

Spillover Effects of Investment in Big Data Analytics in B2B relationships: What is the Role of Human Capital?

Highlights

- The Big Data literature has paid scarce attention to the benefits from the upstream investment in Big Data Analytics (BDA)
- Insights into the role of B2B marketing environment in realizing the benefits of upstream investment in BDA are limited
- We test to what extent access to larger share of local graduates can enhance the benefits of upstream investment in BDA
- We show that retailers with greater access to local graduates are more efficient and benefit more from such investment
- Upstream investment in BDA is positively associated to frontier shifts over time (i.e. technical progress)

Spillover Effects of Investment in Big Data Analytics in B2B relationships: What is the Role of Human Capital?

Abstract

Economy-wide investment in Big Data Analytics (BDA) offers retailers a number of opportunities and while there is some evidence that new technologies have been widely adopted by retailers, it also transpires that many retailers have yet to fully exploit the benefits of BDA. Most research on Big Data and productivity (or performance) tends to focus on internal factors that prevent retailers from fully exploiting their investment in BDA. Research has paid scarce attention to the benefits that can accrue to the focal firm from the upstream investment in BDA and the features of the B2B marketing environment that may hamper (or enhance) these benefits. Unlike the previous literature, the paper tests the extent to which retailers, by having access to larger share of graduate workforce at regionally, can benefit more from inter-industry upstream investment in BDA than retailers located in areas where such workforce is scarce. Using data from ORBIS, KLEMS and QLFS, we show that retailers located in regions with a larger proportion of graduate workforce benefit more from inter-industry upstream investment in BDA as they tend to be more efficient on average. Equally, upstream investment in BDA is positively associated to frontier shifts over time (i.e. technical progress).

Keywords: Frontier, Human Capital, B2B Relationships, Big Data Analytics, Retail.

1. Introduction

In today's increasingly competitive and dynamic business environment, Big Data with its massive volume, variety and velocity provides an array of opportunities to firms (Manyika et al., 2011; Wang, Kung, & Byrd, 2018). However, Big Data per se is not a source of comparative advantage to businesses unless it is exploited to inform both strategic and operational decisions through an integrated set of aggregation, analytics and interpretation techniques (Wang et al., 2018). This set of techniques goes under the label of Big Data Analytics (henceforth, BDA) and comprises all the hardware, software and methodologies that can be used to exploit Big Data.

Several industries in the UK have been identified as key beneficiaries of the widespread adoption of Big Data technologies with retail being one of these. The inclusion of the retail sector in such a list is hardly surprising: first, even if the sector had experienced average annual employment growth rates of about 1% in the decade preceding the 2009-2010 economic recession (Haskel & Sadun, 2008), in reality major retailers in the UK have been increasingly relying on automation and IT to manage key areas of their activities (like the supply chain and logistics) while gradually cutting jobs for nearly a decade now¹. As a result, they are well positioned to benefit from the wider adoption of BDA across the economy. Second, the sector is extremely rich with data as large retailers collect data from millions of customers and marketing channel partners (Bradlow et al., 2017) although evidence suggests that many retailers have yet to fully exploit the benefits of BDA (Germann et al., 2014).

There is an additional benefit that could accrue to retailers from the economy-wide adoption of BDA i.e. increasing productivity levels. The sector accounts for about a third of the UK total factor productivity (TFP) slowdown as documented by Riley (2015). Although several reasons have been put forward to explain such a poor performance of the sector in terms of TFP (as well as labour productivity)², its

¹ According to the British Retail Consortium, major retailers employed 3.0 per cent fewer staff in the third quarter of 2017 compared to the same period in 2016, and total hours worked fell by 4.2 per cent year-on-year (Milliken, 2017).

² Haskel and Sadun (2007, 2008) found that retail's TFP has decreased by 0.4% a year after 1996 when planning regulations were changed in such a way that opening large stores became more expensive. Equally Griffith and Hamgart (2005) suggested that these changes might have had a negative impact on retailers' productivity by preventing them from operating at the minimum efficient scale and hindering the opening of new stores. Griffith et al. (2004) suggest that poor

lackluster performance in terms of productivity growth can be rationalized as being the result of a mix of increasing technical inefficiency and of technological slowdown. Indeed, economic theory suggests that productivity growth can be decomposed into technical change (i.e. improvements in the technology available to the industry as a whole) and technical efficiency change (defined here as changes in the firm's capability of producing the maximum level of output given the current inputs' usage). While there is some evidence that new technologies have been widely adopted by retailers, evidence suggests that many retailers have yet to fully appropriate the benefits of BDA in terms of increasing technical efficiency (Germann et al., 2014; Roh, 2018) suggesting that the latter may drive the productivity slowdown.

Authors have pointed out that the widespread adoption of BDA can help improve technical efficiency among retailers through two channels (see for instance Manyika et al., 2011). The adoption of BDA among retailers may have a direct positive impact on their efficiency as BDA allows them to improve inventory control, storage optimization and pricing. The adoption of BDA across the whole economy can also have an indirect positive impact on retailers' efficiency (and eventually productivity) through their supply chain. According to this view, the upstream (i.e. suppliers') investment in BDA may have a chain effect on retailers and can trigger improvements in technical efficiency through business-to-business (henceforth, B2B) relationships. For example, the investment and use of BDA among suppliers may prompt downstream firms to re-organise their production processes and structures to take advantage of the analytics-induced increased efficiency in running transactions (Criscuolo & Waldron, 2003). Such measures may potentially yield sustained improvements in efficiency due to an improved matching of inventory to customer demand as well as more efficient stock management. Therefore, identifying the factors that may stop retailers from benefitting from the adoption of BDA across the economy is quite an important research area with major managerial implications.

management, a lower-skilled labour force, an inadequately competitive retail market, land regulation/planning differences and the slow adoption of information and communications technology (analytics) are the main factors behind low productivity in the UK retail sector. Also Griffith and Harmgart (2008) suggest that, in addition to regulation, it is necessary to control for the differences in the environment where stores operate as these would affect their fixed costs and their profitability.

Most of the research on BDA and productivity (or performance more generally) tends to focus on internal factors that prevent retailers from fully exploiting their investment in BDA (Wamba et al., 2017). However, research has paid scarce attention to features of the marketing environment where retailers operate that still have the potential of hampering their efficiency levels. This is rather puzzling as it has been argued that the impact of local effects outside the control of the manager should be taken into account when measuring retailers' efficiency (Kamakura, Lenartowicz, & Ratchford, 1996; Fritsch & Changoluisa, 2017) as different environmental conditions faced by retailers can lead to huge differences in technical efficiency, even for retailers with similar intrinsic characteristics (Banker & Morey, 1986). In this paper we focus on a specific aspect of the environment where retailers operate, namely the local availability of graduate workforce. This topic is very relevant for Britain where there is a substantial heterogeneity in the distribution of human capital across the country³ and where it is suggested that the availability of graduate workforce is one of the main constraints retailers face⁴. We suggest that conditions associated with environment (i.e. the availability of graduate workforce locally) are positively associated to positive efficiency (and eventually productivity) growth accrued from the upstream investment in BDA within the UK retail sector. Against this background, our purpose is to quantify the extent to which differences in the availability of graduate workforce across UK can help retailers to benefit from upstream investment in BDA and so help them improve their technical efficiency, experience technical progress and eventually improve productivity growth. Importantly, although the paper focuses on the two components of productivity growth rather than on financial performance, it is important to recall that a number of studies have shown that fast productivity growth is a pre-condition for fast financial performance (Blažková & Dvouletý, 2018) suggesting that both productivity and efficiency have a a major bearing on the financial performance of a firm.

³ The figures and the facts are well known: the average educational attainment in underperforming regions is below the average during the compulsory schooling years. These differentials are reinforced after graduation as graduates tend to relocate to regions with more productive jobs (Duranton and Monastiriotis, 2002; Office for National Statistics, 2019). For instance, in 2000, over 40% more graduates were employed in London than studied there (HM Treasury, 2001).

⁴ Miliken (2017) suggests that shortages of qualified workers in both managerial and sales occupations across the whole British retail sector create a major constraint to the retailers' operations.

The contribution of our paper to the literature is twofold. First, our study examines the influence of upstream investment in BDA on a focal retailer's efficiency growth and technical change. This is a major departure from the extant literature on Big Data as research so far has mostly focused on how intra-firm characteristics (such as cultural, organizational and infrastructural attributes of a firm) influence the capability of a business to benefit from its investment in BDA (Alharthi, Krotov, & Bowman, 2017). At the same time, this is an important research area on two counts: a) from a managerial standpoint it is important to quantify the benefits (in terms of efficiency and productivity growth) of upstream investment in BDA as it allows to modulate the recipient firms' own complementary investments in BDA and b) the literature has not paid much attention as to how characteristics of B2B relationships - such as supply chain partners' investment in BDA (i.e. upstream investment in BDA) and its associated outward spillovers - may influence a retailer's performance (Bradlow et al., 2017). This is an area where new research is needed as the existing research has not managed to find evidence of beneficial effects of the upstream investment in BDA onto downstream partners (Ren et al., 2016; Srinivasan and Swink, 2018).

Second, the previous studies on Big Data have predominantly focused on firm-level effects on the performance of BDA (e.g. Akter et al., 2016; Bradlow et al., 2017) and the role of turbulence in market environments in this process (e.g. Gunasekaran et al., 2018; Johnson, Friend, & Lee, 2017). Though prior research in Big Data has shown the importance of skills for the BDA successful deployment (Wang et al., 2019), no previous paper has focused on the impact that the local availability of skilled workforce has on the benefits that BDA may generate for downstream firms. Still, it is an important research area as the local availability of suitably skilled workforce may be an important factor in helping the successful deployment of BDA at firm level (Gaskell, 2018; Lam et al., 2017). To the best of our knowledge, there is a dearth of research on this topic particularly in the context of the retail sector. While a few empirical studies have found that spillovers generated by investment in technologies similar to BDA can have a positive impact on the TFP of the other non-adopter firms (Van Leeuwen & van der Wiel, 2003; Rincon & Vecchi, 2004), these studies have not considered the environment associated with the local availability of skilled or graduate workforce that the firm operates. In this sense, unlike these studies, our paper also focuses on the

characteristics of the business environment where retailers operate which may support the absorption of spillovers generated by the upstream investment in BDA.

