

RESEARCH

Open Access



# Reaping the benefits of big data in telecom

Jacques Bughin<sup>1,2,3,4\*</sup> 

\*Correspondence:  
Jacques\_bughin@mckinsey.com

<sup>1</sup> McKinsey and Company,  
Avenue Louise, 480,  
1050 Brussels, Belgium  
Full list of author information  
is available at the end of the  
article

## Abstract

We collect big data use cases for a representative sample of telecom companies worldwide and observe a wide and skewed distribution of big data returns, with a few companies reporting large impact for a long tail of telecom companies with limited returns. Using a joint model of adoption and returns to adoption, we find that the skewness of the distribution arises from a few telecom companies being able to follow key big data managerial and organizational practices. We also find that big data returns exhibit economies of scope, decreasing returns to scale, while big data talents are complementary to big data capex investments.

**Keywords:** Information systems, Big data, Organization assets

## Background

Big data in business has been successfully pioneered by large internet platforms facing the task of handling real time data in billions of bytes [10], and [1].<sup>1</sup> The sport industry is a classical example of big data use, as made famous by the book *Moneyball* by Michael Lewis [23], later adapted into a drama movie.<sup>2</sup> In the US alone, the sport tech market has been growing at more than 50 % a year according to various analysts.<sup>3</sup> Another industry compelled to benefit from big data should be the telecom industry. Telecom companies have long had access to extensive bits of data with large base of their subscribers connecting daily to their network and services [32]. By enlarging their voice business to broadband, telecom companies are now capturing more and more data volume (consumers are making more calls and connecting more and more to internet); are benefiting from a larger variety of sources (large use of multiple internet broadband applications) as well as from higher velocity in data generation (for example, smartphone users in the UK tend to do 220 tasks a day, and use their phone 1500 times a week versus

<sup>1</sup> Regarding some insights on how companies such as Ebay or Google are using big data, the reader can refer to among others <http://www.zdnet.com/article/how-ebay-uses-big-data-to-make-you-buy> and <http://www.smartdatacollective.com/bernardmarr/202181/big-data-google>. Hal Varian, by now chief economist at Google is a clear advocate of big data analytics [31].

<sup>2</sup> The *Moneyball* drama movie was released in 2011. The story is about how an Oakland baseball manager who hires a Yale economist to leverage sport analytics for better team scouting and coaching decisions.

<sup>3</sup> For a fresh detail of the market, see Wintergreen [27].

© 2016 The Author(s). This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

only a few calls and short messages in the recent past).<sup>4,5</sup> However, a comprehensive understanding of how big data technologies are adopted, what and how returns are made from them for the telecom industry, is yet to exist. This article aims at filling this gap, using a confidential survey conducted on telecom players worldwide for a Google-McKinsey conference on the state of digitization and big data adoption in the industry [7]. Four core findings emerge from the study.

1. First, big data contribution to total telecom profit is minor, but in line with its relative size of investment in total telecom spent, and generates a productivity impact aligned with other research (Tambe [28], or Bughin [9]).
2. Second, the distribution of returns exhibits a long tail of limited (sometimes even negative) returns with a few companies claiming to make disproportionate returns from big data investments.
3. Third, using a joint model of big data adoption and of returns on adoption, we try to explain key drivers of this variance in big data performance. Consistent with the theory of technology adoption (e.g. [16]) and of organizational capital [5], we find that a large part of the cross-sectional distribution of returns is explained by the types of capabilities of companies to effectively manage big data projects. Further, as found in theory of production [9, 28], we find the existence of returns to scale and scope, and complementary talent to IT architecture investments in big data.
4. We finally test whether big data capabilities are of crucial importance for the financial returns linked to big data projects. We find indeed that mastering capabilities at scale are necessary to generate returns above cost of capital for big data in the telecom industry.

After a quick introduction (“[Background](#)” section of the paper), we discuss the status of big data for the telecom industry in the “[Statistics](#)” section. “[Explaining drivers of big data returns](#)” section is the core of the paper and develops a formal statistical analysis of adoption and returns achieved. The last, “[Conclusions](#)” section summarizes and provides avenue for research.

## Statistics

### Sample

This paper leverages a client-confidential survey reported in Bughin [7]. The online survey was posted in November and December 2014 on the McKinsey Telecom extranet, and was composed of a series of 75 questions addressed to telecom companies worldwide. The survey covers companies’ current financial performance, their product and services mix, their view on their competitive environment, their current management agenda. The survey also includes adoption by companies’ of enterprise digitization programs, including big data use cases. Table 1 provides some high level sample statistics; Table 2 focuses on digitization and big data projects by telecom companies. Data are measuring practices by end of 2014.

<sup>4</sup> Data were collected from Techmark, available <http://www.techmark.co.uk/smartphone-usage-data-uk-2014>.

<sup>5</sup> On top of this, data can be captured real time, and sometimes “ahead of times”—such is for instance Google searches that can anticipate sales; In a recent study [8], it was shown that branded telecom search queries and social media comments could predict future telecom brand sales up to one and two quarters ahead of actual products purchase by new clients such as telephony and broadband access line.

