Firm in the clouds

Supplier market power and technology adoption in cloud computing

[Preliminary: Do not cite without permission.]

Nicholas Emery-Xu[*†]

August 2, 2026

Abstract

I study the effects of technology adoption on firm productivity in the European cloud computing services market, a setting with falling adoption costs and a high concentration of suppliers. I study how firms make technology bundling choices when faced with these adoption frictions, and analyze the effects of competition policies on downstream adoption and welfare. I find that adopting cloud computing increases firm productivity from 0.2% to 1.6%, with heterogeneous effects across sectors and a larger impact for firms that purchase services from multiple cloud providers. To estimate the effects of competition policies to increase supplier switching, I estimate a model of industry dynamics, in which firms produce output and make computing input bundle choices. I find substantial entry and switching costs, though low rates of supplier switching are largely driven by heterogeneity in brand preferences. Simulating the effects of a ban on data egress fees from the EU Data Act and an increase in software interoperability, I find these policies produce annual welfare gains of 0.05% and 0.10%.

[*]UCLA, Department of Economics

[†]I am grateful for the support and feedback of my advisors John Asker, Hugo Hopenhayn, Simon Board, and Will Rafey. I thank (in random order) Michael Rubens, Jackson Mejia, Danial Lashkari, Martin Hackmann, Yangkeun Yun, Bill Whyman, Jiaer Liu, Theodore Naff, Jacopo Bellabio, Jihye Jeon, Jacob Lariviere, and seminar and workshop participants at UCLA, MIT FutureTech, IIOC 2026, BSE Summer Forum Applied IO Workshop 2026, and the Institute for Humane Studies for data provision and useful feedback. I thank several anonymous cloud engineers at a large health insurance firm, and a data scientist at a cloud security firm for answering my questions about adoption. I thank the Institute for Humane Studies, UCLA Lewis L Clarke Fellowship, and Coefficient Giving for financial support. This work used computational and storage services associated with the Hoffman2 Cluster which is operated by the UCLA Office of Advanced Research Computing's Research Technology Group. All remaining errors are my own.

1 Introduction

> *"By 2030: 75% of European businesses should use cloud-edge technologies for their activities."* - European Commission

> *"The French position is to encourage and support European innovation while preventing historical monopolies from recurring."* - Mireille Clapot, MP

General-purpose technologies increase firm productivity, but may be adopted too slowly as firms continue to use older technologies (Collard-Wexler and De Loecker [2015], Chari and Hopenhayn [1991], Bresnahan and Trajtenberg [1995]). Existing literature largely focuses on adoption frictions (Hornbeck et al. [2024]), assuming technology input decisions are independent of supplier market power.

However, many information technologies are provided by a relatively small handful of suppliers, raising concerns that market power could raise input markups or price discriminate against users, producing inefficiently low rates of adoption of these technologies. A recent example is cloud computing, which allows firms to increase productivity by leasing computing inputs over the Internet (Jin and Bai [2022]). The general-purpose nature of the technology, combined with a high concentration of suppliers, has prompted regulatory agencies worldwide to begin to research how to increase supplier competition with the goal of increasing utilization (Competition and Markets Authority [2025h], Korea Fair Trade Commission [2022]).

In this paper, I ask whether supplier market power produces inefficiencies in the rate of technology adoption in the cloud computing industry. To do so, I construct a new model of firm technology adoption in which firms choose whether to purchase cloud computing services from a small set of differentiated suppliers. Adopting cloud computing offers productivity benefits, which are compounded when firms bundle suppliers. Firms face two forms of market power, prices above marginal cost (Byrne et al. [2018]) and switching frictions when changing suppliers. I derive a computationally-tractable estimation strategy to structurally estimate the model parameters. Using the model, I assess the counterfactual impact of competition policies proposed by European and Korean regulators on downstream adoption and total

welfare.[1]

I introduce supplier choice in a model of industry dynamics with technology adoption. Each year, firms choose their set of computing inputs, as well as capital investment decisions in anticipation of their future productivity gains. Firms can choose between on-premise only computing (old technology) or on-premise along with any bundle of cloud technologies from existing providers. Adopting cloud computing increases firm productivity, with additional benefits from bundling multiple providers, and providers are heterogeneous in quality-adjusted variable costs. Finally, firms face fixed costs when first adopting cloud computing and when switching bundles that serve as a form of lock-in to existing technologies.

To model supplier market power, vertically-differentiated cloud computing suppliers compete for short run profits by setting per-unit computing prices and charging data egress fees for moving data between suppliers. After observing the prior period market shares, firms choose egress fees and usage prices to attract new consumers and maximize revenue from existing ones.

The modeled effects that can produce inefficiently low adoption rates are twofold. First, bundling cloud providers creates positive productivity benefits for users that are at best partially priced in by incumbent firms (Bresnahan and Trajtenberg [1995]). In addition, the high concentration on the supply side raises prices above marginal cost and allows suppliers to charge fees for moving data off platform, reducing the benefits from adoption.

I estimate the model using a newly-merged dataset on French firms during the period 2017-2021. I merge Orbis data on firms revenue and costs to a survey on firms' technology usage from Aberdeen, a market research firm. I manually construct a dataset of cloud computing prices and features from a set of first- and third-party websites. My data notably contains data on firms' cloud computing usage by supplier, allowing me to analyze bundling and switching over time. In addition, I observe features of firms' on-premise software stack and distance from cloud datacenters, such that I can construct a rich model of supplier choice.

Using my dataset, I first document overall patterns of usage. I find that adoption has risen rapidly, from around 12% of firms in 2017 to 28% by 2021. Bundling is common, with

[1] [**In progress: Include lit related to supplier market power in firm dynamics literature**]

over 1/2 of cloud-using firms using multiple suppliers in 2021. Larger firms, and firms in more high-technology industries are more likely to adopt cloud computing. Despite rapid adoption, the upstream market remains concentrated - among firms in my sample, over 75% use two suppliers - Microsoft and Amazon.

I estimate my model in two stages. First, I estimate an industry-level model of production, using a control function approach to control for endogenous selection into cloud computing (de Loecker [2013]). I find that adopting cloud computing increases firm productivity by 0.2-1.5%, with the largest gains for high-technology manufacturing firms. I find support for the primary externality: bundling amplifies the productivity effects of cloud computing adoption. Over time, firms that use cloud computing have noticeably higher productivity levels than those that do not, and those that bundle suppliers experience additional productivity growth.

Second, I estimate the adoption, switching, and maintenance costs of using cloud computing using the full-solution approach to estimate a Markov moment equilibrium (Ifrach and Weintraub [2017]). I show that adoption costs are large but falling over time due to global progress in reducing frictions. While switching costs are large, around €3 M per petabyte of data stored, I find that persistence in provider choice is driven by price as well as complementarities between cloud computing and on-premise software.

Using these estimates, I simulate three counterfactual scenarios, based on existing policies to promote cloud computing adoption. First, I simulate an egress fee ban, or ban on charging consumers to move data off of a cloud provider's servers, based on the one implemented in the EU Data Act of 2024, which reduces firms' switching frictions based on their data stock and can lead to increased rates of adoption and bundling. I find that, holding supplier prices fixed, an egress fee ban leads to welfare gains fo 0.02% or €47.9 per year, while allowing suppliers to reoptimize prices increase the gain to 0.05%, or €157.2 per year. Second, I simulate increased complementarity between Microsoft server products and rival cloud platforms, based on concerns about anticompetitive policies raised by the UK Competition and Markets Authority. Such a mandate would increase welfare by 0.01%-0.10% per year, up to €297.5 million. Finally, to study the role of market power directly, I simulate the effects of a horizontal merger between Microsoft's and Google's cloud divisions, resulting in welfare losses of €841.7 million per year, or 0.28%.

Related literature: My paper contributes to several bodies of literature. First, it contributes to the literature on the dynamic productivity effects of firm technology adoption (Rubens [2025], Hornbeck et al. [2024], Humlum [2022], Aw et al. [2011]). I model cloud computing products as having both features of substitutes, as did water and steam power in Hornbeck et al. [2024], and of complements when firms multihome, as in Peters et al. [2017], who find complementarities between product and process innovation. Unlike prior literature, my model focuses on the impact of product differentiation and pricing on the downstream benefit of adoption.[2] I exploit the ability to observe adoption at the supplier level to model firms' costs and benefits of adoption as depending on prices charged by providers and a rich set of product features.[3]

Second, this paper contributes to the literature on the economics of IT adoption, particularly cloud computing, on firm productivity (De Ridder [2024], Brynjolfsson and Hitt [2003], Tambe and Hitt [2012], Bloom et al. [2016], Thompson [2017]), growth (Azar [2021], Barefoot et al. [2018], Bresnahan and Trajtenberg [1995]), and innovation (Besiroglu et al. [2024]). While prior work has focused on the productivity gains from a broad suite of cloud computing technologies (Jin and Bai [2022], DeStefano et al. [2023]), I focus particularly on infrastructure-as-a-service, the leasing of computation and storage.[4] My paper develops a new structural model of cloud adoption, featuring entry, switching, and multihoming decisions as well as provides the first, to my knowledge, empirical estimation of cloud supplier pricing conduct.[5]

Finally, this paper contributes to the literature on the implementation of competition policy in oligopolistic markets for general purpose technologies (Kearns [2024], Igami and Uetake [2020], Goettler and Gordon [2011]).[6] While the majority of these papers focus on policies that increase innovation, I model innovation as exogenous, since the measurable

[2]Gortmaker [2024] focuses on how heterogeneity in open source web software impacts firms' production costs, but does not consider productivity benefits.

[3]More broadly, this paper contributes to a literature of dynamic discrete choice with switching frictions (for examples, see Honka [2014], Shcherbakov [2016], Polyakova [2016], Yeo and Miller [2018], Gowrisankaran and Rysman [2012], Dubé et al. [2009]).

[4]More broadly, this paper is also related to a wider literature on capital adjustment frictions (Asker et al. [2014], Bloom [2009]) and how leasing capital can reduce such frictions (Gavazza [2011]).

[5]Kilcioglu and Rao [2016] build a theoretical model of price and quality competition.