For our empirical analysis we estimate a production frontier on a panel of retailers drawn from ORBIS and referring to the period 2009-2015. To estimate technical efficiency in our panel, we adopt the so-called “frontier” approach to the measurement of technical efficiency where (in)efficiency is computed as the distance from an estimated production frontier for the sector. To estimate the impact that the local availability of graduate workforce has on retailers’ efficiency, we use the one-stage approach to the estimation of the production frontier proposed by Battese and Coelli (1995) that allows to estimate simultaneously the parameters of the production frontier and the size of the impact of a set of observable variables on the efficiency scores’ mean. To capture the different channels of transmission of spillovers from BDA investment, we construct a proxy for these spillovers. The inter-industry effect is captured by means of a weighted BDA industry variable, where the weights capture the intensity of inter-firm transactions (using input-output coefficients of intermediate transactions) by using data from EU KLEMS. Therefore, we account for the lagged impact of the spillovers on efficiency by including various lags of the spillover variables in line with Brynjolfsson and Saunders (2009). This measure is interacted with the 2-digit industry-level share of the workforce with a university degree in a region (sourced from the Quarterly Labour Force Survey – QLFS) and then these are introduced among the determinants of technical efficiency to gauge the extent to which the local availability of graduate workforce can mediate the relationship between upstream investment in BDA and retailers’ efficiency growth.

The rest of the paper is structured as follows. In Section 2, we develop the theoretical background of the study and the hypotheses based on the literature on vertical knowledge spillovers of BDA and human capital. Section 3 includes a short survey on the measurement of technical efficiency in the retail sector as well as our empirical strategy. The main results are presented and discussed in Section 4. Finally, while some discussions are provided in Section 5, some concluding remarks are offered in Section 6.

2. Theoretical Background and Hypotheses

Our analysis is grounded in the theory of knowledge spillovers and their impact on the productivity (or performance more generally) of recipient firms. The starting point of such a theory is that investment activities of firms (i.e. source firms) may generate externalities from which other firms (so called recipient firms) can benefit from in terms of enhanced performance. It is customary to label these externalities as spillovers that are then classified as horizontal and vertical knowledge spillovers (Griliches, 1992). While beneficiaries of the horizontal knowledge spillovers are firms within the same industry the spillovers are generated from, vertical spillovers tend to benefit firms operating in different industries and positions along the supply chain (Chyi, Lai, & Liu, 2012; Ozdemir, Kandemir, & Eng, 2017; Rindfleisch & Moorman, 2003). Several authors have pointed out that vertical spillovers tend to provide novel insights that may trigger organizational innovations among recipient firms (Belberdos, Carree, & Lokshin, 2004; Rindfleisch & Moorman, 2003; Ozdemir et al., 2017). Vertical spillovers are particularly relevant in a supply chain setting: typically, firms involved in B2B relationships tend to belong to a variety of industries which do not share the same technology and therefore less reliant on horizontal spillovers. Indeed, previous research has shown that technological proximity between firms is not important for the absorption of spillovers along the supply chain (Isaksson, Simeth, & Seifert, 2016). In addition, firms that are connected through supply chains tend to be more willing to embed new knowledge acquired through their relationships with suppliers (from whatever industry) into their processes given the fact their performance is conditioned by their partners along the supply chain (Isaksson et al., 2016; Ozdemir et al., 2017).

While benefitting from vertical spillovers does not require co-location with the source firms (Dolfsma & Van Der Eijk, 2016; Isaksson et al., 2016; Rindfleisch, & Moorman, 2001; Ozdemir et al., 2017), the characteristics of the recipient firms' environment may condition their capability to benefit from the vertical spillovers. In the context of the topic of our paper (i.e. the local availability of graduate workforce), it is well known that access to more educated workforce has a positive impact on firm-level performance (whether productivity or efficiency): for instance more qualified workforce allows firms to implement new management practices effectively (Filatotchev et al., 2011; Fu, 2012); also it facilitates the introduction of process innovation that allow firms to benefit from vertical spillovers in a tangible way through improvements in productivity

(Destefanis and Sena, 2005; Higon and Sena, 2014). In summary, all this hints that the local availability of graduate workforce may play an important role in the context of the vertical spillovers from BDA although it is not clear yet whether this is the case (Akter et al., 2016; Bradlow et al., 2017). In the remainder of the section, we will flash out the theoretical arguments supporting our main proposition that the local availability of graduate workforce mediates the relationship between performance and vertical spillovers from upstream investment in BDA.

2.1. Upstream Investment in BDA and the Role of Human Capital

Big Data (or large expanses of data routinely collected by firms through their operations) are typically considered to be an important source of competitive advantage at firm level (Germann et al., 2014) and therefore investments in Information Technology (henceforth, IT) systems capable of exploiting Big Data have dominated the business investment in Information and Communication Technology (henceforth, ICT).

While most of the literature has traditionally focused on knowledge spillovers generated by investment in R&D and ICT⁵, the evidence on spillovers from investment in BDA is still in its infancy. In general, BDA can be considered a “General Purpose Technology” (Zhang et al., 2016), i.e. a technology which can be widely used across different sectors and which can create the conditions for continuous improvement of production techniques (and therefore of productivity) among its adopters when combined with complementary investments such as new work systems, organizational redesign and so on. According to much theoretical and empirical evidence, investment in BDA offers direct benefits for a wide range of business processes; at the same time, upstream investment in BDA can help downstream firms to develop a set of capabilities allowing to process, analyze and interpret large volumes of data that are critical for their retail operations (Grewal, Roggeveen, & Nordfält, 2017; Wang et al., 2019). As a result,

⁵ Stiroh (2002) regresses TFP growth on ICT capital and other controls for 20 US manufacturing industries over the period 1984-1999. Instead he finds that ICT had a negative impact on TFP growth. Haskel and Wallis (2010), using aggregate data for the UK, find no evidence of spillovers from software assets, nor from other intangible assets such as economic competencies and R&D. Similarly, Acharya and Basu (2010) fail to find positive ICT spillovers in a industry-level analysis for 16 OECD countries, but they do find significant spillover effects from domestic and foreign R&D investment. Van Leeuwen and van der Wiel (2003), using a sample of Dutch companies operating in market services, find a positive and significant ICT spillover on labour productivity. Similarly, Severgnini (2010) finds evidence of positive ICT spillovers in a sample of Italian manufacturing firms.

firms can manage their processes more efficiently and, as a consequence, they can increase their technical efficiency.

At the same time, the benefits (in terms of productivity gains) of the investment in BDA can be propagated well beyond its adopters. Upstream adoption of BDA may increase efficiency of transactions among downstream firms. Electronic transfer of payments and invoices, automated inventory replenishment, on-line markets for placing and receiving orders may prompt downstream firms to re-organise their production processes and structures to take advantage of the analytics-induced increased efficiency in running transactions⁶ (Wang et al., 2016). As analytics-based systems for upstream supply chain management are becoming more common, retailers may benefit from their suppliers' investment in BDA in terms of increased productivity. For instance, Rowlett (2001) suggests that in the context of computer networks, their introduction has facilitated the exchange of information between retailers and suppliers which in turn has improved the supply chain management. In this context, the investment in BDA among suppliers may prompt downstream firms to gradually re-organise their production processes to take advantage of the upstream firms' increased efficiency in running transactions which may lead to a reduction of the operational costs relating to information search, inventory management and order processing practices, and so on (Crisuolo & Waldron, 2003; Rincon, Vecchi, & Venturini, 2013). This implies that adoption of a BDA by an individual firm may have a knock-on effect on its B2B customers. In other words, investment in BDA may be beneficial to not only to the investing firms but also all other firms operating in the sector.

Over time, upstream investment in BDA can also enable retailers to develop their own BDA capabilities that may eventually lead to an improvement in technical efficiency. However, internalizing and exploiting BDA related insights from upstream firms requires a firm to progressively absorb new knowledge as it unlearns or relearns the existing knowledge through experience (Fiol & Lyles, 1985). In other words, improvements in technical efficiency will emerge gradually as downstream retailers gradually absorb new knowledge through vertical knowledge spillovers. Therefore, we

⁶ For instance, upstream investment in real-time BDA can motivate retailers to set up flexible arrangements in their service offerings and operational processes to meet the changing market requirements, and match fluctuating supply and demand (Erevelles et al., 2016; Wang et al., 2016).

suggest that:

H1: Upstream investment in BDA has a positive impact on retailers' technical efficiency with a lag.

In spite of all the evidence on the impact of BDA investment on the performance of firms (Wang et al., 2019; Wang et al., 2016), less clear are the factors that can facilitate the absorption of upstream investment in BDA among downstream retailers. The hypothesis that the impact of the upstream investment in ICT on the focal firm depends on a number of additional factors has already been investigated in relation to human capital (Griffith et al., 2004; Vandenbussche, Aghion, & Meghir, 2006). The literature on human capital has shown that access to highly educated workforce is an important condition for firms to benefit from the investment in new technologies (Haller & Siedschlag, 2011; Hsu & Wang, 2010). For instance, previous empirical evidence has suggested that investment in ICT can only generate a positive return to the adopter if it is accompanied by a complementary investment in skilled workforce (Bugamelli & Pagano, 2004; Díaz-Chao, Sainz-González, & Torrent-Sellens, 2015). More importantly, this is true not only for the firms that directly invest in the new technology but also for firms that are exposed to the knowledge spillovers from the upstream firms in the supply chain. For example, graduate workforce is also more likely to assimilate new knowledge from suppliers and use it for commercial ends thanks to their skills and expertise (Cohen and Levinthal, 1990; Subramaniam and Youndt, 2005). This argument translates very easily to the BDA context as workforce with high qualifications are in a good position to learn from the deployment of BDA upstream (Dahl and Pedersen, 2005). For example, the re-organisation of the production process within firms, fostered by computerization, can be considered as the result of learning-by-doing. This learning does not necessarily happen in isolation as workers learn from the experience of other workers (Aghion, 2002). In a related context, Alexander et al. (2005) found that the B2B interactions of supermarkets' employees with equipment suppliers were one of the major sources of knowledge transfer in supermarkets in the 50s.

If this is the case, then, it is possible to argue that features of the local environment which can favor informal exchanges of new information among workers across

different firms should have an impact on the retailers' capabilities of absorbing and benefitting from also upstream investment in BDA and presumably can contribute to explain differences in performance among retailers in different areas even if they share similar characteristics. We focus here on the availability of graduate workforce. From a theoretical standpoint, access to more graduate workforce may have a positive impact on technical efficiency. For instance, more graduate workers can address the requirements of deploying BDA accruing from upstream investment in BDA in a more effective way, and this can enhance a retailer's technical efficiency.

Importantly, there are additional benefits that can accrue to a store by being located close to graduate workforce. Larger density of graduate human capital can foster informal B2B interactions among employees from different firms that in turn may help the sharing of new information, facilitation of informal knowledge transfer across firms, and therefore improvement of operations efficiency and productivity. For instance, the local availability of graduate workforce may affect the productivity gains of local retailers through employee turnovers and B2B interactions among employees of surrounding firms who either have the skills to absorb the new knowledge or are equipped with complementary knowledge that can facilitate the translation of the knowledge spillovers into increasing efficiency once they are embedded into the recipient firms' processes.