**Table 1 Telecom sample statistics. Source: Bughin [7], sample size: 273 telecom executives**

	Statistics (%)
1. Scope	
Business scope	
Fixed and mobile	54
Mobile	31
Fixed	15
Geographic scope	
Mono-country	57
Multi-country	43
HQ location	
Europe	47
Americas	24
Africa	5
Asia/Australia	24
Customer scope	
B2C	79
B2B	71
Own assets scope	
Network coverage	78
Distribution coverage	71
IT coverage	55
Call center/install coverage	45
Cont(/.)	
2. Financial performance	
Revenue (USD)	
Less than 1 billion	29
Less than 5 billion	22
Less than 10 billion	12
More than 10 billion	37
Revenue growth (%)	
Negative	18
<4 %	41
>4 %	41
Cash flow (%)	
Negative	17
<10 %	44
>10 %	39
3. Business priorities	
(% Implementing)	
New innovation	67
Lean cost technique	53
Enterprise process digitization	46
Outside telecom diversification	38
Digital Capabilities upgrade	31
Big data	30
M&A	26
Business model re-definition	25

**Table 2 Telecom big data projects. Source: Bughin [7], sample size: 273 telecom executives**

	Statistics (%)
1. Big data projects	
Big data priority	
Doing it	30
Considering	45
Interested	23
Big data domains when invested	
Sales/marketing	77
Customer care	57
Competition	41
Network	36
Supply chain	30
Big data architecture investments made	
Applications	53
Database	44
Megaservers	22
Big data capabilities in house	
Analytics	78
IT architects	78
Telecom digitization projects	
Digitization priority	
Doing it	46
Considering	32
Interested	19
Digitization domains when launched	
Sales/marketing	76
Customer care	62
Competition	22
IT/network	55
Supply chain	60
Processes digitized	
Admin	44
Book to bill	39
IT as service	44
R&D	42
Network	38
Install	20
Digital investments	
Applications	68
Database	26
Servers	21

The target survey respondent was the CEO, or at least someone from the executive committee. 36 % of answers were from CEO, 64 % from someone from the management. The survey was answered by 273 respondents. Those respondents are managers of companies that stand for 23 % of total global telecom revenues. The sample is biased on the largest telecom companies worldwide as the survey was sent to the top five telecom companies (if existing) by countries, and is slightly over-represented towards European companies.

54 % of the telecom companies are present in both fixed and mobile; for the mono-segment companies, 57 % are mobile operators only. 79 % serve the consumer segments, 71 % serve the B2B segment.

### **Big data adoption in telecom**

#### ***Big data relevance and adoption***

Big data is still in early phase of deployment. Recent business studies claim that about 20 % of companies in all sectors, have been deploying big data, for a total of 70 % considering big data as an important project [20].<sup>6</sup> Gartner [18] reports that 26 % of companies have been testing and implementing Hadoop technology tools.<sup>7</sup> Likewise, we observe in Table 2 that big data is becoming an important topic on the agenda of the telecom executives. About 30 % of companies were in the middle of launching big data projects in multiple functional cases, and another 45 % was actively considering to invest by 2014. As executive initiatives, though, big data ranks only as the 6th management topic in importance against which initiatives were being launched in 2014. Regarding the five most relevant management topics, launching new innovations ranks as the most important topic of concern (for 67 % of telecom companies), followed by the ability to achieve a lean cost structure, by the need to launch enterprise digitization and by the upgrade of telecom capabilities.

#### ***Big data archetypes***

Big data projects have multiple features see Gandomi et al. [17] or Özköse and Gencer. [25]. Regarding IT architecture, big data includes applications, architecture and mega servers which allow the handling of massive flows of unstructured data. In our sample, about 2 out of three telecom companies have adapted their IT architecture to support big data applications. Likewise, 53 % of companies have invested in big data applications by 2014 (See Table 2). Also, big data may also involve multiple use cases. The survey has collected information on five functional domains: (1) sales and marketing (sales, retention, and targeted marketing), (2) customer care, (3) competitive intelligence, (4) network load optimization, and (5) supply chain optimization (logistics and telecom equipment purchase).

The large majority, 77 %, of telecom companies adopting big data, have launched projects in sales & marketing domains. 57 % of companies have used big data for customer care; 41 % did so for competitive intelligence, 36 % for network load optimization and 30 % for supply chain optimization. There is alas a paucity of information as to the mix of big data domains launched by industries; for the spare available data, Bughin [9] reports that 72 % of companies invested in big data had launched big data applications in customer facing domains (marketing and sales, fraud, retention, and services)—for 35 % using big data for competitive insights. Business reports generally support the view that the most often use cases are linked to customer facing domains.<sup>8,9</sup>

<sup>6</sup> See IDG [21].

<sup>7</sup> See Gartner [19].

<sup>8</sup> A review of 50 use case by Phys.org leads to 62 % of use cases linked to customer facing domains. <http://phys.org/news/2012-12-big-survey-potential-business.html#jCp>.

<sup>9</sup> We were not able to find any other data reporting the type of use case frequency in telecoms—some anecdotal evidence of use case mix in favor of marketing analytics, fraud, and network optimization can be found, see <http://blog.fluturasolutions.com/2012/07/3-game-changing-big-data-use-cases-in.html>.

**Table 3** Frequency of bottlenecks encountered in big data projects. Source: Bughin [7], sample

Constraints	Dimensions	Percent encountered (%)
Data	Lack of quality	60
	Lack of permission to use data	37
	Lack of sufficient data	23
Talent	Limited pool to recruit	29
	Limited intelligence as to where and how to apply big data	15
Organization	Big data org too low in organization	57
	Big data IT not synchronized with function	47

***Big data pain points***

Respondents clearly report that big data investments are not easy projects to launch and implement (Table 3). Telecom companies are facing issues around data itself, whether it relates to poor quality or limited permission to use data. Also, finding enough talent pool remains difficult (see Manyika et al. [10]). A surprising finding is the high portion of telecom companies reporting major organizational confusion. Typically, telecom companies complain that big data projects are managed “at too low level in organizations”, or that big data projects are not run smoothly between IT, data analysts, and business. The issue of organizational capital for generating returns from technology can be a big deal, as discussed in Brynjolfsson et al. [5]. Note finally that some other pain-points may be visible in big data projects but, which were not measured by our telecom sample. A data provider, Experian [29], recently published a report highlighting among other extra issues, such as lack of proper funding and poor change management. Experian [29] develops the concept of a data sophistication curve where companies must act upon people, technology and process to be sure of reaping optimal revenue from big data.