[6]More geenrally, this paper also contributes to the literature on the diffusion of GPTs (Humlum [2022], Chari and Hopenhayn [1991].

parts of quality improvements are a result of innovations by semiconductor (e.g. Intel) and hard-disk manufacturers (e.g. Western Digital). Instead, I focus on the implementation of policies that affect pricing. [7] A few theoretical papers focus on the impact of an data egress fee ban on cloud computing markets, finding that such a ban is either welfare reducing (Gans et al. [2023]) or depends on the demand elasticity (Do and Miklós-Thal [2025]). I estimate such a counterfactual and find that a ban is welfare-improving, though suppliers see a welfare decrease.

The rest of the paper proceeds as follows. Section 2 provides background on the cloud industry, focusing on the economic forces driving concentration in cloud computing and regulatory responses. Section 3 describes my dataset, including summary statistics and stylized facts that will drive my modeling decisions. Section 4 presents a structural model of firm production with cloud supplier choice, featuring multihoming and adoption and switching frictions. Sections 5 and 6 describes the my estimation procedure, including both firm production and cloud computing costs, sources of identification, and results, comparing them when possible to existing estimates in the literature or in industry reports. Section 7 considers my three counterfactual scenarios: the first two scenarios consider cases both holding supplier prices fixed and allowing suppliers to respond to the implemented policy. Finally, Section 8 concludes.

2 Background

Given that cloud computing is a technology with wide applicability, the welfare loss to consumers from upstream market power may be large. This section provides background on cloud computing, including adoption and technical details. I then discuss upstream market structure and regulatory responses to increase competition and consumer adoption. Further details are in Appendix E.

[7] Zhuo and Tiew [2025] focuses on the impact of location subsidies on the pricing of semiconductors.

2.1 Background on cloud computing adoption

Cloud computing allows firms to lease computation, storage, and software from cloud providers instead of investing in their on-premise IT services (Asker et al. [2014], Gavazza [2011], DeStefano et al. [2023]). Cloud firms invest in physical computing hardware and lease computing capital, as well as complementary software and services, to firms. This allows cloud providers to take advantage of economies of scale, spreading datacenter costs over a larger set of computing hardware in order to lower prices below that of on-premise computing.[8]

As a result of these productivity gains, cloud adoption is widespread. The European Commission reports that 45.2% of European firms have adopted the technology by 2023 (Eurostat / European Commission [2023]). Likewise, using Census data, Zolas et al. [2021] find that 78.4% of U.S. firms that use IT capital purchased at least one cloud service in 2018. Existing studies have found significant productivity gains from cloud services (DeStefano et al. [2023], Jin and McElheran [2024]).

Cloud computing comes in three primary forms: software-as-a-service (SaaS), platform-as-a-service (PaaS), infrastructure- as-a-service (IaaS). The first is the broadest and encompasses software leasing, such as Adobe Creative Cloud and Microsoft Office 365, hosted in the cloud. PaaS, such as Dropbox or Google Drive, also abstracts away from the underlying hardware by focusing on an application layer such as serverless computing or containers. IaaS is the simplest service, focusing on the leasing of "pure" computation (usually in the form of virtual machines (VMs)) and storage (usually in the form of "blobs"). For more detail, see Appendix E.1.

Since my data allows me to distinguish between IaaS, PaaS, and SaaS adoption, throughout the remainder of this paper, I focus on IaaS for two reasons. First, IaaS is the most popular cloud computing service.[9] Second, pricing is relatively straightforward, with con-

[8]Studies commissioned by cloud providers tend to find large economic impacts. A 2019 study commissioned by Microsoft finds a return on investment of 428% over 3 years North et al. [2019], with fixed migration costs totaling $1.38 million. A 2022 study of AWS finds an ROI of 241% over 3 years, with fixed costs totaling $604,801. Phillips and Owens [2022]. Finally, a case study of an IT services firm in the UK oil and gas industry finds a 37% cost reduction from using Amazon's IaaS relative to on-premise and a 21% reduction in support calls from customers Khajeh-Hosseini et al. [2010]. However, Emeras et al. [2016] find that usage of the University of Luxembourg's in-house computing cluster was less costly than leasing the equivalent compute on Amazon's AWS.

[9]See Brand et al. [2024] Appendix A. According to the UK Competition and Markets Authority, in 2022, Azure's top product was compute, or IaaS virtual machines (Competition and Markets Authority [2025a],

sumers choosing from a menu of compute and storage products and paying a linear price per hour in the case of VMs or gigabyte-month in the case of storage. In addition, a significant majority of consumers are price-takers.[10] This allows me to use supplier-year average list prices since my data do not contain consumer level prices.[11]

2.2 Market structure and regulatory response

Like many information technology markets with large returns to scale, the market for cloud computing services is highly concentrated, potentially allowing suppliers to raise prices above marginal cost and reduce consumer welfare. Worldwide, it is largely dominated by Microsoft, Amazon, and Google. The UK Competition and Markets Authority, for example, found that the HHI of the IaaS market was over 3000 between 2020 and 2023 (Competition and Markets Authority [2025h], p.90). In 2025, the top 3 cloud providers had a 70% market share in the IaaS market (Statista). This is due partly to the high fixed costs of entry (cf. De Ridder [2024]), which requires large datacenter, storage device, and semiconductor investments.[12]

As a result, cloud suppliers are able to price discriminate against consumers in the form of egress fees, or fees to move data out of a cloud service.[13] Figure 1 displays an example of these egress fees for Microsoft Azure. While these transfers have nonzero marginal cost, evidence suggests that firms charge a markup for data transfer (Do and Miklós-Thal [2025], Competition and Markets Authority [2025e], p. 9). In addition to shifting welfare from consumer surplus to producer surplus, such fees can also reduce the positive externality firms derive from multihoming, which I show below is an important driver of productivity gains from adopting IaaS.[14]

17).

[10] In a survey of IaaS users, the UK CMA found that only the largest 5% of firms were able to negotiate a price below list price (Competition and Markets Authority [2025a], p. 8). For further discussion, see Brand et al. [2024].

[11] For the remainder of the paper, I use the terms "IaaS" and "cloud computing" interchangeably.

[12] Google Cloud, for example, did not experience positive operating income until Q1 2023.

[13] Besides the egress charged by firms, firms also incur switching frictions from learning a new platform, including retraining employees, hiring new specialists or consultants for their provider(s) of choice, and rewriting software to use the new platform.

[14] The evidence for whether egress fees hinder multihoming is mixed. A 2022 study by the Korean Fair Trade Commission found that 60.3% of firms cited technological frictions and costs of data migration as a barrier to multihoming (Korea Fair Trade Commission [2022]). On the other hand, the UK CMA found that several mentioned such costs as a minor barrier to using multiple providers (Competition and Markets Authority

Internet Egress (routed via Microsoft Premium Global Network)

Source Continent	First 100 GB/Month	Next 10 TB/Month	Next 40 TB/Month	Next 100 TB/Month	Next 350 TB/Month
From North America, Europe to any destination	Free	$0.087 per GB	$0.083 per GB	$0.07 per GB	$0.05 per GB
From Asia (China excluded), Australia, MEA to any destination	Free	$0.12 per GB	$0.085 per GB	$0.082 per GB	$0.08 per GB
From South America to any destination	Free	$0.181 per GB	$0.175 per GB	$0.17 per GB	$0.16 per GB

Note: 1 TB = 1,000 GB

*Contact us for over 500 TB

Figure 1: Sample of data egress fees charged by Microsoft Azure in 2025.

To improve competition, regulators have honed in on a few remedies - capping data egress fees and mandating software interoperability between IaaS suppliers. Since 2022, regulators in the UK, Japan, South Korea, France, and the European Union have considered banning data egress fees or limiting them to marginal cost (Korea Fair Trade Commission [2022]). In France, the SREN law (Loi Visant à Sécuriser et à Réguler l'Espace Numérique) entered into force on May 21, 2024. The law goes hand-in-hand with the EU Data Act, which will enter into force on January 12, 2027. Both regulations ban egress fees for cloud computing customers. On the other hand, the UK CMA recommends against capping egress fees at marginal cost due to the technical difficulties of monitoring and enforcement and the fact that only a few firms they surveyed mentioned egress fees as a barrier to switching suppliers or exiting (Competition and Markets Authority [2025d] 4, 7-10). In response to the Data Act, all three of the major providers have eliminated data egress costs for exiting customers, but have not changed costs for continuing customers (Competition and Markets Authority [2025g], 44).[15]

A second remedy relates to Microsoft's vertical integration of popular server software–Windows Server and SQL Server–that makes it feasible to foreclose their use on rival cloud platforms. In particular, Microsoft prevents customers from transferring on-premise server

[2025d] pp. 4, 7-10).

[15]In Appendix **??**, I exploit Google's unexpected announcement on January 11, 2024 that it would drop egress fees for exiting customers worldwide to see whether a drop in these fees reduced demand for Google virtual machines in the short run. I estimate a difference-in-differences specification 54 days around the announcement of the effect of the announcement on virtual machine spot prices, finding no significant effects. The estimation window is limited by the fact that both Amazon and Microsoft responded to Google's announcement by also dropping prices in the EU in early March 2024.

software licenses to public cloud services, though Microsoft server software consumers can obtain a discount when purchasing Azure virtual machines using those software through its Auzre Hybrid Benefit program, launched in 2016 (Competition and Markets Authority [2025h]). As a result, the CMA has found the median consumer pays more to use Microsoft server software on Google or Amazon's clouds than on Microsoft's (Competition and Markets Authority [2025f]). This has led for calls to mandate price parity for Microsoft server licenses across cloud platforms.[16]

3 Data

To study the determinants of cloud computing adoption and its impact on firm production, I construct a dataset of firms' inputs, output, and technology usage, as well as cloud supplier prices, features, and costs. To obtain data on both sides of the market, I merge several data sources to create a novel dataset of firms' cloud computing supplier investments, non-IT investments, and prices. In this section, I describe the procedure used to match firms' financial data to technology usage, including data on cloud usage and other computing inputs. Then, I describe how I compute supplier prices, features, and marginal costs. Finally, I describe several stylized facts about cloud adoption and usage that will inform my empirical model. For more details on the data, see Appendix A.

3.1 Dataset construction

To build my dataset, I combine data from Orbis, the Aberdeen CI Technology Database (CITDB), anonymized IaaS usage data from Microsoft Azure, list prices from the three largest cloud providers and CPU quality and cost measures from Passmark. The panel covers firms at the annual level from 2017 to 2021.