This could be particularly relevant for the retail sector that is usually characterized by high labour turnover. It well may be that for retailers which are located in areas with a large density of graduate and skilled workforce employed by other firms and that therefore have access to a better pool of workers, high turnover would facilitate the diffusion of spillovers associated to the investment of BDA and so allow the spread across the sector of organizational changes which can be a pre-condition for increases in total factor productivity and technical efficiency (Agarwal, Ganco & Ziedonis, 2009)⁷. This argument can be extended to the retail sector. Although it has never been formally tested before, this can be of some relevance both for the investment in BDA (as a portion of the benefits from upstream investment in BDA may spread across thanks to direct contacts among employees) and for the retail sector. Retailers operating in local labour markets with greater availability of graduate workforce are likely to

⁷ See Correidora and Rosenkopf (2010) on a similar point.

access workforce with the requisite skills for the absorption of BDA-related knowledge spillovers and ultimately to benefit from the upstream investment in BDA (Gambardella & Giarratana, 2010; Motohashi & Yuan, 2010). Therefore, we suggest that the benefits of upstream investment in BDA on technical efficiency are conditional on the availability of graduate workforce in the local area. In other words:

H2: Upstream investment in BDA has a positive impact on retailers' technical efficiency conditional on the presence of graduate workforce working in retail in the region.

Technical change in an industry is dependent on the diffusion of innovative technologies within the industry which in turn is linked to imitation and network effects (MacVaugh & Schiavone, 2010; Rogers, 2003). For example, Dos Santos and Peffers (1998) show that in the case of electronic commerce, imitation among industry competitors eventually enhanced the technical change. In the context of cross industry ICT spillovers, network effects have also been found to be highly influential on the adoption of a new technology across the industry (Chou, Chuang, & Shao, 2014). The network effect is thus expected to be important within the supply chain in the case of BDA technology adoption and its subsequent influence on technical progress. For instance, retailers with larger supply chains are more likely to adopt BDA themselves following the indirect exposure to benefits associated to the upstream investment in BDA. As an example, consider that the use of BDA can improve assortment planning, anticipation of demand and logistics planning (Aktas and Meng, 2017). In this context, retailers which may not take advantage of upstream investment in BDA may fall behind competitors. Thus, we expect that spillovers from suppliers induce downstream firms such as retailers to adopt more advanced technologies including BDA, and can generate technical change across the industry (Luo, 2007). As a result, we suggest that:

H3: Upstream investment in BDA has a positive impact on retailers' technical progress.

3. The Empirical Strategy and Measurement

3.1. Measurement of Efficiency in Retail Sector

Measuring technical efficiency of the retail sector along with its determinants has become an important research area amid growing concerns about the sector's slow-growing productivity (Barros & Alves, 2004; Lusch, Serpkenci, & Orvis, 1995; Kamakura Lenartowicz, & Ratchford, 1996). Traditionally input-output ratios were used to measure the performance of stores (Vyt, 2008) – although the main limitation of the technique is that the presence of economies of scale cannot be tested. A cursory look at the recent literature in the area shows that frontier techniques have become a very common tool to measure technical efficiency in the sector. In this approach, technical efficiency is computed as the distance from an estimated optimal benchmark (the “production frontier”) which defines the optimal amount of output that can be produced in a sector given the available technology. The closer a firm is to the frontier, the more efficient it is (Farrell, 1957). By estimating an ‘efficient frontier’, a variety of issues such as measuring retailing outlet productivity (Grewal et al. 1999; Ratchford & Brown 1985; Ratchford & Stoops 1988), market efficiency and consumer welfare loss (Donthu & Yoo 1998; Ratchford et al. 1996), sales force efficiency (Boles, Donthu, & Lohtia, 1995; Horsky & Nelson 1996), channel productivity (Bultez & Parsons 1998), and resource allocation (Chebat et al. 1994; Donthu, Hershberger, & Osmonbekov 2005; Luo & Donthu 2005) can be studied.

Two main methodologies have been suggested to estimate the efficient frontier and to compute the distance of the firms from the frontier: a) parametric methods that use econometric analysis to estimate the production frontier and b) non-parametric methods that use linear programming methods to compute the efficient frontier. Both these methods produce a frontier consisting of the set of most efficient performers, thus allowing a comparison to the best performers rather than one to the average performer (as in regression based models).

Previous research measuring the efficiency of the retail sector have employed non-parametric methods to estimate efficiency. For instance, Athanassopoulos and Ballantine (1995) use DEA to compare the efficiency of supermarket chains operating in the United Kingdom. Equally Ket and Chu (2003) assess the performance of 13 supermarkets for 10 years, using a 3 stages approach modelling the transformation of inputs into intermediate outputs and then final outputs. Korhonen and Syrjänen (2004) assess the performance of a supermarket chain in Finland to reallocate the resources

available to the chain among the stores. Sellers-Rubio and Mas-Ruiz (2006) use DEA to estimate the economic efficiency of the supermarket chains in Spain between 1995 and 2001. Barros (2006) analyzes 22 Portuguese hypermarkets and supermarkets using a two-stage procedure where in the first stage DEA is used to compute technical efficiency scores and in the second stage a Tobit model is employed to estimate the determinants of efficiency. Among the other results, he finds that large retail groups are more efficient than small retailers and that market share, number of outlets and location are important determinants of efficiency. Some studies focus on technical efficiency change and its contribution to productivity growth. For instance, Barros and Alves (2004) and Sellers-Rubio and Mas-Ruiz (2007) have computed a Malmquist index and its components to measure technical efficiency change of Portuguese and Spanish supermarket chains, respectively. Barros and Alves (2004) have used the Malmquist index to decompose productivity growth of 47 stores of a Portuguese supermarket chain into pure technical efficiency change; scale efficiency change; and technology change. Wu, Yang and Liang (2006) used a CCR DEA model to examine the performance of the retailing industry in Taiwan between 1998 and 2002. They found that nearly three quarters of the companies analyzed were inefficient during the five-year time span. Moreover, there were only six consistently efficient companies throughout the period under analysis. Perrigot and Barros (2008) analyzed the technical efficiency of the retail industry in France using DEA.

In terms of choice of inputs and outputs of the retail sector, most papers have used profits, revenues and value added as measures of output. Sales are probably the most common output variable (Kamakura et al., 1996; Donthu and Yoo, 1998; Athanassopoulos, 2003; Perrigot and Barros, 2008; Sellers-Rubio and Mas-Ruiz, 2009) while Sellers- Rubio and Mas-Ruiz (2007) select EBITA and Net income as measures of output. As for the input variables, labour is usually proxied by a number of variables associated to the number of employees ranging from the number of full-time equivalent employees (Reardon et al., 1996), the number of man-hours allocated (Kamakura et al., 1996) to the number of full-time or part-time employees (Barros, 2006; Perrigot & Barros, 2008). Some authors have used payroll as a measure of inputs (see for instance Reardon et al. (1996), Thomas et al. (1998)). As for capital, the selling space is often used as a proxy of capital (Kamakura et al. 1996); alternative proxies of capital include investments on assets (Sellers-Rubio and Mas-Ruiz, 2007); assets value (Perrigot and

Barros, 2008; Sellers-Rubio and Mas-Ruiz, 2006; Barros, 2006) and rent costs (Thomas et al., 1998).

As for the determinants of technical efficiency, limited attention has been given by researchers to the impact of environmental variables on technical efficiency (Donthu & Yoo, 1998; Vyt, 2008). There are though a few remarkable exceptions. De Jorge-Moreno (2008) analyses the impact on Spanish hypermarkets' technical efficiency of the devolution of retail planning and regulation to regional governments. Assaf, Barros and Sellers-Rubio, (2011) analyzes the impact of vertical integration, age, geographic expansion, and low price strategy on the cost efficiency of retail stores. The data involve a sample of Spanish retail stores that operate in a highly competitive and dynamic environment. From the results, it is clear that cost efficiency is higher for stores that have: longer years in business, stronger geographical presence, and lower price offerings. Vertical integration, on the other hand, is negatively related to efficiency.

Gauri (2013) measures and compares the inefficiencies of major grocery retailers across various formats and pricing strategies using stochastic frontier a dataset covering 2500 stores across 50 US chains and finds that stores could potentially reduce the proportion of inputs such as selling area, number of checkout counters, number of employees and store features without threatening outcomes (i.e., by holding the output level constant). Xavier, Moutinho and Moreira (2015) measures the technical efficiency of 40 Portuguese clothing stores and finds a positive impact of inventory stocks and the shopping location store on technical efficiency, highlighting the importance of having more merchandise for the customers to select from. Furthermore, the level of staff education also has a positive effect on technical efficiency. Finally, Yu and Ramanathan (2008) have estimated the impact that ownership, legal form and other retail characteristics have on the economic efficiency of 41 UK-based retail chains. However, it is safe to conclude that virtually no study has been carried out which tries to ascertain whether the characteristics of the retailers' external environment, particularly accruing from availability of local graduate workforce, may facilitate the absorption of upstream investment in BDA.

3.2. The Production Frontier Specification, the Datasets and the Variables

The format of our empirical analysis is straightforward enough. First, we apply parametric frontier techniques to compute the technical efficiency indexes for a panel of retailers in between 2009 and 2015. Second, we condition the mean of the technical efficiency scores' distribution on the inter-industry upstream investment in BDA weighted by the coefficients of the input-output table, the regional share of graduate workforce in retail and their interaction so to quantify the impact that the upstream investment in BDA have on the mean of technical inefficiency through the local availability of graduate workforce.

