Big data	...	-0.034 (0.089)	0.039 (0.094)	-0.018 (0.094)
Firm size	...	0.161 [†] (0.096)	0.175 [†] (0.099)	0.097 (0.118)
Extension of Big data (squared)	-0.134* (0.061)	-0.175* (0.089)
Moderating effects				
Extension of Big data * Firm size	-0.161 [†] (0.098)
Extension of Big data (squared) * Firm size	0.125* (0.057)

Industry dummy	Included	Included	Included	Included
Constant	0.807*** (0.211)	0.779** (0.223)	0.872*** (0.229)	0.920*** (0.250)
R-squared	9.24%	12.35%	16.53%	20.55%
VIF	1.28	1.27	1.30	1.59

Note: *** p<0.01, ** p<0.1, * p<0.05, †p<0.10; standard errors are shown in parentheses.

Table 7 Regression models for testing H1 and H2

In Hypothesis 1, we assumed a curvilinear inverted U-shaped relationship between the extension of Big data and business value. Specifically, we posited that as Big data increases, business value increases until business value reaches a maximum; as Big data continues to increase, business value decreases. To verify this hypothesis, we ran Model 3, and found, as hypothesized, an inverted U-shaped effect for the extension of Big data on the business value of Big data. These preliminary results supported the implementation of a three-step procedure to establish the significance of the inverted U-shaped relationship (Lind & Mehlum, 2010) (Table 8).

Test H0: Monotone, H1: Inverted U shaped	Values
<i>First step: Significance of the β coefficients</i>	
Coefficient of the “Extension of Big data (squared)” from Model 3 of Table 7	-0.134*
<i>Second step: Slope Steepness</i>	
Lower bound Interval	-1.445
Lower bound slope	0.426
Lower bound t-value	1.867
Lower bound P> t	0.033
Upper bound Interval	3.140
Upper bound slope	-0.801
Upper bound t-value	-2.196
Upper bound P> t	0.015
<i>Third step: Inflection point range</i>	
Turning point	0.147
Overall test of presence of an Inverse U shaped	
t-value	1.87
P> t	0.033

Note: *** p<0.01, ** p<0.1, * p<0.05, †p<0.10

Table 8 A three-step evaluation of the significance of the inverted U-shape for Model 3

First, we found that the effect was negative and significant, supporting the first step of the procedure. Second, we found that the slope was sufficiently steep at both ends of the data range. We can specify that the slope is sufficiently steep because the slope at the lower bound is positive and significant, and the slope at the upper bound is negative and significant. These results support the second step of the procedure. Third, we found that the inflection point is located well within the data range, thus, supporting the third step of the procedure. To provide further support, we performed an overall test of the presence of an inverted U-shaped relation. The results of this additional test again supported the significant existence of the supposed inverted U-shaped effect. The verification by all these tests indicates unambiguous support for Hypothesis 1.

Concerning Hypothesis 2, we assumed that firm size had a moderating effect on the inverted U-shaped relation between the extension of Big data and the business value of Big data. To test this hypothesis, we ran Model 4 and included the interaction effect of the extension of Big data with firm size, measured as the logarithm of the number of employees, and the same

interaction effect with the squared value of the index of the extension of Big data. The statistically significant interaction effects of the squared value of the index of the extension of Big data and firm size on the business value of Big data support the hypothesized moderating effect.

To provide an adequate interpretation of the moderating effect observed, we plotted the results of Model 3 (Figure 5a) and Model 4 (Figure 5b) to provide a graphical representation of the moderating effect for various variable ranges. To define high and low values, we used a common method that is based on considering values that are one standard deviation above and below the mean. These figures confirm the inverted U-shaped relationship between the extension of Big data and Big data business value, and they also confirm that this relationship is stronger for smaller than for larger firms.

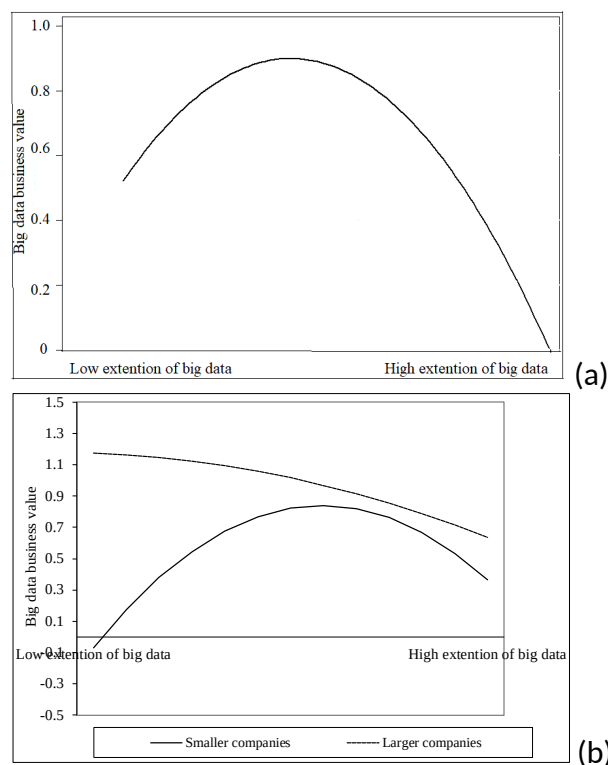


Figure 5 U-shaped effect (Figure 5a) and moderating effect of firm size (Figure 5b) on the inverted U-shaped effect of the extension of Big data on the business value of Big data

4.5. Post hoc analyses

4.5.1. Regression analyses separately considering the three dimensions of Big data

In this study, the extension of Big data is expressed by a synthetic value that considers the three vs. that characterize Big data: velocity, variety and volume. To verify whether our construct for the extension of Big data, which combines the three dimensions, better described our hypotheses than considering each dimension separately, we conducted post hoc analyses (Table 9) considering velocity, variety and volume as three independent variables instead of one variable, i.e., the extension of Big data.

