IARIW 2021

IARIW 2021

Monday 23 – Friday 27 August

Statistisk sentralbyrå
Statistics Norway

Digital Opportunity: How Cloud Computing Changes the Shape of the UK Economy

Oleksii Romanko

(King's College London)

Paper prepared for the 36th IARIW Virtual General Conference
August 23-27, 2021
Session 2: The Potential and Challenges of Big Data and other Alternative Data in the Production of Prices, National Accounts, and Measures of Economic Well-Being
Time: Tuesday, August 24, 2021 [14:00-16:00 CEST]

Digital opportunity: how cloud computing changes the shape of the UK economy

Oleksii Romanko (KCL)[1]

[1]oleksii.romanko@kcl.ac.uk

Abstract. *This paper focuses on the impact of cloud technologies on the productivity of enterprises in the UK. The central belief is that cloud computing gives a boost to company productivity thanks to automation, intangibles and financial channels. We use OLS regressions and Propensity Score Matching and use the FAME dataset which contains micro level data about UK-based enterprises. We combine these data with the cloud usage indicator, constructed with the help of web scraping methods and companies website information. The results show that cloud usage can positively impact on firm productivity, but the impact is not immediate as firms need to learn how to use this new technology. The impacts are higher for smaller firms, consistent with the idea that cloud opens up opportunities for these firms to engage in innovative activities.*
Keywords*: cloud computing, productivity, web scraping.*

1. Introduction

1.1. Context

Rapid changes in the information and technological environment have induced substantial changes in the ways that the modern economy operates. Modern society appears to be going through a fourth industrial revolution (Schwab, 2017), which involves digital, cloud and AI technologies. However, new technologies also gave rise to the intellectualizing of machinery, speed of changes and burden of informational load. It became harder to precisely track those organizational and structural changes in the economy and the impact of the technology. This may have resulted in under accounting of new technologies and the related growth. As Coyle (2017) notes, there is an ongoing replacement of traditional goods and services (books, cinemas, educational institutions) by the digital ones (online books, lectures, YouTube video services, online educational courses). This replacement might lead to underestimation of economic activity, because of shifts from traditionally measured activities to unmeasured and zero-priced ones. This in turn may lead to a considerable measurement gap in the productivity statistics.

1.2. The aim of the study

We aim to add to the literature on explaining the role of intangibles and new technologies on productivity growth. As Haldane (2018) stated, the UK is 'good at R but not at D, where D here includes not only Development, but also Diffusion and Dissemination'. New technologies give an opportunity for firms to execute R&D and implement results of experimentation at higher speed, scale, and lower cost. We focus in this paper on the impact of cloud technologies on the productivity of enterprises in the UK. The central belief is that cloud computing is related to a boost in company productivity thanks to automation, intangibles and financial channels (Acemoglu and Restrepo, 2018, Bloom and Pierri, 2018, Ewens et al. 2018). The Cloud seem to help companies to cut down R&D costs, leverage existing best practices in using ICT, achieve better use of time thanks to automation of internal processes, and get deeper business insight from the existing company data. Cloud technologies also made contributions towards the rise of the giant digital companies of nowadays. Companies like Uber, Airbnb, Netflix and Dropbox would never exist without a possibility to leverage cloud technologies that do not require significant upfront investments and that gives an opportunity to easily scale up firms' businesses. Research suggests that expenditures on cloud services have grown 4.5 times faster than traditional ICT investments since 2009; by 2016, Cloud represented 37.2% of overall ICT infrastructure investment (Forbes, 2017; IDC, 2017).

However, some authors did not observe a significant impact of traditional ICT technologies on productivity (Doms, Dunne and Troske 1997, Morrison 1996, Brynjolffson 1993). The Solow paradox (Solow 1987), suggested that despite rapid development of the ICT industry in the twentieth century, there was no evidence initially of its positive impact on productivity growth. Brynjolfsson and Science (1992) suggested that the impact of ICT and computing is highly heterogeneous and, therefore, might be insignificant in aggregate. Later studies did show positive impacts of ICT at the aggregate and industry levels, (Jorgenson and Stiroh, 1995, 2000), but these were slow to emerge. As cloud technologies are the next development step of previous ICT technologies, this poses the question of whether modern cloud technologies have a positive and economically significant impact on the productivity of enterprises, while it is heterogeneous across industries and firms. While there are some studies on the cloud computing impact (Jin et al., 2017, Dimitrov and Osman 2014), there is insufficient evidence to date about the productivity impact of the Cloud. Haldane (2018) suggests that there is a high dispersion between low-productive and high-productive enterprises in the UK, and the gap is 80% higher then in other counties. High degrees of heterogeneity might be partially explained by new technologies impact on 'winners and losers in the knowledge economy' (Riley, 2019).

1.3. Data and methodology

The paper utilizes data on cloud usage statistics in the UK gathered from 2012 until 2020. The data has been used to find empirical evidence about the positive impact of cloud computing technologies on the productivity of firms. This large-scale study states that the benefit of the usage of cloud technologies is increasing throughout time and is more prominent for small UK companies, consistent with the literature on the nature of cloud computing technologies (Jin et al., 2017). This recent tendency is opposite to what was happening with traditional ICT investments where, due to high upfront costs, larger companies could afford to invest more in traditional ICT infrastructure (Bugamelli, Pagano 2004).

In order to answer the main question about the influence of cloud technologies on the productivity of UK enterprises, the authors used a register of companies that had website addresses. We collected financial indicators of companies' performance, such as income and profit indicators, financial ratios, number of employees, statistics about foreign and own investments. Financial statistics were used in order to assess the productivity performance of the enterprises.

The second part of the data gathering process was to combine these financial statistics with cloud usage statistics. This part of the data gathering process required web scraping techniques and some analysis of the Internet infrastructure. Using metadata that companies leave in their Internet records, it was possible to determine whether the given company uses cloud-related technologies. 20In order to build cloud usage indicators, the authors use a history of DNS records generously provided by SecurityTrails. We started with the data for all companies listed in the financial dataset (FAME) from 2008 till 2018. We classify web hosting providers as cloud and non-cloud ones. In such a way, we build the indicator of cloud usage by assigning one (1) to companies that employ cloud-related vendors for their website hosting, and zero (0) otherwise. We performed a statistical analysis of the data gathered, using regression and propensity score matching models in order to relate the usage of the cloud to the differences between companies. Both models show consistent and positive, economically and statistically significant impacts of cloud technologies on the productivity of firms.

In Section 2 we proceed with an overview of cloud computing, its history and current trends. Section 3 proceeds with a review of the literature on the topic of ICT, cloud computing, their impact on economic activity and the productivity of enterprises. We present our approach on gathering the cloud computing adoption statistics and modelling results in Section 4.

We conclude the paper with a discussion of the importance of cloud computing technology in the further development of the digital economy. Further issues that need to be discussed include the broad scope of network externalities of the technology, ability of firms to have faster learning processes thanks to the technology, absence of fixed irreversible investments, continuous reduction in prices, as well as augmentation of jobs.

2. Cloud computing

2.1. Cloud history

Cloud, according to the US National Institute of Standards and Technology, is a service, that “enables on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. In other words, cloud computing technologies are usually referred to as a set of computing resources, which can be accessed by anyone who wants to use computing power for some period of time. Cloud computing technologies allow for the same computing resources to be used by several different people through a remote internet connection.

Cloud services are considered to be part of the next wave of technological changes arising from ICT innovations during the last century. The Cloud emerged from the early 2010s, after the boom of the World Wide Web. The whole history of preceding technologies that resulted in the cloud computing ecosystem is illustrated in (Figure 1).

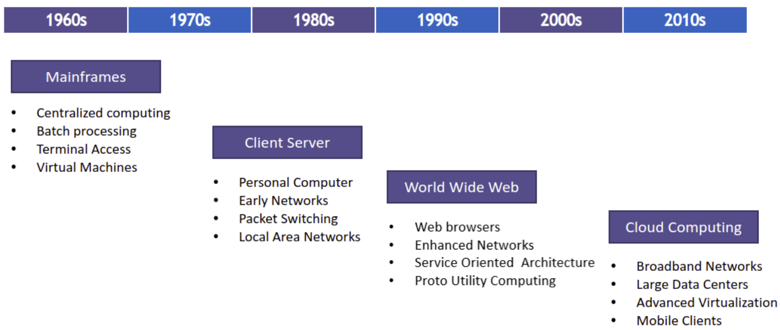

Figure 1. History of the cloud computing (source: own compilation based on Trivedi (2013))

The usage of cloud computing resources is similar in some ways to the standard leasing procedure; however, the crucial difference is in the “sharing economy” part of the cloud services. Several users can utilize the same computing resources, simultaneously or sequentially. Thus, a 100% utilization rate of resources can be achieved. The same “social” and cost-saving characteristics relate not only to cloud computing, but to cloud storage, management, and all other services provided by cloud vendors.

The idea of a shared pool of resources is not new. As we can see from Figure 1, in the early 1970s, there were mainframe computers that were also used by several operators. However, with the emergence of local area networks in 1980 and the internet ecosystem in the 1990s, the concept of computing power aggregation and further sharing of resources became publicly available to the mass of consumers. Thus, the simple form of ”ICT-outsourcing” has been known for decades (Dibbern et al, 2004).

The popularization of personal computers, the development of computer networks and internet technologies were several essential ingredients for the new technology. Cloud computing as a system has gone through substantial changes since initial utilization as a group of interconnected devices . For the Cloud as a concept to be viable, required extensive work by computer engineers. Firstly, problems of parallel computation were solved. The solution of the problem of how to run a given computational problem on a set of distributed resources simultaneously, resulted into grid computing and sequentially into the current architecture of the cloud.

The next problem was to sell computing facilities as a service to clients. In order to do so, the process of metrics collection (time of usage, resources usage, memory usage) was invented. The resulting technologies formed the basis of modern cloud metering facilities.

The last stage was a marketing effort in order to bring technologies to the masses. Marketing efforts resulted in invention of a new type of monetization, pay-as-you-go, combined with service subscription mechanisms. This type of scheme allowed to efficiently minimize costs for the service usage. As a result, these three combined innovations resulted in the modern cloud service facility. After the rise of Amazon Cloud Services (AWS) in 2006 and substantial upgrades of cloud computing services in 2008, the market experienced a new way of accessing complex ICT solutions and services without facing high upfront costs (Bryne and Corrado, 2016). ”Mix and match” solutions (McKenrick, 2011) helped cloud users to efficiently leverage ’best practice’ technologies on the market. This, in turn, enabled efficient learning channels for companies.

There are different deployment models of cloud computing, that companies might follow:

1. Private cloud
2. Community cloud
3. Public Cloud
4. Hybrid Cloud

Private cloud stands for a bespoke infrastructure owned by a single business and offers more controlled access to the IT environment for the business. Public cloud is generally owned by outsourced cloud vendors and gives access to many businesses using pay-per-use models. It could be ideal choice for SMEs with limited budgest and desire for quick and easy deployment of their IT resources. There is also a third type, hybrid cloud, which combines the benefits of the previous ones. It presents more specialized IT solutions that meet specific business needs. Community cloud is a shared infrastructure among several business owners.

Companies that provide cloud computing services (cloud vendors) operate various business models:infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS) and cloud application service (software-as-a-service, SaaS). IaaS providers offer an instant computing infrastructure, that is managed over the internet. The main IaaS services include servers to rent (virtual machines, VMs), cloud storage and databases, and security solutions for networking. PaaS providers also offer additional middleware, development tools, and business intelligence services in addition to IaaS tools. SaaS providers sell access or subscription for cloud based products and solutions, for example cloud CRM systems, Microsoft Office online (Office 365). In comparison to IaaS and PaaS providers, who sell cloud tools, software-as-a-service providers offer a complete solution and manages resources in the background, usually not involving any customer efforts. It is worth noting that the current study is focused mainly on impact of IaaS and SaaS providers.

Although cloud computing services emerged just a decade ago, its rate of progress cannot be overstated. Mind-blowing speed of changes caused significant processes of lagging behind for industries, governments, economies. One of the key benefits of the cloud is an opportunity to pool computing resources across a wide group of enterprises, in order to achieve 'shared' economies of scale as Jin and McElheran (2018) suggest. This enables smaller enterprises to compete with larger ones, that historically were able to operate at a scale and have an advantage to spread fixed costs of ICT across their outputs (Tambe and Hitt 2012, McElheran 2015).

2.2. Current state of the cloud market

The rapid development of the Internet has opened new opportunities for the cloud technologies, business decision makers and governments. The Cybersecurity Ventures forecasts that by 2022 there will be 6 billion Internet users in the world, whereas by now there are around 4.4 billion[1]. According to ONS, in 2018 the UK was ranked third out of all EU countries by the number of internet users, with a rate of 95% [2] (Figure 2).

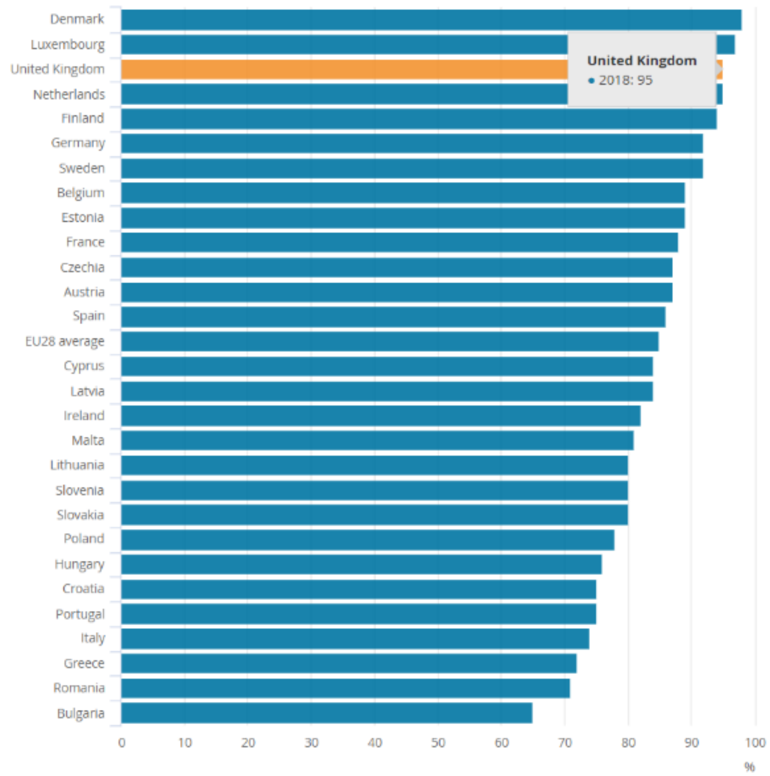

Figure 2. Recent internet users, adults aged 16 to 74 years, EU, 2018. Source: ONS, Internet users, UK

Cloud computing is one of the fastest growing activities that is directly linked to internet penetration. Anticipating the rapid growth, OECD forecasts that global public cloud computing market revenue will reach $331.2 billion in 2022, which is 54.5% higher than the 2019 estimate. It is expected that the cloud market in the UK will be worth around £9 billion by the end of 2020[3].

---

[1]https://cybersecurityventures.com/how-many-internet-users-will-the-world-have-in-2022-and-in-2030/

[2]Office for National Statistics, https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2019

[3]Tech UK "Cloud 2020 and beyond. Unlocking the power of the cloud". July 2019

In general, British private companies have a positive attitude to the cloud transition and 90% of them use cloud technologies to some extent[4]. Public organizations also support cloud adoption at different levels, but their utilization rates are still much lower[5]. According to a survey conducted by CIF, in 2018 nine out of ten UK companies have used at least one cloud service in their IT departments. This private statistic includes the use of any cloud-based application (Office365, Dropbox, WordPress, etc.). Also, the data suggests that cloud-oriented businesses use more of the delivery model than ever before. At the same time, 75% of cloud users utilize two or more services, while 40% utilize more than three. It indicates that there is a room for improvement in terms of cloud usage, as cloud consists of hundreds of useful services targeted to a wide circle of customers.

At the same time, official statistics posted on Eurostat states that only 41.9% of UK businesses adopted cloud technologies in 2018. Comparing with 2016, we might observe an upward trend in cloud adoption, e.g. the historical rate was 36%. There are a number of countries in Western Europe that outperform the UK[6], for example Sweden, Denmark, Norway[7] (see Figure 3).

We should note, that the difference between private (90%) and official statistics (41%) is significant, but there is a simple explanation for that. As we mentioned before, private sources include all cloud-based services used by businesses in calculation of adoption rates, while the official ones reflect the percentage of businesses who use cloud service more extensively.