***Big data returns***

A large literature has been devoted to finding evidence of a positive link between information technology and firm performance. Such a relationship has been confirmed via a meta-analysis of tens of studies by Lim et al. [24] and induces large productivity effects in the range of 5–6 %, for enterprise technologies such as ERP solutions [2], and recently, for big data investments [28] and [9]. Our goal here is not to derive an estimate of the returns to big data in telecom using a technology production function, but rather to use direct self-reported returns at the company level. The main reason is that our scope of work is more in understanding drivers of difference in returns among telecom companies. Keeping in mind the possible caveat of reported returns, the following is visible from our sample.<sup>10</sup>

***Big data as a portion of telecom investment***

Investment in big data are of two types, investment in IT (architecture, data bases such as Apache, Pig, Hadoop etc. as well as applications)—on top of hiring new big data capabilities (e.g. talents for running analytics, or for managing data base and IT architectures

<sup>10</sup> If those reported answers are good indicators of true returns, we should be able to correlate them with variables from theory of firm performance difference. In general, also, variables with large “noise” are not a problem as far as they are dependent variables. If there is further a common drift in the reporting of returns, this will be tackled by the constant term in the return equation.

etc.). An overwhelming amount (more than 90 %) of telecom companies have sought to hire new big data specific resources—as much for the IT function as for the analytics supporting the various big data functional use cases.

Our survey further shows that 1.8 % of workforce, as well as, 3.1 % of telecom capital expenditures have been invested by telecom companies for big data in 2014. At average pro-forma spent of a telecom company, the total spent for big data was respectively 1.1 % labor towards and 0.9 % capital towards, of telecom revenue.

#### ***Big data contribution to telecom profit***

Is there a (perceived) return to big data investments? The average telecom company respondent reports that big data contributes 2.9 % of its total telecom company profit. This reported impact is larger than the share of spent in big data (2 % of revenue spent in total), but slightly lower than the share of capex spent (3.1 %), which would suggest that big data leads to barely the same profitability as other projects in telecom companies.

While we concede that such a reported measure can at best be considered as a noisy marker of the true underlying return, we further observe that such a magnitude of impact on companies' bottom line falls into typical ranges of impact found in the recent literature. Big data impact of 5–6 % of total productivity has been reported in Brynjolfsson et al. [3] for a sample of companies using data analytics. Tambe [28] demonstrates that Hadoop investments yielded a 3 % higher labor productivity for firms that were established in data-intensive industries (as in this case of telecom).

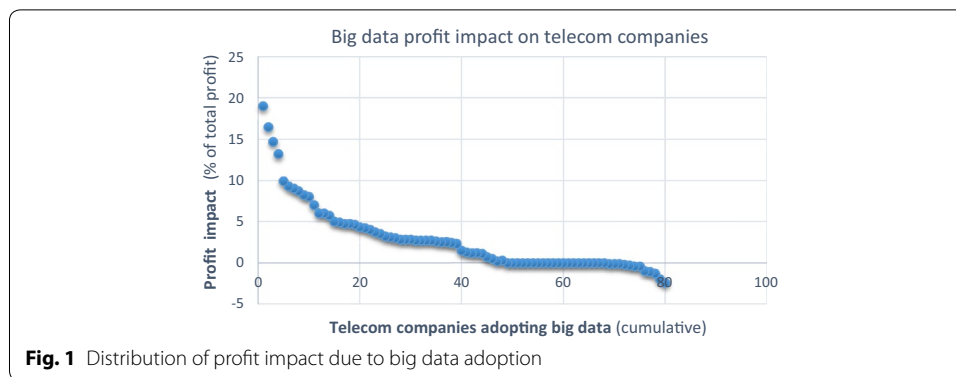
Returns also vary by functional domains: aggregating returns by domains, the largest returns are to be found in projects in: customer service, sales and marketing (average, 4.1 %; median, 2.3 %), followed by, competitive intelligence (3.3 %, median; 1.9 %). On average, networks (1.9 %; median = 1.8 %) and supply chain domain projects (returns of 1.7 %, for a median profit impact of 1 %) contributed the less.

This pattern of use case returns is difficult to confirm given the paucity of data in other research. In Bughin [9], projects targeting application domains, such as customers and competitive intelligence bring slightly more performance than other domains such as supply chain management.<sup>11</sup> Chen et al. [12] emphasize business and competitive intelligence as one key profitable corporate use case.

#### ***The distribution of profit impact of big data***

Another stylized fact of the IT literature is the large variance of IT returns among firms [14]. As Fig. 1 illustrates, we also find a large spread in the profit contribution of big

<sup>11</sup> A few use case examples from the survey can further illustrate the type of projects and returns telecom companies have captured from big data. One telecom company in our sample used multiple big data software to assess the load in real time of heavy video streaming applications, and limit congestion on fiber networks. A predictive model led to identify multiple causes of congestion, and their proactive prevention led to a 15 % reduction in capex planning, while reducing IPTV customer churn by 10 %, via reduction of jitter and packet loss. Another telecom company reported on its use of machine learning based model of voluntary churn, and fraud/bad payments. The data used was a mash up of socio-demographics, all customer touchpoints (online, call and service center, social media), network points' usage, etc. The machine learning model classifies in real time, the most likely churners and non payors, including the types of actions to retain or to ensure payment recovery. About 2 points of churn could be improved technically by big data analytics alone. A major cost to fixed telecom providers is the installation of services at the customer premises. A telecom company built a big data software to help profile homes and anticipate likely faulty or challenging home installs. The home data and predictions were provided real time to all installers with a list of hints, so that they can also check the premises, and ensure first time right installation. The rate of first time right increases by about 11 %, leading to significant lower costs of install.



data projects among telecom companies. This spread also appears at the disaggregated level of each big data use case. We further note that the distribution is rather skewed, with a median impact of big data on profit at 1.4 %, or a 50 % reduction of profit impact compared to the average. Also, 40 % of telecom companies which invested in big data, reported no, even negative, impact on their total profit. Only 5 % claim to have generated a quite material profit impact that is, big data contributes already above 10 % of a telecom company's operating profit.