3.2.1. *The Production Frontier Specification*

Unlike previous papers in this area we use parametric methods to estimate the production frontier for the retail sector. The main advantage of the parametric methods is that unlike deterministic methods (like DEA) they allow to specify a stochastic term so making the estimates less vulnerable to the influence of random events and measurement errors than non-parametric models (Kumbhakar & Lovell, 2000). Estimating the levels of inefficiency within the retail sector is only the first step of our analysis as we are really interested in quantifying the impact that the local availability of graduate workforce has on the distribution of the efficiency scores in the sector. To this purpose, a few production frontier models are available where environmental variables are incorporated directly into the inefficiency components model which is then estimated simultaneously with the parameters of the production frontier. In our study we use the approach suggested by Battese and Coelli (1995) where the mean of the inefficiency component is assumed to be a function of a set of observable variables. Therefore, the production frontier is specified as:

$$y_{kt} = f(x_{kt}, t) + \varepsilon_{kt} \quad (1)$$

where y is the output of the retailer k at time t , x is a set of inputs that affects the shape of the production technology at time t and t is a time trend that measures disembodied technical progress. The error term has two components: u_{kt} and v_{kt} . v_{kt} is the independent and identically distributed random error, assumed to be distributed as a $N(0, \sigma_v^2)$ and to be independent of the u_k . In turn, this is a non-negative random variable assumed to account for technical inefficiency in the production and to be independently distributed as a truncation at zero of the $N(z_{kt}\delta, \sigma_u^2)$ distribution where z_{kt} (or technical

efficiency effects) is a vector of observable variables which may influence the mean of the efficiency scores. Both the parameters of the production frontier and of the technical inefficiency specification are estimated simultaneously by using Maximum Likelihood. The technical efficiency score of a retailer is defined as the ratio between the actual level of output and the maximum level of output that it could be produced given the technology specified by the production frontier. The efficiency score varies between one and zero, where the lower the score, the more inefficient the corresponding retailer is. Various null hypotheses on the parameters of the production frontier and of the inefficiency model can be tested using the Likelihood-Ratio (LR) test, defined as:

$$LR = -2[L(H_0) - L(H_1)] \quad (2)$$

where $L(H_0)$ is the log-likelihood value of the frontier model, as specified by a null hypothesis, H_0 , and $L(H_1)$ is the log-likelihood value of the general frontier model under the alternative hypothesis, H_1 . This test statistic has a chi square distribution where the degrees of freedom are equal to the difference between the number of parameters under the null and the alternative hypotheses.

A central tenet of the frontier approach is that the units under analysis should be homogeneous in terms of the technology they use so to get meaningful frontier estimates. This is particularly relevant for our analysis as the Retail sector includes seven sub-sectors that can be assumed to use technologies with different capital-labour ratios. To control for this source of heterogeneity we add to our basic production frontier specification a set of interactions between the input variables and the dummy variables for each sub-sector. So our frontier specification (assuming a translog functional form) is the following:

$$\ln(y_{k,t}) = \beta_0 + \sum_i^I \beta_i \ln(x_{k,t}) + 0.5 \sum_i^I \sum_j^J \beta_{ij} \ln(x_{i,k,t}) \ln(x_{j,k,t})$$

$$k=1, \dots, K \quad t=1, \dots, T \quad i=1, \dots, I \quad j=1, \dots, J \quad (3)$$

where $\ln(y_{kt})$ is the log of production of the k -th retailer at time t , $\ln(x_{kt})$ is the vector of the log of inputs (capital and labour, respectively) of the k -th retailers at time t and β is a vector of unknown parameters. We allow for the possibility of technical change by introducing in the empirical specification a continuous time trend and its square in the specification. The model for the mean of the technical efficiency effects in the stochastic frontier model is:

$$\mu_u = \alpha_0 + \alpha_1 share_{c,t-1} + \alpha_2 spill_{s,t-1} + \alpha_3 share_{c,t-1} spill_{s,t-1} \quad (4)$$

where α_0 is a constant term. From (4), we can examine the impact of an increase in the upstream investment in BDA on the retailers' mean technical efficiency conditional on the availability of graduate workforce in a region. Afterwards we introduce our variables of interest: the inter-industry measures of upstream investment in BDA (*spill*, measured at the 2-digit industry level – denoted by s), the share of graduate workforce in retail (*share*, measured at regional level – denoted by c) and their interaction. Our expectation is that retailers located in regions with a large availability of graduate workforce will benefit more from upstream investment in BDA and therefore will experience higher technical efficiency. Importantly, the two variables are lagged one period: this has been done to avoid potential endogeneity problems in the regression model and to allow for the possibility that the impact of spillovers is delayed as focal retailers may need to change structures and processes before they can fully benefit from the upstream investment (Brynjolfsson and Hitt, 2003). However, when estimating the efficiency model we also test alternative lag structure for our two variables of interest.

We also control for the retailers' location by introducing a set of regional dummies. These variables are traditionally introduced to capture the impact that additional environmental conditions have on the firms' level of production. Betancourt (2004) suggests that local factors such as infrastructure endowment, level of income, external economies linked to the accessibility of location and so on have an impact on the retailers' level of output. (3) and (4) can then be estimated simultaneously by Maximum Likelihood using the procedure suggested by Battese and Coelli (1995).

3.2.2. The Data-Sets and Some Descriptive Statistics

Our empirical analysis is carried out on a data-set of 48 retailers (NACE REV 2: 47) sourced from ORBIS and covering the period 2009-2015. Spillovers from upstream investment in BDA have been calculated using data from EU KLEMS. The EU KLEMS database collects data on capital services and investment for a number of assets including computer hardware, telecoms equipment, computer software and databases. As mentioned by OECD (2013), ICT investment and the use of analytics for the exploitation of Big Data go hand in hand: indeed, most data-intensive sectors that have invested in the development of infrastructure and systems to improve their analytics capabilities tend to be ICT-intensive as well since they have invested in smart ICT applications so to be able to collect and exploit their data holdings. For our analysis, we focus on investment in computer software and databases that are typically associated to the development of capabilities in analytics. The share of graduate workforce employed by the retail sector in a region has been computed by using information drawn from the Quarterly Labour Force Survey (QLFS), 2009-2015. In constructing our final data set, retailers that did not disclose any data for net sales, employment or net physical capital were excluded from the estimation, as were those companies displaying negative values. We restrict the sample to firms for which there are at least two consecutive years of data available. We have taken care of removing missing observations as well as 1% of the observations on both the upper and lower tails of the data distribution. Finally, all the data have been weighted to make them representative of the whole population. We use value added as a measure of firm output that is equal to turnover net of material costs; as for inputs, we use the number of full-time equivalent employees while capital input is measured as total tangible assets by book value, recorded annually following Sellers-Rubio and Mas-Ruiz (2007). A full description of the datasets and of how the variables have been constructed is in the Data Appendix. Table 1 shows the mean and the standard deviation of the inputs and the output in our sample. Table 2 reports the investment in BDA by industry and year as sourced from EU KLEMS.

4. The Empirical Results

The Maximum Likelihood estimates of our production frontier are reported in Table 3. The result of the Likelihood Ratio test (reported at the bottom of the table) indicates that the frontier model is a significant improvement over the standard average

production function. In the specification of the production frontier, we control for the sub-sectors that belong to the retail industry⁸ through a set of dummy variables; in addition we also allow for the possibility that each sub-sector uses capital and labour in different proportions by interacting the dummy variables for each sub-sector with capital and labour. We have tested whether these two sets of dummy are significant and the LR tests suggest they are not implying that the all retailers (independently of the sub-sector they belong to) tend to use similar proportions of capital and labour.

The estimates are generally statistically different from zero at the 5% level. We have mean-corrected our sample data before the estimation so that we can interpret the 1st order coefficients listed in Table 3 as the partial output elasticities evaluated at the sample mean. The figures show that the output elasticity associated to labour is smaller than the equivalent elasticity for capital suggesting that across the sector output may be less responsive to increases of labour than to increases of capital. In this case a 10% increase of the stock of capital implies a 34% increase in value added. This is consistent with the fact that retailers in Britain are under-capitalized due to regulatory restrictions and therefore the increase in added value following an increase in the stock of capital can be rather substantial (Haskel and Sadun, 2008). Returns to scale are not constant across the sector. The estimates of the production frontier show there is no significant neutral technical change across the sector.

The estimated coefficients of the z 's variables are listed at the end of Table 3. As mentioned above the z 's variables are all lagged one period. This is in line with the empirical literature on ICT suggesting the adoption of ICT is followed by a period of learning where firms learn how to profitably exploit the investment in ICT (Brynjolfsson and Hitt, 2003). As a result, we would expect the effect of the spillover variable on technical efficiency to be delayed and in our case we use a one-year lag to capture this delayed effect. Recall that a positive coefficient associated to one of these variables implies that the variable has a negative impact on the mean of the retailers' technical efficiency. From these results, we can see that our variables of interest are all significant. In terms of the interpretation of the coefficients, the parameters measuring the direct effect of the upstream investment in BDA and the share of graduate workforce

⁸ These include sub-sector 471, 472, 473, 474, 475, 476, 477, 478, 479.

do not have a meaningful interpretation in our context as the two variables never take the value of zero. Indeed, the partial derivative of technical inefficiency with respect to the upstream investment in BDA depends on the values of the share of graduate workforce in the county. To understand how its changes affect the partial derivative of efficiency with respect to the spillovers variable, consider the following two cases. For retailers located in regions where on average 80% of sector's workforce is made of graduates, an increase of 1% of the upstream investment in BDA generated in the area will imply a 0.38% increase of technical efficiency. On the contrary, for retailers located in regions where on average only 1% of the sector's workforce is made of graduates, an increase of 1% of the local upstream investment in BDA will improve the mean technical efficiency by 0.35%. Although the size of the impact on the average technical efficiency of an increase in upstream investment in BDA is not very large, the difference of the values of the two partial derivatives is though significant. In other words that upstream investment in BDA increase downstream retailers' technical efficiency on average if these are located in a region where a larger share of graduate workforce is employed in the retail sector.

Several hypothesis tests have been carried out on the estimated production frontiers. Table 4 summarizes the main results. We have tested whether the sector experiences technical progress; in this case the null hypothesis is that the time trend and its squared value are simultaneously equal to zero. The results from the test suggest that the null hypothesis has to be rejected. We have also tested whether each sector has a translog production function against the alternative Cobb-Douglas. In this case the null hypothesis is $\beta_{ij} = 0$. In this case the null hypothesis is rejected. Also we have tested whether the explanatory variables in our technical efficiency effects model are significant. For this test, the null hypothesis is formulated as $H_0 : \delta_z = 0$, where $z = 0, 1, 2, \dots, 3$; also in this case the null hypothesis is strongly rejected by the data. Finally, we have tested whether the regional dummies are significant in the inefficiency model and the results suggest that they are (i.e. the null hypothesis that they are equal to zero can be rejected).

In the model presented above, we have assumed that the spillovers from the upstream investment in BDA are only associated to the mean technical efficiency but do not have

any bearings on the movements of the frontier over time. Of course, whether this is really the case should be tested statistically. Therefore, we decide to estimate a new specification of the production frontier where the spillover variable is interacted with the time trend. The new estimates of the production frontier are presented in Table 5 (for the spillover variable). The estimates of the efficiency models are consistent with the main specification presented above although the output elasticities of labour and capital differ from those obtained from the main specification.

Whether each variable has any bearing on the movements of the frontier over time has been tested separately and in both cases the null hypotheses have been rejected. In the first case, the formal significance test is a LR test whose value is presented at the bottom of table 5. The model presented in Table 5 has been compared to a specification of the production frontier where there is no neutral technical change (both in level and squared) and the spillover variable (both in level and interacted with the trend) is not included in the model. The result of the test show that the null hypothesis has to be rejected suggesting that the variable of interest affects the position of the frontier over time. In other words, spillovers from upstream BDA investment can shift the position of the frontier over time.