<i>Dependent variable</i> <i>Independent variables</i>	Business value of Big data
---	-----------------------------------

Model	M5	M6	M7	M8	M9	M10	M11	M12	M13
<i>Direct effect</i>									
Velocity	0.243** *	0.284**	0.341*
	(0.073)	(0.142)	(0.194)
Velocity squared	...	0.019	0.062
	...	(0.040)	(0.087)
Variety	0.097	0.514	0.499	...
	(0.090)	(0.427)	(0.421)	...
Variety squared	-0.019	-0.020	...
	(0.019)	(0.019)	...
Volume	0.056	0.543 [†]	0.677*
	(0.106)	(0.296)	(0.348)
Volume squared	-0.151*	-0.227**
	(0.074)	(0.106)
Firm size	0.138	0.140	0.149	0.138	0.136	0.144	0.118	-0.220	-0.358
	(0.088)	(0.090)	(0.095)	(0.100)	(0.107)	(0.108)	(0.103)	(0.340)	(0.288)
<i>Moderating effect</i>									
Velocity * Firm size	0.116
	(0.319)
Variety * Firm size	-0.666	...
	(0.425)	...
Volume * Firm size	-0.609**
	(0.283)
Velocity squared * Firm size	0.037
	(0.058)
Variety squared * Firm size	0.021	...
	(0.017)	...
Volume squared * Firm size	0.160**
	(0.077)
<i>Control variables</i>									
Industry dummy	Include d	Include d	Include d	Include d	Include d	Include d	Include d	Include d	Included
Constant	0.811** *	0.783** *	0.778** *	1.161*	0.784** *	1.292** *	0.747** *	1.157* *	1.539***
	(0.211)	(0.236)	(0.217)	(0.479)	(0.224)	(0.364)	(0.246)	(0.478)	(0.462)
R-squared	18.80%	18.80%	13.40%	14.70%	12.60%	16.10%	19.00%	20.20%	19.10%
VIF	1.30	2.08	1.27	4.77	1.31	2.98	2.97	9.95	9.43

Note: *** p<0.01, ** p<0.1, * p<0.05, [†]p<0.10; standard errors are shown in parentheses.

Table 9 Regression models of the direct and U-shaped effects of each of the three vs. dimensions of the variable “Extension of Big data” on Big data business value

Specifically, in Model 5 and Model 6, we tested the direct and squared effects of velocity on the business value of Big data; in Model 7 and Model 8, we tested the effects of variety, and in Models 9 and 10, we tested the effects of volume. From Model 11 to Model 13, we investigated whether there is a stronger curvilinear inverted U-shaped relationship between the extension of Big data and the business model for smaller companies than for larger companies. Overall, the results from Model 5 and Model 6 show that velocity has a positive and statistically significant effect on the business value of Big data, but this effect is not curvilinear. Instead, these results demonstrate that, in isolation, the variety of Big data alone does not influence the Big data business value linearly or in a U-shaped way, while the

volume has an inverted U-shaped effect on the business value achieved by Big data, according to the results from Model 10.

As previously discussed, we followed the three-step procedure to establish the significance of the inverted U-shaped relation (Lind & Mehlum, 2010) with volume as an independent variable, and the results support all three steps in the procedure. To provide further support, we ran the overall test of the presence of an inverted U-shaped relationship. This additional test supports the significant existence of an inverted U-shaped effect for the volume dimension (Table 10).

<i>Test H0: Monotone, H1: Inverted U shaped</i>	Values
<i>First step</i>	
Coefficient of the “Extension of Big data (squared)” from Model 10 of Table 9	-0.151*
<i>Second step</i>	
Lower bound Interval	-1.102
Lower bound slope	0.876
Lower bound t-value	1.925
Lower bound P> t	0.029
Upper bound Interval	4.577
Upper bound slope	-0.840
Upper bound t-value	-2.105
Upper bound P> t	0.020
<i>Third step</i>	
Turning point	1.798
Overall test of presence of an Inverse U shaped	
t-value	1.92
P> t	0.029

Note: *** p<0.01, ** p<0.1, * p<0.05, †p<0.10

Table 10 A three-step evaluation of the significance of the inverted U-shapes for Model 10

In addition, the results from Models 11 to 13 demonstrate no influence of company size on the relationship between velocity, in Model 11, or variety, in Model 12, and the business value of Big data. However, for the volume dimension, in Model 13, size plays a role, and we discovered a statistically significant impact of the squared value of volume and company size.

4.5.2. Regression analyses separately considering the four dimensions of Big data business value

In this study, the business value of Big data is based on four main components of value: transactional value, strategic value, transformational value and informational value. To verify whether the combination of the four dimensions better describes our hypotheses than does each dimension separately, we conducted another post hoc analysis (Table 11).