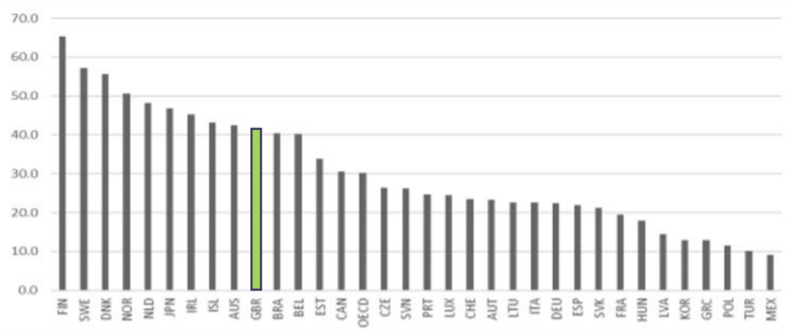

Figure 3. Cloud usage as a percentage of enterprises with ten or more persons employed. Source: OECD

According to the CIF research, 82% of UK enterprises state that cloud is dominant factor for them (Figure 4). Such an increase in the consumption of cloud services affects the amount of the budget allocated to the local ICT infrastructure. Cloud infrastructure already accounts for a large share of the IT budget (19%) and is expected to widen the gap between cloud and traditional on-premises infrastructure in the coming years (Figure 5).

[4] https://www.bain.com/insights/the-secret-to-more-cloud-adoption-in-europe-more-supply/
[5] Tech UK "Cloud 2020 and beyond. Unlocking the power of the cloud". July 2019
[6] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cloud_computing_-_statistics_on_the_use_by_enterprises
[7] ICT Access and Usage by Businesses Database, December 2018 http://oe.cd/bus

Figure 4. Cloud adoption rates and the number of cloud services used by busi-nesess. Source: CIF

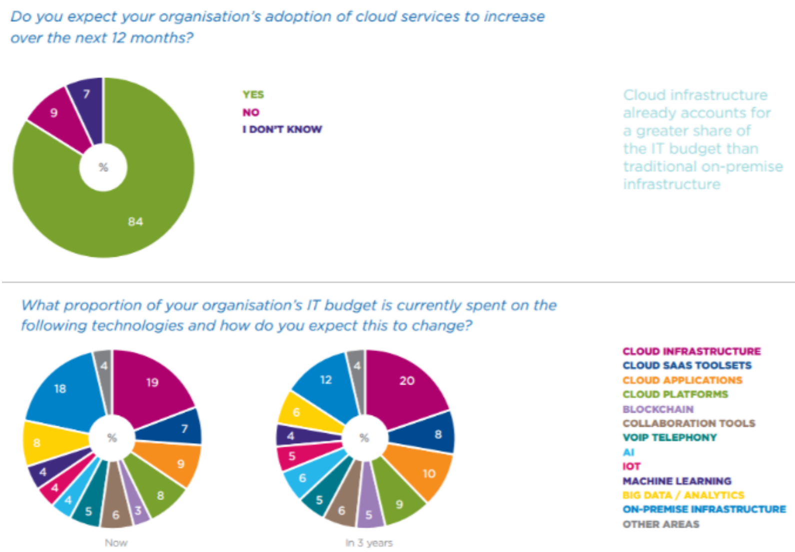

Figure 5. Busines expectations about future adoption rates and distribution of IT budget spending. Source: CIF

UK companies recognize agility, flexibility and scalability as major benefits of cloud utilization. The location of data centers is another important factor for businesses. Clients pay attention to data protection and compliance with data governance restrictions. It is worth noting that physical proximity lowers the latency rates (or increases time to get the data over the internet). Currently there are five major locations of data centers In the UK t (Figure 6).

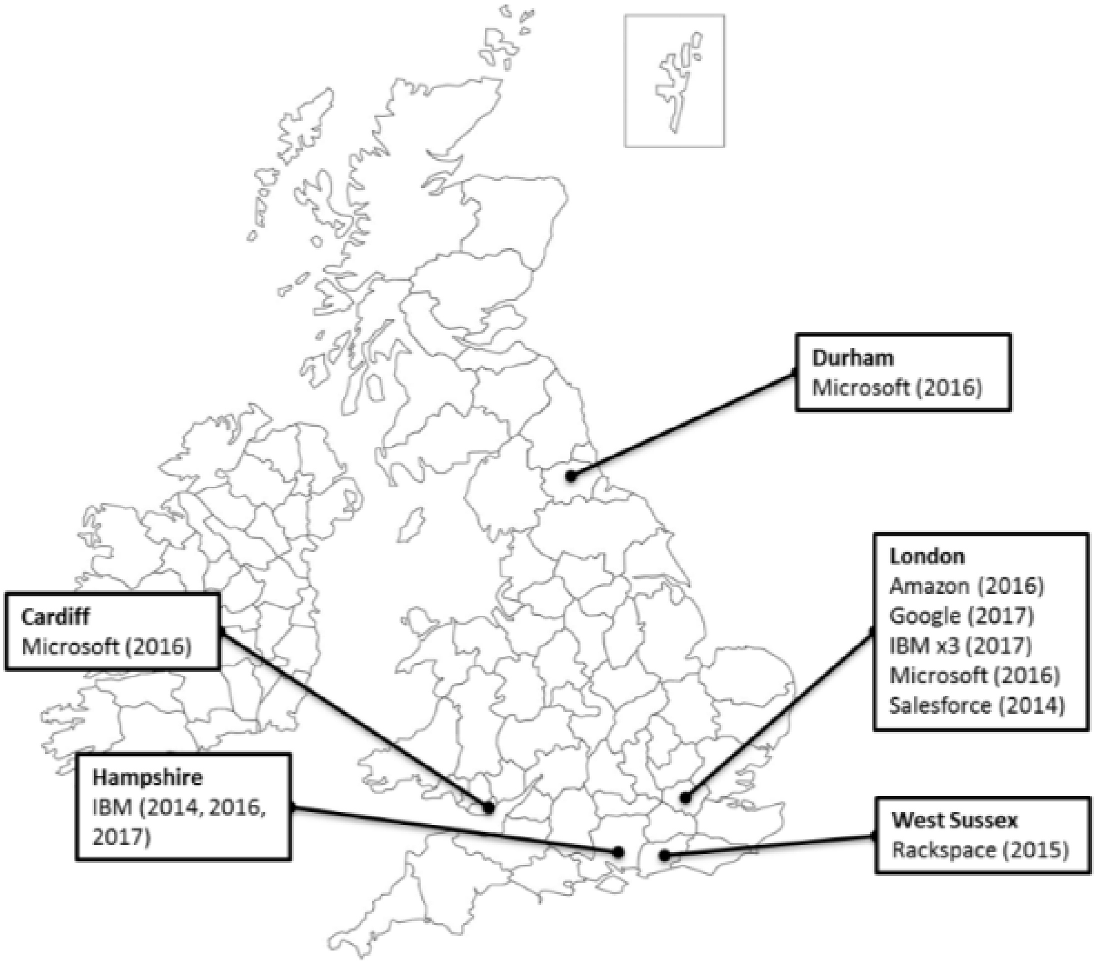

Figure 6. Map of the UK data centers by main cloud providers Source: Coyle et al. (2018)

As we can see from Figure 6, among the main UK providers are Microsoft, Amazon, IBM, Google, Salesforce and Rackspace. Also we can highlight Alibaba, Oracle and SAP, which are presented both on the global market and in the UK. The main services that cloud vendors provide are data storage, computer processing and communication in and out of the data centers.

Amazon AWS was the first global company that had entered the European market. In November 2007, the company opened their first data center in Dublin ((Coyle et al., 2018)). European expansion continued by opening data centers in Frankfurt (October 2014), London (December 2016), Paris (December 2017) and Stockholm (2018). Now AWS is present in four European regions, while it has five centers in North America, one in South America and eight in APAC region.

Microsoft started providing cloud services in Dublin from July 2009. Further expansion affected Amsterdam, London, Durham and Cardiff in 2016. As a result, the company operates now in eight regions in Europe. As was mentioned above, there are four centers located in the UK, two in France and Germany,with additional ones in Ireland and the Netherlands. Starting from 2010, Google also entered Europe by investing in a Belgium data center. After this, Google continued its expansion in Finland (2011), Dublin (2012, 2016), London (2017), Frankfurt (2017) and the Netherlands (2018). The company states, that the data center in London was aimed to reduce latency rates in the UK by 40-80%.

IBM is currently present in three countries: UK, Frankfurt and Paris. The UK locates six data centers, which started functioning in 2014 and all of them are in London. The first data center in Chessington (London), was registered to have space for 150 racks, 4,000 physical nodes, 15,000 servers and a floor space of 10,000 square feet. Salesforce established its first data center in London in October 2014. Another two data centers were opened in Frankfurt and Paris in 2015. In 2018 Salesforce acquired MuleSoft with the aim to boost PaaS products. Gartner calculates that the global public cloud service market has reaches $258 billion in 2020 and is expected to grow by 18 % in 2021 [8]. Figure 7 provides a comparison of predictions for growth for major types of cloud computing business models.

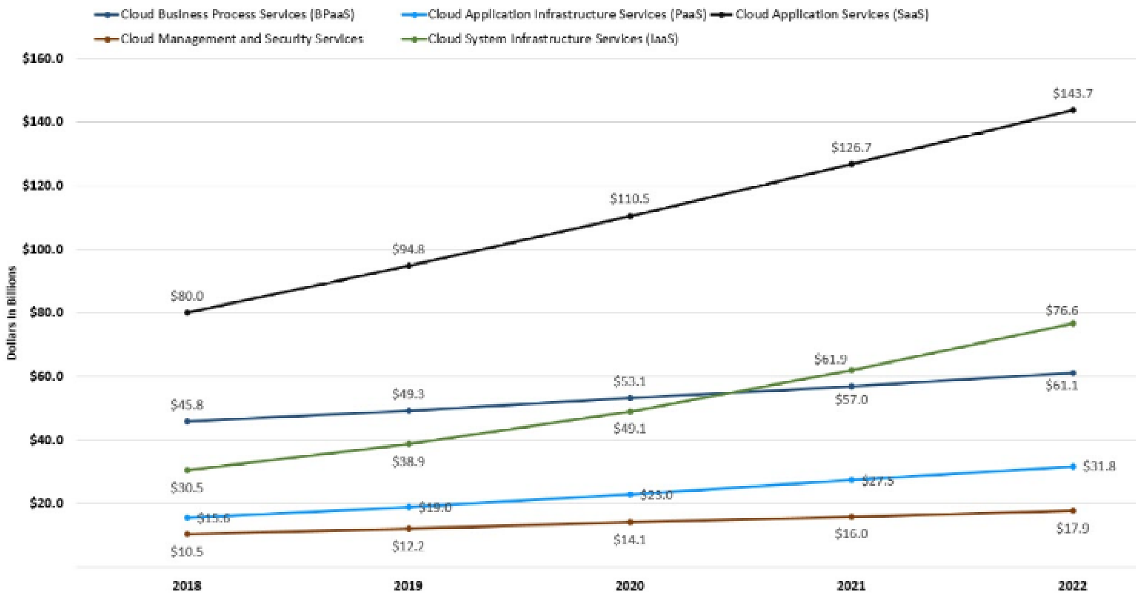

Figure 7. Worldwide Public Cloud Service Revenue Forecast, 2018 - 2022 (Billions of U.S. Dollars). Source: Gartner

2.3. Example of Cloud computing services and cloud computing platform

In this section we ilustrate the typical services that are offered by a majority of infrastrusture-as-a-service or platform-as-a-service cloud providers. We use Amazon Web Services (AWS) as an example of the cloud providers, that provides comprehensive cloud services in the UK since 2016. In general, 'the cloud' is a set of different services and architectural solutions. However, it is possible to broadly classify all of the cloud services into 4 parts: computing machinery and computing services; cloud storage and databases; load measurement and development operations (devops) tools; and security management services. We discuss a few of examples of popular AWS cloud products.

EC2 or 'elastic compute cloud' is an example of computing services, provided by Amazon. It is a virtual machine (VM) service introduced by the cloud provider in March

[8]https://www.gartner.com/en/newsroom/press-releases/2020-11-17-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-grow-18-percent-in-2021

2006. The concept of a virtual machine, as an abstraction layer over computing hardware resources, emerged several decades ago. This technology allows you to run several different operating systems with their own independent share of resources on top of one physical machine. It means that on one physical computer you can have several virtual computers. Using this concept, Amazon allowed their users to choose among one of several bundles of computing power, depending on users‘ needs. The service is called 'elastic' because the basic computation facilities allow to be rapidly scaled up to 40%, based on the utilization demand. The concept of renting of VMs is similar to leasing, however you lease only a part of physical machine. Netlix, Amazon Alexa are among a few of the examples of companies whose business was built on top of the EC2 services.

Lambda function is another example of a computing service. It is a serverless computing facility introduced by Amazon in November 2014. Lambda function is simply a piece of code written in some programming language (java, javascript, python,etc.), that can be executed automatically or on user demand in the cloud. In comparison to EC2 services, where users constantly uses computing resources, lambda service utilizes computing resources only in a short period of time of the code execution. Thus, lambda services enable the achievement of 100% utilization of physical machines, and saves money for customers, as there is no idle time when this resource is not utilized, at all.

Cloudwatch is a part of Amazon services to control usage and metrics at scale. It is a part of the various load management and development operations tools that makes the process of deploying and controlling resources in the cloud much easier. Cloudwatch is a good example of the metering facility, that is a crucial ingredient for any cloud (Figure 2.1). It includes several layers of usage statistics, hardware and traffic load, utilization charts and security information in order to have sufficient level of control over the cloud resources.

Identity and Access management (IAM) is a cloud security management service that allows managing security and preventing vulnerabilities of cloud resources at a scale. It allows to organize single sign on opportunities for users, which means that uses can have one profile and one set of credentials in order to have access to a preconfigured pool of heterogeneous resources. The IAM tool is a very important part of the cloud, which is generally underused by the broad public of customers, leading to security leaks and cybersecurity problems. However, it is worth noting that security services and practices, offered by Amazon, can exceed the majority of custom practices at companies. Thus, cloud usage is generally associated with raising the cybersecurty protection for a company.

Dynamo DB is a proprietary non-relational database, provided by Amazon since January 2012. It serves as an example of cloud databases facilities. It enables fast and simple data management and is one of the simplest databases with efficient web interface, that simplifies database management and development of programmes. S3 is a cloud storage introduced by Amazon in March 2006. It enables users to store different types of data in the cloud, similarly to OneDrive or Google Drive services. Coursera, Netflix and Edx use it to store their videos.

Having explored the basic set of products available on any cloud platform, we now give some examples of heterogeneous products, that are built on top of the basic products. These include various AI tools and ready to use architectural solutions, that simplify development. AWS Well-architected is a knowledge platform for democratized computing with efficient knowledge sharing. It was created 'to help cloud architects build secure, high-performing, resilient, and efficient infrastructure for their applications' (AWS Well-Architectured documentation). The platform enables spreading the best and most efficient use cases and best practices for all AWS customers, as well as providing connections to AWS partners that can help to implement cloud solutions for companies. Amazon Connect is a cloud – based automated customer support service which uses Artificial Intelligence to manage telephone calls campaigns. It is one of 200+ services[9], that are built on top of the basic Amazon solutions, such as EC2 and S3.

2.4. Conclusions

In this chapter we covered the general information about cloud computing, its history and trends. We explained, why cloud computing technologies share most of the properties of the conventional ICT. Cloud computing emerged as a result of the continuous development of the computing industry and computer networks.

Starting from the early 2010s, cloud computing started to win broad markets and customers. After several price declines, and introduction of more cloud data centers in the UK, it started to be extensively used by businesses in the UK. The vast amount of services that are being built on top of basic cloud resources, enhance the variety of the cloud usage and makes solutions affordable for businesses.

Today the cloud can be deemed as a general purpose technology, whose value grows with the amount of users that use the technology, experiment with it and implement new solutions on top of it. Cloud allows users to save cost and development time and avoid irreversible expenses. Cloud providers compete to give the best price and quality for their services.

[9] https://aws.amazon.com/what-is-aws/

3. Literature review

Cloud computing technologies have many similar characteristics to their traditional ICT predecessors. Both modern and traditional technologies impact firms' structure, and lead to coordination and communication costs. Both technologies are associated with price reductions and changes in the quality of outputs. Monopolization of the market and increase of producers' market power due to lock-in effects can be viewed as an important concern for policymakers.

However, there is a lot of evidence about the growing heterogeneity of new technology impacts on the economy. Due to the complicated nature of these technologies, various impact channels appear to divide the market into digital "winners and losers". Moreover, there is evidence that ICT intensive sectors benefit the most from ICT technologies. This means that ICT technologies will drive the polarization more, creating a broader gap between high productive, successful companies and low-productive unsuccessful ones. This digital divide is already visible in the UK economy, and the difference is predicted to grow. This is another concern for the country and its policymakers. One aspect of both traditional and modern ICT based technologies is the knowledge economy transition, whereby knowledge becomes the most valuable part of the modern economy.