4.1. Technical Efficiency Scores

Based on the estimates of the production frontier reported in Table 3, we can then compute the technical efficiency scores. Table 6 shows the average yearly technical efficiency scores computed using the formula suggested by Battese and Coelli (1995). There are two important issues to notice. First, the average efficiency scores tend to be rather mid-range across the sector. On average, these vary between 0.59 and 0.99. In other words, the worst performers could have produced the same amount of output using 41% less inputs while the best performers could have produced the same amount of output using 1% less inputs. Figure 1 shows the smoothed estimate of the probability density function of the efficiency scores across all the years in the sample and confirms that the bulk of the distribution is between 0.4 and 0.9 suggesting that there is a number of retailers who are not particularly good at combining capital and labour in an efficient way. This is an important result as it shows that some of the poor performance of the sector in terms of productivity can be attributed to the slow growth of technical efficiency. Second, the distribution of the scores shows there is a large difference

between the best performer and the worst performer every year. For instance in 2013, the efficiency score of the best performer was 0.93 while the score of the worst performer was 0.010. In other words, there is a number of poor performers in the sector. In addition the composition of the lower tail is rather persistent over time. Indeed further tests have shown that retailers that are therefore tend to stay in the tail throughout our sample suggesting they find difficult to readjust the mix of capital and labour their use in their production process.

4.2. Additional Tests

In this section, we also test whether the mechanism we have identified for the transmission of spillovers from upstream investments in BDA applies to workforce with alternative qualifications i.e. A-level or vocational specifications. In the main analysis, we have focused on a specific transmission mechanism that relies on the local availability of workforce with a graduate qualification in the retail industry. In this part of the paper, we test the possibility that alternative - such as A-levels and vocational qualifications (like BTEC or NVQ qualifications) – may be relevant as well. Therefore, we test whether there is a relationship between technical efficiency, spillovers from upstream investment in BDA and local share of workforce with A-levels (or vocational qualification). The results of the new efficiency models are shown in Table 7 and suggest that the presence of local workforce with A-levels or with vocational qualifications does facilitate the transmission of BDA spillovers to technical efficiency. The variables of interest are significant in both models. As for the direction of the impact of the two sets of variables on technical efficiency, we conduct the same experiment with the coefficients as we did with the main specification. Therefore, we start by assuming there are regions where on average 80% of the retail workforce is made of workers with vocational (A-level) qualifications. In this case, an increase of 1% of the upstream investment in BDA generated in the area will imply a 0.37% (1.9% in the case of A-level qualifications) increase of technical efficiency. On the contrary, for retailers located in regions where on average only 1% of the sector's workforce has a vocational (A-level) qualification, an increase of 1% of the local upstream investment in BDA will decrease the mean technical efficiency by 0.13% (0.20% in the case of the A-level qualification). In other words, there exists a non-linearity in the relationship between spillovers from upstream investment in BDA and technical efficiency which is linked to the qualifications of the local workforce available to the sector. Once the

share of the retail workforce with qualifications (of either type) increases, then the negative impact of the spillovers on average technical efficiency will be reversed suggesting that potential agglomeration effect may “kick in” once the threshold share is reached⁹. In another set of robustness tests, we test whether the main results are robust to the introduction of time-varying regional variables like the log of the regional GVA and the log of the regional consumption expenditure. The new results (not reported here) show that the two new control variables are not significant suggesting that the main findings still hold. Finally, we test whether alternative lag structure of the two variables of interest are significant. More specifically, we introduce we test whether the regional share of the graduate retail workforce in level and its interaction with the spillover variables is significant in the efficiency model. Our results (not shown here) indicate this is not the case. We also test whether the main result is robust to the introduction of the spillover variable in level: the main result still holds in this case.

5. Discussions

5.1. Theoretical Implications

What conclusions can we draw from these results? First, our findings confirm that upstream investment in BDA has a positive impact on retailers’ technical efficiency with a lag. This finding is consistent with the previous empirical evidence showing how spillovers from other organizations can improve the technical efficiency of a focal firm (Fritsch & Slavtchev, 2011). Our findings are in line with the B2B marketing research observing that upstream adoption of innovations induces downstream firms to search for new skills (Siamagka et al., 2015) so that they can progressively absorb new knowledge generated upstream (Fiol & Lyles, 1985) and invest in complementary technologies that can improve their performance (Rindfleisch & Moorman, 2003; Belberdos et al., 2004). Importantly, it is not only the technical efficiency component of productivity growth that improves; indeed, our results show that spillovers from upstream investment in BDA are associated to positive technical change as it possibly induces focal firms to adopt more advanced technologies to run their operations. In the context of cross industry ICT spillovers, previous research has shown that the adoption

⁹ The threshold share is 9% for A-level qualifications and 35% for vocational qualifications.

of a new technology across industries is facilitated by the network effects (Chou et al., 2014) and this effect can explain our own finding in the context of technical change.

In the context of the BDA literature, these findings are quite important: previous research in the area has mostly focused on the impact of intra-firm characteristics on the BDA deployment using the resource-based view or dynamic capability approach (Wamba et al., 2017; Erevelles et al., 2016); however our paper suggests that BDA capabilities embedded within other firms may have an impact on downstream performance through B2B connections. In addition, we have shown that the quality of the pool of workers a retailer has access to matters for its performance in terms of productivity growth. This finding is consistent with the existing literature on BDA investment suggesting that access to qualified workforce is key to the successful exploitation of Big Data technologies (Bughin et al., 2018). In particular, the finding of our study suggests that access to local graduate workforce is a pre-condition for the development of the human capital capability which allows to benefit from upstream investment in BDA (Kang & Snell, 2009). In this sense, our study is consistent with the theory of localized spillovers, which emphasize the role of geographical density of human capital in the absorption of spillovers by firms (Jaffe, 1989; Audretsch & Feldman, 1996). Such environmental influences associated with access to qualified workforce have not received sufficient attention in the previous empirical work on B2B relationships as studies have rather focused on the effect of supply chain partners and relationships in exploiting the benefits of BDA (Wang et al., 2016). The results of this study have increased our understanding of how upstream investment in BDA can improve firms' performance (Akter et al., 2016; Bradlow et al., 2017). Our results show that the exploitation of the upstream investment in technology, and in particular BDA infrastructure and methodologies, is predominantly contingent on the skills available in the local area.

5.2. Implications for Practice

We have identified an additional channel through which the distribution of graduate workforce across a country matters. In other words, the downstream appropriation of benefits from the upstream investment in BDA is conditioned by the existence of local graduate workforce (and not necessarily equipped with analytics skills). The implication is that retailers need to consider the availability of graduate local workforce

in a local area when deciding where to locate centers that manage their supply chain. In this context, if firms operate in areas in which graduate workforce is scarce, it would be important to develop strategies to employ workforce outside the local area. Furthermore, it would be essential for firms to develop strategies to promote sharing of good practice from other firms. Particularly, organizing collaborative workshops and conferences with suppliers can support information exchange between employees with complementary knowledge and skills. Similarly, upstream and downstream firms can develop collaborative product and service strategies through sharing of Big Data with the goal of improving their performance.

Last but not the least, retailers need to be aware of the fact that upstream investment in BDA can improve their own productivity growth. This opens up many opportunities to retail managers who may be interested in supporting the development of the BDA infrastructure and capabilities through supplier relationships. For example, managers can encourage suppliers to invest in BDA by developing innovative strategies for sharing Big Data with them. This would, in turn, help retailers to improve their service provision, inventory management and logistics.

5.3. Limitations and Future Research Directions

Our research has focused only on upstream investment in BDA. In other words, we have not analyzed how both horizontal and downstream investment in BDA affect the performance of retailers. As such, future research can investigate how investment in BDA by different types of firms in the supply chain, which may have direct or indirect B2B relationships with a retailer, may influence its performance. For instance, previous studies in B2B marketing suggests that while horizontally positioned firms may provide more supplementary knowledge and insights, vertically connected firms such as the ones operating at diverse positions of the value chain can provide more complementary knowledge and skills to one another (Rindfleisch & Moorman, 2003). This may have implications on the type of BDA-related insights and skills that can be gained from these firms, and their eventual influence on the performance of firms.

In addition, it can be argued that internal factors mediate the relationship between upstream investment in BDA, the availability of the graduate local workforce and the performance of focal firms. One example is the learning process in the workplace: for

instance, it can be argued that formal learning and informal learning may mediate the relationship among our variables of interest differently (Enos, Kehrhahn, & Bell, 2003). Similarly, future research can examine potential moderating effect of local institutions, such as local governments' support to universities and research institutions in the promotion of BDA skills and knowledge in their geographical area. Likewise, future studies could examine whether retailers' types of market-oriented behaviors - including proactive or responsive market orientations - may influence the relationship between technical efficiency and local availability of graduate workforce. Finally, future research can also consider the impact of the strength of relational ties among local retailers on their performance and how relational ties mediate the relationship between upstream investment in BDA, local access to human capital and performance. For example, Granovetter's (1973) strength of weak ties theory can be tested to see whether retailers perform better when their employees have weaker or stronger relational ties with the upstream investors in BDA.

6. Conclusions

This paper has analyzed the impact of upstream investment in BDA on the technical efficiency and technical progress of British retailers. It has also examined the extent to which retailers located in regions with greater graduate workforce can benefit more from upstream investment in BDA than retail firms located in regions with fewer graduate workforce. Our empirical analysis shows the positive effect of upstream investment in BDA in the technical efficiency of retailers over time as well as in their technical progress. We also find that retailers located in regions with a larger proportion of graduate workforce tend to benefit more (in terms of technical efficiency growth) from inter-industry upstream investment in BDA. Additional tests show these results are robust to different specifications of the production technology; in addition, we find there exists a non-linearity in the relationship between spillovers from upstream investment in BDA and technical efficiency which is linked to the qualifications of the local workforce available to the sector: a small increase in the proportion of the local workforce with either A-levels or vocational qualifications in the sector may lead to a negative impact of upstream investment in BDA on technical efficiency. However, if the increase in the proportion of the workforce is sufficiently large, then the relationship between average technical efficiency and upstream investment in BDA becomes positive.