<i>Dependent variable</i> <i>Independent variables</i>	Transactional value				Strategic value			
	<i>Model</i>	<i>M14</i>	<i>M15</i>	<i>M16</i>	<i>M17</i>	<i>M18</i>	<i>M19</i>	<i>M20</i>
<i>Direct effect</i>								
Extension of Big data	...	-0.268*	-0.179	-0.294*	...	-0.002	0.034	0.023
	...	(0.126)	(0.125)	(0.118)	...	(0.111)	(0.135)	(0.137)
Firm size (log number of employees)	...	0.219†	0.236†	0.026	...	0.308*	0.315*	0.290*
	...	(0.117)	(0.119)	(0.139)	...	(0.118)	(0.120)	(0.145)
Extension of Big data (squared)	-0.163	-0.304*	-0.066	-0.086
	(0.102)	(0.121)	(0.079)	(0.113)

<i>Moderating effect</i>								
Extension of Big data * Firm size (log number of employees)	-0.189†	-0.003
	(0.109)	(0.142)
Extension of Big data (squared) * Firm size (log number of employees)	0.269** *	0.027
				(0.074)	(0.072)
Industry dummy	Include d	Include d	Include d	Include d	Include d	Included	Included	Included
Constant	0.406	0.380	0.493	0.674*	0.521*	0.463†	0.510†	0.536†
	(0.316)	(0.307)	(0.309)	(0.315)	(0.243)	(0.242)	(0.256)	(0.285)
R-squared	12.20%	19.90%	23.60%	34.40%	8.40%	14.70%	15.30%	15.40%
VIF	1.28	1.27	1.30	1.59	1.28	1.27	1.30	1.59

Note: *** p<0.01, ** p<0.1, * p<0.05, † p<0.10; standard errors are shown in parentheses.

Table 11 Regression models of the U-shaped effects of “extension of Big data” on transactional and strategic value

<i>Dependent variable</i> <i>Independent variables</i>	Transformational value				Informational value			
<i>Model</i>	<i>M22</i>	<i>M23</i>	<i>M24</i>	<i>M25</i>	<i>M26</i>	<i>M37</i>	<i>M28</i>	<i>M29</i>
<i>Direct effect</i>								
Extension of Big data	...	0.0861 (0.116)	0.187 (0.125)	0.141 (0.131)	...	0.0470 (0.119)	0.115 (0.133)	0.056 (0.133)
Firm size (log number of employees)	...	0.0996 (0.118)	0.119 (0.117)	0.075 (0.145)	...	0.018 (0.128)	0.031 (0.134)	-0.003 (0.144)
Extension of Big data (squared)	-0.183* (0.073)	-0.195† (0.107)	-0.124† (0.072)	-0.113 (0.100)
<i>Moderating effects</i>								
Extension of Big data * Firm size (log number of employees)	-0.173 (0.128)	-0.282† (0.142)
Extension of Big data (squared) * Firm size (log number of employees)	0.092† (0.067)	0.111 (0.071)
Industry dummy	Include d	Include d	Include d	Include d	Include d	Include d	Include d	Include d
Constant	0.906* ** (0.300)	0.883* ** (0.310)	1.011** * (0.321)	1.019* ** (0.349)	1.396** * (0.229)	1.390** * (0.236)	1.476** * (0.235)	1.451** * (0.262)
R-squared	9.70%	11.30%	16.00%	17.70%	9.90%	10.20%	12.80%	17.20%
VIF	1.28	1.27	1.30	1.59	1.28	1.27	1.30	1.59

Note: *** p<0.01, ** p<0.1, * p<0.05, †p<0.10; standard errors are shown in parentheses.

Table 11 (continued) Regression models of the U-shaped effects of “Extension of Big data” on transformational and informational value

This post hoc analysis separately considers four dependent variables, namely, transactional value, strategic value, transformational value and informational value, instead of the single business value of Big data. Specifically, from Model 14 to Model 17, we performed a regression with transactional value as the dependent variable; from Model 18 to Model 21, strategic value was the dependent variable; from Model 22 to Model 25, the dependent variable was transformational value; and finally, from Model 26 to Model 29, the dependent variable was informational value. For each set of the four models, the first model included the independent variables as control variables. The second set of four models included the control variables and the first-order variables of size and the extension of Big data. The third set of the four models included, in addition to the control variables and first-order variables, the squared value of the extension of Big data to verify the existence of an inverted U-shaped effect between the extension of Big data and each dimension of Big data business value. Finally, the fourth set of the four models included the interaction effect of the squared value of Big data extension with company size to verify the different relationships according to company size.

Specifically, the results of Model 16, Model 20, Model 24 and Model 28 suggest that the supposed inverted U-shaped relationship between the extension of Big data and the dependent variable is verified for the transformational and informational dimensions of Big data business value; in contrast, for the other two dimensions, this effect does not emerge.

As previously noted, we followed a three-step procedure to establish the significance of the inverted U-shaped relationship (Lind & Mehlum, 2010) found in Model 24 and Model 28, and the results support all three steps of the procedure. To provide further support, we ran an overall test for the presence of an inverted U-shaped trend. This additional test again supported the significant existence of an inverted U-shaped effect for the transformational value and the informational value of Big data business value (Table 12).

<i>Test H0: Monotone, H1: Inverted U shaped</i>	Values	
	Model 24	Model 28
<i>First step</i>		
Coefficient of the “Extension of Big data (squared)” from Model 24 of Table 11	-0.183*	...
Coefficient of the “Extension of Big data (squared)” from Model 28 of Table 11	...	-0.124 [†]
<i>Second step</i>		
Lower bound Interval	-1.445	-1.445
Lower bound slope	0.715	0.473
Lower bound t-value	2.374	1.555
Lower bound P> t	0.010	0.062
Upper bound Interval	3.140	3.140
Upper bound slope	-0.961	-0.662
Upper bound t-value	-2.409	-1.697
Upper bound P> t	0.009	0.047
<i>Third step</i>		
Turning point	0.511	0.466
Overall test of presence of an Inverse U shaped		
t-value	2.37	1.55
P> t	0.010	0.062

Note: *** p<0.01, ** p<0.1, * p<0.05, [†]p<0.10

Table 12 A three-step evaluation of the significance of the inverted U-shapes for Model 24 and Model 28

Finally, in Model 17, Model 21, Model 25 and Model 29, we found that the curvilinear, inverted U-shaped relation between the extension of Big data and business value is stronger for smaller companies than for larger firms in the case of transactional value and transformational value, but in the other two cases, there was no difference according to the dimension of the business value of Big data.