Clearly, this distribution obeys to a power law, as often found for new technology adoption impact on profit by firms—see for instance Bughin [6] for enterprise 2.0 technologies. Finally, the queue of the distribution seem to exhibit a “fat tail”.<sup>12</sup>

### Explaining drivers of big data returns

We now aim to explain the variance among companies. Let us denote by  $BDIMPACT_{i,p}$ , the extent to which the big data project of type  $p$  ( $p = 1, \dots, 5$ ) contributes to the profit of the  $i$ th firm ( $i = 1, \dots, 273$ ).

### Theory and hypotheses

#### Adoption factors

Multiple theories have been developed to explain difference in firm technology investment returns. The first theory relates to the *optimal speed* of adoption in technology. In turn, different adoption patterns lead to observed different returns by firms, at a certain time  $t$ .

Game-theoretical models of technology adoption confirm the typical S-curve of technology diffusion, arising from the distribution of companies' perceived uncertainty and mostly, their competitive rivalry [16]. The second-theory is linked to path-dependency of technologies [2] show that CRM and SCM returns are dependent on firm having successfully invested early in ERP system. Most of big data investments rely on the extent to which a company has been engaged in its digitization process

<sup>12</sup> A rank fit linear regression of the form:  $z = a - b \cdot r + u$  can be estimated via least squares with  $u$  is a random term;  $r$  is the log of the rank of a company in terms of its reported impact,  $y$ , of big data on profit, and we note  $z = \log(1 + y) = y$  (when  $y$  is small). The estimated line produces the following results:  $z = 7.4 - 4.2 \cdot r$ , or a power law exponent of  $1 + (1/4.2) = 1.3$  (with R-square of 98 %, and both coefficients statistically significant at 5 %). At rank one, the regression leads to 7.4 % impact (versus 20 % reported); at the median rank,  $r = 40$ , the regression leads an impact of roughly 0.7 %, while at last rank  $r = 80$ , the impact is  $-0.6$  % (versus  $-2.5$  % reported).

[30].<sup>13</sup> Table 2 provides statistics as to how digitization has spread among telecom companies in our sample. The general picture emerging is that telecom enterprise digitization is ahead of big data. Let us then denote by  $DIG_{i,e} = 1$  (respectively,  $BD_{i,e} = 1$ ) when the  $i$ th telecom company has invested in digital technologies (respectively in big data);  $= 0$ , otherwise and we note  $F' = (BD, DIG)$ . We postulate:

*H1: [DIGITIZATION] Big data adoption accelerates with extensive digitization:  $\delta(BD)/(DIG) > 0$*

*H2: [EARLY ADOPTION] Early big data adopters generate more returns to big data:  $(BDIMPACT)/(BD) > 0$*

### Internal firm effects

Multiple theories predict differences in firm returns based on internal capabilities of companies. The first theory is the one linked to organizational capital, that is companies may have to ensure adequate mix of resources and adjust their internal organization to reap all the benefits of their new IT investments [5, 28]. Table 3 suggests that organizational bottlenecks remain material for the telecom industry. Anecdotal business evidence highlights the same issue; companies such as Harrah's or Macy's in the US have ensured that big data is an important business unit, under the leadership of a specific leadership in the executive team. [13] emphasizes the importance of big data to be managed high enough in organization hierarchies in order to ensure it is perceived as a strategic mandate for companies.

The second theory is linked to the firm technical production function. Based on production form, investment productivity can be dependent on labor productivity, can be enhanced by scale or scope. Using a translog function, Bughin [7] for instance highlights strong complementarity between big data labor and big data specific capital investments.

Another theory is the one of dynamic capabilities, that is, the firm's capability to use its resources more adequately than others in a high turbulent environment. [34] states, that even if capabilities converge among all firms, permanent difference in firm performance may be visible as the result of different timing and cost of deploying those capabilities. In our sample, a proxy of timing and cost of big data capabilities is the number of big data use cases, used by companies. Likewise, we have already reported that a set of telecom companies have chosen to launch big data projects without major change in either their existing IT infrastructure, or in their existing set of applications and/or servers. Those companies believe that they can win time, and already benefit from big data by squeezing "some juices" from existing IT investments [11].<sup>14</sup>

Let us define the following variables.  $ORG_i$  is the organizational fit variable, with  $ORG_i$  ( $ORG_i = 0, 0.5, 1$ ) and  $ORG_i = 0$  means that both vertical and horizontal organizational barriers are absent. Likewise,  $ARCH_i$  ( $ARCH_i = 0,33 \%, 66 \%, 1$ ) = 1 when the

<sup>13</sup> An illustrative case is the recently launch of the telecom tech company, GiffGaff, as a mobile data virtual network operator in the UK. GiffGaff has digitized its customer care extensively, with care resolution being crowd-sourced by its users via a large array of video posts available for the community subscribers; In turn, this user generated content has hence been mined cleverly via big data analytics by GiffGaff to reward best customers, etc. See Lithium (2011), "Giffgaff case study: using community to build an entirely new kind of company", available at <https://www.lithium.com/pdfs/casestudies/Lithium-giffgaff-Case-Study.pdf>.