References

- Acharya, R., & Basu, S. (2010). ICT and TFP growth: intangible capital or productive externalities?, Industry Canada Working Paper 2010-1.
- Agarwal, R., Ganco, M., & Ziedonis, R.H. (2009). Reputations for toughness in patent enforcement: implications for knowledge spillovers via inventor mobility. *Strategic Management Journal*, 30(13), 1349-1374.
- Aghion, P. (2002). Schumpeterian growth theory and the dynamics of income inequality. *Econometrica*, 70(3), 855-882.
- Aktas, E., & Meng, Y. (2017). An exploration of Big Data practices in retail sector. *Logistics*, 1(12), 1-28.
- Akter, S., Wamba, S.W., Gunasekaran, A., Dubey, R., & Childe, S.J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment?. *International Journal of Production Economics*, 182, 113–131.
- Alexander, A., Shaw, G., & Curthe, L. (2005). Promoting retail innovation: knowledge flows during the emergence of self- service and supermarket retailing in Britain. *Environment and Planning*, A37, 805–821.
- Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to Big Data. *Business Horizons*, 60(3), 285–292.
- Assaf, A. G., Barros, C. and Sellers-Rubio, R. (2011), Efficiency determinants in retail stores: a Bayesian framework, *Omega*, 39, 283-292.
- Athanassopoulos, A. (2003). Strategic Groups, Frontier Benchmarking and Performance Differences: Evidence from the UK Retail Grocery Industry. *Journal of Management Studies*, 40(4), 921-953.
- Athanassopoulos, A. D., & Ballantine, J. A. (1995). Ratio and frontier analysis for assessing corporate performance: evidence from the grocery industry in the UK. *Journal of the Operational Research Society*, 46, 427–440.
- Audretsch, D. B., & Feldman, M. P. (1996). R&D Spillovers and the geography of innovation and production. *American Economic Review*, 86(3), 630-640.
- Banker, R.D., & Morey, R.C. (1986). Efficiency Analysis for exogeneously fixed inputs and outputs. *Operations Research*, 34(4), 513-521.
- Barros, C.P. (2006). Efficiency measurement among hypermarkets and supermarkets and the identification of the efficiency drivers: a case study. *International Journal of Retail & Distribution Management*, 34(2), 135–154.

- Barros, C.P., & Alves, C. (2004). An empirical analysis of productivity growth in a Portuguese retail chain using Malmquist productivity index. *Journal of Retailing and Consumer Services*, 11(5), 269–278.
- Battese, G., & Coelli, T.J. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics*, 20, 325-32.
- Belberdos, R.M., Carree, M. & Lokshin, B. (2004). Cooperative R&D and firm performance. *Research Policy*, 33, 1477-1492.
- Betancourt, R. (2004). *The Economics of Retailing and Distribution*. Edward Elgar Publishing, Ltd.
- Blažková, I., & Dvouletý, O. (2018). The causes of firm performance variation in the Czech food processing industry in the context of the outlier effect. *Management Research Review*, 41(8), 968-986.
- Boles, J., Donthu, N., & Lohtia, R., (1995). Sales person evaluation using relative performance efficiency: the application of data envelopment analysis. *Journal of Personal Selling and Sales Management*, 15, 31 - 39.
- Bradlow, E.T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of Big Data and predictive analytics in retailing. *Journal of Retailing*, 93(1), 79-95.
- Brynjolfsson, E., & Hitt, L. (2003). Computing productivity: Firm-level evidence. *The Review of Economics and Statistics*, 85(4), 793-808.
- Brynjolfsson, E., & Saunders, A. (2009). *Wired for innovation: How information technology is reshaping the economy*. MIT Press, Cambridge, MA.
- Bugamelli, M., & Pagano, P. (2004). Barriers to investment in ICT. *Applied Economics*, 36(20), 2275-2286.
- Bughin, J., Hazan, E., Lund, S., Dahlstrom, P., Wiesinger, A., & Subramaniam, A. (2018). Skills shift automation and the future of the workforce. McKinsey&Company. McKinseu Global Institute, Discussion Paper, May 2018.
- Bultez, A., & Parsons, L. (1998). Channel productivity: in the small and in the large (Editorial). *International Journal of Research in Marketing*, 15(5), 383–400.
- Chebat, J., Ierre, F., Arnon, K., & Sholomom, T. (1994). Strategic auditing of human and financial resource allocation in marketing: An empirical study using data envelopment analysis. *Journal of Business Research*, 31, 197-208.
- Chou, Y-C., Chuang, H.H-C., & Shao, B.B.M. (2014). The impacts of information technology on total factor productivity: A look at externalities and innovations. *International Journal of Production Economics*, 158, 290–299.

- Chyi, Y. L., Lai, Y. M., & Liu, W. H. (2012). Knowledge spillovers and firm performance in the high-technology industrial cluster. *Research Policy*, 41(3), 556-564.
- Cohen, W.M., & Levinthal, D.A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), 128-152.
- Corredoira, R.A., & Rosenkopf, L. (2010). Should auld acquaintance be forgot? The reverse transfer of knowledge through mobility ties. *Strategic Management Journal*, 31(2), 159-181.
- Criscuolo, C., & Waldron, K. (2003). E-Commerce and firm productivity. *Economic Trends*, 600, 52-57.
- Dahl, M.S., & Pedersen, C.Ø.R. (2005). Knowledge flows through informal contacts in industrial clusters: myth or reality?, *Research Policy*, 33(10), 1673-1686.
- De Jorge-Moreno (2008). Efficiency and regulation in the Spanish hypermarket retail trade: A cross section approach. *International Journal of Retail & Distribution Management*, 36(1), 71-88.
- Destefanis, S. and Sena, V. (2005) Productivity, spillovers and human capital: an analysis for the Italian firm data, https://www.researchgate.net/publication/228423274_Productivity_spillovers_and_human_capital_an_analysis_for_the_Italian_firm_data.
- Díaz-Chao, A., Sainz-González, J., & Torrent-Sellens, J. (2015). ICT, innovation, and firm productivity: New evidence from small local firms. *Journal of Business Research*, 68(7), 1439-1444.
- Dolfsma, W. and Van Der Eijk, R. (2016). Distances in organizations: Innovation in an R&D lab. *British Journal of Management*, 27(2), 271-286.
- Donthu, N., Hershberger, E.K., & Osmonbekov, T. (2005). Benchmarking marketing productivity using data envelopment analysis. *Journal of Business Research*, 58, 1474–1482.
- Donthu, N., & Yoo, B. (1998). Retail productivity assessment using data envelopment analysis. *Journal of Retailing*, 74(1), 89–105.
- Dos Santos, B.L., & Peffers, K. (1998). Competitor and vendor influence on the adoption of innovative applications in electronic commerce. *Journal of Information and Management*, 34(3), 175 – 184.
- Duranton, G., & Monastiriotis, V. (2002). Mind the gaps: The evolution of regional inequalities in the UK 1982-1997. *Journal of Regional Science*, 42(2), 219-56.
- Enos, M.D., Kehrhahn, M.T., & Bell, A. (2003). Informal learning and the transfer of learning: How managers develop proficiency. *Human Resource Development Quarterly*, 14(4), 369-387.

- Erevelles, S., Fukawa, & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A*, 120, 3, 253-81.
- Filatotchev, I., Liu, X., Lu, J., & Wright, M. (2011) Knowledge spillovers through human mobility across national borders: Evidence from Zhongguancun science park in China. *Research Policy*, 40 (3). 453-462.
- Fiol, C.M. & Lyles, M.A. (1985). Organizational learning. *The Academy of Management Review*, 10(4), 803-813.
- Fritsch, M., & Changoluisa, J. (2017). New business formation and the productivity of manufacturing incumbents: Effects and mechanisms. *Journal of Business Venturing*, 32(3), 237-259.
- Fritsch, M., & Slavtchev, V. (2011). Determinants of the efficiency of regional innovation systems. *Regional Studies*, 45(7), 905-918.
- Fu, X. (2012). Foreign direct investment and managerial knowledge spillovers through the diffusion of management practices. *Journal of Management Studies*, 49(5), 970-999.
- Gambardella, A. & Giarratana, M.S. (2010). Localized knowledge spillovers and skill-based performance. *Strategic Entrepreneurship Journal*, 4, 323–339.
- Gaskell, A. (2018). Organizations striving to close the data science skills gap. *Forbes*, Jun 18, 2018.
- Gauri, D.K. (2013), Benchmarking retail productivity considering retail pricing and format strategy, *Journal of Retailing*, 89, 1-14.
- Germann, F., Lilien, G.L., Fiedler, L., & Kraus, M. (2014). Do retailers benefit from deploying customer analytics? *Journal of Retailing*, 90(4), 587–593.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.
- Griffith R, & Harmgart, H. (2005). Retail productivity, IFS Working Papers W05/07, Institute for Fiscal Studies.
- Griffith, R., & Harmgart, H. (2008). Supermarkets and Planning regulations, CEPR DP 6713.
- Griffith, R., Redding, S. & Van Reenen, J. (2004). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. *The Review of Economics and Statistics*, MIT Press, 86(4), 883-895, November.

- Griliches Z. (1992). The search for R&D spillovers. *Scandinavian Journal of Economics*, 94(0), S29-47.
- Grewal, D., Levy, M., Mehrotra, A., & Sharma, A. (1999). Planning merchandising decisions to account for regional and product assortment differences. *Journal of Retailing*, 75(3), 405–24.
- Grewal, D. Roggeveen, A.L., & Nordfält, J. (2017). The future of retailing. *Journal of Retailing*, 93 (1), 1-6.
- Gunasekaran, A. and Yusuf, Y.Y. and Adeleye, E.O., & Papadopoulos, T. (2018). Agile Manufacturing Practices: The Role of Big Data and Business Analytics with Multiple Case Studies. *International Journal of Production Research*, 56(1-2), pp. 385-397.
- Haller, S.A. & Siedschlag, I. (2011). Determinants of ICT adoption: Evidence from firm-level data. *Applied Economics*, 43(26), 3775-3788.
- Haskel, J., & Sadun, R. (2007). Entry regulation and Productivity: Evidence form the Uk retail sector, mimeo.
- Haskel, J., & Sadun, R. (2008). Regulation and UK retailing productivity: Evidence from micro data, mimeo. *Economica*, 79(315), 425-448.
- Haskel, J., & Wallis, G. (2010). Public support for innovation, intangible investment and productivity growth in the UK market sector” paper at <http://www.ceriba.org.uk/bin/view/CERIBA/PublicSupportCeriba>
- Higon, D. & Sena, V. (2014). Productivity, R&D Spillovers and Educational Attainment, *Oxford Bulletin of Economics and Statistics*, 1, 14-37.
- HM Treasury (2001). *Productivity in the UK: The regional dimension*, London, UK.
- Horsky, D., & Nelson, P. (1996). Evaluation of salesforce size and productivity through Efficient Frontier Benchmarking. *Marketing Science*, 15(4), 301-320.
- Hsu, L-C., & Wang, C-H. (2010). Clarifying the effect of intellectual capital on performance: The mediating role of dynamic capability. *British Journal of Management*, 1(2), 1-27.
- Isaksson, O.H.D., Simeth, M., & Seifert, R.W. (2016). Knowledge spillovers in the supply chain: Evidence from the high tech sectors. *Research Policy*, 45(3), 699-706.
- Jaffe, A. (1989). Real effect of academic research. *American Economic Review*, 79, 957-70.
- Johnson, J.S., Friend, S.B., & Lee, H.S. (2017). Big Data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process. *Journal of Product Innovation Management*, 34, 640–658.