In Table 13, we provide a summary of the post hoc analyses conducted in this section of the study.

Relationship >>	H1 supported?	H2 supported?
Velocity squared >> business value of Big data	No	n.a.
Velocity squared * firm size >> business value of Big data	n.a.	No
Variety squared >> business value of Big data	No	n.a.
Variety squared * firm size >> business value of Big data	n.a.	No
Volume squared >> business value of Big data	Yes	n.a.
Volume squared * firm size >> business value of Big data	n.a.	Yes
Extension of Big data >> transactional value	No	n.a.
Extension of Big data * firm size >> transactional value	n.a.	Yes
Extension of Big data >> strategic value	No	n.a.
Extension of Big data * firm size >> strategic value	n.a.	No
Extension of Big data >> transformational value	Yes	n.a.
Extension of Big data * firm size >> transformational value	n.a.	Yes
Extension of Big data >> informational value	Yes	n.a.

Extension of Big data * firm size >> informational value	n.a.	No
--	------	----

Note: ">>" stands for "influences" which means that the independent variable on the left of the symbol influences the dependent variable on the right of the symbol; "n.a." stands for "not applicable".

Table 13. Summary of the post hoc analyses

5. Discussion

Compared to previous studies that explored the link between Big data initiatives and business value and assumed a linear, positive relationship, we investigated the hypothesis that a TMGT effect plays an important role in explaining the business value potential of Big data investments. The research results have several contributions, both theoretical and practical. We also acknowledge some limitations that could be starting points for further research.

5.1. Theoretical contributions

While the body of literature supporting a positive relationship between Big data and performance is constantly growing (e.g., Grover et al., 2018; Mikalef et al., 2020), the effects of diminishing returns, or the TMGT effect, have been largely disregarded, despite the literature having previously hinted at a complex relationship between IT investments and business value (e.g., Khallaf et al., 2017; Lin & Shao, 2006; Mitra, 2005). We contribute to unravelling the complex relationships between IT and value with an empirical investigation of Big data investments. Our empirical results offer three main contributions in this regard: (1) the existence of an inverted U-shaped relationship between Big Data extension and business value; (2) the differential effects of Big data extension on the four dimensions of business value; and (3) the influence of firm size in shaping this quadratic curve. The explanations of these three main contributions follow.

5.1.1. Existence of an inverted U-shaped relationship

Our findings corroborate the hypothesis that organizations manifest limits in their capacity to benefit from Big data investments. Specifically, we observe a TMGT effect for Big data extension, and this effect eventually becomes negative for business value. While an expanding stream of research (e.g., Raguseo et al., 2020; Wamba et al., 2020) has provided evidence that diminishing returns manifest for Big data investments, little is known about this relationship. Despite the attention that Big data has received as an idiosyncratic IT investment, this study represents a unique analysis that provides clear confirmation of a curvilinear and inverted U-shaped relationship between Big data extension and business value.

In this context, we can only propose hypotheses on the reasons why such a mechanism manifests and whether it is idiosyncratic to Big data, but we rely on existing theories to explore these topics further. Previous studies on absorptive capacity in the context of Big data (Božič and Dimowski, 2019) hinted at the importance of organizations developing adequate capabilities to synthesize and leverage acquired knowledge. In other words, too much Big data may limit a firm's ability to transform the generated insights into business performance improvements if the necessary support capabilities (e.g., technological, human, or relational) are lacking. This result is aligned with those of recent studies (Raguseo, 2018) suggesting that firms require a certain level of organizational readiness to profit from Big data considering knowledge integration mechanisms and product innovativeness (Tsai et al., 2015). Excessive investment would then lead to an organizational incapacity to profit from the opportunity, resulting in adaptation periods or additional decision-making, which may negatively affect

transformational business value. In summary, we showed that the extension of Big data could lead to the erosion of business value.

We can now say that the problem is not Big data itself but rather too much of it, that “Big” data are not always better than “small” data and that more data does not always mean better value for the organization. We are the first to empirically show that the TMGT effect characterizes Big data investments. As far as we measured Big data extension as a combination of the three vs. of Volume, Velocity and Variety, we now know that firms can be overwhelmed from too much volume of data, they cannot keep pace with too rapid streams of data and data diversity can become unmanageable. If the overload of information has already been investigated (Clark et al., 2006; A. R. Lee et al., 2016; Schroeder & Benbasat, 1975; Zhang et al., 2016), we apply the analysis in the Big data domain.

In summary, with our study, we added an original piece to the body of knowledge about inverted U-shaped relationships, and we successfully challenged the largely diffused assumption of a linear relationship among variables in MIS.

5.1.2. Differential effects on the dimensions of business value

More specifically, our results demonstrate that the relationship between the extension of Big data and business value varies depending on the dimensions of business value. On the four dimensions of value—transactional, transformational, strategic and informational—the inverted U-shape is specifically verified for transformational and informational values. These two dimensions of business value are the main sources of the overall inverted U-shape that we demonstrated for Big data business value as a whole.