Firm level financial data. I use data from the Orbis dataset for France from 2015-2021, including revenue, capital stock, materials expenditures, labor, and the wage bill. Due to French financial reporting requirements, Orbis covers a large fraction of total output of both public and private firms, limiting sampling biases. As such, data on France specifically has

[16]For further discussion of the regulatory environment in the EU and France, see Appendix E.2.

been widely used in the literature on firm-level productivity (e.g. Asker et al. [2014], Bloom et al. [2016]). While I only observe data on computing inputs for a subset of firms, I use the full Orbis sample to construct industry-location-year average wages and industry-year average depreciation rates at the 4-digit NAICS code level.

Firm technology adoption. I use the Aberdeen CITDB to obtain firms' cloud vendor choices, data storage and other IT hardware investments, and vendor choice for Microsoft server technologies. Aberdeen is a market research firm that seeks to obtain data on firms' IT usage, hardware, and software investments in order to sell to IT vendors, which it obtains largely through interviews, supplemented with interpolation and web-scraping. It has been used extensively on research studying the impacts of IT inputs on firm performance (Brynjolfsson and Hitt [2003]), including cloud computing (Bloom and Pierri [2018]). For a full description of the methodology, see Appendix A.1. The data is reported at the establishment level. To obtain firm-level data, I sum investment data across domestic establishments. Cloud computing usage and server vendor are reported at the extensive margin. I assume a firm uses a particular cloud provider if any of its establishments is observed having used the provider.[17]

Cloud pricing. Cloud pricing data come from web scraping the archived VM pricing pages on Wayback Machine for Microsoft and Google. Amazon's data is not archived directly, so I use archives of EC2Instances.info Contributors [2025] to construct my panel. I scrape pricing for virtual machines at all datacenters for a given provider nearest to an adopting firm's headquarters at a quarterly level and then average across quarters. Since I do not observe which datacenters firms are using, I follow the assumptions in Wang et al. [2019] and assume that firms use the datacenter closest to their headquarters at the time of entry.[18] Figure 12 in Appendix A.7 provides maps of suppliers' datacenter locations in 2017 and 2021. I also scrape data egress fees from Microsoft and Google's sites. As Amazon's data is not archived between 2015 and 2025, I use current prices and price change announcements on the AWS blog to reconstruct the price history (Amazon Web Services [2025]).

[17] See Appendix A.3 for a table of the IaaS technologies I observe for each provider.

[18] Some firms may choose datacenters that are instead near a majority of their consumers. See Wang et al. [2019] for a discussion of the attenuation bias that this produces. Demirer et al. [2024] provides evidence using internal data from a major cloud provider that datacenter choices are persistent, even within a supplier, and that 98% of firms headquartered in the US or EU use only datacenters in that region.

As there are a large number of SKUs across virtual machines and storage, I make two simplifying assumptions about the prices firms face. First, I assume that firms respond to VM prices over storage prices. Prior research has shown that firms that use IaaS use more computation than storage (Demirer et al. [2024]). In addition, Amazon's storage prices are not fully archived, making it difficult to reconstruct an accurate price history. Second, to deal with the high number of SKUs offered by each supplier, I take the weighted average of the price per vCPU hour at each datacenter-quarter-year, where the weights come from an anonymized representative sample of 5,000 customers of the monthly virtual machine usage at a Microsoft Azure datacenter in 2017 and again in 2019 (Cortez et al. [2017]).[19] For Microsoft Azure, I match prices to usage exactly. For Google and Amazon, I match each product to the nearest product in the Azure trace data, using fuzzy matching on product features such as core count and memory. Summary statistics, discussion of the match procedure, and further details are presented in Appendix A.6.

Other data sources. I use a number of auxiliary datasets to add control variables and deflate quantities. The data on datacenter locations for each cloud provider comes primarily from `datacenters.com` (Datacenters.com [2025]). For the remaining datacenters, I manually compile data from the colocation partners of the major cloud providers, and then obtain the latitude and longitude from `datacenters.com`. Second, I data from PassMark Software [2026a] to compute CPU quality, which I take to be the weighted average benchmark score of CPUs in a datacenter, where weights again come from Cortez et al. [2017]. Passmark also provides data on CPU prices to calibrate marginal costs. Third, data used to deflate the prices of inputs and outputs comes from Eurostat, with the exception of the price index for computer hard drive storage, which comes from the US Bureau of Labor Statistics (BLS).

Matching and cleaning. As the downstream firm datasets do not contain a common firm identifier, I merge them using a combination of exact and string-distance metrics based on firm name and other identifying information, such as website URL and zipcode. The match procedure is described in Appendix A.2. Using this procedure, I am able to match 24.2% of firms in my Orbis sample. To clean the data, I replace missing revenue and employ-

[19] Much of the variation in pricing across SKUs comes from the number of vCPUs offered on the processor. Dividing by vCPUs reduces the pricing variance.

ment observations in the Orbis data with data from Aberdeen, which in turn sources these data from Dun & Bradstreet, and missing depreciation rates and wages with their averages at the industry (NACE 4-digit)-region (NUTS3)-year level. I deflate nominal values for computational investments, capital, materials, and output using industry-level Eurostat price deflators. I deflate storage investments using data from the US Bureau of Labor Statistics (McCallum and of Labor Statistics [2024]). Finally, I use the deflated book value of capital as well as Orbis depreciation data to construct a measure investment. If depreciation data is omitted, I use the average annual depreciation rate constructed from the entire Orbis sample.[20]

Summary statistics for the matched sample are displayed in Table 13. Prior research has demonstrated that the response rate to Aberdeen is increasing in firm size (Bloom et al. [2016]). In Appendix A.4 I compare the full Orbis sample to my matched subsample.

Table 1: Summary Statistics

Variable	Observations	Mean	25th Percentile	Median	75th Percentile
Revenue	1,038,913	20.43	1.00	2.07	6.06
Labor	1,038,913	67.42	6.00	14.00	32.00
Capital	1,027,282	28.35	0.11	0.38	1.36
Materials	704,949	19.45	0.63	1.90	6.83
Computation	1,038,913	73.45	2.00	9.00	25.00
Firm Age	1,038,913	22.74	11.00	20.00	31.00
IaaS Usage	1,038,913	0.10	0.00	0.00	0.00

Notes: The units for revenue, capital, and materials are in millions of euros. The units for labor and computation are in units. The unit for firm age is years.

3.2 Cloud adoption patterns

To motivate my empirical model below, I present a series of stylized facts regarding cloud computing usage in France. The first two motivate my modeling choices. In particular, I demonstrate that the market is rapidly expanding, with a near doubling in usage over my sample period. Second, I show that the upstream market is highly concentrated, dominated

[20] I use book values of capital rather than the perpetual inventory method due to data limitations, as Orbis only includes a full sample dating 10 years prior to the download date (January 2024). Details on the constriction of capital stocks and investment flows are in Appendix A.5.

by 3 large suppliers. The second two facts motivate recently-proposed pro-competitive remedies. First, I show that egress fees are negatively associated with firms' decision to switch suppliers. Second, I show that firms that use Microsoft server products have a greater probability of using Microsoft cloud products relative to those of other suppliers, indicating product complementarities.

Fact 1: Cloud computing usage is rising rapidly, and purchasing from multiple suppliers is common. In Figure 2, I plot the fraction of firms in my sample using cloud computing across all years.[21] Growth has been rapid, increasing from 12.4% of firms in 2017 to 27.9% in 2021, with a rapid increase during the COVID-19 pandemic. In addition, a large minority of firms are using two or even three cloud providers. On average, 45.95% of cloud users in my sample use more than one cloud provider, and 23.28% use all three major cloud providers.[22] As a result, several non-leading cloud providers, including Google and IBM, have sought to position themselves as the optimal second cloud for consumers (Competition and Markets Authority [2025a], 23).

Industry case studies show that firms experience strong complementarities from using multiple cloud providers. Multihoming allows firms to choose suppliers on whose virtual machines software or tasks run most efficiently. For example, in 2018, Capital One began using both Amazon and Google clouds - leveraging Amazon for its reliable infrastructure and Google for its data analytics software (MultiCloud.Tips [n.d.]). In addition, firms may use multiple providers to enhance reliability in the case of outages, as in the case of Adobe. However, multihoming has downsides. Besides the switching costs mentioned above, managing a workflow across multiple providers also requires additional costs - fixed costs of ensuring software compatibility and human capital to manage a workflow across multiple providers. For both of these reasons, one should expect to see larger firms be more likely to adopt multiple providers. Table 2 shows that multihoming firms tend to be larger than single-cloud users, with a higher intensity of IT usage, as measured by software investment.

Fact 2: Supply of cloud services is highly concentrated The market for cloud services is

[21] Aberdeen only tracks IaaS usage in 2016 for establishments that survived until 2020. To avoid this survivorship bias, I begin the figure in 2017.

[22] In the UK, the rates of multihoming among cloud users was 7.04% in 2020 and 7.50% in 2021 (Competition and Markets Authority [2025b], 5). On the other hand, Flexera finds that 86% of surveyed firms use more than one public cloud provider (Flexera [2020], 16.