- Kamakura, W.A., Lenartowicz, T., & Ratchford, B.T. (1996). Productivity assessment of multiple retail outlets. *Journal of Retailing*, 72(4), 333–356.
- Kang, S. & Snell, S. (2009). Intellectual capital architectures and ambidextrous learning: A framework for human resource management. *Journal of Management Studies*, 46(1), 65-92.
- Ket, H. T. and Chu, S. (2003). Retail productivity and scale economies at the firm level: a DEA approach. *Omega*, 31, 75–82.
- Kodde, D.A., & Palm, F.C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, 54(5), 1243-1248.
- Korhonen, P., & Syrjänen, M.J. (2004). Resource allocation based on efficiency analysis. *Management Science*, 50(8), 1134–1144.
- Kumbhakar, S.C., & Lovell, C.A.K. (2000). *Stochastic Frontier Analysis*, Cambridge, UK: Cambridge University Press.
- Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth*, 11(2), 97–127.
- Van Leeuwen, G., & van der Wiel, H. (2003). Do analytics spillover matter: Evidence from Dutch firm-level data, CPB Discussion Paper No 26.
- Lam, S.K., Sleep, S., Hennig-Thurau, T., Sridhar, S., & Saboo, A.R. (2017). Leveraging frontline employees' small data and firm-level Big Data in frontline management: An absorptive capacity perspective. *Journal of Service Research*, 20(1) 12-28.
- Luo, Y. (2007). The independent and interactive roles of procedural, distributive, and interactional justice in strategic alliances. *Academy of Management Journal*, 50(3), 644-664.
- Luo X., & Donthu, N. (2005). Assessing advertising media spending inefficiencies in generating sales. *Journal of Business Research*, 58, 28–36.
- Lusch, R.F., Serpkenci, R.R., & Orvis, B. (1995). Determinants of retail store performance: a partial examination of selected elements of retailer conduct. In Grant, K., & Walker, I. (Eds.), *World Marketing Congress*, vol. 7, pp. 95–104.
- MacVaugh, J., & Schiavone, F. (2010). Limits to the diffusion of innovation: A literature review and integrative model. *European Journal of Innovation Management*, 13(2), 197 – 221.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byer, A.H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey & Institute, May 2011.
- Milliken, D. (2017). UK retailers cut jobs at fastest rate since 2008 – BRC. Reuters, October 26, 2017.

Motohashi, K. & Yuan, Y. (2010). Productivity impact of technology spillover from multinationals to local firms: Comparing China's automobile and electronics industries. *Research Policy*, 39, 790–798.

OECD (2013), "Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"", *OECD Digital Economy Papers*, No. 222, OECD Publishing, Paris.

Office for National Statistics (2018). Human capital estimates, UK: 2004 to 2017: National, regional, occupational and length of employment human capital stock estimates for the UK, considering factors like education and demographics. Office for National Statistics, October 1, 2018.

Ozdemir, S., Kandemir, D., & Eng, T-Y. (2017). The role of horizontal and vertical new product alliances in responsive and proactive market orientations and performance of industrial manufacturing firms. *Industrial Marketing Management*, 64, 25-35.

Perrigot, R., & Barros, C.P. (2008). Technical efficiency of French retailers. *Journal of Retailing and Consumers Services*, 15, 296-305.

Ratchford, B.T., Agrawal, J., Grimm, P.E., & Srinivasan, N. (1996). Toward understanding the measurement of market efficiency. *Journal of Public Policy and Marketing*, 15, 167–184.

Ratchford, B.T., & Brown, J.R. (1985). A study of productivity changes in Food retailing. *Marketing Science*, 4(4), 292-311.

Ratchford, B.T., & Stoops, G.T. (1988). A model and measurement approach for studying retail productivity. *Journal of Retailing*, 64(3), 241-263.

Reardon, J., Hasty, R., & Coe, B. (1996). The effect of information technology on productivity in retailing. *Journal of Retailing*, 72(4), 445-461.

Ren, J.S., Wamba, S. F., Akter, S., Rameshwar, D., & Stephen, C. (2016). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, 55(17), 5011-5026.

Riley, R. (2015). The UK Productivity Puzzle 2008-2013: Evidence From British Businesses. National Institute of Economic and Social Research, NIESR Discussion Paper No: 450.

Rincon, A., & Vecchi, M. (2004). The dynamic impact of ICT spillovers on companies' productivity performance, National Institute of Economic and Social Research, NIESR Discussion Paper No: 245.

Rincon, A., Vecchi, M., & Venturini, F. (2013). ICT as a general purpose technology: Spillovers, absorptive capacity and productivity performance. National Institute of Economic and Social Research, Discussion PaperNo.: 416.

- Rindfleisch, A., & Moorman, C. (2003). Interfirm cooperation and customer orientation. *Journal of Marketing Research*, 40(4), 421-436.
- Rindfleisch, A. & Moorman, C. (2001). The acquisition and utilization of information in new product alliances: A strength-of-ties perspective. *Journal of Marketing*, 65(2), 1-18.
- Rogers, E.M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Roh, L. (2018). The retail apocalypse is not happening: Why retailers that innovate will never go extinct. *Forbes*, Sep 28, 2018.
- Rowlatt, A. (2001). Measuring e-commerce: Developments in the United Kingdom. *Economic Trends*, 575, 30-36.
- Sellers-Rubio, R., & Mas-Ruiz, F. (2006). Economic efficiency in supermarkets: evidences in Spain. *International Journal of Retail & Distribution Management*, 34(2), 155–171.
- Sellers-Rubio, R., & Mas-Ruiz, F. (2007). An empirical analysis of productivity growth in retail services: evidence from Spain. *International Journal of Service Industry Management*, 18(1), 52-69.
- Sellers-Rubio, R., & Mas-Ruiz, F. (2009). Technical efficiency in the retail food industry: the influence of inventory investment, wage levels, and age of the firm. *European Journal of Marketing*, 43(5/6), 652-669.
- Severgnini, B., (2010). Is ICT a jack-in-the-box? A counterfactual approach for identifying productivity spillovers, Copenhagen Business School, mimeo.
- Siamagka, N-T., Christodoulides, Birkbeck, G., Michaelidou, N., & Valvi, A. (2015). Determinants of social media adoption by B2B organizations. *Industrial Marketing Management*, 51, 89- 99.
- Srinivasan, R., & Swink, M. (2018). An investigation of visibility and flexibility as complements to supply chain analytics: An organizational information processing theory perspective. *Production and Operations Management*, 27(10), 1849–1867.
- Subramaniam, M., & Youndt, M.A. (2005). The influence of intellectual capital on the types of innovative capabilities. *Academy of Management Journal*, 48(3), 450–463.
- Stiroh, K.J. (2002). Are ICT spillovers driving the new economy? *The Review of Income and Wealth*, 48(1), 33-57.
- Thomas, R.R., Barr, R.S., Cron, W.L., & Slocum Jr., J.W. (1998). A process for evaluating retail store efficiency: A restranalytised DEA approach. *International Journal of Research in Marketing*, 15(5), 487-503.
- Xavier, J. M., Moutinho, V.F. and Moreira, A.C., (2015). An empirical examination of performance in the clothing retailing industry: A case study, *Journal of Retailing and*

Consumer Services, 25, 96-105.

Yu, W., & Ramanathan, R. (2008). An assessment of operational efficiencies in the UK retail sector. *International Journal of Retail & Distribution Management*, 36(11), 861–882.

Vyt, D. (2008). Retail network performance evaluation: A DEA approach considering retailers' geo-marketing. *The International Review of Retail, Distribution and Consumer Research*, 18(2), 235-253.

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

Wang, G., Gunasekaran, A., Ngai, E.W.T., & Papadopoulos, T. (2016). Big Data Analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98-110.

Wang, Y., Kung, L.A., & Byrd, T.A. (2018). Big Data Analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.

Wang, Y., Kung, L.A., Gupta, S., & Ozdemir, S. (2019). Leveraging big data analytics to improve quality of care in healthcare organizations: A configurational perspective. *British Journal of Management*, 30(2), 362-388.

Wu, D. D., Yang, Z., & Liang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. *Expert Systems with Applications*, 31(1), 108-115.

Zhang, Y., Guangquan, Z., Chen, H., Porter, A.L., Zhu, D., & Lu, J. (2016). Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological Forecasting & Social Change*, 105, 179–191.

A1. DATA APPENDIX

The data used for the analysis come from the ORBIS database which is a commercial data based provided by Bureau van Dijk and have been used extensively in empirical work involving productivity analysis. They cover the entire population of registered UK firms sourced from Companies House. Information such as name, registered address, firm type, and industry code are available for all firms. Availability of financial information varies substantially across firms. We limit the analysis to firms classified as belonging to the industry NACE-REV 2 47.

All monetary variables obtained from ORBIS are expressed in thousands of British Pound Sterling and are deflated using the ONS deflators. We use value added as a measure of firm output that is equal to turnover net of material costs. ORBIS provides the number of full-time equivalent employees. Capital input is measured as total tangible assets by book value, recorded annually. Tangible assets include land and buildings, fixtures and fittings, plants and vehicles, and other tangible assets (including ICT assets).

To trace inter-industry flows of spillovers we use the procedure first suggested by Terleckyj (1974). The main idea of this method is to use information from the input-output table to identify the upstream and the downstream sectors. The procedure uses the 2-digit industry series on BDA investment, weighted by input-output intermediate transactions' coefficients, defined as ratio between the flow of intermediate inputs sold by industry f to industry j and the gross output of the selling sector. Data on industry-level investment in BDA have been drawn from EU KLEMS 2009-2015. More specifically, we use the EU KLEMS series on investment in software and databases. Input-output intermediate transactions' coefficients are taken from the OECD I-O output table at benchmark years.

The regional share of the graduate workforce over the total workforce in each sub-retail sector has been calculated using the data from the Quarterly Labour Force Survey (QLFS). The QLFS is a sample survey of households living at private addresses in Great Britain. The population covered is all people resident in private households, all persons resident in National Health Service accommodation and young people living

away from the parental home in a student hall of residence or similar institution during term time. The sample design currently consists of about 55,000 responding households in Great Britain every quarter, representing about 0.2% of the population. Its main purpose is to provide information on the UK labour market which can then be used to develop, manage and evaluate labour market policies. The survey asks each respondent the highest qualification attained, which industry he/she works for and the region of the main job. Importantly the list of regions respondents can choose from is more granular than the list of Government Office Regions (GOR) allowing to focus on regions that are smaller than the Government Office Regions. The three questions combined together allows us to compute the share of graduates in the retail sector by region as well as the share of retail workers with a vocational qualification/A-level in the retail sector by region.