The relevance of the negative effect of “too much” Big data on transformational value may be explained by the produced effects of investments on the organizational capacities and structure; notably, instead of expanding a firm’s capabilities, levels of complexity or performance discontinuities are introduced that decrease overall performance. The finding concerning the effects on informational value may reflect the early evidence that information overload, originating at the individual level, could lead to poor decisions and other negative effects (Clark et al., 2006; A. R. Lee et al., 2016; Schroeder & Benbasat, 1975; Zhang et al., 2016). We acknowledge that this effect may be reinforced by other contextual factors, including organizational inertia in adapting to the changes associated with Big data investment. Indeed, the change management literature is rich in studies on the risks of negative impacts that change may have on performance (e.g., Rieley & Clarkson, 2001)

In contrast, the transactional and strategic dimensions of business value are not inverted U-shaped by the Big data extension. This means that firms that target Big data investments mainly for these two dimensions of business value could not suffer from the TMGT effect. Nonetheless, as far as these explorations emerge from the post hoc analysis results, we remain cautious in our conclusions, as the design of our research did not primarily target the test of these hypotheses.

5.1.3. Influence of firm size

We confirmed that the curvilinear, inverted U-shaped relation between the extension of Big data and business value is stronger for smaller companies than for larger companies. Smaller companies appear more sensitive to Big data investment and benefit marginally more than larger organizations. The flexibility (Ebner et al., 2014) and simpler structures of small companies provide a clear advantage compared over larger firms, which suffer from greater organizational inertia (Eisenhardt & Martin, 2000). Furthermore, the integration of different developing cycles and Big data technologies in existing infrastructure may contribute to the high rigidity of large organizations in benefitting from Big data investments, resulting in

flatter slopes of the Big data value curve (Figure 5b). The greater agility of smaller organizations may, thus, have competitive relevance, enabling these firms to innovate at a faster pace than larger companies and to more quickly occupy exploitable markets. This result provides an indirect confirmation of the findings of previous studies on Big data organizational readiness; in these studies, the low structural preparedness of smaller firms was observed, and this finding was related to the technical, managerial and data management capabilities of these firms (Raguseo et al., 2018). Because their structural components are inadequate, small firms may prefer to keep Big data initiatives and the supporting infrastructure separate from the existing IT function structure, which may not be suited to meet current business needs of these firms (Haffke et al., 2017). In this context, studies of innovation and organizational ambidexterity may note “bimodal operations” in which the organization may leverage the existing infrastructure and process to support the exploitation of current systems, the evolution of these systems, and the use of new Big data technology for the exploration of and experimentation with new products and markets (Haffke et al., 2017; Pettey, 2016).

5.2. Practical implications

Our results have relevant practical implications. First, we were able to demonstrate that, beyond the inflection point, additional Big data investments may be less productive or counterproductive, producing negative returns and project failure. This result directly implies that managers should consider closely monitoring their investments and adopting a pace of innovation compatible with the capacity of their organization. However, we cannot provide further insight into where the inflection point resides, i.e., where the business value reaches a turning point. Contingency, in this case, seems to dictate and determine this location. Hence, firms should be cautious when considering their investments in Big data and evaluate the value created by each investment. As the investment approaches its turning point, companies should stop investing in Big data initiatives as soon as their return on investments shrinks toward zero because the next investment could destroy, not create, value. The inverted U-shaped curve is the additive combination of the revenues and costs of Big data; firms should monitor marginal revenues and costs of Big data. As marginal revenues gradually decrease toward zero, while marginal costs should steadily increase, the most effective technique would be to monitor the revenues. As soon as the revenue growth from a new Big data investment is not large enough, it may be the moment to stop further investments in Big data.

However, the relatively flat curve we observed for large organizations (Figure 5b) suggests that the inertia of large organizations may protect them from experiencing sudden downturns in their business value generated by the TMGT effect. Additionally, this finding seems to suggest that a viable strategy for Big data initiatives, at least in the short term, may be to extend beyond the limits of the current IT structure or provide the necessary context to enable the safe exploration of new product and market opportunities. Small organizations, thus, appear to benefit from their flexibility in adapting to changing market conditions; although, while limited resources may hinder their capacity to invest in Big data, they are more adaptable and capable of seizing novel business opportunities. At the same time, once smaller companies go past the turning point, the extent of value erosion increases sharply, meaning that firm survival could be at stake at the first Big data investment overshoot. This red alert for smaller companies should encourage them to assess, more closely than bigger companies, each Big data investment to verify that they are not surpassing the turning point of the U-shaped curve.

5.3. Limitations

As with all research, the present study presents several limitations. While we tried our best to assess the size of Big data investments in terms of their extension, different measures could be used to complement our analysis and explore more contextual factors that may contribute to the U-shaped curvilinear relationship. In particular, as far as we mobilized only the size of the firm as the unique moderator, other moderating variables could complement or replace our approach.

Furthermore, in our analysis, we excluded companies with no or little Big data, which allowed us to retain the studied curve but possibly skewed the analysis for firms marginally using Big data. Further studies may consider including companies with no or little Big data, which may result in a richer analysis of the shape of the curve than that provided in this study (Pierce & Aguinis, 2013). In-depth case studies could complement our survey approach, which limited the fine-grained understanding of the mechanisms that cause the observed curves. Hence, we cannot provide examples of excessive investments in Big data or of the conditions that could minimize value erosion. Finally, even if we highlighted the scarcity of longitudinal studies in Big data research (e.g., Raguseo et al., 2020), our cross-sectional design did not contribute to filling this gap. A longitudinal study would be able to see the shaping of the business value curve with the progressive extension of the Big data in firms, allowing us to precisely define the inflection point of the inverted U-shaped curve.