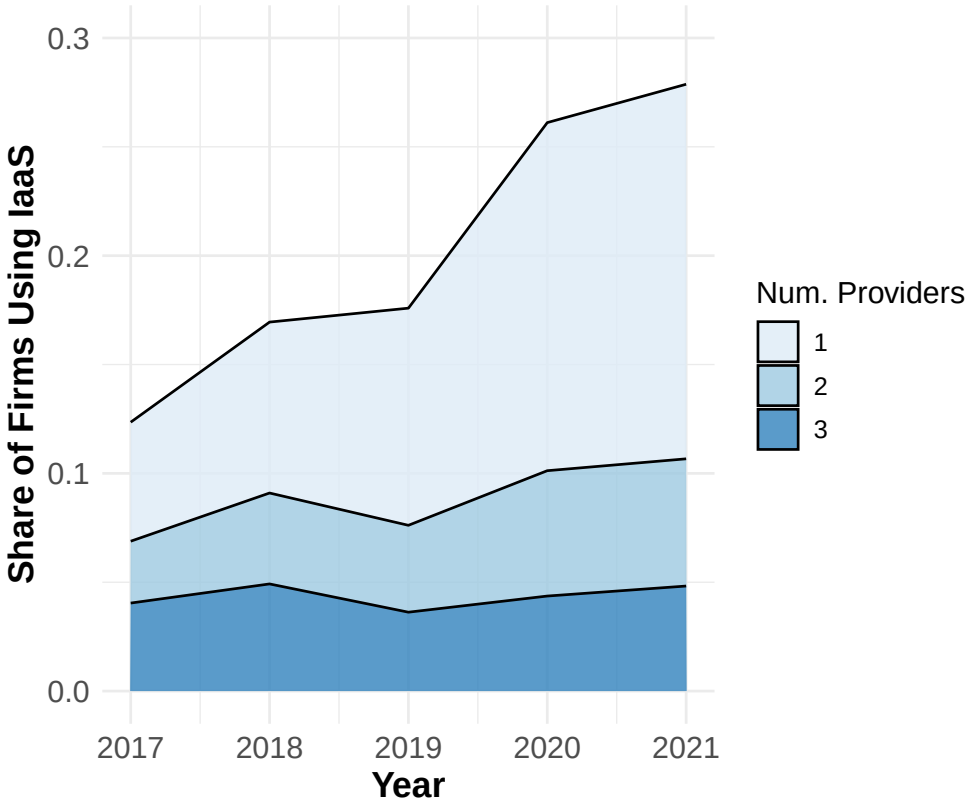


Figure 2: Growth of IaaS usage over time

Table 2: Comparison of single- and multihoming firms

	Med. Revenue	Med. Computers	Med. Employees	Med. Software Intensity
Single supplier	5.48	26	30	0.00397
Multiple suppliers	45.4	363	252	0.00518

Notes: Values computed among cloud adopters. Revenue is in million nominal euros. Software intensity is nominal software investment over revenue.

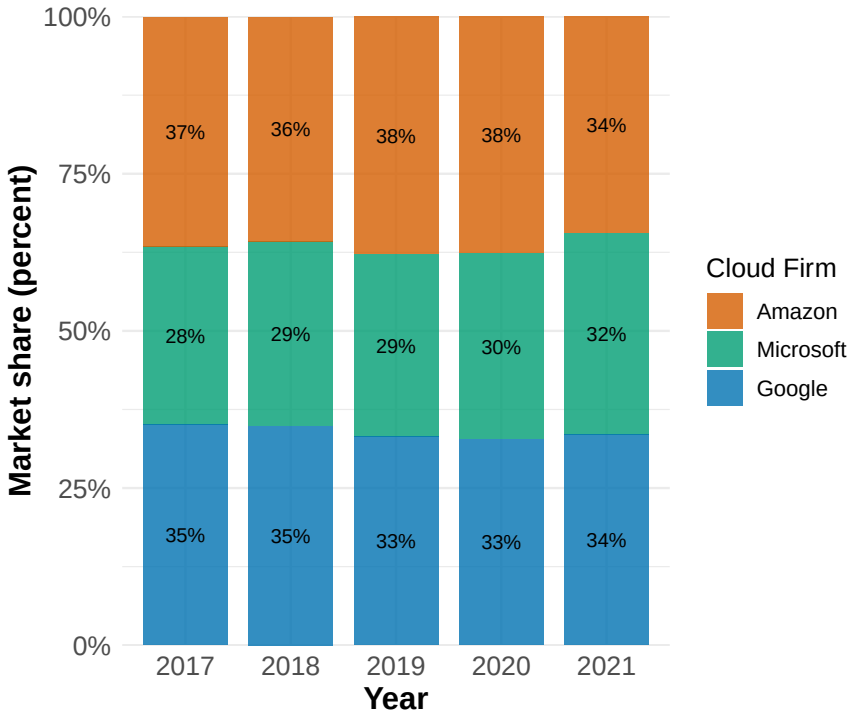


Figure 3: Cloud provider market share

Shares are weighted by firm software investments. For multihoming firms, consumption is divided evenly among suppliers used.

dominated by three major firms - Amazon, Microsoft, and Google, with smaller providers such as IBM and Alibaba targeting specific industries. As of 2025, the three largest firms control 70% of the IaaS market, with Alibaba and IBM controlling another 18% (Statista [2025]). Figure 3 displays the market shares of the three largest suppliers in our data, weighted by firm software investments.[23] Market shares are roughly constant over the sample period, with Microsoft slowly gaining market share at the expense of Amazon, such that by 2025 it has 27% market share compared to 20% for Amazon (Statista [2025]).[24]

The reasons for this dominance are numerous. As the first firm to offer cloud computing in 2006, Amazon has enjoyed a first-mover advantage that has persisted due to continued quality improvements as well as the lock-in effects highlighted above. As a result, the suppliers have enjoyed significant profits. Figure 4 displays the quarterly operating income, in real USD, of each supplier's cloud segment, starting in Q1 2013 for Amazon and Microsoft and Q4 2020

[23] Aberdeen CITDB also contains data on IBM IaaS usage for a small number of firms, but none are matched to my Orbis sample.

[24] If I do not weight by software investments, Microsoft has the most individual consumers, with a 54%-65% market share depending on the year.

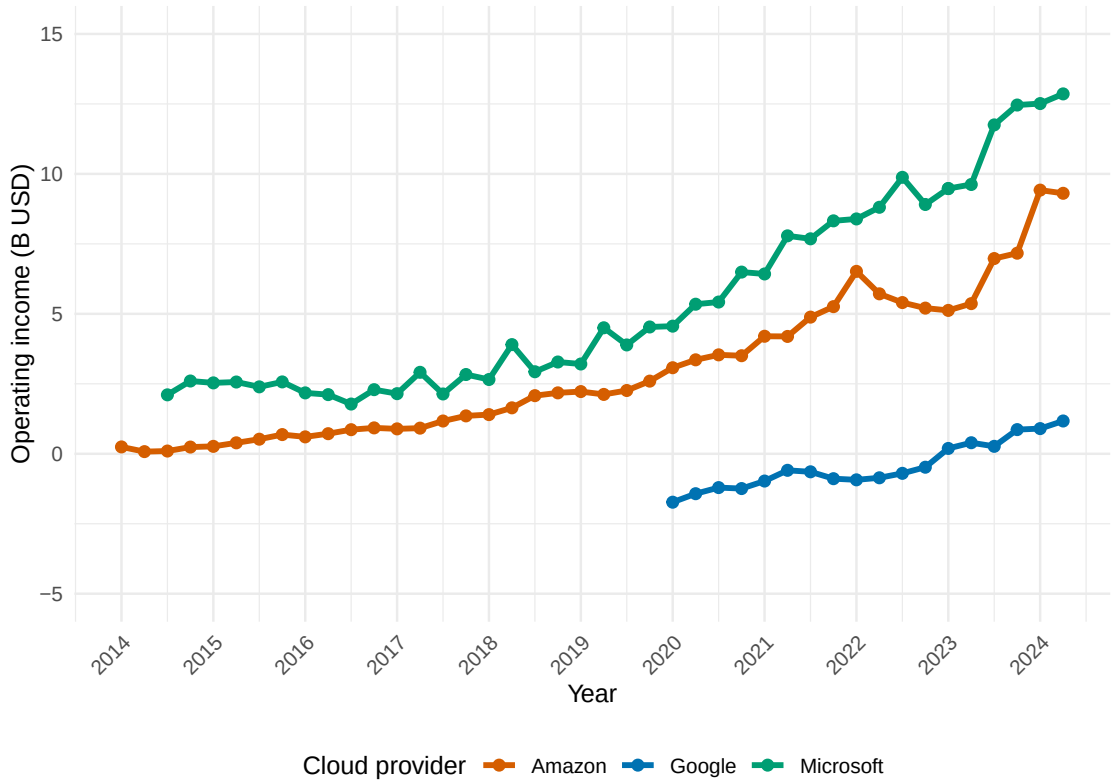


Figure 4: Operating income of the cloud divisions of the 3 largest cloud suppliers (Q1 2014-Q1 2024)

for Google.[25] While Google posted operating losses for the first few quarters for which data is available, by 2023, all three firms enjoy significant margins.

Fact 3: Cloud supplier choice is persistent During my period of study, innovation in the cloud computing industry has been rapid. All three firms significantly increased the number of regional datacenters (see 12 in the appendix). Likewise, the number of virtual machine offerings increased rapidly between 2017 and 2021, including the addition of specialized processors, such as Google's tensor processing units (TPUs) for machine learning, which launched in 2018. Despite this innovation, the of switching cloud supplier bundles was low, averaging 6.57% of cloud users in my sample.[26]

To study the determinants of switching provider bundles, I run the following probit

[25]Note that this segment incorporates both IaaS and PaaS products.

[26]A 2023 Ofcom report found that approximately 20% of firms have ever switched cloud providers (Ofcom [2023].

Table 3: Correlates of supplier switching

Variable	Model 1	Model 2
Intercept (ζ_0)	11.498 (0.452)***	11.510 (0.453)***
Log egress fee (ζ_1)	-0.816 (0.024)***	-0.816 (0.024)***
Log capital (ζ_2)		0.003 (0.022)
Log capital × log egress fee (ζ_3)		0.000 (0.005)
Log VM price (ζ_4)	0.136 (0.044)**	0.136 (0.044)***
Log VM quality (ζ_5)	-9.448 (0.297)***	-9.455 (0.297)***
Log Azure × Microsoft server usage (ζ_6)	0.021 (0.018)	0.020 (0.018)
Observations	80,830	80,830
Year FE	✓	✓
Industry FE	✓	✓
Log likelihood	-20321.86	-20316.56

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions estimated via a probit model with year and industry fixed effects.

regression on the subset of firms that could incur an egress fee:[27]

$$\mathbb{P}[b_{ft} \neq b_{ft-1}] = \zeta_0 + \zeta_1 e_{ft}^{TB} + \zeta_2 k_{ft} + \zeta_3 e_{ft}^{TB} \cdot \log(K_{ft}) + \mathbf{X}'_{ft}\boldsymbol{\zeta}_{4:6} + \tau_i + \tau_t + \epsilon_{ft}$$

where the outcome is the probability that a firm f switches from a cloud computing bundle b_{ft}, where $b \in \{\text{Amazon}, \text{Microsoft}, ..., \{\text{Amazon}, \text{Microsoft}\}, ...\}$, e_{ft}^{TB} is the log of the cost faced by a firm to migrate 100 terabytes of data in millions of real euros. To control for the fact that fees scale with firm data storage, I use k_{ft}, the log of the firm's real capital stock. To control for observed factors affecting switching, I include $\mathbf{X}'_{ft}$ as a vector of controls - the average price per core-hour of b_{ft-1} in real euros, the average log VM quality, and an indicator for whether a firm used Azure and Microsoft Server products in period $t-1$. I also include industry and year fixed effects. Results are displayed in Table 3.