Figure 1. Kernel density plot – Technical efficiency scores

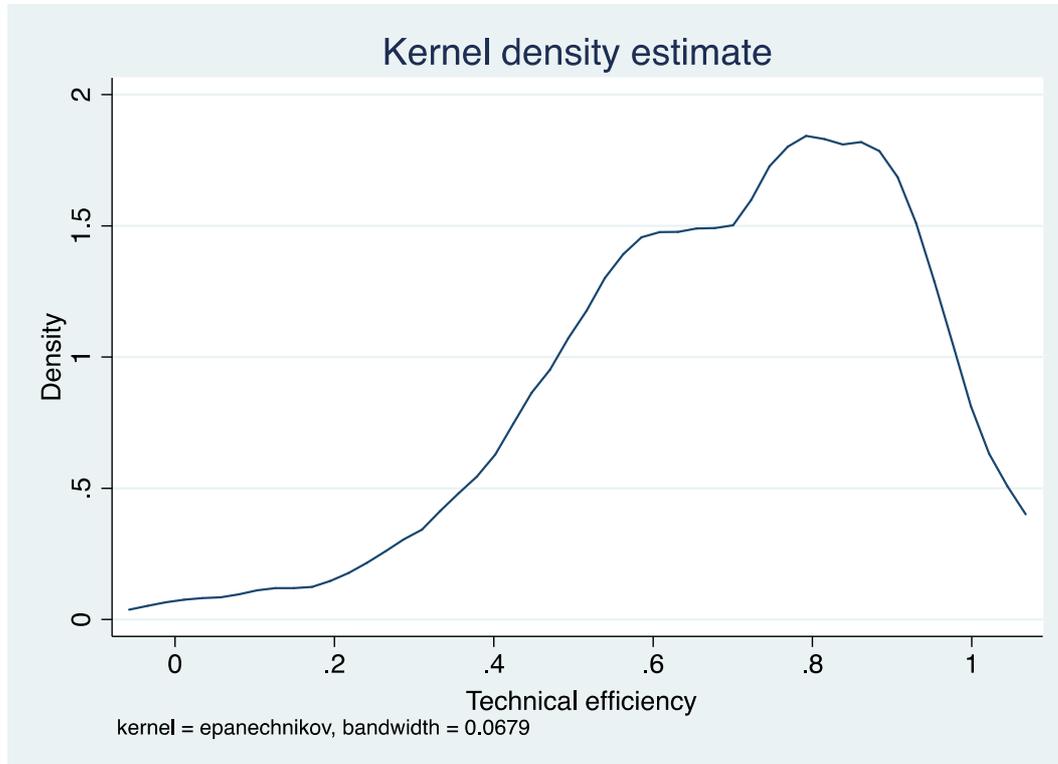


Table 1. *Sample means of inputs and outputs (by year) - Retail*

	Turnover	Labour	Capital
2009	585610	49656	3115987
2010	6096075	49898	3204197
2011	6492015	52067	3452975
2012	5486439	42566	2944582
2013	478075	37322	2476627
2014	4579616	35158	2301271
2015	4044888	32198	1990724

Note: Source: ORBIS. Definitions of variables can be found in the Data Appendix. Monetary figures in the table have been deflated (using the deflators detailed in the text) and are measured in thousands pounds. Labour is measured as number of employees.

Table 2. Investment in software and BDA by sector and year

	2009	2010	2011	2012	2013	2014	2015
Food products, beverages and tobacco	406.00	421.00	409.00	376.00	447.00	425.00	405.00
Textiles, wearing apparel, leather and related products	49.00	50.00	54.00	52.00	52.00	50.00	40.00
Wood and paper products; printing and reproduction of recorded media	313.00	276.00	297.00	313.00	324.00	333.00	324.00
Coke and refined petroleum products	56.00	66.00	83.00	75.00	82.00	83.00	74.00
Chemicals and chemical products	345.00	287.00	312.00	303.00	304.00	315.00	345.00
Rubber and plastics products, and other non-metallic mineral products	250.00	243.00	252.00	269.00	264.00	290.00	290.00
Basic metals and fabricated metal products, except machinery and equipment	297.00	303.00	344.00	331.00	346.00	345.00	361.00
Electrical and optical equipment	207.00	250.00	233.00	275.00	268.00	274.00	285.00
Machinery and equipment n.e.c.	439.00	422.00	451.00	434.00	452.00	456.00	445.00
Transport equipment	504.00	576.00	586.00	648.00	600.00	707.00	689.00
Other manufacturing; repair and installation of machinery and equipment	422.00	421.00	405.00	435.00	463.00	464.00	453.00
ELECTRICITY, GAS AND WATER SUPPLY	886.00	933.00	984.00	1048.00	1164.00	1191.00	1376.00
CONSTRUCTION	386.00	407.00	371.00	405.00	404.00	419.00	433.00
WHOLESALE AND RETAIL TRADE; REPAIR OF MOTOR VEHICLES AND MOTORCYCLES	3476.00	3523.00	4016.00	4323.00	4628.00	5363.00	5130.00
ACCOMMODATION AND FOOD SERVICE ACTIVITIES	202.00	206.00	199.00	207.00	237.00	253.00	249.00
INFORMATION AND COMMUNICATION	4307.00	4310.00	4680.00	4926.00	5141.00	5118.00	4781.00
FINANCIAL AND INSURANCE ACTIVITIES	2995.00	3224.00	3400.00	3834.00	3526.00	3721.00	3847.00
REAL ESTATE ACTIVITIES	260.00	265.00	276.00	322.00	345.00	329.00	376.00
PROFESSIONAL, SCIENTIFIC, TECHNICAL, ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES	3711.00	3987.00	4108.00	4325.00	4565.00	4703.00	4926.00

Source: EU KLEMS, 2009-2015.

Table 3. Production Frontier estimates – Maximum Likelihood estimator

<i>Variable</i>	<i>Coefficient</i>	<i>t-ratio</i>
Labour (log)	0.24	6.56
Capital (log)	0.34	7.26
Time trend	-0.015	-0.25
Squared time trend	0.0037	0.81
Squared Labour (log)	0.16	1.21
Squared Capital (log)	-0.0099	-0.09
Labour (log) * Capital (log)	-0.068	-0.27
Constant	0.42	1.71
<i>Inefficiency Model</i>		
Spillovers from upstream investment in BDA (lagged one period)	0.0035	2.15
Proportion of graduate workforce in retail over the total employment in retail by region (lagged one period)	-1.04	-3.59
Interaction between the two variables above	0.00046	3.51
Variance parameters		
Variance inefficiency	1.17	
Variance stochastic term	0.060	
<i>LR test</i>	381.01**	

Note: Maximum Likelihood estimates of the parameters from the simultaneous estimation of the production frontier and of the inefficiency model as in Battese and Coelli (1995). The dependent variable is the log of the value added. The excluded year in the inefficiency model is 2009. The inefficiency model includes regional dummies. The test statistic has a mixed chi-square distribution and so critical values at 5% significance level are sourced from Kodde and Palm (1986).

Table 4. Hypothesis tests.

<i>Null Hypothesis</i>	<i>Likelihood Ratio Test</i>	<i>Decision</i>
<i>Test 1:</i> <i>Null Hypothesis:</i> <i>There is no technical progress.</i>	85.72**	Reject Null
<i>Test 1:</i> <i>Null Hypothesis:</i> <i>Cobb-Douglas Functional form.</i>	178.74**	Reject Null
<i>Test 3:</i> <i>Null Hypothesis:</i> <i>No efficiency effects (=0)</i>	293.32**	Reject Null
<i>Test 4:</i> <i>Null Hypothesis:</i> <i>The regional dummies are not significant in the efficiency model.</i>	115.7**	Reject Null

Note: Hypotheses tests on the parameters of the production frontier (Test 1 and Test 3) and of the inefficiency model (Test 3 and Test 4). Likelihood Ratio test is computed by using (2) in the text. The test statistic has a chi-square distribution.

Table 5. Production Frontier estimates - Maximum Likelihood estimator

<i>Variable</i>	<i>Coefficient</i>	<i>t-ratio</i>
Labour (log)	0.95	13.94
Capital (log)	-0.011	-0.17
Time trend	0.51	2.30
Squared time trend	0.025	2.34
Spillovers from upstream investment in BDA	0.0032	4.54
Spillovers from upstream investment in BDA * time trend	-0.00056	-2.42
Squared Labour (log)	-0.143	-2.03
Squared Capital (log)	-0.065	-1.17
Labour (log) * Capital (log)	0.22	1.76
Constant	-0.39	-0.21
<i>Inefficiency Model</i>		
Spillovers from upstream investment in BDA (lagged one period)	0.0042	2.20
Proportion of graduate workforce in retail over the total employment in retail by region (lagged one period)	-1.44	-3.37
Interaction between the two variables above	0.0007	1.67
Variance parameters		
Variance inefficiency	0.60	
Variance stochastic term	0.022	
<i>LR test</i>	133.10**	

Note: Maximum Likelihood estimates of the parameters from the simultaneous estimation of the production frontier and of the inefficiency model as in Battese and Coelli (1995). The dependent variable is the log of the value added. The excluded year in the inefficiency model is 2009. The inefficiency model includes regional dummies. The LR test has a chi-square distribution and it has been calculated under the null hypothesis that there is no technical change (including neutral technical change and technical change interacted with the upstream spillovers from BDA) in the frontier model.

Table 6. *Technical Efficiency score (mean values by year).*

<i>Year</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>
2009	0.69	0.30	0.99
2010	0.72	0.44	0.99
2011	0.79	0.48	0.99
2012	0.66	0.20	0.99
2013	0.63	0.010	0.99
2014	0.59	0.055	0.97
2015	0.73	0.34	0.99

Table 7. Efficiency Models – Alternative qualifications.

Model 1		
Spillovers from upstream investment in BDA (lagged one period)	-0.0032	-3.02
Proportion of retail workforce with vocational qualifications over the total employment in retail by region (lagged one period)	-9.97	-5.81
Interaction between the two variables above	0.0087	5.74
<i>LR test</i>	689.59	
Model 2		
Spillovers from upstream investment in BDA (lagged one period)	-0.0023	-2.38
Proportion of retail workforce with A-levels over the total employment in retail by region (lagged one period)	-33.75	-7.31
Interaction between the two variables above	0.027	7.25
<i>LR test</i>	551.76	

Note: The LR test has a chi-square distribution and it has been calculated under the null hypothesis that the three variables listed above are equal to zero.