<sup>14</sup> As mentioned by one referee, there is a wrong notion that big data is the replacement of existing data capabilities. As we highlight in this article, big data is rather a set of major platform IT investments that allows data analysts to perform new forms of analyses not in use before, due to lack of data to empower those new analyses. Companies which are not adapting their IT investment to create this analytic boost will then be making a "paradigm" mistake.

ith telecom company has invested in all three building blocks of (1) big data architecture, (2) new applications and (3) new mega-servers.  $CASE_i$  ( $CASE_i = 0.20\%, 40, \dots, 1$ ) measures the diversity of big data use cases launched by the  $i$ th telecom company.  $CASE_i = 1$  when the  $i$ th company has invested extensively in all types of big data use cases among the five functional domains. Finally, we have computed  $BDVIT_i$  and  $BDHC_i$  as respectively the amount of IT investment and human resources invested in big data by the  $i$ th telecom company. Using the above, we postulate the following hypotheses:

*H3: [DYNAMIC CAPABILITIES] Big data returns are enhanced by new IT architecture and scope of big data investments:  $\delta (BDIMPACT)/\delta (CASE) > 0$  and  $\delta (BDIMPACT)/\delta (ARCH) > 0$*

*H4: [COMPLEMENTARITY] Big data returns exhibit complementarity between labor and capital:  $(BDIMPACT)/\delta (BDVIT)(BDHC) > 0$*

*H5: [ORGANIZATIONAL CAPITAL] Big data returns benefit from adequate organizational changes:  $(BDIMPACT)/\delta (ORG) < 0$*

Table 4 provides the summary statistics regarding the new additional variables created to test our set of H1-H5 hypotheses. It shows that a telecom company (1) has been investing on 2 out of 5 big data use case domains ( $CASE = 38\%$ ); (2) has been “half way” in investing in the spectrum of big data investments ( $ARCH = 42\%$ ), and typically suffers more than not from organizational misfit linked to big data ( $ORG = 63\%$ ).

#### The econometric model

We postulate a function in log form, so as to account for the skewed distribution of most of the variables, as in the generic Eq. (1), for each use case,  $p$ :

$$\text{LOG}(1 + BDIMPACT)_{i,p} = \text{cst} + \text{LOG}(F', M')_{i,p} + u_{i,p} \quad (1)$$

where  $\text{cst}$  is a constant term to be estimated,  $u$  is a disturbance term assumed to be normally distributed across companies and  $F' = F'(BD, DIG)$ , and  $M' = M'(BDIT, BDHC, ORG, ARCH, CASE)$  are two vectors of regressors.

#### Controlling for possible biases

Estimating (1) above by a traditional OLS procedure may be subject to two biases, one linked to sample self-selection; the other, to endogeneity issues. Concerning the later, ex post returns are most probably correlated with ex ante returns, which themselves influence

**Table 4 Sample statistics for testing H1–H5. Source: authors’ computation based on sample in (Bughin [7]) across all use cases**

Variables	Average (%)	Standard deviation (%)	Coeff variation (%)
DIG	46.0	38.9	84.6
BD	30.0	28.0	93.3
ARCH	42.3	20.0	47.3
ORG	63.5	31.0	48.8
BDIT	0.9	1.2	133.3
BDHT	1.1	1.0	90.9
CASE	38.3	31.0	89.8

the decision to adopt big data (see above as per game-theoretical models of adoption). Concerning the former, telecom companies which are early adopters of technology may also be getting more returns to big data, because they are generally more efficient than average. This sample selection has been found to be significant, in the case of big data [9].

Accordingly, we are thus made endogenous, the variable, BD, and we postulate the following equation:

$$(BD)_{i,p} = cst + b \cdot (DIG) + c \cdot (SPILLOVER)_{i,p} + r_{i,p} \quad (2)$$

where  $cst$  is a constant,  $b$  and  $c$  are parameters to be estimated,  $r$  is a disturbance term assumed to be randomly distributed. We expect  $b > 0$  as per H2 above.  $SPILLOVER_{i,p}$  is the proportion of other telecom companies ( $j \neq i$ ) which have invested as well in project type  $p$  of big data. This variable captures the rivalry effect as per game theoretical models of adoption [22]. Another rationale is that  $SPILLOVER$  captures a common external factor market (e.g. [28]). Hence, we should expect  $c > 0$ .

Technology diffusions are known to exhibit more or less a S-curve, so one would be enclined to transform the Eq. (2) in a logistic equation. Technically, however, Heckman [27] shows that the inverse of the Mills ratio,  $CBD_{i,p}$ , must be measured directly in probit form from the regression (2), as the accurate procedure to control for the selection bias in the Eq. (1).

### Final regression system

Using the above variables to expand the vectors  $F'$  and  $M'$ , Eq. (2) is substituted into Eq. (1) to provide Eq. (3) hereafter:

$$\begin{aligned} \text{LOG}(1 + \text{BDIMPACT})_{i,p} = & cst + s \cdot \text{CBD}_{i,p} + g \cdot \text{LOG}(1 + \text{ARCH})_{i,p} + h \cdot \text{LOG}(1 + \text{BDIT})_{i,p} \\ & + j \cdot \text{LOG}(1 + \text{BDHC})_{i,p} + k \cdot [\text{LOG}(1 + \text{BDIT}) * \text{LOG}(1 + \text{BDHC})]_{i,p} \\ & + m \cdot \text{LOG}(1 + \text{CASE})_i + n \cdot \text{LOG}(1 + \text{ORG})_i + w_{i,p} \end{aligned} \quad (3)$$

with  $cst$  is a constant;  $w$  is a random term,  $s, \dots, n$  are parameters to be estimated, and  $s > 0$ . Based on set of hypotheses H1–H5, we expect:  $h, j > 0$  would imply returns to scale;  $h, j > 1$ , would entail *increasing* returns to scale.  $k > 0$  would confirm the existence of complements between big data IT and talent resources;  $m > 0$ , would be a proof of returns to diversity, and  $n < 0$  implies organizational misfit.

We further add a few telecom control variables in each Eqs. (2) and (3). We control for telecom company's headquarter location, LOC; customer segment presence (B2B, B2C), as well as product segments (MOB for mobile and FIX, for fixed telecom products). We also control for size (as measured by revenues, REV). In practice, we first experimented with clustering telecom features. We were able to define three clusters, C1, C2, C3, where C1 = European and US integrated fixed and mobile operators; C2 = European mobile operators, and C3 for all others. We find indeed that LOC, B2B and REV are all positively associated with C1 and are thus rather redundant variables. C3 is used as the default in the reported regressions.<sup>15</sup>

<sup>15</sup> Remember that the variables tacking  $M'$  in Eq. (3) are expressed in the form  $\log(1 + M')$ , as well as is the dependent variable (returns), in order to account for the skewed distribution of each variable. We can approximate  $M'$  as  $M' = \log(1 + M')$  for  $M'$  being small; hence the coefficients estimates in Eq. (3) may be loosely interpreted as marginal effect on the total profit impact arising from big data.