The two most distinguishing differences between modern and traditional ICT effects highlight the importance of our study. Firstly, in contrast with conventional technology, which leads to the intensification of ICT capital and ICT capital deepening, cloud technologies require less traditional ICT capital expenditures by creating a "shared ICT" environment, where several firms can share same hardware resources. Cloud computing thus creates a less capital-intensive but more skill intensive production environment, as the new technology requires highly skilled specialists whose value grows. Secondly, the traditional ICT effect was positively correlated with the size of the company. Big companies had a bigger effect on the utilization of ICT technologies and had higher propensity to use ICT in their work environment. The cloud computing effect appears to be reversed. It was shown that Cloud expenditures increase survival rates for small companies, and decrease survival for large ones (Jin and McElheran, 2018).

3.1. Impact of traditional ICT: productivity

A great deal of research was devoted to understanding the impact of the first wave of ICT on the economy, mostly during the mid 1990s - early 2000s periods. The main conclusion was that the impact of ICT investments was seen long after the initial investments had been made. However, significant benefits from ICT did emerge, mostly arising in the decade from 1995-2005.

In the first half of the 1990s, the initial impact of ICT technologies was shown to be insignificant. There is a famous quote regarding the lack of visible benefits, 'you can see the computer age everywhere but in the productivity statistics' (Solow, 1987). Despite the fact that expenditures on ICT were growing, there was no visible productivity increase associated with those investments. Stanley and Roach (1987) showed that increases in computer investment over the previous decade did not increase productivity for white collar workers. Berndt and Morrison (1995) and Morrison (2000) stated that the gross marginal product of 'high tech capital' was less than its cost as seemingly labor saving investments turned into expenses for increased professional labour demand. The belief that new technologies would drive productivity growth did not find any significant evidence. The general dominating thought was that ICT is a sand castle, as more than 40 years of growth in ICT research and development and rapidly increasing investments into ICT did not produce any economically significant results. This economic puzzle was called a 'productivity puzzle' after the above famous statement of Robert Solow.

Only during the second half of the 1990s did economists present contrary findings. Brynjolfsson and Hitt (1996) and Lichtenberg (1995) found a significant positive connection between ICT investments and productivity, while Greenan and Mairesse (1996) suggested increases in output, and productivity increases in government activities at the process level were found by Mukhopadhyay et al. (1997). Jorgenson and Stiroh (1995), Oliner and Sichel (1994) and Stiroh and Jorgenson (1999) suggested that technical progress in computing facilities positively contributed to the real output growth in the US. Lichtenberg and Lehr (1996) and Dewan and Kraemer (2000) found positive effects at the country level as well. The initial productivity paradox was considered to be resolved (Dedrick et al., 2003), as positive impact of ICT were found in firm level studies and later in aggregate ones (Jorgenson and Stiroh, 2000; Oliner and Sichel, 2000; Timmer et al., 2018).

Overall, ICT technologies were found to have positive and significant impact on all aggregate levels of the economy, however the micro level impact was assessed earlier. It was shown that ICT positively influences firms' output and a reverse causality was unrealistic. Despite the fact that the initial studies failed to find any significant impact of ICT, further research supported the positive impact hypothesis. Two possible issues that influenced the research on ICT productivity are the significant delay between investment and results of ICT as well as the data granularity issues (Brynjolfsson and Hitt, 1996). Presumably, it takes a long journey for any technology: starting with productivity increases at the process level, then expanding it's effect on the firm level and finally - industry and country level.

A major lesson learned from the research was that ICT investments may cause structural changes inside the enterprise, driving additional organizational changes and complementary intangible investments inside the firm. A large literature indicates that ICT leads to greater cost savings in business coordination. Brynjolfsson et al. (1994) found that ICT reduces the level of vertical integration for the company by reducing transaction and coordination costs. Thus, companies can shift from hierarchies to flat organizational structures and better coordinate business activity. DeStefano et al. (2018), Hitt (1999), Gurbaxani and Whang (1991) and Clemons and Row (1992) also found that an increase in information technology capital is related to the decline in average firm size and reduction in vertical integration.

Bresnahan et al. (2002) found that ICT investments lead to higher decentralization processes among enterprises as new technologies allow for better work distribution. A 'computer mediated transactions' concept was extensively discussed by Varian et al. (2004). They facilitate the collection of additional information about customers and their behavioural patterns, allowing more advanced price discrimination strategies. Computer transactions also increase visibility, decrease costs for information retrieval and time to make decisions, operation efficiency, and fraud detection, thus allowing for more flexible and cost effective business models.

While allowing for unprecedented transaction costs savings across the value chains, ICT goods are known to become less expensive and thus more affordable by business through time. Decreasing prices of ICT goods have led to substantial factor substitution for other production inputs (Chwelos et al., 2010; Stiroh, 2002; Lin and Shao, 2006; Dewan and Min, 1997; Timmer et al., 2018). The long term ICT impact in many of these studies was initially related mainly to an input efficiency enhancing mechanism. Less was known about the total factor productivity effect.

However, it was shown later that effects of ICT extend beyond input usage (Bosworth and Triplett, 2007). At the firm level there are positive impacts on total factor productivity due to network externalities and ICT-driven innovations such as new business processes and more efficient supply chains. (Chou and Shao (2014); Kim and Narasimhan (2002); Kim et al. (2011); Brynjolfsson and Hitt (2000), Brynjolfsson and Hitt (2003)). As a result, ICT exerts a positive improvement in production processes through the ICT capital deepening (Oliner and Sichel (2000); Jorgenson et al. (2008), Stiroh (2002)), and input substitution processes (Chou et al., 2014). Timmer et al. (2018) discuss productivity growth across EU countries, accompanied by higher labour productivity (especially visible in ICT intensive industries), higher ICT capital shares and lower labour shares across major industries, as well as steadily growing demand for skilled workers.

3.2. Impact of traditional ICT: intangibles

As the ICT impact extends far beyond invested resources, we next consider impact of technologies on intangibles of the company. Jorgenson et al. (2006) noticed that firms using ICT in a creative and innovative way dominate in productivity growth in the US. Effects are even greater when investments are aimed not on cost reduction but on economic growth (Mithas et al., 2012). This makes ICT a creative instrument that can be used for creative development and innovation, not just as another instrument for cost reduction and automation.

In addition, as mentioned above, ICT effects are likely to be observed after a considerable amount of time spent on adjustment of processes, involving a learning curve (Mithas et al., 2012). This fact is supported by evidence by Brynjolfsson and Yang (1997) who found that, on average, for 1$ of direct ICT investment there is 9$ of additional intangible investments that are needed in order for ICT investments to be effective. Lee et al. (2005) also noted that there is some minimal level of ICT capital stock needed and some minimum level of accumulated ICT expertise in order for an ICT impact to be evident. As noted by Nicholas Garr (2003)[10], it is not enough to just 'invest' in ICT, but it takes considerable time and effort to make those investments work. As everyone invests in ICT, investment alone does not give a competitive advantage to firms, but additional expenditures and innovation in work processes and integration divides companies into 'winners and losers'.

Bresnahan et al. (2002) also suggested a strong relationship between the level of ICT investments and investments in human capital. A significant amount of work suggests that investments in technology equipment, demand for skilled workers and 'knowledge capital' are strongly connected (Berndt et al., 1992; Berman et al., 1994; Autor and Krueger, 1998). Krueger (1993) argued that ICT is a skill-based technology and its value strongly depends on skill levels available to the firm or country.

[10] https://hbr.org/2003/05/it-doesnt-matter

Chou et al. (2014) noted a platform aspect of ICT that enhances the technological level of processes. According to Miozzo et al. (2006), a platform technology is characterized by rapidly falling costs, plentiful supply (sometimes unlimited, you can never run out of html' (Varian et al., 2004)), and numerous applications to products and processes. Also. ICT complements innovations and generate new synergies (Varian et al., 2004).

The role of ICT as a 'general purpose technology' that facilitates new innovations was widely recognized in the literature (Bresnahan and Trajtenberg, 1995; Helpman, 1998; Varian et al., 2004). General purpose technologies (like electricity or steam in the previous century) are used to build and create new inventions that use these technologies as a basis. The GPT technology is combined with other technologies, with a specific unique combination that creates a new value. Once new technology becomes available, it can be used as an input into new technological inventions.

Schumpeter (1934) uses a term 'combinatorial innovation' to point out that new inventions are coming in waves or 'clusters': once a new component becomes available, it opens a way for a branch of new inventions. '... as soon as the various kinds of social resistance to something that is fundamentally new and untried have been overcome, it is much easier not only to do the same thing again but also to do similar things in different directions, so that a first success will always produce a cluster.' (p 142). The invention of wireless transmission is a good example of combinatorial innovation: once invented, it gave birth to a variety of new technologies built on top of the initial concept - radio, television, mobile telephones, Bluetooth, WiFi, wireless internet protocols (WAP, GPRS, 3G, etc.). The development of internet networks gave birth to the World Wide Web, cloud computing, new messaging communication standards and social networks, and blockchain technologies. We should note that unlike radio, which required more than 100 years to build new inventions on top of it, it took approximately a decade only for network technologies to cause a new wave of technological revolution (the 3rd digital revolution). In comparison, just several years were needed for cloud technologies to enable new business models, create new business giants, such as Uber, Airbnb, Amazon, Netflix, and give spread to new technologies such as Artificial Intelligence, the Internet of Things, etc.

Thus, the modern ICT economy should be viewed as a network economy with many network externalities, including a greater propensity for new innovations and new opportunities, emergence of new businesses etc.. Such network effects benefit both purchasers and stakeholders of the technology (Chou et al., 2014). Communication systems and software are good examples of ICT externalities that have immense network effects (Brynjolfsson and Hitt, 1996; Shy and Oz, 2001).

Overall, there are two types of network effects, the direct and indirect. Direct network effects are observed for technologies whose values are directly enlarged with a number of additional users, according to Hajji et al. (2012) (for example, email or social networks). Yet, at the same time there is also the indirect network effect in ICT. One example of indirect network effect is the interconnection between the ICT diffusion and telecommunication infrastructure (Kauffman and Techatassanasoontorn (2009), Oxley and Yeung (2001), Robison and Crenshaw (2002), Dedrick et al. (2013)), as more computer usage leads to higher usage of internet and vice versa. Another great example of the indirect network effect is Amazon Cloud. As more users came to use the cloud, the more user-generated documentation appeared on the internet (Stack Overflow recommendations and best practices, etc.). This in turn indirectly influenced ease of learning and integration of new cloud vendors on the market driving general interest and pushing Amazon to faster innovation and development of their products.

Varian et al. (2004) described a simple intuitive framework beyond network effects of technologies. The general network effect can be described as a dynamic system with three equilibrium points, two unstable and one stable. Figure 8, which for simplicity assumes perfectly elastic supply curve, illustrates these equilibria.

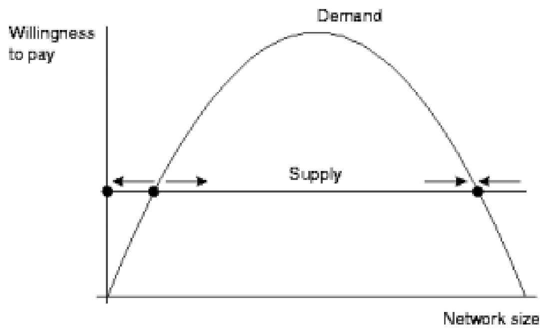

Figure 8. Network effects of a good. Source: (Varian, 2004).

The first stable equilibrium point is a zero sized network of product. When the network size is less than some critical point, a quantity of product sold is decreasing (the willingness to pay is lower than the price for the good) and network benefits cannot play a significant role in popularizing the product. When the number of users reaches above a critical mass (2nd equilibrium), then a positive feedback mechanism (network externalities) increases the value of the product and its demand. The demand will gradually decrease as the size of the network grows because of selling to consumers with progressively lower willingness to pay. However, if the product does not succeed gathering the critical mass of its users, it will gradually fall back to the zero-demand point. This simple model elegantly describes the general demand dynamics in modern network economies.

In summary, ICT requires investment in complementary, often intangible, investments. Brynjolfsson and Yang (1997) found that the market value of big, established firms increases 10 times more than direct information technology investments. This finding serves as an illustration that direct ICT investments are usually complemented by investments into additional software, skills and specialists, new business processes and additional organizational transformations. Complementary investments are required to create an important environment where invested technologies may work and bring utility to the firm. Otherwise, the positive potential impact of new technologies and practices may be restrained by organizational immaturity, lack of skills and specialists to deal with the technology, restrictions of the legacy software etc.

3.3. A new technology: Cloud computing impact

Recent research highlights some changes in the way traditional ICT impacts productivity. DeStefano et al. (2018) showed that government incentives to invest in traditional ICT restricts companies from experiments and hinders development, resulting in lower cloud adoption rates. Similarly, Jin and McElheran (2018) empirically confirmed that usage of traditional ICT capital is associated with greater likelihood of failure for young firms but they argue that this effect is the opposite for modern ICT services, such as cloud). This evidence suggests there are structural changes in the way modern ICT technologies impact the economy and individual firms. New ICT technologies provide 'ready to use' solutions, that have better speed, scalability and modularity of ICT services, compared to traditional ICT. Another considerable difference is an observed ICT capital decrease in firms (Coyle et al., 2018), as opposed to ICT capital deepening in the previous ICT era, related to the growing use of cloud services that are classified as current expenditures of these companies.

However, because of the tight connection of the previous computer era (3rd industrial revolution) to the cloud and AI era (4th industrial revolution), the main paths of impact and causalities remain the same. AI and Cloud technologies as the major part of ICT innovation nowadays have all common characteristics with any technology in the IT sphere. There are several aspects that are present in any IT technology and in cloud as well: the ability to drive complementary organizational changes and increase the productivity by reducing costs and enabling firms to increase output quality in the form of new products or through intangibles.

Cloud computing helps to restructure companies expenses and helps to utilize powerful computing resources, that are scalable and provided on demand. Usage of cloud computing technologies is inevitably changing the shape of an enterprise. AI models build on top of the Cloud which could be seen as the next general-purpose technology. AI could possibly replace humans in many areas which will lead to dramatic organizational changes inside companies. Thus, Cloud technologies bring a combination of technical and organizational innovations to companies.

Cloud computing influences productivity of the company through three main channels. The automation channel reduces labour and transaction costs due to decrease of manual work through process automation and AI. The intangibles channel increases the efficiency of the work, quality and diversity of the end product due to increased knowledge and expertise. The financial channel includes the reduction in the financial risks connected with minimization of upfront investments needed for cloud technologies, fine-grained control over expenses because of smart metering and pay as you go payment schemes, greater security control, economy of scale and scalability. As a result of these influences, there is a substitution effect from traditional IT capital to external IT services such as cloud solutions etc. (Jin and McElheran, 2018).

Acemoglu and Restrepo (2018) characterizes automation as the general increase of the task amount, that can be fully or partially automated, which means lower, or no human labour involved in the process. The Cloud and related technologies (Artificial Intelligence, Internet of Things) enable automated information retrieval and processing, which speeds up the business decisions, reduces uncertainty and complexity for decision makers (Wang et al., 2014). As Levy (2000) suggested, computers are most likely to replace workers that perform mostly rule-based decisions, while helping people who perform non procedural cognitive tasks to be more productive. Examples of job automation and job augmentation are medicine (Wang et al., 2011), physics (Sevior et al., 2010), autonomous vehicles (Yadan, 2019) and conversational AI (Mead, 2017).

The intangible channel of the cloud computing impact is mostly visible though democratized computing changes. In comparison to 'privileged computing' i.e. computing available only for limited group of companies, which was how ICT was perceived in the past (Bloom and Pierri, 2018). Thanks to the democratised nature of cloud computing, it facilitates easier information sharing in terms of communication (Bloom and Pierri, 2018; OECD, 2015) and knowledge sharing (Mohamed and Pillutla, 2014).

It is also important to mention, that because of the simplicity of the knowledge and experience acquisition, the intangible impact of the cloud is even more beneficial to young and small firms enabling more dynamic development of the economy (DeStefano et al., 2018). In comparison, large firms leverage geographical dispersion opportunities and ability to adaptively scale their business (Bloom and Pierri, 2018; OECD, 2015). Jin and McElheran (2018) suggest that cloud and related services provide the means for young firms to achieve better performance before they will learn about their needs and achieve a scale of their own.

The financial side of the cloud computing impact is one of the main arguments for decision makers who consider using the Cloud. By allowing them to avoid irreversible costs of acquiring expensive hardware, and by "renting" it from cloud vendors, cloud adoption gives companies a flexibility and freedom of experimentation, in the face of business uncertainty, so it reduces financial risks Decker et al. (2014). Due to pay as you go schemes of payment, it is possible to change expenses frequently, efficiently scale up and scale down usage of hardware resources.

The cloud gives businesses a fast access to powerful computing resources and ability to change the usage of resources according to changes in their demand Jin and McElheran (2018). This 'demand non-rigidity' has direct economic impacts on businesses, allowing them greater ability to adjust to changing market circumstances DeStefano et al. (2018). The lower need for capital and equipment investments can provide an opportunity to invest more in R&D and marketing (OECD, 2015; Columbus, 2013), thus facilitating faster development of the firm. As financial barriers are lower due to the cloud technologies, investors could change their investing behaviour by providing smaller amounts to more firms, thus, acquiring more diversified portfolios (Ewens et al., 2018).