Finally, industry was mobilized as a control variable, and hence, we did not investigate it further. It was not an objective of this research project. We explored our data in the post hoc analysis. The industry distribution appeared too heterogeneous, compared to the size of our sample, to give us any chance to explore the influence of the industries on the inverted U-shaped relationships. Nonetheless, our post hoc analysis was rich in insights into the dimensions of business value. However, these emerging results should be considered explorative, as our study was not designed with these post hoc objectives in mind. Further research, including the expansion of the size of the sample, is hence necessary to better ground our preliminary findings and conclusions.

With our study, we likely created more questions than answers. However, we believe it is necessary to advance the debate concerning investment in IT and business value. While curvilinear models are more complex than their linear counterparts, studies and assessments of these models may provide improved understandings of digital investments and effectively support management decision-making (Hastie & Dawes, 2001). The choice to apply a linear model or a different model should depend on the specific objectives of the application (Einhorn & Hogarth, 1975), and the possible effects of the different modeling methods beyond considering the mere fit indices. More generally, we believe it is necessary to further our understanding of the TMGT effect and the observed related impacts. We invite MIS scholars to question the basic linear assumption when developing and testing hypotheses.

6. Conclusion

This study enriches the debate on the business value of Big data investments and provides robust evidence that organizations face limits in profiting from Big data investments. Specifically, we provide details on this relationship and demonstrate the existence of a curvilinear, inverted U-shaped relation between the extension of Big data and business value. We observed that as Big data extension increases, business value increases, but past a certain point, value decreases. While this turning point seems determined by contextual and contingent factors, we observed significant moderation based on firm size. In particular, smaller organizations have greater sensitivity to Big data investments than do larger organizations, suggesting that the structural characteristics of small companies may improve their adaptability in dynamic contexts characterized by Big data innovations. Our findings

also have relevant implications for practice. From a managerial perspective, our results suggest that practitioners must cautiously evaluate the extent of their current investments in Big data technologies in relation to the capacity of their organizations to adapt to the structural changes the corresponding innovations demand.

Funding source acknowledgment

The paper is part of an FP7 European research project funded through a Marie Curie Intra European Fellowship. This funding source had no involvement in the study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

References

- Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.
- Amankwah-Amoah, J. (2016). Emerging economies, emerging challenges: Mobilising and capturing value from big data. *Technological Forecasting and Social Change*, 110(Supplement C), 167–174. <https://doi.org/10.1016/j.techfore.2015.10.022>
- Asay, M. (2017). 85% of big data projects fail, but your developers can help yours succeed. *TechRepublic*. <https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/>
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94. <https://doi.org/10.1007/BF02723327>
- Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, 36(3), 421–458. JSTOR. <https://doi.org/10.2307/2393203>
- Bughin, J. (2016). Big data, Big bang? *Journal of Big data*, 3(1), 2. <https://doi.org/10.1186/s40537-015-0014-3>
- Cadwalladr, C., & Graham-Harrison, E. (2018). The Cambridge analytica files. *The Guardian*, 21, 6–7.
- Cappa, F., Oriani, R., Peruffo, E., & McCarthy, I. (2021). Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*. *Journal of Product Innovation Management*, 38(1), 49–67. <https://doi.org/10.1111/jpim.12545>
- Cass, D. (1965). Optimum Growth in an Aggregative Model of Capital Accumulation. *The Review of Economic Studies*, 32(3), 233–240. JSTOR. <https://doi.org/10.2307/2295827>
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67–80.
- Chen, H., Chiang, R., & Storey, V. (2012). Business Intelligence and Analytics: From Big data to Big Impact. *Management Information Systems Quarterly*, 36(4), 1165–1188.
- Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64–73.
- Clark, B. H., Abela, A. V., & Ambler, T. (2006). An Information Processing Model of Marketing Performance Measurement. *Journal of Marketing Theory & Practice*, 14(3), 191–208. <https://doi.org/10.2753/MTP1069-6679140302>

- Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, 30(1), 44-57.
- Cox, M., & Ellsworth, D. (1997). Managing big data for scientific visualization. *ACM Siggraph*, 97, 21-38.
- Davis, F., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982-1003.
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19(4), 9-30.
- Dong, J. Q., & Yang, C.-H. (2020). Business value of big data analytics: A systems-theoretic approach and empirical test. *Information & Management*, 57(1), 103124. <https://doi.org/10.1016/j.im.2018.11.001>
- Ebner, K., Bühnen, T., & Urbach, N. (2014). Think Big with Big data: Identifying Suitable Big data Strategies in Corporate Environments. 2014 47th Hawaii International Conference on System Sciences, 3748-3757. <https://doi.org/10.1109/HICSS.2014.466>
- Einhorn, H. J., & Hogarth, R. M. (1975). Unit Weighting Schemes for Decision Making. *Organizational Behavior and Human Performance*, 13(2), 171-192.
- Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic Capabilities: What Are They? *Strategic Management Journal*, 21(10/11), 1105-1121.
- Fleishman, E. A. (1998). Patterns of leadership behavior related to employee grievances and turnover: Some post hoc reflections. *Personnel Psychology*, 51(4), 825-834. <https://doi.org/10.1111/j.1744-6570.1998.tb00740.x>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
- Gomes, J. F. S., de Weerd-Nederhof, P. C., Pearson, A. W., & Cunha, M. P. (2003). Is more always better? An exploration of the differential effects of functional integration on performance in new product development. *Technovation*, 23(3), 185-191. [https://doi.org/10.1016/S0166-4972\(01\)00107-9](https://doi.org/10.1016/S0166-4972(01)00107-9)
- Goold, M. (1999). The Growth Imperative. *Long Range Planning*, 32(1), 127-129. [https://doi.org/10.1016/S0024-6301\(98\)00133-2](https://doi.org/10.1016/S0024-6301(98)00133-2)
- Gregor, S., Martin, M., Fernandez, W., Stern, S., & Vitale, M. (2006). The transformational dimension in the realization of business value from information technology. *The Journal of Strategic Information Systems*, 15(3), 249-270. <https://doi.org/10.1016/j.jsis.2006.04.001>
- Grover, V., Chiang, R. H. L., Ting-Peng Liang, & Dongsong Zhang. (2018). Creating Strategic Business Value from Big data Analytics: A Research Framework. *Journal of Management Information Systems*, 35(2), 388-423. <https://doi.org/10.1080/07421222.2018.1451951>
- Haans, R. F. J., Pieters, C., & He, Z.-L. (2016). Thinking about U: Theorizing and testing U- and inverted U-shaped relationships in strategy research. *Strategic Management Journal*, 37(7), 1177-1195. <https://doi.org/10.1002/smj.2399>
- Haffke, I., Kalgovas, B. J., & Benlian, A. (2017). Options for Transforming the IT Function Using Bimodal IT. *MIS Q. Executive*.