## Results

The equation system (2), (3) is estimated for each of the five big data use cases. We leverage a SUR technique among all the various use cases of returns from adoption [33] as the big data use case returns are not independent as originating from the same telecom company. Results are presented in both Table 5 and 6.

**Table 5 Telecom big data adoption rate**

Variables (t-stat)	Use cases				
	Sales and marketing	Customer care	Network	Supply chain	Competitive intelligence
Constant	8.1 % (2.23)	−0.6 % (−3.11)	−0.8 % (−1.03), NS	0.4 % (0.98), NS	−3.7 % (−2.92)
C1	2.6 % (3.21)	1.2 % (1.03), NS	4.7 % (3.17)	1.9 % (1.97)	−2.2 % (−2.35)
C2	1.4 % (4.56)	2.3 % (2.02)	−2.0 % (−1.56), NS	0.6 % (2.23)	0.3 % (2.90)
DIG	20.2 % (3.78)	24.0 % (4.02)	14.4 % (2.18)	10.7 % (2.84)	21.2 % (1.97)
SPILLOVER	23.4 % (4.55)	32.0 % (2.79)	41.7 % (3.71)	33.4 % (1.99)	44.0 % (2.49)
Pseudo R2	45.5 %	37.60 %	29.20 %	46 %	51 %

Probit model estimation  
NS non significant coefficient

**Table 6 Explaining telecom big data returns**

Variables (t-stat)	Use cases				
	Sales and marketing	Customer care	Network	Supply chain	Competitive intelligence
Constant	−1.0 % (−4.06)	−0.2 % (−1.49), NS	−0.9 % (−3.84)	−1.7 % (0.98), NS	−1.3 % (−2.2)
C1	1.1 % (2.17)	2.4 % (2.45), NS	3.2 % (1.99)	0.9 % (1.97)	−0.2 % (−1.36), NS
C2	−0.2 % (−0.55), NS	0.4 % (3.61)	−1.7 % (−1.63), NS	−0.6 % (2.23)	0.3 % (2.90)
CBD	1.0 % (3.00)	0.7 % (2.24)	0.6 % (2.67)	1.4 % (2.84)	0.8 % (1.97)
ARCH	5.9 % (2.12)	4.7 % (1.96)	4.0 % (3.71)	2.7 % (3.07)	1.9 % (2.06)
BDIT	23.3 % (4.06)	19.60 % (4.14)	35.20 % (2.29)	31 % (2.15)	45 % (1.99)
BDHT	34.2 % (3.78)	41.2 % (2.79)	45.2 % (2.18)	12.8 % (2.84)	35.8 % (1.97)
BDIT*BDHT	197.20 (4.55)	112.80 (2.32)	188.90 (1.97)	97.70 (3.04)	67.60 (0.78), NS
CASE	1.9 % (3.78)	1.3 % (1.02), NS	1.9 % (2.01)	3.0 % (2.45)	4.7 % (3.01)
ORG	−3.2 % (−4.55)	−2.1 % (−2.65)	−4.0 % (3.14)	−1.1 % (−2.34)	−4.5 % (−2.98)
R <sup>2</sup>	55 %	61 %	49 %	62 %	44 %

NS non-significant coefficient estimate  
SURE regression estimates

### ***Big data adoption probit estimates***

We first discuss results linked to big data adoption in Table 5.

The general cross-sectional fit is good, with a variance explained (as measured by the pseudo-R<sup>2</sup>) in the range of 30 % for big data network use case, and more than 50 % for competitive intelligence. All coefficients linked to digitization adoption and imitation are statistically significant. Further, we find:

1. First, big data contribution to total telecom profit is minor, but in line with its relative size of investment in total telecom spent, and generates a productivity impact aligned with other research (Tambe [26], or Bughin [10]).
2. Second, the distribution of returns exhibits a long tail of limited (sometimes even negative) returns with a few companies claiming to make disproportionate returns from big data investments.
3. Third, using a joint model of big data adoption and of returns on adoption, we try to explain key drivers of this variance in big data performance. Consistent with the theory of technology adoption [e.g. [20] and of organizational capital [4], we find that a large part of the cross-sectional distribution of returns is explained by the types of capabilities of companies to effectively manage big data projects. Further, as found in theory of production [10, 26], we find the existence of returns to scale and scope, and complementary talent to IT architecture investments in big data.
4. We finally test whether big data capabilities are of crucial importance for the financial returns linked to big data projects. We find indeed that mastering capabilities at scale are necessary to generate returns above cost of capital for big data in the telecom industry.

### ***Big data contribution to profit regression estimates***

We then use the fitted results of the probit Eq. (2) to compute the Mills ratio of the propensity to adopt and we then include this variable as additional regressor in Eq. (3). Results are presented in Table 6; the results are then used to illustrate the difference in big data contribution to total telecom profit, between the bottom quartile telecom companies and the top quartile, 25 %, of telecom companies, for each big data use case (Table 7).

Table 6 demonstrates that goodness-of-fit is relatively good, between 49 % (for network) to 62 % (supply chain) of variance explained in use case contribution to profit. Further, we find that:

1. The inverse of the Mills ratio CBD is systematically positive, confirming H2. More crucially, the effect ranges from 1 to 1.4 % of total profit contribution; this stands for a rather significant figure, as the contribution to profit of big data is oscillating between 1.7 and 4.2 % of profit impact, depending on use case.
2. On one hand, returns to big data are positively related to the extent spent on big data talent and architecture, but also,  $0 < h_p, j_p < 1$ , for most of the use cases, implying *marginally* decreasing returns to scale; on the other hand we see returns to *quality* of spent to be positive,  $h > 0$ , that is H3 is confirmed that it pays off to be investing in the *right* domains of big data (architecture, servers and applications).