In both specifications, I compute an average marginal effect of -0.11, indicating that a 1% increase in egress fees reduces the probability of switching by 0.11%. Likewise, the coefficients on price and quality are significant and of the correct sign. This suggests that, controlling for observable product characteristics, higher switching costs are more likely to

[27] That is, firms using cloud computing in period t or $t-1$.

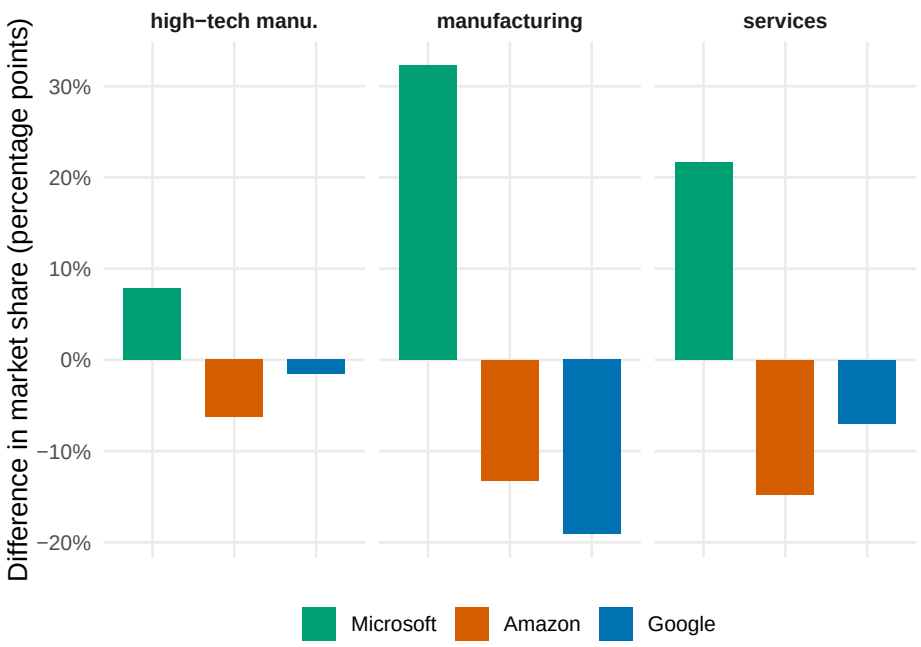


Figure 5: Difference in cloud supplier market share by Microsoft server adoption status

deter cloud users from switching suppliers.

Fact 4: Server supplier predicts cloud supplier choice.

The dominant cloud suppliers are all large, multi-product firms that offer complementary goods to their cloud services. Arguably the most important of these is Microsoft, which is the dominant firm in the server software market and offers lower prices as well as other interoperability measures between its server and cloud products, as discussed above. In addition, Microsoft sales representatives often attempt to sell Azure to existing customers.[28] Figure 5 displays firm-level cloud supplier choice shares for firms with and without Microsoft Server products for firms in high-technology manufacturing (sectors with mean R&D expenditures above 5%), low-technology manufacturing, and services. Across sectors, firms that are using Microsoft Server products are much more likely to use Microsoft Azure as well.

[28]I thank Bill Whyman and the director of cloud infrastructure at a health insurance firm for independently making this observation.

3.3 Implications of empirical facts

The above four facts inform my choices in modeling cost heterogeneity in cloud adoption. The first fact suggests that the fixed costs of adoption are rapidly falling, prompting a model of firms' dynamic decisions as a nonstationary equilibrium. In addition, the frequency of multihoming suggests that firms derive significant benefits from complementarities between cloud suppliers' products. I rationalize this by allowing firms to derive additional productivity benefits from multihoming while facing additional variable costs from the organizational complexities of integrating multiple suppliers. Facts 3 and 4 suggest that the potential barriers to competition– egress fees and complementarities between IT vendors' products–are associated with consumers' choice of cloud supplier, highlighting the potential for welfare gains from policies that increase price competition among suppliers, analyzed in Section 7.

4 Model

To estimate the impact of upstream concentration on firms' cloud adoption and usage decisions, I develop a model of firm production with computational input choices. Firms produce output using a Cobb-Douglas production technology, and choose capital investment and computing bundles each period to maximize the net present value of current and future profit flows. Cloud adoption increases the future path of productivity, and multihoming generates additional productivity benefits. In the short run, firms pay fixed costs of adoption or bundle switching. Firms are price-takers, choosing the suppliers with the best price and feature set to minimize variable costs of using cloud computing. To study supplier pricing incentives, I construct a model of Betrand competition in which suppliers observe the vector of market shares and choose prices and data egress fees to maximize profit.

4.1 Firm production and cloud computing demand

I first describe how cloud computing affects firms' productivity and output. Firms make a discrete choice of a bundle of computing inputs $b_{ft} \in \mathcal{B}$. Firms can choose to use on-premise computing only, $b_{ft} = \varnothing$, or on-premise computing and any number of cloud providers in my

sample: $b_{ft} \in \{\text{Amazon}, \text{Google}, \text{Microsoft}, \{\text{Amazon}, \text{Google}\}, ...\}$. Including the on-premise-only bundle, $b_{ft} = \varnothing$, there are 8 discrete choices: $|\mathcal{B}| = 8$. Adopting cloud computing affects firm productivity, $\Omega_{ft}(b_{ft})$. In addition to a bundle of computing providers, firms choose static inputs of labor and materials, L_{ft} and M_{ft}, as well as a dynamic capital input, K_{ft}. In each industry $i \in \mathcal{I}$, firms produce output according to the following Cobb-Douglas production technology:[29]

$$Y_{ft} = \Omega_{ft}(b_{ft}(\mathbf{X}_{fbt}))L_{ft}^{\beta_l} M_{ft}^{\beta_m} K_{ft}^{\beta_k} \tag{1}$$

where Y_{ft} is revenue; β_X for $X \in \{L, M, K\}$ are the output elasticities; and Ω_{ft} depends on the chosen IT input bundle through its feature vector $\mathbf{X}_{fbt}$. Throughout the remainder of this paper, I will focus on the above revenue production function and its associated notion of productivity, TFPR.[30] Assuming that firms take input prices p_{Xt} as given, I can use the firm's cost minimization assumption to solve for optimal L, M and rewrite its flow profit net of cloud computing costs in terms of the state variables (Ω_{ft}, K_{ft}) as

$$\pi_{ft}(\Omega_{ft}, K_{ft}) = \lambda \Omega_{ft}^{1/(1-\beta_L-\beta_M)} K_{ft}^{\beta_K/(1-\beta_L-\beta_M)} - C^K(I_{ft}, K_{ft})\epsilon_{ft}^K - \kappa(b_{ft})D_{ft} \tag{2}$$

$$\text{where } \lambda = (1-\beta_L-\beta_M)\left(\frac{\beta_L}{p_{Lt}}\right)^{\beta_L/(1-\beta_L-\beta_M)}\left(\frac{\beta_M}{p_{Mt}}\right)^{\beta_M/(1-\beta_L-\beta_M)} \tag{3}$$

where I_{ft} is the period t investment in K_{ft}, and $\epsilon_{ft}^K \sim \mathcal{N}(0, \sigma_K)$ is an i.i.d. shock to the cost of investment incurred by a firm, rationalizing within period variation in investment (cf. Gowrisankaran and Schmidt-Dengler [2024]). $C^K(\cdot)$ is a convex adjustment cost function parameterized as as follows:[31]

$$C^K(I_{ft}, K_{ft}) = p_t^K I_{ft} + C_1^K K_{ft}\left(\frac{I_{ft}}{K_{ft}}\right)^2 \tag{4}$$

[29] I employ a Cobb-Douglas specification in order to simplify estimation of the dynamic model. As well, a Cobb-Douglas production function is the first-order approximation of any general production function. In Appendix D.2.1, I estimate a translog production function, finding similar results.

[30] See Foster et al. [2008] for a discussion of the difference between TFPQ and TFPR and the limitations of the latter.

[31] Some models Cooper and Haltiwanger [2006], Bloom [2009] allow for a fixed adjustment cost as well, though Asker et al. [2014] finds it is precisely 0 in the French manufacturing sector.

Next, $\kappa(b_{ft})$ refers to the total cost of cloud computing adoption and maintenance, which are scaled by a firm's data stock, D_{ft}. Finally, I assume K_{ft} evolves according to the law of motion

$$K_{ft} = (1 - \delta_{it})K_{ft-1} + I_{ft} \tag{5}$$

for industry-year depreciation rate δ_{it}.

Productivity process. Next, I let ω_{ft} follow the industry-specific law of motion given by the AR(1) process:

$$\omega_{ft} = g(\omega_{ft-1}, b_{ft}(\mathbf{X}_{bft}), \mathbb{1}\{t = 2020\}) + \nu_{ft} \tag{6}$$

where $\nu_{ft} \sim \mathcal{N}(0, \sigma_i^{\omega})$ is an i.i.d. shock. I assume that a cloud computing bundle choice affects firm productivity according to the function $g(\cdot)$, which is dependent on past productivity, cloud bundle choice, and a COVID-19 shock. In my empirical specification, I find that only the number of providers, $\#j$, is statistically significant across sectors, such that I can parametrize the productivity process as

$$\omega_{ft} = \gamma_0 + \gamma_1\omega_{ft-1} + \gamma_2\omega_{ft-1}^2 + \gamma_3\omega_{ft-1}^3 \tag{7}$$

$$+ \gamma_4 \cdot \#j + \gamma_5\mathbb{1}\{t = 2020\} + \nu_{ft} \tag{8}$$

This specification follows existing models of firm input adoption, such as Peters et al. [2017], Aw et al. [2011]. While richer interactions between productivity and cloud usage are possible, I choose this simpler specification to keep the model tractable for my relatively short panel of data (cf. Humlum [2022]).