- Hair, J. F., Black, W., Babin, B., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate Data Analysis* (7th edition). Pearson.
- Hastie, R., & Dawes, R. M. (2001). *Rational Choice in an Uncertain World: The Psychology of Judgement and Decision Making*. SAGE.
- Illich, I. (1973). *Tools for Conviviality*. Harper & Row.
- Kamioka, T., & Tapanainen, T. (2014). Organizational Use of Big data and Competitive Advantage-Exploration of Antecedents. PACIS, 2014, 18th.
- Khallaf, A., Omran, M. A., & Zakaria, T. (2017). Explaining the inconsistent results of the impact of information technology investments on firm performance: A longitudinal analysis. *Journal of Accounting & Organizational Change*, 13(3), 359–380. <https://doi.org/10.1108/JAOC-11-2015-0086>
- Lee, A. R., Son, S.-M., & Kim, K. K. (2016). Information and communication technology overload and social networking service fatigue: A stress perspective. *Computers in Human Behavior*, 55, 51–61. <https://doi.org/10.1016/j.chb.2015.08.011>
- Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293–303. <https://doi.org/10.1016/j.bushor.2017.01.004>
- Lehrer, C., Wieneke, A., Vom Brocke, J., Jung, R., & Seidel, S. (2018). How Big data Analytics Enables Service Innovation: Materiality, Affordance, and the Individualization of Service. *Journal of Management Information Systems*, 35(2), 424–460. <https://doi.org/10.1080/07421222.2018.1451953>
- Lin, W. T., & Shao, B. B. M. (2006). The business value of information technology and inputs substitution: The productivity paradox revisited. *Decision Support Systems*, 42(2), 493–507. <https://doi.org/10.1016/j.dss.2005.10.011>
- Lind, J. T., & Mehlum, H. (2010). With or Without U? The Appropriate Test for a U-Shaped Relationship*. *Oxford Bulletin of Economics and Statistics*, 72(1), 109–118. <https://doi.org/10.1111/j.1468-0084.2009.00569.x>
- Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, 24(3), 149–157. <https://doi.org/10.1016/j.jsis.2015.08.002>
- Love, P. E. D., Irani, Z., Standing, C., Lin, C., & Burn, J. M. (2005). The enigma of evaluation: Benefits, costs and risks of IT in Australian small–medium-sized enterprises. *Information & Management*, 42(7), 947–964. <https://doi.org/10.1016/j.im.2004.10.004>
- Lynch, C. (2008). Big data: How do your data grow? *Nature*, 455(7209), 28–29. <https://doi.org/10.1038/455028a>
- Mackenzie, J. S. (1899). The Idea of Progress. *International Journal of Ethics*, 9(2), 195–213.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, H. B. (2011). Big data: The next frontier for innovation, competition, and productivity (p. 156). McKinsey & Company.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The Management Revolution. *Harvard Business Review*, October. <https://hbr.org/2012/10/big-data-the-management-revolution>
- Medhora, S. (2019, February 18). Over 2000 people died after receiving Centrelink robo-debt notice, figures reveal. In *Triple j*. <https://www.abc.net.au/triplej/programs/hack/2030-people-have-died-after-receiving-centrelink-robodebt-notice/10821272>