**Table 7 Breaking down telecom big data contribution to profit between top and bottom 2 % of telecom companies. Source: from other tables, author's own computation**

Gap	Use cases					
Top 25 %/ bottom 25 %	Sales and marketing (%)	Customer care (%)	Network (%)	Supply chain (%)	Competitive intel- ligence (%)	Weighted average (%)
Gap of which	2.7	2.8	2.5	1.6	2.9	2.6
C1, C2	0.2	0.9	−0.1	−0.1	0.1	0.3
ARCH	0.7	0.6	0.5	0.3	0.2	0.5
BDIT, BDHT	0.8	0.7	1.0	0.6	0.8	0.8
CASE	0.4	0.3	0.4	0.6	1.0	0.5
ORG	0.5	0.35	0.66	0.2	0.7	0.5

3. H4 is also broadly confirmed: investing in big data talent resources is complementary to investing in big data IT architecture as found elsewhere in Bughin [10]. This effect is relatively large, but inelastic at the mean. The complementary effect is not statistically significant when it relates to competitive intelligence. We also find  $m > 0$ , that is, returns to scope for every domain of big data use case except customer care.
4. Last but not least, H5 is also verified, with evidence of large negative effect of organization misfit on returns from big data,  $n < 0$ .

#### **Comparison between bottom and top quartile impact of big data**

In order to get a sense of how those factors explain the distribution of big data contribution to profit, Table 7 computes the difference between moving from the bottom 25 % to top 25 % of telecom companies in terms of profit impact. Three main messages are visible:

1. Moving from the bottom to the top 25 % would lead to an uplift of 2.6 points of percentage of profit contribution, or a rather significant effect.
2. All drivers, except control effects, always increase profit contribution, when moving from bottom to top 25 %
3. While we witness some differences in the drivers of profit uplift by use case (for example, network use case uplift is relatively more driven by amount of investment in big data than for other cases) the underlying message is that investments in appropriate big data investments and organizational fit are material factors to big data contribution to profit. This is a crucial point as the set of bottlenecks, as highlighted in Table 3, remains large in telecom companies. In fact, we observe that too many companies still ignore data quality problems, do not allow for synchronization of skills for big data and do not take the right investment approach to big data. While finding solutions to all bottlenecks is clearly out of scope of this paper, a few important tips may already improve the situation for telecom companies and companies all large. For instance, we see that data quality is more ensured by companies placing data quality requirements into the corporate risk function. We note as well that many companies have also reorganized their structure to include a chief data officer at the management board, who is coordinating data across all key functions. Finally,

best practice companies have a chief analytic officer position in charge with HR to develop the new data culture mindset and training [10].

### Is investing in big data value creating?

Based on the above results, we now provide a rough-cut perspective on the value creation potential of big data, that is, we raise the question as whether the *risk-adjusted* returns to big data are large enough to improve the fate of telecom companies. In general, the message in the management and ICT literature is that the risk-adjusted return of IT investment is very close to the cost of capital for the average firms, see Im et al. (2001); only a few companies are disproportionately making more than their cost of capital. We find roughly the same pattern regarding big data:

1. The average profit of telecom companies was 8 % of revenue by 2014 in our sample. This implies that the return of big data capitalisable investment,  $ROCI = (8 \% * 2.9 \%) / 3.2 \% = 7.25 \%$ . This value is about the same ballpark level as the weighted cost of capital of the telecom industry, in the range of 7 % by 2014.<sup>16</sup> Otherwise stated, the risk-adjusted returns are just meeting the competitive rate of our telecom companies.<sup>17</sup>
2. Are they some companies doing better? The distribution of ROCI can be further estimated as follows. Referring to Fig. 1, the average company cuts off at about 30 % of our sample, with the later generate above average ROCI, while the other lower part of distribution, roughly 70 % of telecom companies, generating dilutive returns to telecom companies. The head of the distribution includes 18 % of telecom companies which have managed to generate contribution to profit above 5 %, for a contribution of 7.5 % of profit from big data, while spending 1.5 times more. We can compute that those companies generate a ROCI level of 13 %, or twice the cost of capital. Applying same method, the top 5 % of companies reaches a ROCI, above 20 %, *three* times the benchmark of the cost of capital.

### Conclusions

This article has looked at the returns generated by five big data use cases applied in the telecom industry. The article finds evidence that big data projects generate positive contribution, but also with only a few providing additional risk—adjusted, market value from big data.

Using a probit model of big data adoption as well as a regression model returns to adoption, we find that big data returns can be enhanced significantly to the extent that companies obey to a few managerial and organization practices. Among others, we find evidence for the following five recommendations:

<sup>16</sup> We did not compute the cost of capital for our telecom companies, even if we have access to the full accounts and 64 % of the companies in the sample are publicly listed. As indication, we rather use the cost of capital as computed and maintained by A. Damodaran, at the New-York School of Business, for 2014; see: [http://people.stern.nyu.edu/adamodar/New\\_Home\\_Page/data.html](http://people.stern.nyu.edu/adamodar/New_Home_Page/data.html).

<sup>17</sup> Some may argue that the cost of capital may be higher given the risky nature of IT investment; Dewan et al. [15] even come to conclusion that 30 % of return to IT covers higher risk, For a capital market point of view, this risk is likely non-systematic, so we chose not to include this as part of the weighted cost of capital.

1. Big data adoption should be earlier than late;
2. When spending money on big data use cases, money should first be spent on the right IT architecture to reap the complementary benefit of big data talents. This notion is not one of replacing data capabilities, but rather to make sure that data scientists have the right IT architecture, tools and data to deliver their insights;<sup>18</sup>
3. Big data should be put high enough in organization to allow traction, and favor end-to-end synchronization between users and big data producers within an organization;
4. Companies should invest in multiple use cases.
5. Digitization helps adopting big data, provided that big data projects are well managed.