Computing bundle choice. Firms also face a dynamic decision to adopt a bundle of cloud computing providers b_{ft} or only adopt on-premise computing, $b_{ft} = \varnothing$. As with other capital goods, K_{ft}, firms make an adoption decision in period t, which is realized in period $t + 1$. Adopting and using cloud computing incurs both fixed and variable costs. Fixed costs are divided into adoption costs, κ_t^E, which are incurred in period t when a firm

adopts cloud computing, and switching costs, κ^S, which are incurred when a firm switches cloud computing bundles or exits the cloud. The one-period time to build assumption and fixed costs represent the one-time costs of hiring, software rewriting, and data migration.[32] I allow the cost of entry to vary to account for the development of cloud migration tools and expansion of the cloud migration specialist labor force over time.[33] To reduce model complexity, I fix switching costs since I do not observe high variation in switching rates over time. For each period using a cloud bundle, firms incur a maintenance cost $\kappa_M(\mathbf{X}_{fbt})$ which varies with bundle characteristics $\mathbf{X}_{fbt}$. These costs represent the variable costs of deploying virtual machines and storage, in addition to the ongoing costs of employing software engineers with the relevant human capital and of ensuring software compatibility across multiple suppliers. For example, a firm using multiple providers is likely to have higher costs of ensuring continual compatibility of software across cloud services, as well as data egress fees for ongiong data transfers between providers. Finally, I assume that these costs scale with a firm's stock of data, a complementary good to computing (Brand et al. [2024]). [34] To maintain tractability for estimation, I assume that data is part of a firm's total capital stock and, as such, there exists a one-to-one mapping between capital and data, such that I can write data as $D_{ft}(K_{ft})$.

Using these assumptions, I can write the cloud computing usage costs as follows:

$$\kappa_t = \underbrace{\kappa_t^E \cdot \mathbb{1}\{b_{ft+1} \neq b_{ft} | b_{ft} = \varnothing\}}_{\text{Entry cost}} + \underbrace{\kappa^S \cdot \mathbb{1}\{b_{ft+1} \neq b_{ft} | b_{ft} \neq \varnothing\}}_{\text{Switching friction}} + \underbrace{\kappa_1^M(\boldsymbol{\phi}_b' \mathbf{X}_{bft}) \cdot \mathbb{1}\{b_{ft} \neq \varnothing\}}_{\text{Maintenance cost}} \tag{9}$$

where entry costs are parametrized as:

[32] While individual software developers may create VM instances for very little fixed costs, migration of existing firm workloads is costly. Since the Aberdeen data is sourced from interviews regarding firm production processes to sell to IT hardware and software providers, it is likely capturing the latter use case.

[33] Each of the cloud providers in my data offers an internal migration tool. In addition, firms can use third party migration software such as Cortex or Carbonite Migrate.

[34] While Aberdeen contains a number of variables related to firms' IT capital investments, I choose to use their data on storage investment rather than PC investment to proxy for investments in computation and data storage. Computer storage devices are relatively homogeneous (Igami and Uetake [2020]), while PCs are likely to differ on many other features in addition to raw computing power, which would introduce measurement error.

$$\kappa_{ft}^E = \exp(\kappa_0^E + \kappa_1^E \cdot t) + \epsilon_{ft}^E \tag{10}$$

where $\epsilon_{ft}^E \sim Gumbel(0, \sigma^E)$. The common component of the entry cost follows the standard functional form assumption in other models of firm adoption of technologies with rapidly falling costs (cf. Rubens [2025], Humlum [2022]). Denoting $server_{ft} \in \{0, 1\}$ as an indicator for whether firm f is using Microsoft server products in year t and P_{bft} as the mean price per core hour of a virtual machine in bundle b, maintenance costs are parametrized as

$$\kappa_{ft}^M = \alpha_0 + \alpha_1 \underbrace{P_{bft}}_{\substack{\text{Price} \\ \text{per core hour}}} + \alpha_2 \cdot \underbrace{\#j \cdot \mathbb{1}\{\#j > 1\}}_{\text{Multihoming}} + \alpha_3 \underbrace{\log(q_{bft})}_{\text{VM quality}} \tag{11}$$

$$+ \alpha_4 \underbrace{\mathbb{1}\{\text{Microsoft} \in b_{ft} \cap server_{ft} = 1\}}_{\substack{\text{Microsoft} \\ \text{complementarity}}} + \underbrace{\sum_{j \in b} \xi_j}_{\substack{\text{Vertical} \\ \text{differentiation}}} + \epsilon_{ft}^b \tag{12}$$

These maintenance cost shifters capture the baseline costs of using cloud computing relative to on-premise as well as how these costs vary with cloud computing prices, multihoming choices, virtual machine quality, complementarities with on-premise software stacks, and unobserved vertical differentiation.[35] Finally, $\epsilon_{ft}^b \sim Gumbel(0, \sigma^b)$ is an i.i.d. shock that firms in each period in which they are using cloud computing.[36]

Each period, then, firms observe their current productivity level and, conditional on their current capital and bundle choices, make capital investment and bundle decisions, which are realized in the next period. I assume that firms can perfectly forecast exogenous aggregate states - non-IT input prices, Microsoft server adoption decisions, and the growth rate in VM quality, which I assume is exogenously related to Moore's law.[37] For cloud computing prices, chosen by suppliers, I assume that firms forecast prices using an AR(1) decision rule,

[35] I model Microsoft server adoption status as an exogenous AR(1) process to account for this complementarity, as described in Appendix A.8.

[36] It is well known that one cannot identify both maintenance costs and unobserved scrap values using only firm exit rates (Aguirregabiria and Suzuki [2014]). I choose to normalize scrap values to 0 because cloud provider features are more naturally modeled as maintenance cost shifters and not scrap value shifters.

[37] This is an upper bound on progress, since cloud firms choose to continue offer instance types based on older CPUs. I abstract away from quality choice in the model, however.

described below. Them, using $\mathbf{P}_t$ to represent the vector of input prices and quality, I can write the payoff relevant state as $s_{ft} = (\Omega_{ft}, K_{ft}, b_{ft}, server_{ft}, \mathbf{P}_t) \in \mathcal{S}$. I can then write firm value functions as follows:

$$V_{ft}(s_{ft}) = \max_{I_{ft}, b_{ft}} \pi_t(s_{ft}; \boldsymbol{\beta}, \boldsymbol{\gamma}) \tag{13}$$

$$- C^K(I_{ft}, K_{ft})\epsilon^K_{ft} - D_{ft} \cdot (\kappa^E_t + \epsilon^b_{ft}) - D_{ft} \cdot (\kappa^M_{ft}(s_{ft}; \boldsymbol{\alpha}) + \epsilon^b_{ft}) \tag{14}$$

$$+ \beta \mathbb{E}_t[V_{ft+1}(s_{ft+1})|s_{ft}] \tag{15}$$

where β is the discount factor. The time subscript t is due to the fact that profits π_{ft} will vary over time due to falling adoption costs and changes in input prices. In each period, each firm 1) draws relevant cost shocks for investment and cloud entry and maintenance costs, $\epsilon^K, \epsilon^E, \epsilon^b$, and decides its level of capital investment, I_{ft}, computing bundle to maximize the net present value of future profits.[38]

4.2 Supplier competition

Upstream, cloud computing suppliers $j \in$ {Amazon, Microsoft, Google} engage in Bertrand price-setting of virtual machine prices and data egress fees in the next period. Due to adjustment frictions, these prices are be a function of the current state, s_t. Before downstream firms make their investment choices in each period, upstream firms choose prices p_{jt} to maximize their value function:

$$W_{jt} = \max_{P_{jt}, E_{jt}} \mathbb{E}_t[\pi^S_{jt} + W_{jt+1}(s_{t+1})|s_t] \tag{16}$$

$$\text{where } \pi^S_{jt} = (P_{jt} - C^{vm}_{jt})Q_{jt} + (E_{jt} - C^{egress})Q^{exit}_{jt+1} \tag{17}$$

[38] The reasons for including a cost shock on maintenance costs are threefold. First, since I estimate the model on broad industries yet only observe investment cost at the annual level, it is likely there is meaningful empirical variation in the costs of investment for firms. Second, Gowrisankaran and Schmidt-Dengler [2024] demonstrate that the algorithm has speed improvements over transitional methods of finding firms' investment policy functions. Third, including functional form assumptions on the cost shocks permits me to write capital choice probabilities in closed form (Appendix B.2), allowing me to take analytical derivatives when solving suppliers' pricing problems (Appendix B.3).

The price P_{jt} refers to the average price per core-hour supplier j sets in a given year. While a richer model would consider the full menu of virtual machine prices, in practice, publicly-available virtual machine usage data suggests that variance in prices per vCPU-hour is rather small.[39] In addition, egress fees E_{jt} are set to deter consumer exit. In reality, I observe few nominal price changes in egress fees during my sample period, which are declining purely as a function of the inflation rate. Therefore, for computational tractability, I abstract away from egress fee choices by assuming suppliers always choose the observed egress fee schedule and then choose P_{jt} to maximize profit.

It is natural to assume that suppliers are forward looking - in a market with rapid expansion, they are likely to price below their static marginal cost plus markup in order to grow their installed base. To preserve computational tractability, I represent suppliers' preferences for investing over harvesting using a semi-reduced-form conduct parameter, ζ_{jt} (cf. Bresnahan [1982]). That is, taking supplier j's first-order conditions, we have

$$Q_{jt} + (P_{jt} - C_{jt}^{vm})\frac{\partial Q_{jt}}{\partial P_{jt}} + \beta\frac{\partial \mathbb{E}_t[W_{jt+1}(s_{t+1})]}{\partial P_{jt}} = 0 \tag{18}$$

Rearranging, we have

$$P_{jt} - C_{jt}^{vm} = \underbrace{\frac{Q_{jt}}{-\partial Q_{jt}/\partial P_{jt}}}_{\text{Markup}} + \underbrace{\beta\frac{\partial \mathbb{E}_t[W_{jt+1}(s_{t+1})]}{\partial P_{jt}}}_{\text{Harvesting incentive}} \tag{19}$$

Finally, we can rearrange to solve for P_{jt}:

$$P_{jt} = C_{jt}^{vm} + \zeta_{jt} \cdot \frac{Q_{jt}}{-\partial Q_{jt}/\partial P_{jt}} \tag{20}$$

$$\text{where } \zeta_{jt} = \frac{\beta}{Q_{jt}} \cdot \frac{\partial \mathbb{E}_t[W_{jt+1}(s_{t+1})]}{\partial P_{jt}} \tag{21}$$

Thus, ζ_{jt} represents a measure of "invest vs. harvest" dynamics. If $\zeta \geqslant 1$, the supplier is focused on harvesting the existing user base, while if $\zeta < 1$, the supplier is setting prices below what static Bertrand would imply in order to expand its customer base. In the em-

[39] See Appendix A.6.1 for a full analysis.

pirical application, this parameter will be estimated, allowing me to preserve computational tractability on the supply side via a simple Bertrand solve each period while allowing for more realistic supplier pricing than static Bertrand would imply.