Jin and McElheran (2018) argue that the financial impact is mostly related to complementary investments under uncertainty. Uncertainty regarding investment opportunities constitutes a big part of the picture about cloud computing benefits, especially when executives lack information about their profit opportunities (Jovanovic, 1982). This relates to an earlier literature, e.g., Dixit and Pindyck (2012) explained that firms tend to underinvest in the face of uncertainty, in order to be flexible (see also Jovanovic, 1982). With cloud services, firms can benefit by generating real options from their low cost experiments (Kerr et al., 2014; Thomke, 2003). Firms experiment with their processes, customers and partners thanks to cloud technologies (Palmer, 2012).

Guiso and Parigi (1999) and Bloom et al. (2007) also provide evidence on investment delays that accompany uncertainty. Given that new firms face the highest business uncertainty (Knight, 1921), cloud provides disproportionately high benefits for young companies, by providing a flexible way to control expenses, avoid irreversible investments and learn how to become more efficient (Palmer, 2012). Jin and McElheran (2018) predict that firms most influenced by cloud sectors are those where ICT knowledge is of a particular value, learning is difficult and the risk of survival is higher (leaving less space for mistakes when competition is high and profit margin is low).

Although, Ewens et al. (2018) argue that this flexibility is most beneficial in ICT intensive services, there is an increasing demand for cloud-related services more broadly, including data collection, storage, analysis and communication (Columbus, 2013). Kerr et al. (2014) describes how manufacturers that create their own ICT products, still can benefit from the cloud technologies by experimenting with types of ICT technologies and available standard solutions before building their own customized ones. Moreover, certain parts of the industry heavily relies on computer-aided design (CAD), which became more accessible thanks to software as a service (SaaS) solutions. Additionally, cloud solutions offer a great instrument for standardization of products and interconnections between vendors and partners. The consumer electronics market also became dependent on the cloud solutions, as its ability to quickly rent or scale ICT infrastructure is highly valued for business growth and experimentation (Jin and McElheran, 2018).

In general, cloud allows for greater flexibility, as cloud gives more financial freedom to the company (Jin and McElheran, 2018). As a result, there are lower entry barriers and a greater competition on markets (OECD, 2015; Etro, 2009). As an example, European Commission (2017) predicted, that between 2018 and 2020, there will be 1.6 million jobs, 303 000 new businesses and 449 EUR billion of revenue created due to the cloud computing adoption and usage.

The most immense effect of cloud, however, is still waiting for us in the future. Cloud as a general-purpose technology drives new businesses (Etro, 2009) and new future technological innovations, that uses cloud as a basis. Thus, the role of cloud computing technologies may be vastly underestimated as the real effect is the multiplicative effect of future innovations.

Nevertheless, certain tendencies may be seen even today. Cloud has already opened a wide area of opportunities for Internet of Things technologies, that will bring further digitalization of tools and mechanisms that citizens and workers will use in their everyday life. IoT devices are already changing the landscape of marketing, healthcare and defense industries. Cloud and IoT are tightly connected to the new product of cloud technologies - artificial intelligence, which gained its growth thanks to networks, big data and available cloud computations. Cloud enabled faster growth and implementation of CRM and ERP systems that facilitate easier and more precise business management and decision making. Cloud gave rise to many start-ups, new ways of digital experimentation and learning.

In the next section, we proceed with the study of the aggregate impact of cloud computing on the productivity of the firm. The main hypothesis supporting the rationale for such framing of the study is that the cloud technologies are the fourth wave of the industrial revolution (Schwab, 2017). After several years of the development of new technologies we should see emerging productivity effect of it, that will undoubtedly result in the next economic boom. We should draw a parallel with the third industrial revolution (Greenwood, 1997), which started with the boom of investment and overall productivity slowdown, continued with emerging evidence of its positive impact on the economy, and ending with dotcom boom which in turn placed ICT as one of the strategic determinants of economic growth.

4. Methodology and data description

4.1. Cloud usage indicator

Data about cloud usage and companies expenses are sparse in the UK. The Office for National Statistics carries out an e-commerce survey that covers some broad general questions, for instance, whether firms use cloud technologies. These surveys have no precise financial statistics of expenditures and investments in the cloud technologies. Moreover, a study based on a sample of approximately 2,000 respondents may not be sufficienty representative. In addition, each survey contains a different subset of companies in consecutive years so it is not possible to track firm-level performance changes related to the cloud usage. There is also publicly available information for public service companies and organisations (e.g. NHS, Transport for London), which contains monthly data about cloud expenses on services and specialists. However, the sample is also restricted, and the specific features of government organisations, such as lack of competition, means this data is of limited use.

It is theoretically possible to utilise factual data about cloud usage based on direct statistics gathered by private companies (Amazon Web Services, Google Cloud and other cloud providers). However, access to this type of data is difficult because private companies rarely disclose such statistics due to internal policies and external obligations.

Therefore, we decided to generate our own dataset by gathering cloud usage statistics using companies' website metadata. Every website has DNS metadata left as a publicly available footprint. The Domain name service (DNS) is a register of companies' sites along with physical addresses of actual servers (Internet Protocol addresses or IPs of the server). For simplicity, we can compare the DNS to the address book of companies, where we can get information about the physical location of an entity by its name. Domain name service records can be considered as key-value storage, where the key is a company website (or website name). A value contains essential information about the website location (physical IP address of the site, website owner, name of a vendor who hosts a company website, last year of ownership). This data is a publicly available record that contains the information needed in order to access a website. Apparently, the site is not directly accessible by the website address (like www.company.com), but only through the IPv4/Ipv6 address, whose purpose is to identify the address of the server, where the web site is hosted (e.g. 172.16.254.1 according to Ali (2012)).

In general, every company chooses a way to host their website. Companies can build the website using their servers, employ vendors to host a website for them (hosting providers) or use cloud-related infrastructures [11]. The first option is to buy the server equipment, connect it to the Internet and set it up in order to host a website. In this case, the company is classified as a non-cloud user. Option two is to use a third-party hosting provider, which means to rent a ready-to-go server with pre-installed software. This case would be also classified as a non-cloud deployment. However, the second option may not work with companies that require either non-standard or advanced web services (such as a complex website system with a content delivery network, cloud storage or distributed architecture).

Alternatively a company may go with the third option, namely cloud hosting services. Either by using its ICT resources or by obtaining help from a cloud consultancy firm, it

[11] https://startbloggingonline.com/how-to-host-a-website/

Table 1. Cloud information

Parameter	Value
Name	King's College London
Domain	kcl.ac.uk
Domain registered on	2007-01-01
Expires on	2020-01-01
Last update	2018-11-24
DNS	*.kcl.ac.uk, *ja.net
Servers	Apache, Varnish
Additional services	Outlook.com Mailchimp

is possible to build its cloud web deployment. Such deployment opens a door of opportunities, including more efficient data storage and processing, information retrieval, cloud computing and artificial intelligence facilities.

Therefore, a name mentioned in the DNS record can be a name of the company itself, traditional web hosting vendor, or cloud technologies vendor. When the firm hosts its website using any of the three options, the DNS registry is updated with the new information. An assumption is that if a company is spotted in using a cloud vendor to host their website, there is a certain probability that it also uses other cloud-related technologies. The primary rationale beyond this assumption is that it is suboptimal for a company to use cloud deployment for a website without using any other cloud-related technologies. If the company registers its website with cloud providers, for example, Amazon or Azure, they will use some additional cloud services as well (at least virtual machines to host their website).

Table 1 contains an example of what is possible to retrieve from the DNS registry. From the information gathered over the Internet, one can state that kcl.ac.uk was registered as a domain name in 1970, its DNS record was registered in 2003, and the last update to DNS records was made in 2018. For managing web page requests, King's College London uses its internal servers (with Ubuntu, Apache server and Varnish cache service). It is also linked to one of the servers of Janet Network, which is an educational network and cloud services provider. Probably, King's uses Outlook.com as a cloud service provided by Janet Network. It also utilises MailChimp for creating email campaigns. From the information mentioned above, one can infer that King's College London uses cloud services. This inference was confirmed by King's College London IT services department.

We track the changes in companies' web hosting infrastructures and their cloud usage patterns over time. We check whether the company hosts the website using cloud vendor services (SaaS or PaaS providers). We also add useful information about the usage of modern technologies by utilising additional website metadata. The available information discloses additional services that the company uses such as Gmail, Outlook, Salesforce, SharePoint and other technologies.

In order to build a cloud usage indicator, we used a history of DNS records generously provided by SecurityTrails. This cybersecurity company gathers DNS historical metadata for more than 3 billion websites worldwide. The data contains detailed information about the usage of specific web hosting servers and providers at a certain period of time. We constructed cloud usage statistics for all companies listed in the FAME financial dataset from 2008 till 2020, who have a website (see a detailed description of the FAME dataset in the Section 4.2). We parsed DNS records for these companies and extracted names of cloud vendors that hosted a website for the company's benefit.

Then we perform an automated search in Google using its Python API in order to find a website of the hosting provider given its name. We used Python and web parsing libraries (Scrapy) in order to web scrape the main page from all websites of hosting providers. We gathered a dataset containing all textual information from the company's web page.

We also used an additional step in order to track the history of web hosting providers. Since a number of providers started with offering conventional hosting services, and moved to the cloud provision later, we utilize WaybackMachine[12] - a platform which provides historical snapshots of websites. The platform contains snapshots of websites with the frequency of several times a year for less popular websites, and up to several times a week for websites with a lot of visitors. We scrape sites of hosting providers on a yearly basis, aiming for a snapshot made at the middle of each year. If the snapshot is not available, we search for the closest available date.

We utilise web parsing and natural language processing tools in order to divide all hosting providers into two groups: cloud providers and others. Using a cloud-related bag of words, we identified whether a given page contains information about the cloud-related services provided by the company. We classified the company as a cloud provider if there are more cloud-related keywords (Figure 9) rather than words not related to the cloud (Figure 10) on its webpage. If a particular year is absent in the data, we perform interpolation: if the provider was offering cloud services in previous and subsequent years, we assume that they were using cloud in the intervening year. If there was no cloud services offered in previous or next year, we label provider as non cloud, as it is not possible to estimate the start date when the provider started using the cloud. Despite a logical assumption that businesses will keep their modernized offerings once introduced, there are examples when companies started to offer cloud services, but subsequently removed them from their offer list (for example a2hosting started to offer cloud vps in 2015 but removed the service later, concentrating on traditional web hosting as a core product).

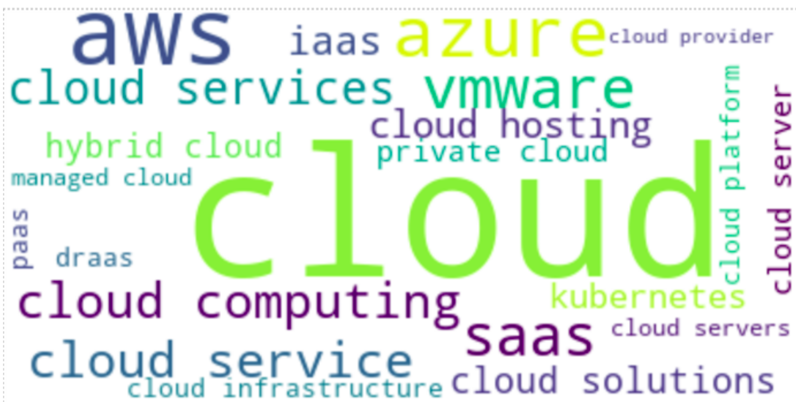

Figure 9. Cloud related words on companies' websites

[12]https://archive.org/web/

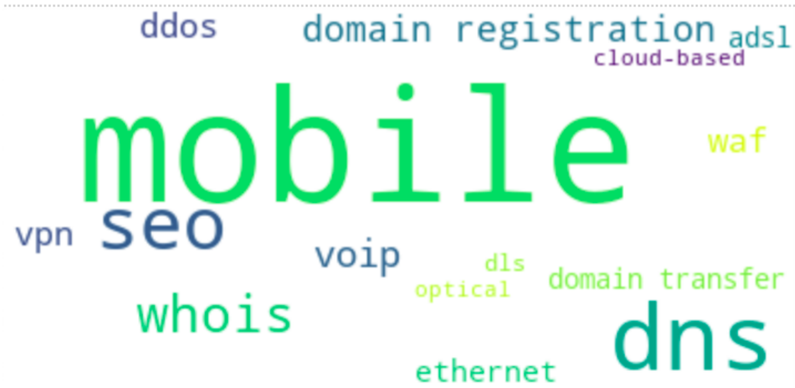

Figure 10. Non cloud related words on companies' websites

Chart 11 suggests that the number of providers that offered cloud services was steadily increasing starting from 2008, and now constitutes about 40 per cent of the total number of service providers. While the general trend is upward, we can compare it with the trends of cloud usage (Figure 16 below) to see that supply growth of cloud services did not fully correlate with the cloud usage in the UK. Nevertheless the results suggest a pronounced shift to cloud technologies (despite once being considered as a 'buzzword' technology).

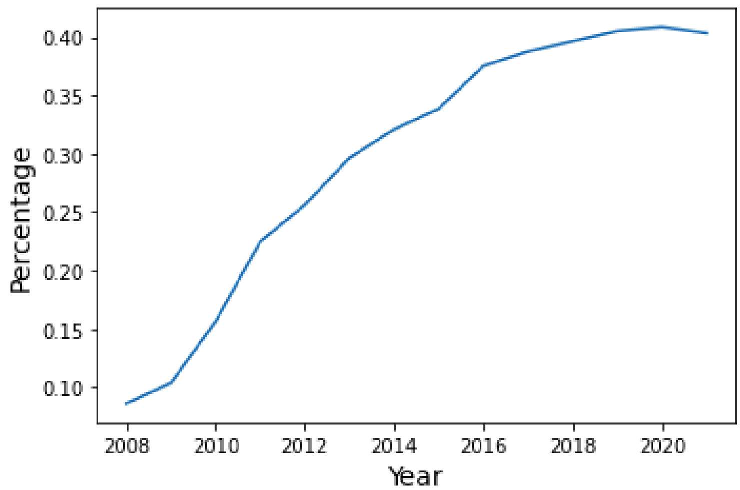

Figure 11. Percentage of providers offering cloud services by year

The last step of the indicator construction is to assign the value one to companies that use cloud hosting providers for more than 6 months a year and zero otherwise. Our central hypothesis is that such an indicator of a website placement can be a good proxy for a company's cloud usage, as the average cost of hosting services offered by cloud vendors is usually higher than the cost offered by conventional hosting providers. Thus, if a company chooses to use a cloud provider then it is willing to pay a higher price for the opportunity to use additional cloud-related services (such as virtual machines, cloud storage or cloud security opportunities). From another point of view, the most significant incentive, from the very beginning of the cloud usage process, is to transfer a website hosting to the cloud [13].

As we only track companies who have a website, this limitation can introduce a downward bias in the productivity estimates. Also the process of gathering the cloud indicator imparts some potential sample biases. Firstly, some websites of cloud vendors block web scraping, thus more advanced web scraping techniques are needed. Additionally, sometimes websites of cloud providers could be erroneously picked using Google search (for example, Wikipedia page or some general information page about the company are picked instead of the company website). Based on the manual verification of the top 200 providers, the misclassification rate is 10 percent (6 percent false positive cloud classifications, 4 percent of false negative classifications). However, the distribution of cloud users follow Pareto's law: the top two hundred web hosting providers serve approximately 97 percent of websites. Thus, after manual verification and correction of the first 200 cloud providers, the final error is estimated to be 0.3 per cent on the sample of more than 10000 hosting providers.

The third limitation is the precision of the cloud statistics gathered. Unfortunately, the indicator would not cover companies using only private cloud, or all other services but cloud hosting. Sometimes companies switch to the cloud out of curiosity, just to try cloud functionality, or simply because of the cloud providers' successful marketing campaign. However, we try to mitigate these issues by assigning cloud indicators only to those companies that use cloud during a period of 6 months and longer. Moreover, non-efficient cloud users in the sample can only reduce the magnitude of productivity estimations, as non-efficient firms that "played and failed" with technology will diminish the overall results and significance.

As a result, we obtain a proxy of cloud usage which is theoretically justifiable. Although the proxy is subject to minor imperfections (some cloud users can still use conventional web hosting providers, as well as some non-cloud users can use cloud providers), this should be considered as non-frequent edge cases that are not Pareto-efficient.

4.2. Financial and firms performance indicators

In order to assess the impact of cloud usage on the productivity of firms, we use the Financial Analysis Made Easy (FAME) dataset, that covers a population of businesses in the UK and derives information from Companies House records. FAME is a commercial dataset provided by Bureau van Dijk[14]. The dataset contains over two million active companies and about a million of so-called 'inactive' ones, who belong to one of the following categories: dissolved, liquidated, entered receivership or declared non-trading.