Obeying those principles might reverse the curse of limited risk-adjusted return from big data.

#### Author details

<sup>1</sup> McKinsey and Company, Avenue Louise, 480, 1050 Brussels, Belgium. <sup>2</sup> McKinsey Global Institute, Avenue Louise, 480, 1050 Brussels, Belgium. <sup>3</sup> Free University of Brussels, Brussels, Belgium. <sup>4</sup> Applied Economics, Katholiek Universiteit, Leuven, Louvain, Belgium.

Received: 11 March 2016 Accepted: 6 July 2016

Published online: 28 July 2016

#### References

1. Amatriain X. Beyond data: from user information to business value through personalized recommendations and consumer science. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management 2013 Oct 27 (pp. 2201–2208). ACM.
2. Aral S, Brynjolfsson E, Wu DJ. Which came first, it or productivity? Virtuous cycle of investment and use in enterprise systems. *Virtuous Cycle of Investment and Use in Enterprise Systems*. 2006.
3. Brynjolfsson E, Hitt L, Kim H. How does data-driven decision-making affect firm performance?, MIT—Sloan School of Management. 2011.
4. Brynjolfsson E, Hitt L. Paradox lost? Firm-level evidence on the returns to information systems spending. *Manage Sci*. 1996;42(4):541–58.
5. Brynjolfsson E, Hitt LM, Yang S. Intangible assets: computers and organizational capital. *Brooking Pap Econ Act*. 2002;2002(1):137–81.
6. Bughin J. The power law of enterprise 2.0. In: Lee I, editor. *Encyclopedia of E-business development and management in the global economy*. IGI Global edition: Hershey; 2010. p. 2010.
7. Bughin J. *Relaunching TMT: results from the 2014 extranet survey*. London: McKinsey Tele-communications, Media, and Technology; 2015.
8. Bughin J. Google searches and twitter mood: nowcasting telecom sales performance. *Econ Res Electron Netw*. 2015;16(12):87–105.
9. Bughin J. Big data, big bang. *J Big data*. 2016;3(2):1–14.
10. Bughin J, Chui M, Manyika J. Clouds, big data, and smart assets: ten tech-enabled business trends to watch. *McKinsey*. 2010;56(1):75–86.
11. Chen CP, Zhang CY. Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Inf Sci*. 2014;275:314–47.
12. Chen H, Chiang RH, Storey VC. Business intelligence and analytics: from big data to big impact. *MIS Q*. 2012;36(4):1165–88.
13. Davenport T. *Big data at work: dispelling the myths, uncovering the opportunities*. Brighton: Harvard Business Review Press; 2014.
14. Dedrick Jason, Gurbaxani Vijay, Kraemer Kenneth L. Information technology and economic performance: a critical review of the empirical evidence. *ACM Comput Surv*. 2003;35(1):1–28.
15. Dewan S, Shi C, Gurbaxani V. Investigating the risk-return relationship of information technology investment: firm-level empirical analysis. *Manage Sci*. 2007;53(12):1829–42.
16. Fuentelsaz L, Gomez J, Polo Y. Intrafirm diffusion of new technologies: an empirical application. *Res Policy*. 2003;32(4):533–51.

---

<sup>18</sup> We thank a referee for highlighting this point.

17. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage*. 2015;35(2):137–44.
18. Gartner Survey Analysis. Hadoop adoption drivers and challenges. 2015. Available at <http://www.gartner.com/document/3051617>.
19. Hannan TH, McDowell JM. Rival precedence and the dynamics of technology adoption: an empirical analysis. *Economica*. 1987;54:155–71.
20. IDG. IDG enterprise big data research. 2014. Available at <http://www.idgenterprise.com/report/big-data>.
21. Im KS, Dow KE, Grover V. Research report: a reexamination of IT investment and the market value of the firm—an event study methodology. *Inform Syst Res*. 2001;12(1):103–17.
22. Karshenas M, Stoneman PL. Rank, stock, order, and epidemic effects in the diffusion of new process technologies: an empirical model. *Rand J Econ*. 1993;24(4):503–28.
23. Lewis M. *Moneyball: The art of winning an unfair game*. New York city: WW Norton & Company; 2004.
24. Lim J, Richardson VJ, Roberts TL. Information technology investment and firm performance: a meta-analysis. In: *Proceedings of the 37th Annual Hawaii International Conference on system Sciences*. IEEE; 2004.
25. Özköse H, Ari ES, Gencer C. Yesterday, today and tomorrow of big data. *Proc Soc Behav Sci*. 2015;195:1042–50.
26. Research Wintergreen. Sport analytics: market shares, market strategies and market forecasts, 2015–2021. Lexington: Winter Green Research Inc; 2015.
27. Heckman J. Sample Selection Bias as a Specification Error. *Econometrica*. 1979;47(1):153–61.
28. Tambe P. Big data investment, skills, and firm value. *Manage Sci*. 2014;60(6):1452–69.
29. Experian. The 2016 global data management benchmark report. Available at <https://www.cdq.com/white-papers/2016>.
30. Varian HR. Big data: new tricks for econometrics. *J Econ Perspect*. 2014;28(2):3–27.
31. Wiemker J. The digital transformation of the telecommunication industry: A qualitative benchmark study in the telecommunication industry to identify success factors for a new business model approach which best leverages digital technologies to improve customer interaction.
32. Yongjun H, Ming F, Shengyong D, Yongbing F. Big Data Development Strategy for Telecom Operators. *Telecommun Sci*. 2013;3:002.
33. Zellner A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J Am Stat Assoc*. 1962;57(298):348–68.
34. Zott C. Dynamic capabilities and the emergence of intraindustry differential firm performance: insights from a simulation study. *Strateg Manag J*. 2003;24(2):97–125.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---