A few additional comments are in order. By focusing on price setting, I abstract away from dynamic investment decisions in improved products, such as new datacenter capacity (Pan Fang and Greenstein [2025]), or the development of custom processors, such as Google's tensor processing units (TPUs) or Amazon's Graviton series. While a full treatment of these impacts is beyond the scope of this paper, I argue that Bertrand price setting accurately captures short run changes in virtual machine pricing. Prior research has found, for example, that Amazon and Microsoft responded rather quickly to Google's entry into the market by cutting prices (Byrne et al. [2018]).

4.3 Timing and equilibrium

Finally, I describe the timing of the game and my equilibrium concept. With many firms selling heterogeneous products, for any individual firm, tracking the states of all competitors is infeasible. Given the number of downstream firms, solving for a MPE using Ericson and Pakes [1995] is computationally infeasible. Therefore, I adopt the Markov moment equilibrium (MME) concept from Ifrach and Weintraub [2017] as a simplifying assumption. In such an equilibrium, firms keep track of their own state, $\tilde{s}_{ft} = (\Omega_{ft}, K_{ft}, b_{ft}, server_{ft})$, as well as moments of the aggregate industry state. The payoff relevant states include the exogenous price changes in industry level wages, materials price, capital cost, and VM quality index $(P_i^l, P_i^m, P_i^k, \log(q_j))$ as well as the endogenous price of each supplier, P_t^b. I make the simplifying assumption that that firms forecast a simple AR(1) pricing rule (cf. Garcia-Osipenko et al. [2025], Hornbeck et al. [2024]), which I find works well empirically.[40]:

$$P_{jt+1} = \rho_0^j + \rho_1^j P_{jt} \tag{22}$$

[40]I set a convergence criterion that the maximum forecast error must be less than 0.1%. Because dynamic pricing games with switching frictions generally do not have unique equilibria (cf. Farrell and Klemperer [2007]), I begin each iteration using the empirical estimates for the AR(1) transition parameters during estimation. During the counterfactuals, I begin each outer iteration at the values obtained during estimation.

Thus, downstream firms best respond to their own state as well as their perceived price forecast for each supplier. Finally, given the falling entry costs, this setting is nonstationary. Therefore, assume that the industry transitions from $t_0 = 2017$ to some period T, after which the industry is in a steady state. Thus, firms' strategies depend on the initial state, s_0. Throughout, I focus on symmetric pure strategies. The timing of the game is as follows:

1. The game begins in period $t = 2017$, at an initial industry state s_0.
2. Suppliers observe $\mathbf{s}_t$ and choose P_{jt}.
3. The idiosyncratic shocks σ^E, σ^B are realized and downstream firms choose b_{ft+1}
4. Investment cost shocks σ^K are realized and downstream firms choose investment I_{ft}.
5. Firms and suppliers receive per-period profits net of any relevant costs.
6. State transitions are realized, and the game proceeds to period $t + 1$.

Discussion of assumptions. While the AR(1) price forecast assumption is made for computational tractability purposes, I argue that that there is reason to believe that it approximates well firms' actual pricing forecasts. First, virtual machine prices stay relatively stable over time: the AR(1) persistence term for each supplier in the data is between 0.90 and 0.97. Second, prior research shows that firms have difficulty optimizing cloud costs (Brand et al. [2024]), with 65% (73%) of firms sampled by Flexera ranking "optimizing the existing use of cloud" as one of their primary cloud initiatives in 2019 (2020)(Flexera [2020, 2019]). Given the complexity of optimal VM provisioning, it is unlikely firms form sophisticated beliefs about the future evolution of prices.

Second, the nonstationarity assumption is due to the fact that the cloud computing industry is rapidly growing during my sample period, with double-digit CAGR.

5 Estimation

I estimate the model, report the results in Section 6, and provide details in Appendix B. First, I use a control function approach to estimate firms' production function with endogenous

technology choice (cf. de Loecker [2013], Ackerberg et al. [2015]). Firms' choices reveal the output elasticities of inputs as well as the productivity gains from using cloud computing. Second, given production function coefficients, I estimate my remaining cost parameters using a method of simulated moments (McFadden [1989], Pakes and Pollard [1989]). Firms' investments and bundle choices recover entry and switching costs as well as brand specific maintenance cost shifters. Table 4 lists parameters and associated sources of identification.

5.1 Estimation: Production function

To estimate the productivity gains from cloud computing, I first estimate a production function for each industry. While my sample size is sufficient to estimate production at the NACE 2-digit or even 4-digit level, this greatly increases the consumer type space and therefore computational burden of solving the dynamic model. Therefore, I follow Peters et al. [2017] and divide the manufacturing sector into high-technology and low-technology manufacturing. I define high-tech manufacturing as any 4-digit NACE code for which R&D expenditures are above 5% of gross value added, where values are taken from (Galindo-Rueda and Verger [2016]). Since services are a large portion of the economy, I also include services as a third industry.

Second, since I do not observe cloud computing usage at the intensive margin in my data, to avoid measurement error, I follow Peters et al. [2017] and assume technology choice is an ordinal variable (cf. Mairesse et al. [2005]).[41]

To estimate the production function, I follow the approach from Ackerberg et al. [2015] and non-IT investment as a proxy variable to control for the simultaneity bias of productivity and investment choices. To construct non-IT investment, I first use data from Orbis to construct firms' total capital investment. Orbis reports firm level depreciation values for **[In progress]** percent of firms. For the remaining firms, I use firms' time series of capital stocks, along with industry-year level depreciation rates, to compute investment. I construct these depreciation rates from the Orbis data as $\frac{\text{Depreciation}_{ft}}{K_{ft}}$ and then take the mean. I then

[41] One could use model assumptions to back out IaaS spending using firms' cost minimization equation and current on-premise computing usage. However, this involves making strong assumptions about computing capital utilization rates, which are known to very widely by type of computer as well as firm and industry (Brynjolfsson and Hitt [1996], Brand et al. [2024]).

Table 4: List of parameters and estimation strategy

Parameter	Source of identification
Production function parameters (GMM)	
$\boldsymbol{\beta}_x^i$ (Output elasticities of inputs)	Within-industry output and input covariation
$\boldsymbol{\gamma}_{0:3}^i$ (Productivity persistence)	Within-firm autocorrelation in investment
γ_4^i (Impact of cloud adoption on productivity)	Within-industry covariation of investment and cloud adoption
γ_5^i (Impact of COVID-19 on productivity)	Within-industry covariation of investment and 2020 year indicator
Bundle cost parameters (MSM)	
κ_t^E (Cloud adoption costs)	Annual share of cloud users
κ^S (Switching cost)	Autocorrelation of bundle choices among cloud users
α_0 (Maintenance cost intercept)	Exit rate from cloud computing
α_1 (Virtual machine price cost shifter)	Supplier-tear price variation among new entrants
α_2 (Multihoming cost shifter)	Share of cloud users that multihome
α_3 (Quality cost shifter)	Covariance of average quality and bundle choice among new entrants
α_4 (Azure-MS server cost shifter)	Covariance of Azure and Microsoft server usage among new entrants
ξ_b (Vertical differentiation)	Market shares of suppliers among new entrants
σ^E (Variance of entry costs)	Size premium of new cloud entrants
σ^b (Variance of maintenance costs)	Size premium of multihoming firms among cloud users
Investment parameters (MSM)	
C_1^K (Convex adjustment cost)	Industry level average investment rates
σ^K (Variance of investment costs)	Industry level variation in investment rates
Conduct parameter (MSM)	
ζ (Supplier conduct)	Supplier pricing vs. implied markups
Remaining parameters (calibrated)	
β (Discount rate)	Calibrated to 0.94 (Bloom [2009])
C_{jt}^{vm} (VM marginal cost)	Calibrated (depr. + electricity costs)
C_{jt}^{e} (egress marginal cost)	Calibrated (Boudreau [2022])

This table summarizes the steps of my estimation procedure, detailed below. Each step lists the parameter, definition, and source of identification. For the first step, production function parameters are identified via Ackerberg et al. [2015], with the de Loecker [2013] correction for identifying the productivity impact of endogenous firm technology choice. For the second step, parameters are estimated via method of simulated moments (MSM). Finally, several moments are calibrated from existing data sources. An informal description of target moments are listed in column 2.

apply the perpetual inventory method to construct annual investment flows. Details are in Appendix A.5.

Three comments are in order about identification in my specific setting. First, to satisfy the monotonicity assumption of investment in productivity, I subtract PC, server, and storage investment from total investment to recover investment in non-IT capital. This is necessary because, in France, cloud services are expensed as materials costs, while on-premise computing is expensed as capital (Dos Santos and Ovsyannikova [2023]). Thus, a firm may reduce on-premise computing investment in response to IaaS adoption. If computing is a relatively large share of a firm's capital stock, this could violate the assumption that I_{ft} is monotonically increasing in $\tilde{\Omega}_{ft}$. By restricting my proxy variable to be non-IT investment, I ensure that monotonicity is preserved.[42]

Second, I need to control for the endogeneity of firms' cloud computing choices in productivity. It is likely, for example, that more productive firms are more likely to select into cloud usage. As well, finding a suitable instrument is difficult, due to most instruments likely also affecting other IT investments.[43] To control for this selection effect, I exploit the timing assumptions of my productivity process. In particular, because I assume that firms make cloud adoption decisions one period in advance, it is orthogonal to firms' current period productivity shock. I follow de Loecker [2013] and exploit the following moment condition to identify γ_4:[44]

[42]There are potentially two issues with this proxy variable. First, using capital investment as a proxy creates a selection on firms with positive investment. In my sample, 80.3% of high-tech manufacturing firms, 77.3% of low-tech manufacturing firms, and 69.9% of services firms have positive investment rates. As a robustness check, I use the Levinsohn and Petrin [2003] estimator with materials as my proxy variable. Second, Aberdeen estimates IT investment expenditures using data on IT stocks collected from interviews. Given the opacity of their modeling procedure, this may introduce measurement error. As a robustness check, I use total investment as my proxy variable. Both results are presented in Appendix D and are largely consistent with my primary specification.