For the purpose of the productivity estimation and comparison of productivity measures among enterprises that use cloud and those firms that do not utilise cloud technologies, we use a measure of output per worker. We use a turnover per worker as the current productivity indicator, deflated to 2008 prices using deflators for 2 digit industry SIC codes [15]. As for employment statistics, FAME provides a full-time equivalent number of

[13] https://www.intechnic.com/blog/why-your-website-hosting-should-be-in-the-cloud/

[14] https://www.bvdinfo.com/en-gb

[15] We use ONS experimental 2 digit SIC industry deflators with 2018 as a base year. We should note that for some broad SIC sections ONS provides class and subclass deflators, which translates to 4digit sic code level. For example, section 11 'Manufacture of beverages' is subdivided into subclasses '11.01-11.06 - Alcoholic beverages' and '11.07 - Soft drinks'. The data is a mixure of product and implied industry-level deflators and is not seasonally adjusted.

employees. We restricted the sample to include enterprises that have a website and used it as an id to link cloud usage statistics with financial indicators of a company (information is gathered yearly). The financial information available includes income and profit indicators as well as gross output statistics, liquidity, turnover and other financial ratios, number of employees, and statistics about foreign and own investments. The data gathered was used to assess the financial profile of the company and measure the performance of the company.

It should be mentioned that not all financial and employment indicators are available for all firms in the dataset. The smallest firms are required to submit only basic balance sheet information to Companies House (such as shareholders funds, total assets, etc.) so we have substantially less information about them. In comparison, FAME provides rich financial and employment information on larger companies. Eberhardt et al. (2010) raised a concern that the indicators non-response rate is a function of firm size, and it may cause some bias in productivity estimations. However, we perform detailed representativeness checks suggesting that FAME can be reasonably used for such calculations (see next section).

We also construct a company age variable, as the number of years between establishment year and the current reporting year. We also construct the average level of cloud adopters by region and industry in order to account for external market factors (general level of cloud adoption), which can impact an adoption by a specific company. As discussed previously, network effects should be accounted in every technology.

The final dataset added is an internet broadband speed data, collected and provided by Offcom (https://www.ofcom.org.uk/), the UK regulator of broadband, TV, home phone and mobile services, universal postal service. In particular, Offcom provides 'United Nations' report about postcode level broadband speeds, as well as yearly reports about region level broadband statistics. We incorporate both 2017 postcode level and yearly region level statistics into our dataset. We match 85% of the data by exact postcode match and 15% by using aggregated average 3-digit postcode values. It is important to include this variable, as availability and speed of broadband connections is a prerequisite for many (if not most) cloud adoption cases.

4.3. Data analysis

In order to form the cloud usage indicator, we used a subset of companies that have a website. This limitation is connected with the specifics of the data gathering process. The current method to determine cloud usage relies on the DNS parsing process, which requires a website address. Consequently, the current study is limited to businesses which have their web sites. Such limitation can potentially lead to sampling bias. However, we believe that absence of the website is substantially decreasing the propensity of cloud usage. As a result, sampling bias is minimal (but is subject to future tests).

The subset also contained inactive companies and companies with an unknown status (19532 companies), which were filtered out together with top and bottom five percentiles of variables distributions in order to avoid the impact of outliers.[16]. The final subset contains 43,588 firms over the period from 2008 till 2020 or 183,450 firm-year observations (see the map of companies on Figure 12). There are indicators of cloud usage, financial

[16]Restricting to just removing the top and bottom percentiles yields similar results, but still leaves a log tail in the distributions

indicators, employment statistics and other important indicators of business performance contained in the dataset. See Table 2 for details.

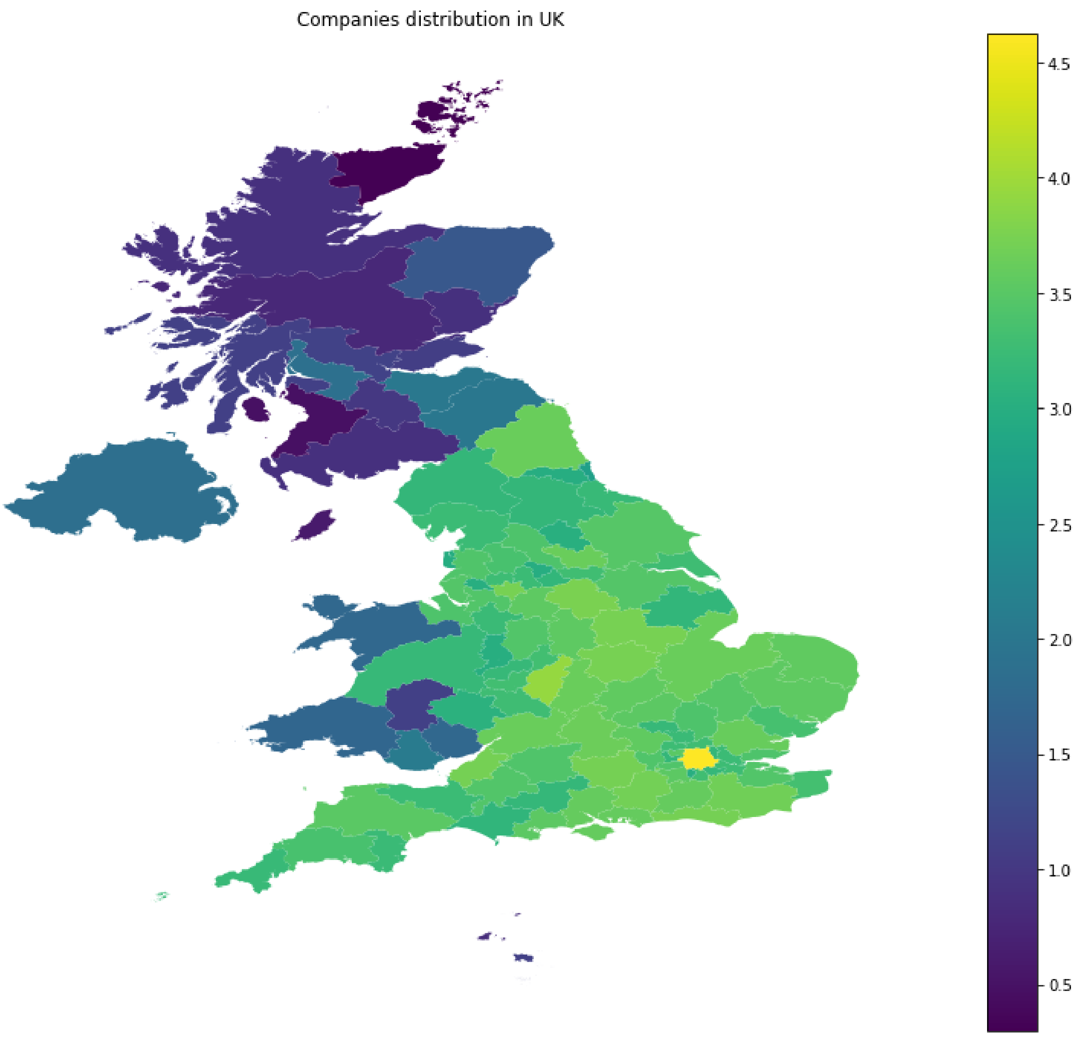

Figure 12. Number of companies in the FAME subset. Source: own compilation

Table 2. Cloud variables

Variable	Description	Measure
Company name	Official company name	Text
Industry (SIC code, 2007)	Standard industrial classification of economic activities	5 digit number of primary economic activity
Region	Region of the registered office.	Text field, icludes 13 regions, like 'Wales', 'North East England','London', etc
Year	Year of the data	2008-2020
Number of employees	Number of employees, each consecutive year (2008-2018)	Cardinal number
Fixed assets	Fixed long term assets of the company (property, plant and equipment, etc)	Cardinal number, th GBP
Turnover	Turnover of the company (sum of all its total sales for a given year), each consecutive year	Cardinal number ,th GBP
Cloud usage indicator	Cloud indicator for each year (2008-2020)	1 if company was using cloud-based domains, 0 otherwise
Broadband speed	Postcode level broadband speed measured by the speed of download	Cardinal number, Mbit/s

Statistical characteristics of the data are displayed in the Table 3. We report initial variables as well as variables used in our estimations, namely l - logged number of employees in the company, k - logged fixed assets of the company (per employee), g is an age of the company, f is a broadband connection speed, h - historical usage of cloud facilities, c - indicator of the start of the cloud usage, p - logged output per worker, n - log of the number of cloud users by industry. More detailed description on the variables constructed will be provided in the next section.

Table 3. Descriptive statistics

	count	mean	std	min	50%	max
turnover_	183,450	40,648.51	784,363.89	-1,280.63	9,733.07	142,235,744.00
number_of_employees	183,450	142.30	346.04	1.00	60.00	9,060.00
broadbandspeed	183,450	33.56	24.43	0.30	30.90	800.90
historical_cloud	183,450	0.40	0.49	0.00	0.00	1.00
age	183,450	22.95	19.27	1.00	18.00	163.00
fixed_assets_	183,450	13,728.94	235,434.18	0.00	1,029.30	51,834,000.00
p	183,450	1.05	0.77	-0.85	0.87	9.59
k	183,450	0.36	0.55	0.00	0.16	5.82
l	183,450	3.95	1.47	0.69	4.11	9.11
g	183,450	2.90	0.76	0.69	2.94	5.10
f	183,450	-3.32	0.72	-6.69	-3.46	-0.26
n	183,450	0.23	0.64	0.00	0.29	0.53
c	183,450	0.10	0.30	0.00	0.00	1.00
h	183,450	1.30	2.13	0.00	0.00	12.00

To check how representative is our sample, we provide a detailed analysis and comparison of the gathered sample with ONS Business Population Estimates (ONS BPE)[17] and cloud estimates from the ONS E-Commerce Survey[18].

A comparison of the number of companies represented in our dataset and the ONS BPE dataset shows the persistently proportionate amount of companies over each year. The proportion of businesses represented each year is about 4% from the corresponding year in ONS BPE.

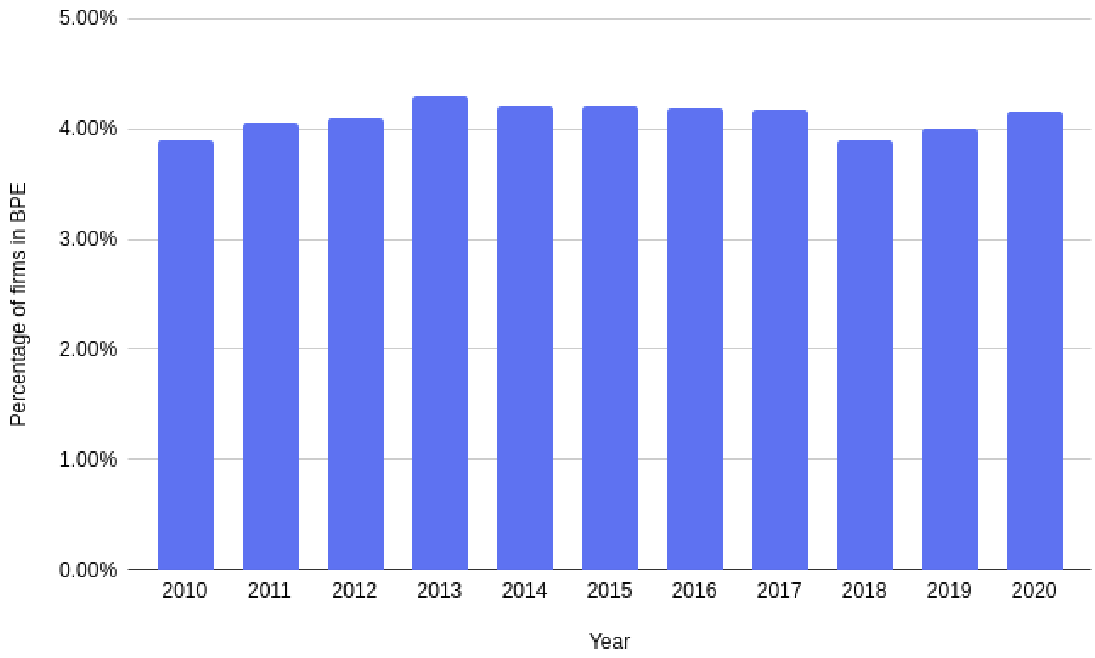

Figure 13. Number of companies in the used subset of FAME,as percentage of entities in BPE. Source: own compilation

An industry comparison, however, shows slight biases in the number of companies represented in the FAME subset in comparison to the BPE population (Figure 14). For 2-digit industry groups, construction, wholesale and retail trade, and the information and communication industries are over-represented in FAME, with all other industries under-represented.

[17]https://www.gov.uk/government/statistics/business-population-estimates-2018
[18]https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/ecommercesurvey

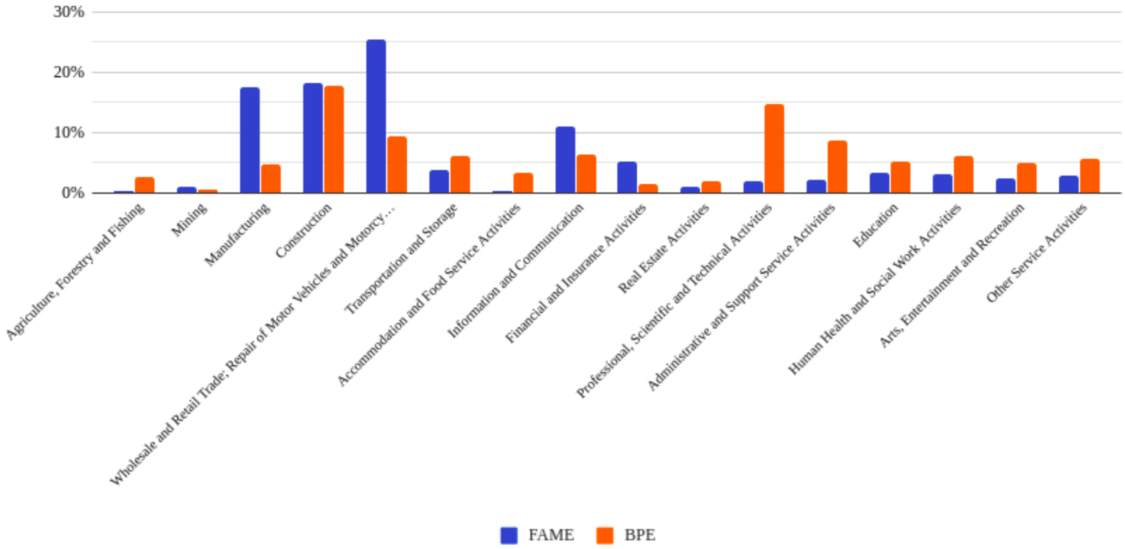

Figure 14. Comparison between FAME subset and BPE database by industries. Source: own compilation

The number of businesses by company size cohorts adds additional information about possible biases in our subset. Large firms and firms with unknown size are slightly over-represented in our sample (see Figure 15). This might be a result of companies selection, as website usage is more probable among larger companies[19].

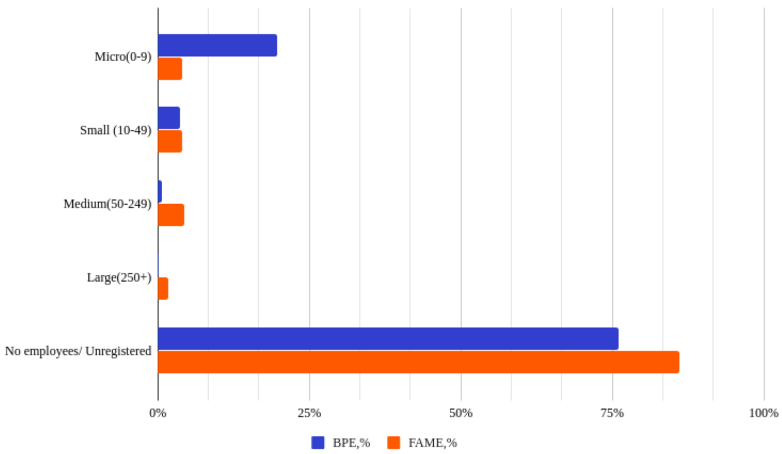

Figure 15. Comparison between FAME subset and BPE firm size cohorts. Source: own compilation

The representation of the cloud data in our sample is of particular importance in order to enhance the quality of the research. As for the total number of companies in the dataset, we would be interested in the number of cloud adopters. The following figures represent the number of companies that have adopted cloud technologies (as mentioned in the previous section).