- Melville, N., Kraemer, K., & Gurbaxani, V. (2004). REVIEW: INFORMATION TECHNOLOGY AND ORGANIZATIONAL PERFORMANCE: AN INTEGRATIVE MODEL OF IT BUSINESS VALUE. *MIS Quarterly*, 28(2), 283–322.
- Mikalef, P., Pappas, I. O., Krogstie, J., & Pavlou, P. A. (2019). Big data and business analytics: A research agenda for realizing business value. *Information & Management*, 103237. <https://doi.org/10.1016/j.im.2019.103237>
- Mikalef, P., Pappas, I. O., Krogstie, J., & Pavlou, P. A. (2020). Big data and business analytics: A research agenda for realizing business value. *Information & Management*, 57(1), 103237. <https://doi.org/10.1016/j.im.2019.103237>
- Mithas, S., Tafti, A., Bardhan, I., & Goh, J. M. (2012). Information Technology and Firm Profitability: Mechanisms and Empirical Evidence. *MIS Quarterly*, 36(1), 205–224.
- Mitra, S. (2005). Information Technology as an Enabler of Growth in Firms: An Empirical Assessment. *Journal of Management Information Systems*, 22(2), 279–300. JSTOR.
- Müller, O., Fay, M., & Vom Brocke, J. (2018). The Effect of Big data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems*, 35(2), 488–509. <https://doi.org/10.1080/07421222.2018.1451955>
- Navanath, R. (2020, March 20). Big data Market 2020 Growth Analysis, Opportunities, Trends, Developments and Forecast to 2026. *Skyline Gazette*. <https://skyline-gazette.com/2020/03/20/big-data-market-2020-growth-analysis-opportunities-trends-developments-and-forecast-to-2026/>
- NewVantage Partners. (2019). Big data and AI Executive Survey 2019 (p. 16). NewVantage Partners.
- Nisbet, R. (1994). *History of the Idea of Progress* (2nd edition). Routledge.
- Nonninger, L. (2018, June 2). Here's why China is concerned about Tencent and Alibaba's credit scoring efforts. *Business Insider*. <https://www.businessinsider.com/china-tencent-and-alibabas-new-credit-scoring-solution-2018-2>
- Oliveira, T., & Martins, M. F. (2011). Literature review of information technology adoption models at firm level. *Electronic Journal of Information Systems Evaluation*, 14(1), 110.
- Pettey, C. (2016). Busting Bimodal Myths. [//www.gartner.com/smarterwithgartner/busting-bimodal-myths/](http://www.gartner.com/smarterwithgartner/busting-bimodal-myths/)
- Pierce, J. R., & Aguinis, H. (2013). The Too-Much-of-a-Good-Thing Effect in Management. *Journal of Management*, 39(2), 313–338. <https://doi.org/10.1177/0149206311410060>
- Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, 12(4), 531–544. <https://doi.org/10.1177/014920638601200408>
- Qian, G., & Li, L. (2003). Profitability of Small- and Medium-Sized Enterprises in High-Tech Industries: The Case of the Biotechnology Industry. *Strategic Management Journal*, 24(9), 881–887. JSTOR.
- Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management*, 38(1), 187–195. <https://doi.org/10.1016/j.ijinfomgt.2017.07.008>

- Raguseo, E., Pigni, F., & Piccoli, G. (2018). Conceptualization, Operationalization, and Validation of the Digital Data Stream Readiness Index. *Journal of Global Information Management (JGIM)*, 26(4), 92–112. <https://doi.org/10.4018/JGIM.2018100106>
- Raguseo, E., Vitari, C., & Pigni, F. (2020). Profiting from big data analytics: The moderating roles of industry concentration and firm size. *International Journal of Production Economics*, 229, 107758. <https://doi.org/10.1016/j.ijpe.2020.107758>
- Ren, S. J., Wamba, S. F., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, 55, 5011–5026. <https://doi.org/10.1080/00207543.2016.1154209>
- Rich, J. T. (2013). THE GROWTH IMPERATIVE. *Journal of Business Strategy*. <https://doi.org/10.1108/eb039992>
- Richardson, S., Petter, S., & Carter, M. (2021). Five Ethical Issues in the Big Data Analytics Age. *Communications of the Association for Information Systems*, 49(1). <https://doi.org/10.17705/1CAIS.04918>
- Rieley, J., & Clarkson, I. (2001). The impact of change on performance. *Journal of Change Management*, 2(2), 160–172. <https://doi.org/10.1080/714042499>
- Schroeder, R. G., & Benbasat, I. (1975). An Experimental Evaluation of the Relationship of Uncertainty in the Environment to Information Used by Decision Makers*. *Decision Sciences*, 6(3), 556–567. <https://doi.org/10.1111/j.1540-5915.1975.tb01043.x>
- Steelman, Z. R., Havakhor, T., Sabherwal, R., & Sabherwal, S. (2019). Performance Consequences of Information Technology Investments: Implications of Emphasizing New or Current Information Technologies. *Information Systems Research*, 30(1), 204–218. <https://doi.org/10.1287/isre.2018.0798>
- Street, C. T., & Meister, D. B. (2004). Small Business Growth and Internal Transparency: The Role of Information Systems. *MIS Quarterly*, 28(3), 473–506. JSTOR. <https://doi.org/10.2307/25148647>
- Taylor, F. W. (2012). *The Principles of Scientific Management*.
- Tiwana, A., Wang, J., Keil, M., & Ahluwalia, P. (2007). The Bounded Rationality Bias in Managerial Valuation of Real Options: Theory and Evidence from IT Projects*. *Decision Sciences*, 38(1), 157–181. <https://doi.org/10.1111/j.1540-5915.2007.00152.x>
- Tsai, K.-H., Liao, Y.-C., & Hsu, T. T. (2015). Does the use of knowledge integration mechanisms enhance product innovativeness? *Industrial Marketing Management*, 46, 214–223. <https://doi.org/10.1016/j.indmarman.2015.02.030>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance Of Information Technology: Toward A Unified View. *MIS Quarterly*, 27(3), 425–478.
- Wamba, S. F., Dubey, R., Gunasekaran, A., & Akter, S. (2020). The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism. *International Journal of Production Economics*, 222, 107498. <https://doi.org/10.1016/j.ijpe.2019.09.019>
- Wolff, J. (2016). Perverse Effects in Defense of Computer Systems: When More Is Less. *Journal of Management Information Systems*, 33(2), 597–620. <https://doi.org/10.1080/07421222.2016.1205934>

Zhang, S., Zhao, L., Lu, Y., & Yang, J. (2016). Do you get tired of socializing? An empirical explanation of discontinuous usage behaviour in social network services. *Information & Management*, 53(7), 904–914. <https://doi.org/10.1016/j.im.2016.03.006>