[19]https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/ecommerceandictactivity/2017

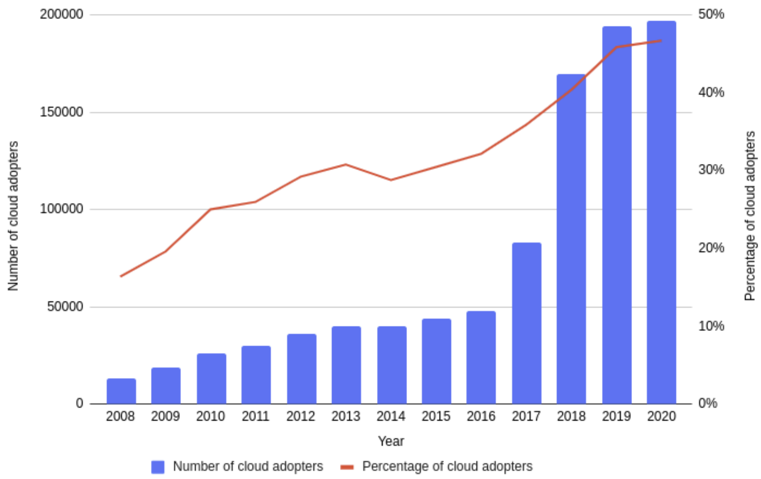

Figure 16. Adoption of cloud technologies

Next, we provide a comparison between cloud statistics represented by the ONS E-commerce survey and our constructed sample. Differences in cloud usage by firm size bands are similar between both data sources. Because the ONS E-commerce survey used a wider definition of cloud services usage, it has higher percentage estimates. The cloud indicator used in the current study refers to companies that used more advanced data science tools (such as migration of their websites to the cloud hosting). Thus adoption rates are lower (Figure 17 and Figure 18).

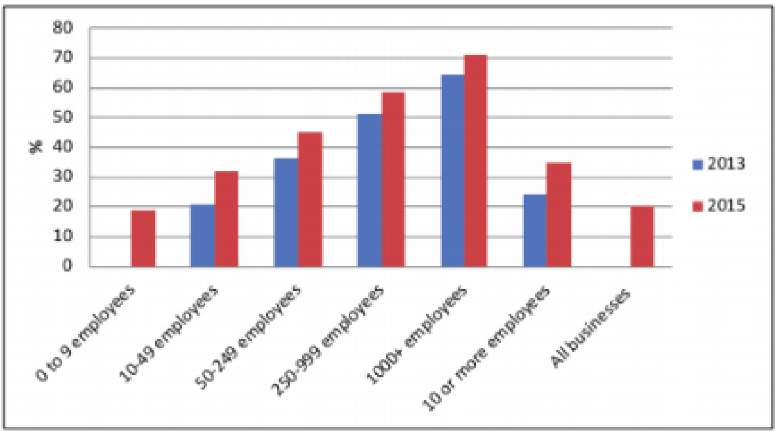

Figure 17. Cloud usage. ONS E-commerse survey by firm size bands. Source: own compilation

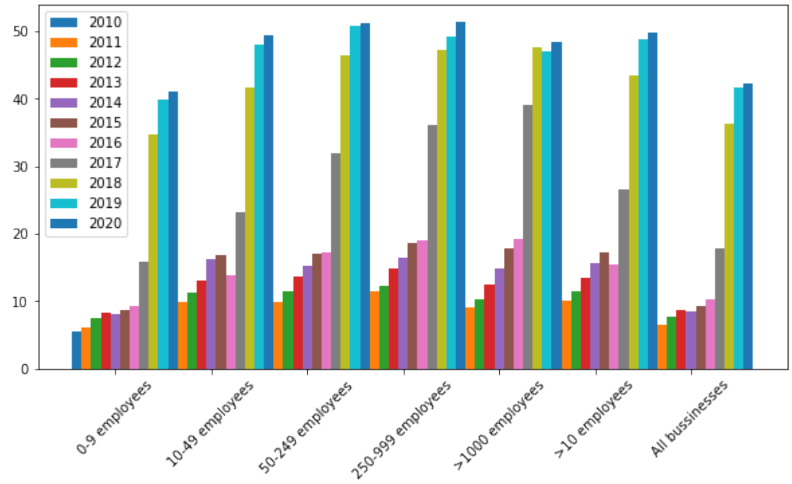

Figure 18. Cloud usage. Our sample constructed sample. Source: own compilation

It is obvious from the data that there was a major change in cloud adoption after 2016. This coincides with Amazon and Microsoft increasing their investments in the UK cloud and introduced a number of new UK cloud data centers (see subsection 2.2). The difference between cloud and non-cloud adopters can also be of significant importance.

As the first stage of analysis, we look at some descriptive statistics of our data. The data reveals that cloud adopters have a persistently higher return on capital employed than non-cloud adopters. An interesting fact is that companies that did not use cloud, have had higher fixed assets before 2016 and substantially lower fixed assets since 2016. Also, non-cloud adopters had higher operating profits before the 2016 cloud adoption and lower operating profit after the 2016 cloud adoption. These two facts may suggest that patterns that were true before 2016 (big companies had a higher tendency to incorporate new technologies) were reversed. After 2016, smaller companies tend to adopt technologies more eagerly.

The map of cloud adoption (year 2018) is presented in Figure 19. Cambridge, Jersey, Aberdeen regions demonstrate the highest rates of cloud adoption. Kirkaldy, Romford, Canterbury regions have the lowest adoption rates.

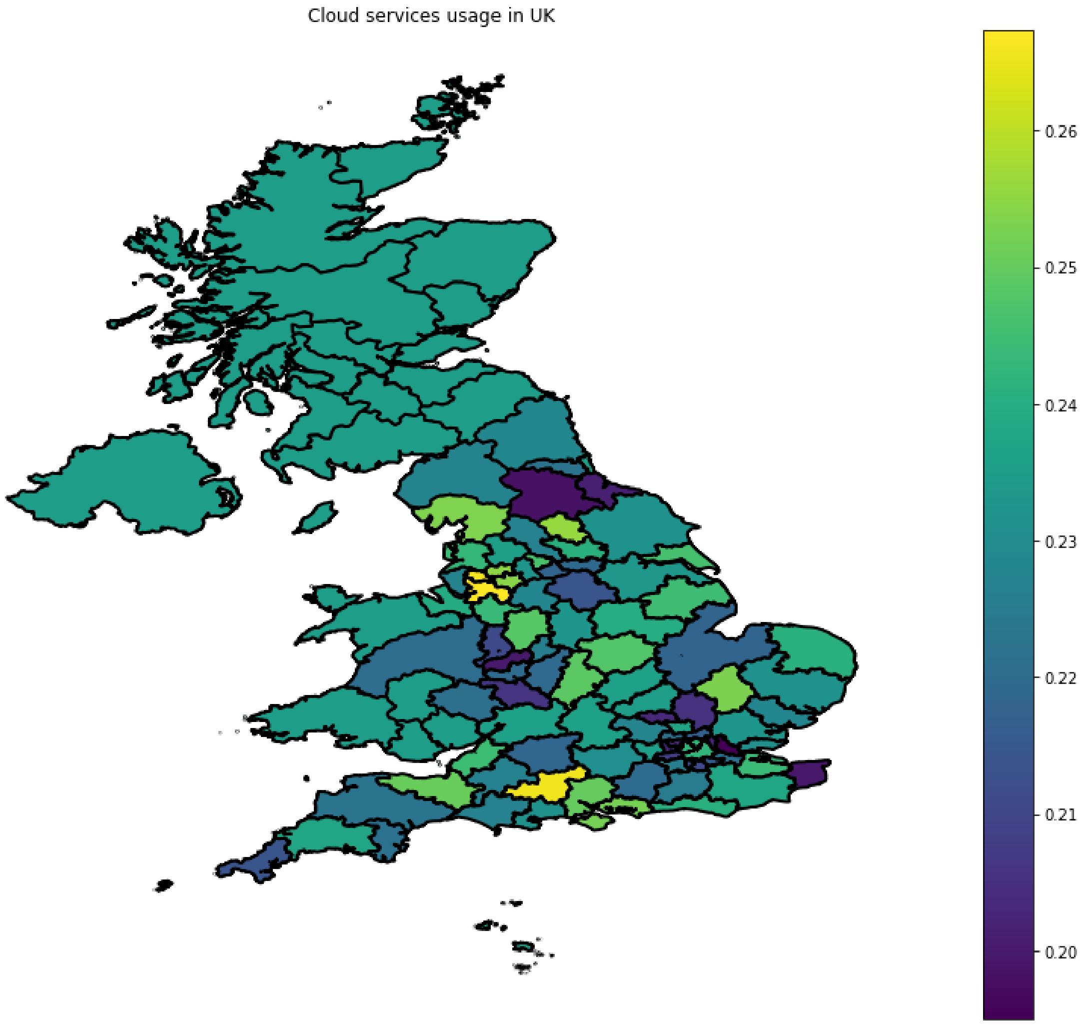

Figure 19. Cloud adoption map. Source: own compilation

Cloud adopters have a higher turnover in all consecutive years after 2016. The final observation is that companies using cloud services pay fewer dividends. This observation is in line with the findings of Brynjolfsson and Yang (1997) showing that ICT investments positively influence the value of the firm and, according to the Tax-Effect Hypothesis (Al-Malkawi et al., 2010), higher value firms are associated with lower dividends. The changing patterns due to the 2016-2017 cloud adoption through the majority of variables can also be observed. For example, 2017 cloud adopters had steadily higher profit or loss before taxation. However, the effect before 2017 is quite the opposite. Shareholders funds indicator for cloud adopters of 2016-2017 is higher than for the other years. Cloud adopters also have a lower liquidity ratio. This observation can be explained by the fact that cloud expenses are current expenses and they can potentially lower the liquidity ratio of the company in a significant way.

We investigated if there was a correlation between cloud adoption indicators. As we mentioned above, each firm can proceed with using cloud services the following year, or it can stop using the cloud and switch to its own architecture or traditional vendors. An examination of correlations across time suggests that if the firm starts using the cloud at some point in time, it is generally more likely that the firm will proceed with cloud usage for at least the next 2-3 years.

5. Regression analysis and propensity score matching

5.1. Regressions on Pooled Data

As discussed previously, there are several factors that account for a company's decision to adopt cloud. The size of the firm (number of employees) influences the cloud adoption, as small firms are more eager to quickly pursue new trends and adopt new technologies (Archibugi et al., 2012). Large and old firms are believed to have a lower agility and speed in adopting new technologies due to the lag effect of current technologies and so - called 'Legacy software'.

On the other hand, there are network effects of the technology: if more firms in the same region or industry adopted new technology, it is more probable that the positive network effects and 'jungle' radio will influence the decision of other players. Proximity to the good internet connection also matters, as it is a prerequisite for using the cloud technologies (DeStefano et al., 2020). As cloud adoption is an R&D activity to some extent, the firm should have resources available to adopt the technology and retrain personnel, so firms with better financial situation and positive financial shocks would have higher propensity to utilize new technologies. With this said, we are also including variables which explain company's financial endowment as a background factor (better endowed companies are more able to pursue R&D technology).

To estimate the difference in productivity between cloud and non-cloud users, we start with logistic regression to analyse the main factors influencing the cloud adoption variable. We must quote a famous aphorism of statistician George Box here: "All models are wrong but some are useful". Thus, we proceed with the intuition that there could be no 'perfect' model which would explain all complexities of the reality, however some models can be less wrong then others. We neglect panel structure for this and next subsection, in order to provide a baseline model and to understand correlations between our variables of interest. We start with pooled dataset, which contains all yearly firm level observations.

We try to model cloud adoption probability according to the formula

$$Pr(C) = a_0 + a_1 * l + a_2 * k + a3 * g + a_4 * f + a_5 * h + a_6 * p \tag{1}$$

where $Pr(C)$ is a probability of using the cloud in the current year, l is a number of employees in the company, k are the fixed assets of the company (per employee), deflated using 2 level SIC deflators from ONS [20] g is an age of the company (number of trading years since registration day till the current year) f is a broadband connection speed based on the Offcom's data [21]. h - historical usage of cloud facilities(amount of previous years when company used cloud),

We use logit model on top of the yearly pooled data and report results in the next table (Table 4)

[20] https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/experimentalindustrydeflatorsuknonseasonallyadjusted

[21] https://www.ofcom.org.uk/

Table 4. Results of logistic regression

	cloud
cloud	
k	0.047^{***}
	(0.007)
l	-0.005
	(0.003)
g	-0.059^{***}
	(0.006)
f	0.072^{***}
	(0.005)
h	0.382^{***}
	(0.003)
BIC	206100

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Several meaningful patterns in these regressions can be observed:

- Capital per employee (k) has a positive sign in the regression, which goes along expectations as discussed previously,
- The number of employees (l) acts as an essential variable, which has a negative impact on the probability of cloud adoption: a higher number of employees means lower probability of cloud adoption for the enterprise. This fact supports evidence from Jin and McElheran (2018) confirming that cloud technologies bring more value to small firms as they enable freedom of experimentation under limitations of uncertainty,
- Age of the firm (g) has a negative impact on the cloud adoption score, as it is believed that older firms have a lower ability to adopt quickly to new technologies,
- broadband speed (f,download speed in Mbits per second, logged) provides higher probability for the cloud usage, as availability of good internet speed connection is an important prerequisite for the cloud,
- cloud usage in previous years strongly increases the probability of cloud usage for the current year,

In order to explore how cloud adoption impacts productivity, we adopt a simple Cobb Douglas production function framework which regresses on capital per employee, number of employees as a measure of economies of scale, measures of cloud usage and control variables (broadband access and age of firms). We also include time, industry and region dummies to control for unobserved factors. Our estimating equation is given by:

$$p_{it} = a_0 + a_1 * l_{it} + a_2 * k_{it} + a3 * g_{it} + a_4 * f_{it} + a_5 * c_{it} + a_6 * h_{it} \tag{2}$$

where p is a productivity per worker, l is a number of employees in the company, k are fixed assets per employee (a substitute for capital per worker), g is an age of the company (number of trading years since registration day till current year), f is a broadband connection speed based on Offcom's data, c is an indicator when the company started to use cloud services, h - historical usage of cloud facilities(amount of previous years when company used cloud, 0 in the first year of the cloud usage), i, t are entity and time subscripts.

Table 5. Regression results

	(1)	(2)	(3)
	p	p	p
	b/se	b/se	b/se
k	0.230***	0.230***	0.230***
	(0.006)	(0.006)	(0.006)
l	-0.076***	-0.077***	-0.077***
	(0.002)	(0.002)	(0.002)
g	0.038***	0.038***	0.038***
	(0.003)	(0.003)	(0.003)
f	0.058***	0.058***	0.058***
	(0.003)	(0.003)	(0.003)
n	0.005**	0.005**	0.005**
	(0.002)	(0.002)	(0.002)
c	-0.035***		-0.022***
	(0.006)		(0.006)
h		0.010***	0.009***
		(0.001)	(0.001)
nobs	140072	140072	140072
R-sqr	0.185	0.185	0.185
dfres	140031	140031	140030
BIC	294812.43	294732.55	294733.24

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: all regressions include year, industry and region dummies

The first regression (i) includes logged capital per employee(k), logged number of employees (l), logged age of the company, internet download speed in the specific post code (f), number of firms in the industry that uses cloud, indicator of start of the cloud usage (c). Results suggest that when the company starts to use cloud, it bears cloud adoption costs, that lower the productivity in this year. Costs are associated with the cloud migration efforts, research and development routines and adaptation, managerial changes and retraining, etc (see section 3.3).

The coefficient of the capital per worker is positive and significant, if not a little lower than commonly found in productivity regressions. The coefficient on number of employees suggests decreasing returns to scale, but these are not too large. The coefficient on age is positive so older firms have higher productivity. Broadband speed shows a consistently positive impact on productivity, suggesting almost 6 per cent productivity increase per every 1 per cent of additional broadband speed availability. While this measure shows that access to IT facilities and the Network is a crucial prerequisite for a successful company, we should mention a potential endogeneity bias associated with the measure. Speed of the broadband may be associated with the price of the office space in the area as well as general business areas agglomeration effects. Thus, the coefficient might be biased due to self selection.

Network effects of the cloud computing technology are suggested to have a positive impact on the productivity as well. We add a quantity of years when company used cloud up to the present year (h, levels) into the regression (column 2). Every additional year of cloud usage adds 1 per cent to the productivity on average. Regression (3) includes the start of the cloud usage indicator as well as number of years of experience with the cloud technology. From this regression we can argue that the first year of cloud usage would be arguably associated with transitions as firms learn to use the technology. The positive coefficient on h indicates that continued use of cloud would compensate initial expenses, and the breakeven period in terms of seeing productivity increases would be 2-3 years. This finding is consistent with other estimations, suggesting that break even point happens after 1-2 years [22]. The finding is also consistent with the technological diffusion models, where firm face adoption costs when considering a new technology or process in place.

Table 6 provides detailed distribution of cloud effects by firm size bands. As we can see, the effect from starting using the cloud is consistently negative and is higher for smaller firms. Possible explanations can encompass higher implementation risks for small companies, lower accounting and metering capabilities inside of the firm in order to ensure the efficiency of the implementation process. The negative impact is lower for large firms because of longer planning horizons and thus better control over the implementation risks. In comparison, the yearly effect from the cloud usage is negative for large firms and very tiny for medium firms (which is consistent with findings from Jin and McElheran (2018)). Instead, as suggested by literature, micro and small firms face the highest yearly benefits from using the cloud. As explained in section 3.3, cloud provides experimentation opportunities, freedom to scale up and save considerable upfront investments (associated with the standard ICT equipment) that often makes a crucial difference for small companies. The cloud provides a positive productivity effect of 2.2% for micro sized enterprises, 1% for small, 0.3% for medium and negative impact for large enterprises.

[22]https://www.delltechnologies.com/asset/sk-sk/services/consulting/industry-market/h15537-the-roi-of-private-cloud-wp.pdf

Table 6. Impact of the cloud usage by firm size

	micro	small	medium	large
	p	p	p	p
k	0.223***	0.224***	0.195***	0.393***
	(0.014)	(0.010)	(0.007)	(0.014)
l	0.063***	-0.104***	-0.145***	-0.079***
	(0.011)	(0.008)	(0.005)	(0.015)
g	0.072***	0.017***	0.016***	0.065***
	(0.008)	(0.005)	(0.003)	(0.007)
	(0.441)	(0.249)	(0.170)	(0.291)
f	0.100***	0.052***	0.040***	0.020**
	(0.008)	(0.005)	(0.003)	(0.008)
n	0.011**	0.019***	0.005**	-0.021***
	(0.004)	(0.003)	(0.002)	(0.004)
c	-0.027	-0.019	-0.017*	-0.039
	(0.018)	(0.012)	(0.007)	(0.023)
h	0.022***	0.009***	0.003**	-0.026***
	(0.003)	(0.002)	(0.001)	(0.004)
nobs	26620	47633	57226	8593
R-sqr	0.186	0.243	0.169	0.198
dfres	26579	47592	57184	8555
BIC	68991.25	107551.72	85367.57	12859.36

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: all regressions include year, industry and region dummies

Table 7 reveals important industry level features of cloud usage effect. We find that retail and construction industries have insignificant or negative impact of the cloud technologies, but manufacturing (man), education and health (ed& health), ICT and professional services (ICT) and entertainment industries face positive and significant consequences of the cloud technology usage. Cloud provides the highest benefit for the ICT and professional services, education and health sectors, where information and tools that help to process it are one of the most crucial assets of the business.

Table 7. Regressions by industry

	man	retail	ICT	ed&health	constr	entert
k	0.478***	0.318***	0.250***	0.101***	0.328***	0.103***
	(0.013)	(0.020)	(0.008)	(0.011)	(0.034)	(0.014)
l	-0.119***	-0.174***	-0.039***	-0.039***	-0.026***	-0.015**
	(0.003)	(0.005)	(0.003)	(0.002)	(0.008)	(0.005)
g	0.042***	0.068***	0.051***	-0.024***	0.042**	0.010
	(0.004)	(0.008)	(0.005)	(0.004)	(0.013)	(0.008)
f	0.007	0.066***	0.041***	0.054***	0.090***	0.052***
	(0.004)	(0.008)	(0.005)	(0.005)	(0.012)	(0.009)
n	0.008	0.008	-0.003	-0.007	0.032	-0.015
	(0.007)	(0.012)	(0.007)	(0.006)	(0.019)	(0.011)
c	-0.021	-0.020	-0.001	-0.004	0.002	-0.028
	(0.014)	(0.024)	(0.014)	(0.011)	(0.035)	(0.022)
h	0.006**	-0.006	0.013***	0.015***	-0.013*	0.009**
	(0.002)	(0.004)	(0.002)	(0.002)	(0.006)	(0.003)
nobs	28127	22982	47959	15370	7861	13269
R-sqr	0.098	0.255	0.072	0.124	0.067	0.115
dfres	28093	22949	47923	15338	7828	13237
BIC	44236.54	57931.54	101870.14	12161.41	18240.83	25587.56

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

However, ordinary OLS estimation does not account for the selection bias, as companies select technologies according to the future profit or productivity projections, that are unknown for the researchers. In such a way, we may also presume that estimation coefficients may be biased due to unobserved shocks, productivity, financial and managerial spillovers. Group selection bias may occur if some internal or exogenous characteristics determine firm's choice of whether adopt cloud technology or not (selection bias, Rubin, 2015). We therefore consider an alternative estimation method in the next section.

5.2. Standard Propensity Score Matching Model

As a consequence, we proceed with an alternative methodology in order to reduce selection bias and assess the difference in the output of the firms which use cloud technologies. Using a propensity score matching (PSM) framework, we try to assess the difference between cloud and non-cloud users in terms of the output per worker. Propensity score matching estimators are widely used in the productivity estimation literature (Rosenbaum and Rubin, 1983; Wagner, 2002). The framework is intended to reduce estimation bias in observational studies, where the researcher has no control over the design of the experiment (in comparison to randomized trials where the researcher has full control over the experiment). PSM helps to understand the effect of cloud computing on the 'treated' group of companies. In our case, the treated group is the group that used cloud computing hosting. The untreated group consists of companies that did not use cloud hosting. The technique helps to ensure that the average characteristics of the treatment and control group are similar in order to accurately calculate the treatment effect (White and Sabarwal, 2014).

We start by introducing notation $C = 0, 1$ - a set of 'treatment' introduced in an observational experiment; $c \in C, c = 1$ would stand for the case when the company has adopted the cloud (so this firm would belong to the treatment group); $c = 0$ would mean that a given company did not adopt the cloud. Y^c is the outcome of the experiment, measured in productivity per worker. As only one of the outcomes is observed for each specific company, the PSM model tries to assess the alternative scenario of 'what would happen' if the specific company did not adopt the cloud technology. This assessment is achieved by finding a 'non-cloud' pair to every 'cloud user' company, which would have all similar characteristics but a cloud variable.

The algorithm of the PSM framework can be generalized in three steps:

1. Create a propensity score for every firm observation, by assessing a 'propensity' of the firm to adopt cloud ($\pi_i = P(C = 1|X)$, where C is a 'treatment', X - covariates). This is done by performing a logit regression with the cloud variable and a set of covariates. The score shows the probability that a given firm adopts the cloud.
2. Using propensity scores, the algorithm performs matching of cloud and non-cloud users across the space of different characteristic of cloud users available in the dataset. As noted in the original paper, Rosenbaum and Rubin (1983), proper variable selection may be necessary in order to accurately estimate propensity scores. As originally suggested, model structure should be induced by the underlying theory, however, if there are covariates that are interconnected, then proper selection of cross terms and variable lags should be performed. Austin (2011), Morgan and Todd (2008) recommended to include higher-order moments and interactions between covariates in order to account for interconnections between them. Rubin (2001) recommended choosing covariates according to the theory and prior research (but without using observed outcomes).
3. Using the bootstrapped dataset, the average treatment effect is estimated by comparing the mean in the two groups of cloud adopters. Treatment effect equals to $E(Y^1 - Y^0) = \sum_i (Y_i^1 - Y_i^0)$, where Y^1 is an outcome if treated (if cloud technologies were adopted), Y^0 is an outcome if untreated (if cloud technologies were not adopted), i subscript stands for the number of matches created by the PSM algorithm.

The validity of the approach relies on several assumptions:

- Stable Unit Treatment Value Assumption (SUTVA). It is assumed that 'treatment' assignment for a given case is not dependent upon outcomes for any other units. In other words, cloud adoption decision should be randomly assigned among the units and the cloud adoption decision of one firm should not influence the decision of another firm. However, there could be an indirect influence of one firm on another, through market mechanism and competition, which cannot be stated explicitly.
- Positivity Assumption. $P(C = c|X = x) > 0$ for all $c \in C$ and $x \in X$ (where X is a set of covariates and C is treatment options). It means that for every level of the covariate (say every company size and every financial aspect), there should be at least 1 'treated' and 1 'untreated' case. Without this condition, it would be impossible to directly compare different companies because of non-similar company background. However, if we have both cloud and non-cloud adopters across

all range of firm characteristics, it would be possible to find a similar 'non-cloud' case to every cloud adopter, in such a way making a comparison to be more 'fair'.

- Ignorability /Conditional independence Assumption (or no unmeasured confounders). For every $x \in X$, treatment assignment is independent of the potential outcome, $Y^1, Y^0 \perp C|X$. In other words, there are no exogenous variables (confounders) that simultaneously affect the treatment decision and the outcome.
- Consistency assumption, $Y = Y^c$ if $c \in C$. The outcome of treatment is equivalent to the observed outcome (so all potential outcomes of the treatment are observed in the data).

We start by running a logit regression on cloud variable and all other explanatory factors in the model.

$$Pr(C) = a_0 + a_1 * l + a_2 * k + a3 * g + a_4 * f + a_5 * h + a_6 * p + c_{t-1} + l_{t-1} + k_{t-1} + year$$

where l, k, g, f, h, p, s are the same parameters as in 2 and $c_{t-1}, l_{t-1}, k_{t-1}$ are added to account for the historical performance of the company as well as historical cloud usage. We add a year variable to match observations in the same or close years. After model construction, we proceed with estimating a propensity score, which is $\hat{Pr(C)}$. Then we perform a propensity score matching to create matching pairs. In order to minimise the distance between covariates in matched groups, we used the nearest neighbour matching algorithm, as in the original paper by Rosenbaum and Rubin (1983), where for every cloud adopter i we search for a pair j with the closest absolute distance between propensity scores

$$d(i, j) = min_j|e(X_i) - e(X_j)|$$

where $e()$ is a propensity score.

We should note that there are several other matching algorithms: caliper matching (Cochran and Rubin, 1973), where the matched pair is found within pre-specified distance, $d(i, j) = min_j(|e(X_i) - e(X_J) < b|)$; Mahalonobis metric matching (Rosenbaum and Rubin, 1985), where $d(i, j) = min_j(D_{ij})$ and $D_{ij} = (V_i^\top - V_j^\top)^\top S^{-1}(V_i^\top - V_j^\top)$, $V = (X, E(x))$ and S is a covariance matrix of the new vector for the control group.; Mahalonobis caliper matching (Guo et al., 2006) where $d(i, j) = min_j(D_{ij} < b)$; genetic matching (Diamond and Sekhon, 2013) which is similar to Mahalonobis metric matching but $D_i j = (V_i^\top - V_j^\top)^\top W S^{-1}(V_i^\top - V_j^\top)$ and W is a weight matrix found using genetic sampling methods.

As suggested by Stuart (2010), we use the 'optimal' matching approach, which allows previously matched pairs to be changed when making a current match, to find well-matched pairs that would have a minimum average distance between matches. The alternative faster method is 'greedy' search, where the pair is found using first- good candidate and the change of pair is not allowed after the match (Rosenbaum, 1989).

Some existing alternatives do not produce exact pairs. These include stratification methods that classify the entire sample into strata, based on percentiles Schafer and Kang (2009); full matching (Hansen, 2004) where one treatment unit is matched to several controls or vice versa; kernel matching which combines both matching and outcome analysis in one procedure (Heckman 1997). However, it was noted by Steiner and Cook (2013) that proper selection of covariates is generally more important than the method used for matching. If the functional form of propensity score function could be complex and estimation can potentially require a significant number of covariates, a non-parametric data-driven approach is preferred (Lee et al., 2009). The ultimate goal of these models is achieving a proper balance of covariates (Austin, 2011; Stuart, 2010). Model fit or significance is not a primary interest, although the quality of the model would affect the final balance of the covariates.

Based on only about 25 per cent of cloud adopters in the sample, we used 1 to 3 matching, i.e. 3 non-cloud adopters were matched to 1 cloud adopting company. We use matching without replacement, along with the SUTVA assumption. Alternatives include matching with replacement, in order to balance treated and untreated records in the dataset, however, the difference in results of the models should not be significant (Steiner and Cook, 2013).

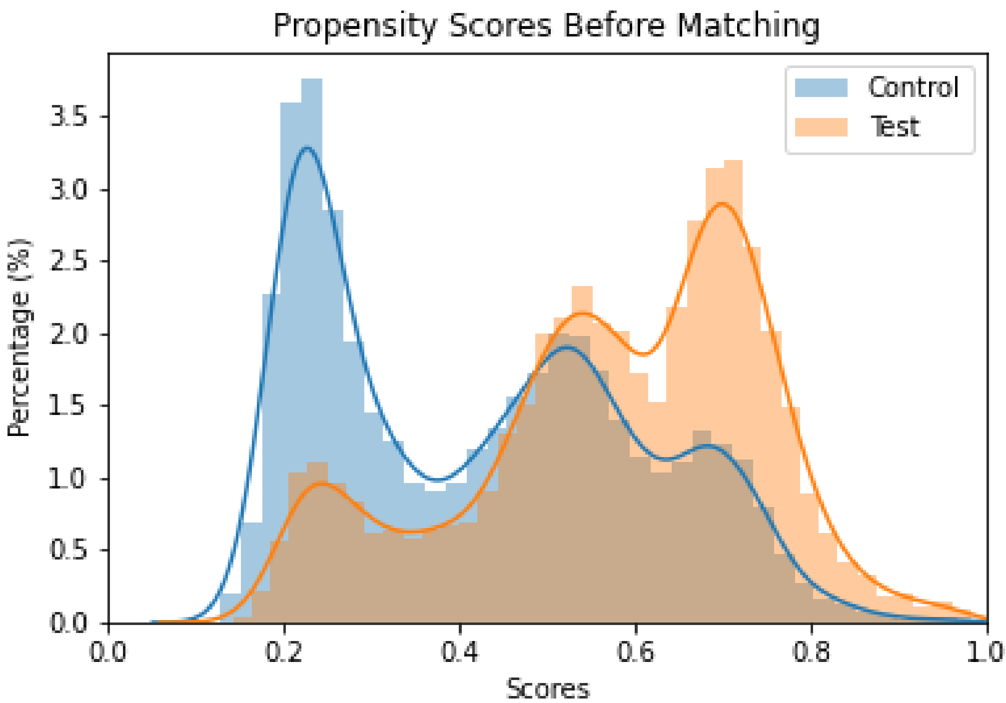

Figure 20. Propensity scores before matching

The PSM algorithm validates the Positivity Assumption showing that there is a separability in the data (average accuracy – 66 per cent, suggesting that covariates used have an impact on the outcome, which suggests that balancing is needed). From the other side, There is no variable level at which the treatment variable would be deterministic. It means that at each level of every variable, we have both treated and untreated samples that is a crucial prerequisite for having an adequate PSM model.

The algorithm aims to balance treated and untreated samples. Therefore, there is no statistical difference between the covariates in the two groups. The following graph shows that the distribution of each covariate after matching is not statistically different, according to the t-test.

var	ks_after
k	0.073
l	0.052
f	0.023
g	0.078
numeric_region	0.051

The next step, according to Pan and Bai (2015) is an outcome analysis. There are several ways this analysis may be executed. As noted by Rosenbaum and Rubin (1985), it is recommended to use a regression on top of matched data, in order to control for remaining selection bias due to covariates that were not perfectly balanced by the algorithm. If a matched subset contains several control units, weights should be created for proper balancing (assigning 1 to treated units and proportion for each untreated unit within the matched subset as suggested by Stuart et al. (2011)).

$$Y = \beta_0 + \beta_1 a + \beta_2 x_1 + .. + \beta_{q+1} x_N + \epsilon$$

and Average Treatment for Treated (ATT) = $\hat{\beta_1}$. It is also possible to conduct analysis on the entire original dataset after matching, in order to obtain ATT or ATE (Average Treatment Effect).

$$ATT = \sum s(n_{s1})(Y_{s1} - Y_{s0})/N_1 or ATT = \sum s(n_{s1})(\hat{\beta_{1a}})/N_1$$

$$ATE = \sum s(n_s)(Y_{s1} - Y_{s0})/N or ATE = \sum s(n_s)(\hat{\beta_{1a}})/N$$

where s denote a matched subset,n_{s1}- number of treated samples within the subset, n_s - number of all samples in the subset, N_1 - number of treated samples overall, Y_i - outcome in either treated or untreated case, $\hat{\beta_{1a}}$ - OLS estimate for the matched subgroup.We proceed with the regression on top of matched data.

In order to obtain initial results, we used the data for the year 2018 only. According to the PSM method, the resulting difference in sales per employee demonstrates that cloud usage increase sales per employee ratio by 1.2 per cent. The result is significant at 99 per cent confidence level (Table 8).

Table 8. Propensity Score Matching results

Dep. var	No. of treat	No. of control	ATT	Standart error	t-value	p-value
Outp. p. worker	9594	19635	1.20%	0.46	2.63	0.01

As an additional exercise, we also tried to perform the PSM matching on the whole dataset, by including a year as a covariate (see Table 9).

Table 9. Propensity Score Matching results with the whole dataset

Dep. var	No. of treat	No. of control	ATT	Standart error	t-value	p-value
Outp. p. worker	62142	114129	0.65%	0.007	9.53	0.001

Dividing by firm cohorts, we see the same robust trends among firm sizes (Table 10). As in our previous exercise, micro and small firms would benefit the most, and large firms have no significant effect from using the cloud technology. The margin of the effect has retained almost the same level, suggesting that there are no controlled confounders within our dataset that would induce a bias on the observed cloud effect coefficient.

Table 10. Propensity Score Matching results by firm size cohort

Group	Estimate	Standard error	t-value	p-value
Micro	0.8%	0.002	4.46	0.0003
Small	1%	0.001	7.28	0.0004
Medium	0.4%	0.001	3.96	0.0001
Large	0.08%	0.002	0.48	0.63

6. Conclusion

This paper presents one of the first sources of data about the cloud usage in the UK that has a long time element for a large cross section of companies. Web scraping techniques were used as a tool to obtain data on the usage of the cloud technologies. Our regression results support the notion that the use of cloud services has a positive impact on firm's productivity. However, this impact is not instantaneous, and the data suggests it takes some time for the positive impact to emerge. Firms that used cloud technologies earlier enjoy higher benefit than later adopters.

We present evidence that cloud computing technologies are mostly useful for smaller companies that face high fixed costs in investing in traditional ICT hardware and software. Cloud facilitates experimentation and drives down R&D costs, providing companies with fast and cheap ways to learn about their needs while avoiding irreversible investments. As a result, lower entry barriers drive creation and expansion of businesses and increase production. Our results suggest a negative impact of using cloud for the largest firms. This might be due to more coordination issues for these companies.

We envisage that cloud computing should inevitably inherit some of the specific performance drivers from traditional ICT, such as being skill biased and generating substantial spillovers. Further research is required to investigate these aspects. For example, matching our cloud indicator to skills data derived from job platform data should shed some light on the skill requirements of these new technologies. It would also be useful to investigate links with intangible capital.

Some unobserved and unaccounted covariates, such as internal firm processes, education of CEO,etc., may influence a firm's decision to adopt cloud. Future analysis plans to use an advanced propensity score matching for multilevel data with time varying-treatments to check the robustness of our results.

References

Acemoglu, D. and P. Restrepo (2018, 01). Artificial intelligence, automation and work. *SSRN Electronic Journal 1*, 1–10.

Al-Malkawi, H.-A., M. Rafferty, and R. Pillai (2010, 01). Dividend policy: A review of theories and empirical evidence. *International Bulletin of Business Administration 9*, 171–200.

Ali, A. N. A. (2012). Comparison study between ipv4 & ipv6. *International Journal of Computer Science Issues (IJCSI) 9*(3), 314.

Archibugi, D., A. Filippetti, and M. Frenz (2012, 04). The impact of the economic crisis on innovation: Evidence from europe. *Technological Forecasting and Social Change 80*.

Austin, P. (2011, 05). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research 46*, 399–424.

Autor, D., L. F. K. and A. B. Krueger (1998, 01). Computing inequality: Have computers changed the labor market? *Quarterly Journal of Economics*, 1169–1213.

Berman, E., J. Bound, and Z. Griliches (1994). Changes in the demand for skilled labor within us manufacturing: evidence from the annual survey of manufactures. *The Quarterly Journal of Economics 109*(2), 367–397.

Berndt, E., C. Morrison, and L. Rosenblum (1992, 04). High-tech capital formation and labor composition in u.s. manufacturing industries: An exploratory analysis.

Berndt, E. R. and C. J. Morrison (1995, 01). Hightech capital formation and economic performance in u.s. manufacturing industries: An exploratory analysis. *Journal of Econometrics*, 9–43.

Bloom, N., S. Bond, and J. Reenen (2007, 03). Uncertainty and investment dynamics. *Wiley-Blackwell: Review of Economic Studies 74*.

Bloom, N. and N. Pierri (2018, 08). Cloud computing is helping smaller, newer firms compete.

Bosworth, B. P. and J. E. Triplett (2007). Is the 21st century productivity expansion still in services? and what should be done about it? *Mimeographed document*.

Bresnahan, T., E. Brynjolfsson, and L. Hitt (2002, 02). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics 117*, 339–376.

Bresnahan, T. F. and M. Trajtenberg (1995). General purpose technologies 'engines of growth'? *Journal of econometrics 65*(1), 83–108.

Brynjolfsson, E. and L. Hitt (1996, 01). Paradox lost? firm-level evidence on the returns to information systems spending. *Management Science*, 541–582.

Brynjolfsson, E. and L. Hitt (2000, 01). Computing productivity: Are computers pulling their weight? *Mimeo, MIT and Wharton*.

Brynjolfsson, E. and L. M. Hitt (2003). Computing productivity: Firm-level evidence. *Review of economics and statistics 85*(4), 793–808.

Brynjolfsson, E., T. Malone, V. Gurbaxani, and A. Kambil (1994, 01). Does information technology lead to smaller firms? *Management Science*, 1628–1644.

Brynjolfsson, E. and S. Science (1992, 02). The productivity of information technology : review and assessment.

Brynjolfsson, E. and S. Yang (1997, 01). The intangible benefits and costs of computer investments: Evidence from financial markets. *Proceedings of the International Conference on Information Systems*.

Chou, Y.-C., H. H.-C. Chuang, and B. B. Shao (2014). The impacts of information technology on total factor productivity: A look at externalities and innovations. *International Journal of Production Economics 158*, 290–299.

Chou, Y.-C. and B. B. Shao (2014). Total factor productivity growth in information technology services industries: A multi-theoretical perspective. *Decision Support Systems 62*, 106–118.

Chwelos, P., R. Ramirez, K. L. Kraemer, and N. P. Melville (2010). Research note—does technological progress alter the nature of information technology as a production input? new evidence and new results. *Information Systems Research 21*(2), 392–408.

Clemons, E. K. and M. C. Row (1992, 01). Information technology and industrial cooperation: The changing economics of coordination and ownership. *Journal of Management Information Systems*, 9–28.

Cochran, W. G. and D. B. Rubin (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 35*(4), 417–446.

Columbus, L. (2013). Making cloud computing pay. *Forbes*.

Commission, E. (2017). *Measuring the economic impact of cloud computing in Europe, Digital Single Market*. European Commission.

Coyle, D., D. Nguyen, et al. (2018). Cloud computing and national accounting. *Economic Statistics Centre of Excellence (ESCoE) Discussion Paper 19*.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014, 08). The role of entrepreneurship in us job creation and economic dynamism. *The Journal of Economic Perspectives 28*.

Dedrick, J., V. Gurbaxani, and K. L. Kraemer (2003). Information technology and economic performance: A critical review of the empirical evidence. *ACM Computing Surveys (CSUR) 35*(1), 1–28.

Dedrick, J., K. L. Kraemer, and E. Shih (2013). Information technology and productivity in developed and developing countries. *Journal of Management Information Systems 30*(1), 97–122.

DeStefano, T., R. Kneller, and J. Timmis (2018, 09). Broadband infrastructure, ict use and firm performance: Evidence for uk firms. *Journal of Economic Behavior & Organization 155*.

DeStefano, T., R. Kneller, and J. Timmis (2020). Cloud computing and firm growth.

Dewan, S. and K. Kraemer (2000). The impacts of information technology on total factor productivity: A look at externalities and innovations. *Management Science*.

Dewan, S. and C.-k. Min (1997). The substitution of information technology for other factors of production: A firm level analysis. *Management science 43*(12), 1660–1675.

Diamond, A. and J. S. Sekhon (2013, July). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *The Review of Economics and Statistics 95*(3), 932–945.

Dixit, A. and R. Pindyck (2012, 07). *Investment under Uncertainty*.

Eberhardt, M., C. Helmers, et al. (2010). Untested assumptions and data slicing: A critical review of firm-level production function estimators. *Department of Economics, University of Oxford*.

Etro, F. (2009, 01). The economic impact of cloud computing on business creation, employment and output in europe. an application of the endogenous market structures approach to a gpt innovation. *Review of Business and Economics LIV*, 179–208.

Ewens, M., R. Nanda, and M. Rhodes-Kropf (2018, 03). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics 128*.

Greenan, N. and J. Mairesse (1996, 01). Computers and productivity in france: Some evidence. *NBER*.

Greenwood, J. (1997). *The third industrial revolution: technology, productivity, and income inequality*. Number 435. American Enterprise Institute.

Guiso, L. and G. Parigi (1999, 02). Investment and demand uncertainty. *The Quarterly Journal of Economics 114*, 185–227.

Guo, S., R. Barth, and C. Gibbons (2006, 04). Propensity score matching strategies for evaluating substance abuse services for child welfare clients. *Children and Youth Services Review 28*, 357–383.

Gurbaxani, V. and S. Whang (1991, 01). The impact of information systems on organizations and markets. *Communications of the ACM*, 59–73.

Hajji, A., R. Pellerin, P.-M. Léger, A. Gharbi, and G. Babin (2012). Dynamic pricing models for erp systems under network externality. *International journal of production economics 135*(2), 708–715.

Hansen, B. (2004, 02). Full matching in an observational study of coaching for the sat. *Journal of the American Statistical Association 99*, 609–618.

Helpman, E. (1998). *General purpose technologies and economic growth*. MIT press.

Hitt, L. M. (1999, 06). Information technology and firm boundaries: Evidence from panel data. *Information Systems Research*, 134–149.

Jin, W. and K. McElheran (2018, 01). Economies before scale: Survival and performance of young plants in the age of cloud computing. *SSRN Electronic Journal*.

Jorgenson, D. W., M. S. Ho, and K. J. Stiroh (2006). The sources of the second surge of us productivity and implications for the future. *Federal Reserve Bank of New York, mimeo*.

Jorgenson, D. W., M. S. Ho, and K. J. Stiroh (2008). A retrospective look at the us productivity growth resurgence. *Journal of Economic perspectives 22*(1), 3–24.

Jorgenson, D. W. and K. Stiroh (1995, 03). Computers and growth. *Journal of Economics of Innovation and New Technology*, 295–316.

Jorgenson, D. W. and K. J. Stiroh (2000, May). U.s. economic growth at the industry level. *American Economic Review 90*(2), 161–167.

Jovanovic, B. (1982, 02). Selection and evolution of industry. *Econometrica 50*, 649–670.

Kauffman, R. J. and A. A. Techatassanasoontorn (2009). Understanding early diffusion of digital wireless phones. *Telecommunications Policy 33*(8), 432–450.

Kerr, W., R. Nanda, and M. Rhodes-Kropf (2014, 08). Entrepreneurship as experimentation. *The Journal of Economic Perspectives 28*.

Kim, G., B. Shin, K. K. Kim, and H. G. Lee (2011). It capabilities, process-oriented dynamic capabilities, and firm financial performance. *Journal of the association for information systems 12*(7), 1.

Kim, S. W. and R. Narasimhan (2002). Information system utilization in supply chain integration efforts. *International Journal of Production Research 40*(18), 4585–4609.

Knight, F. (1921, 01). Risk, uncertainty and profit.

Krueger, A. B. (1993). How computers have changed the wage structure: evidence from microdata, 1984–1989. *The Quarterly Journal of Economics 108*(1), 33–60.

Lee, B., J. Lessler, and E. Stuart (2009, 11). Improving propensity score weighting using machine learning. *Statistics in medicine 29*, 337–46.

Lee, S.-Y. T., R. Gholami, and T. Y. Tong (2005). Time series analysis in the assessment of ict impact at the aggregate level–lessons and implications for the new economy. *Information & Management 42*(7), 1009–1022.

Levy, A. (2000, 12). Why does capital structure choice vary with macroeconomic conditions?

Lichtenberg, F. and B. Lehr (1996, 01). Computer use and productivity growth in federal government agencies. *Columbia - Graduate School of Business, Papers*.

Lichtenberg, F. R. (1995). The output contributions of computer equipment and personnel: A firm-level analysis. *Economics of innovation and new technology 3*(3-4), 201–218.

Lin, W. T. and B. B. Shao (2006). The business value of information technology and inputs substitution: the productivity paradox revisited. *Decision support systems 42*(2), 493–507.

Mead, R. (2017, Jan). Semio: Developing a cloud-based platform for multimodal conversational ai in social robotics. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 291–292.

Miozzo, M., V. Walsh, et al. (2006). *International competitiveness and technological change*. Oxford University Press on Demand.

Mithas, S., A. Tafti, I. Bardhan, and J. M. Goh (2012). Information technology and firm profitability: mechanisms and empirical evidence. *Mis Quarterly*, 205–224.

Mohamed, M. and S. Pillutla (2014, 08). Cloud computing: A collaborative green platform for the knowledge society. *VINE 44*.

Morgan, S. and J. Todd (2008, 12). A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociological Methodology 38*, 231 – 281.

Morrison, C. (2000, 01). Assessing the productivity of information technology equipment in u.s. manufacturing industries. *The Review of Economics and Statistics 79*, 471–481.

Mukhopadhyay, T., S. Rajiv, and K. Srinivasan (1997, 12). Information technology impact on process output and quality. *Management Science 43*, 1645–1659.

OECD (2015). *OECD digital economy Outlook 2015*. Organisation for Economic Co-operation and Development.

Oliner, S. and D. Sichel (1994, 02). Computers and output growth revisited: How big is the puzzle? *Brookings Papers on Economic Activity 25*, 273–334.

Oliner, S. D. and D. E. Sichel (2000). The resurgence of growth in the late 1990s: is information technology the story? *Journal of economic perspectives 14*(4), 3–22.

Oxley, J. E. and B. Yeung (2001). E-commerce readiness: Institutional environment and international competitiveness. *Journal of International Business Studies 32*(4), 705–723.

Palmer, M. (2012). Cloud computing cuts start-up costs. *Financial Times 29*.

Pan, W. and H. Bai (2015). *Propensity score analysis: Fundamentals and developments*. Guilford Publications.

Robison, K. K. and E. M. Crenshaw (2002). Post-industrial transformations and cyber-space: A cross-national analysis of internet development. *Social Science Research 31*(3), 334–363.

Rosenbaum, P. (1989, 12). Optimal matching for observational studies. *Journal of The American Statistical Association - J AMER STATIST ASSN 84*, 1024–1032.

Rosenbaum, P. and D. Rubin (1983, 04). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

Rosenbaum, P. R. and D. B. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician 39*(1), 33–38.
Rubin, D. (2001, 12). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology 2*, 169–188.
Schafer, J. and J. Kang (2009, 01). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological methods 13*, 279–313.
Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle (1912/1934). *Transaction Publishers.–1982.–January 1*, 244.
Schwab, K. (2017). *The fourth industrial revolution*. Currency.
Sevior, M., T. Fifield, and N. Katayama (2010, 05). Belle monte-carlo production on the amazon ec2 cloud. *Journal of Physics: Conference Series 219*, 012003.
Shy, O. and S. Oz (2001). *The economics of network industries*. Cambridge university press.
Solow, R. M. (1987). We'd better watch out. *New York Times Book Review 36*.
Stanley, M. and S. Roach (1987). America's technology dilemma: A profile of the information economy. *Special Economic Study (April 22)*.
Steiner, P. and D. Cook (2013, 01). Matching and propensity scores. *The Oxford Handbook of Quantitative Methods*, 237–259.
Stiroh, K. and D. Jorgenson (1999, 02). Information technology and growth. *American Economic Review 89*, 109–115.
Stiroh, K. J. (2002). Are ict spillovers driving the new economy? *Review of Income and Wealth 48*(1), 33–57.
Stuart, E., G. King, K. Imai, and D. Ho (2011, 06). Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software 42*.
Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science 25*(1), 1 – 21.
Thomke, S. (2003, 01). *experimentation matters unlocking the potential of new technologies for innovation*.
Timmer, M., R. Inklaar, M. O'Mahony, and B. Van Ark (2018). *Economic Growth in Europe: A Comparative Industry Perspective*. Princeton University Press.
Trivedi, H. (2013). Cloud computing adoption model for governments and large enterprises.
Varian, H. R., J. Farrell, C. Shapiro, et al. (2004). *The economics of information technology: An introduction*. Cambridge University Press.
Wagner, J. (2002). The causal effects of exports on firm size and labor productivity: first evidence from a matching approach. *Economics Letters 77*(2), 287–292.
Wang, C., Z. Bi, and L. Xu (2014, 05). Iot and cloud computing in automation of assembly modeling systems. *Industrial Informatics, IEEE Transactions on 10*, 1426–1434.
Wang, H., Y. Ma, G. Pratx, and L. Xing (2011, 09). Toward real-time monte carlo simulation using a commercial cloud computing infrastructure. *Physics in medicine and biology 56*, N175–81.
White, H. and S. Sabarwal (2014). Quasi-experimental design and methods. *Methodological briefs: impact evaluation 8*, 1–16.
Yadan, L. (2019, 01). From transportation infrastructure to green infrastructure — adaptable future roads in autonomous urbanism. *Landscape Architecture Frontiers 7*, 92.