[43]Though see DeStefano et al. [2023] for an exception.

[44]Two issues arise with this assumption. First, it may be that firms, especially smaller firms, upgrade to cloud computing within a year of making an adoption decision. As a robustness check, I instead assume that $\mathbb{E}[b_{ft-1}v_{ft}] = 0$ and productivity gains are only realized one period after adoption is observed. As a second check, I examine whether firms experience gains from learning-by-doing (Brand et al. [2024]) and use the number of years using cloud computing as my independent variable. Second, measurement error Aberdeen's data collection process may attenuate the effects of cloud adoption (cf. Demirer et al. [2024], Brynjolfsson and Hitt [2003]). To address this, I use the data vendor's own "install likelihood score," which details the probability that the technology in question is actually used by an establishment, as a propensity score to weight the cloud adoption variable.

$$\mathbb{E}[b_{ft} v_{ft}] = 0 \tag{23}$$

Identification of $\boldsymbol{\gamma}_{-4}$ follows the standard timing assumptions:

$$\mathbb{E}[v_{ft} | I_{ft}] = 0 \tag{24}$$

where I_{ft} consists of lagged flexible inputs and current and lagged capital stock.

Third, I control for the impact of input and output quality on productivity.[45] I make the assumption that firm input and output prices are increasing in (unobserved) product quality. In the baseline specification, I use firm level wages, P^l_{ft}, to proxy for unobserved worker quality. I observe the total wage bill for a majority of firms. For the remaining firms, I impute the wage at the year-region-industry level.[46] To estimate the input price control function $H(p^l_{ft})$, I estimate a third order polynomial in the first stage of my control function approach:

$$\tilde{\omega}_{ft}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = \hat{\phi}_{ft} - f(\mathbf{x}_{ft}; \boldsymbol{\beta}) - H(p^l_{ft}; \boldsymbol{\vartheta}) \tag{25}$$

where $\boldsymbol{\beta}, \boldsymbol{\vartheta}$ are the production function and input control function parameters, respectively; $\hat{\phi}(\cdot)$ is predicted output; and $\mathbf{x}_{ft}$ is a vector of inputs, including b_{ft}.

Finally, to target the production function parameters $\boldsymbol{\beta}$, I use a lag of the inputs, Z_{ft}, as my instruments.

5.2 Estimation: Cost parameters

Given estimates of the production function parameters, I estimate the remaining parameters via the method of simulated moments (MSM). The common component of the entry cost κ^E_t is identified from the annual fraction of firms that adopt cloud computing. The variance

[45] See De Loecker et al. [2016] for a discussion.

[46] To reduce the size of the state space in my dynamic model, I only include wages as a proxy for input quality. In Appendix D, I include the additional quality proxies from De Loecker et al. [2016], namely export status and firm market share at the industry-region-year level.

of entry cost shocks, σ^E, is identified from the size premium of adopters over non-adopters (Arcidiacono and Miller [2011]). To see the intuition for this, note that, given the assumption on the form of adoption cost shocks, the log odds ratio that a firm adopts some cloud computing bundle is given by

$$\log\left(\frac{\mathbb{P}_t[b_{ft+1}\neq\varnothing|b_{ft}\neq\varnothing]}{1-\mathbb{P}_t[b_{ft+1}\neq\varnothing|b_{ft}\neq\varnothing]}\right)=-\frac{\kappa_t^E}{\sigma^E}+\frac{1}{\sigma^E}\cdot\left(\frac{\beta}{D_{ft}}\max_{b_{ft+1}}\mathbb{E}_tV_{t+1}(b_{ft+1},\tilde{s}_{ft+1})-\frac{\beta}{D_{ft}}\mathbb{E}_tV_{t+1}(\varnothing,\tilde{s}_{ft+1})\right) \tag{26}$$

As long as profits and the data stock D_{ft} do not perfectly scale with size, then σ^E is identified by the size premium of adopters. Next, I identify switching costs κ^S using the autocorrelation of cloud bundle choices among cloud users. To identify the level of maintenance costs α_0, I use the exit rate from cloud computing among cloud using firms. Likewise, the additional multihoming cost α_2 is identified by the rate of multihoming among cloud users. For the remaining moments, I focus on new cloud adopters to separate firm preference heterogeneity from switching costs (cf. Heckman et al. [1981]). To target α_1, I use the covariance of price and bundle choice among new cloud entrants and among new entrants without Microsoft server products. To identify α_3, I use the covariance of choosing a bundle with Azure and using Microsoft server products. To target unobserved vertical quality, $\xi_{\text{Microsoft}}, \xi_{\text{Amazon}}$, I use the market shares of Amazon and Microsoft among new entrants not using Microsoft server products.[47] Finally, to identify the variance in maintenance cost shocks σ^b, I use the size premium of multihoming firms among cloud users, following analogous reasoning to the identification of σ^E.

To identify the variance of capital investment costs, I need variation in investment for firms in similar states and therefore use the variance in the ratio I_{ft}/K_{ft}.

To estimate these parameters $\boldsymbol{\theta}=\{\kappa_0^E,\kappa_1^E,\boldsymbol{\alpha}_{0:5},\boldsymbol{\xi}_j,C_1^K,\sigma^K,\sigma^E,\sigma^b,\zeta\}$ I use a full solution approach. At each guess of the parameter vector $\boldsymbol{\theta}$, for each industry, I solve for firm investment and technology choice policy functions at the steady state period T sufficiently far in the future, when entry costs are low enough such that cloud computing has fully diffused

[47] The author's interview with a chief cloud architect who has worked in the finance and healthcare industries suggests that Google has the lowest quality cloud platform among the major providers, so to separately identify α_0 and ξ_j, I normalize ξ_{Google} to 0.

and firms are in a steady state. I set $T = 2051$. I then use backwards induction to compute downstream firms' investment and computing bundle choices and upstream firms' pricing strategies. I then draw an initial distribution of states, $F(s_0)$, simulate the model forward and compare firms' guessed pricing kernel with the true one and repeat until convergence, at which point I compare the observed and simulated moments. To make this step computationally feasible, I follow Gortmaker [2024] and first use Bayesian optimization, a derivative-free global approach, to obtain an initial estimate. I then employ a trust-region refinement with numeric derivatives to obtain the final estimate. In addition, I hold upstream prices fixed during the first stage and only run full MME optimization during the trust region phase.

To estimate the optimal model parameters, I first aggregate the industry-level predictions to the country level by taking a weighted average, as in Asker et al. [2014]:

$$\boldsymbol{\Psi}(\boldsymbol{\theta}) = \frac{1}{\sum_i N_i} \sum_i N_i \boldsymbol{\Psi}_i(\boldsymbol{\theta}) \tag{27}$$

where N_i denotes the number of firms in industry $i \in \{\text{Manufacturing, Services, Software}\}$. To target each parameter, I simulate statistics, which I average over firm-industry-years and form simulated statistics. Then the moments are matched to the data moments $\hat{\boldsymbol{\Psi}}$. To reduce the effect of outlier moments, I estimate parameters by minimizing the logarithm of a criterion function given by the usual quadratic form, with weighting matrix $\mathbf{W}$:

$$\mathcal{Q}(\boldsymbol{\theta}) = (\boldsymbol{\Psi}(\boldsymbol{\theta}) - \hat{\boldsymbol{\Psi}})' \mathbf{W} (\boldsymbol{\Psi}(\boldsymbol{\theta}) - \hat{\boldsymbol{\Psi}})$$

where I set $\mathbf{W}$ is a diagonal weight matrix with values set such that all moments are roughly on the same scale. In particular, moments with a value above 1 get a weight of 0.1, while all other moments get a weight of 1.[48] Standard errors are calculated using moment vectors from the same boostrapped samples. For each industry, I discretize Ω_i to 20 grid points and K_i to 20 grid points, evenly spaced between the 1st and 99th percentile of the data.[49] Finally, I assume the COVID-19 productivity shock is unanticipated so I simulate the model twice for each parameter guess - once for 2017-2019 without the shock and once

[48] Computational details are in Appendix C.2.

[49] Prior versions of the project estimated the model with 30-point grids for Ω_i and K_i, producing similar parameter estimates. I choose the smaller grids for computational tractability.

for 2020-T with it. Further details are in Appendix C.2, and computational details are presented in Appendix C.3.

6 Estimation results

To study the impact of cloud computing adoption on firm productivity, I present the results of my production function estimation. Using those results, I can then recover the cost parameters. I find small but significant effects of cloud adoption in all industries, with larger productivity gains in manufacturing and among firms that multihome. I find that adopting Microsoft or Amazon leads to lower costs than adopting Google; that firms are fairly sensitive to price; and that there is a strong complementarity between Microsoft's cloud and on-premise server products. In addition, I compare my estimates to prior research and industry case studies on the benefits and/or costs of adopting cloud computing.

6.1 Production function parameters

Production function coefficients are presented in Table 5. Results are significant for most coefficients and reasonable, with services having a lower capital share than manufacturing.

Table 5: Production function estimates

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Labor (β_l)	0.114 (0.094)	0.286 (0.048)***	0.267 (0.038)***
Materials (β_m)	0.617 (0.051)***	0.480 (0.079)***	0.706 (0.310)*
Capital (β_k)	0.084 (0.027)***	0.085 (0.017)***	0.034 (0.015)*
Wage (ϑ_w)	0.139 (0.067)*	0.285 (0.046)***	0.148 (0.112)
Observations	14,893	37,948	153,230
Firms	4,877	12,791	70,321

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] wage control for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

Next, I present coefficients for the productivity process in Table 6. The coefficient on cloud computing adoption is small yet significant across all specifications. Thus, services

Table 6: Productivity evolution estimates

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Productivity intercept (γ_0)	0.160 (0.022)***	0.154 (0.018)***	0.140 (0.003)***
Linear AR(1) term (γ_1)	0.719 (0.050)***	0.686 (0.044)***	0.499 (0.011)***
Quadratic AR(1) term (γ_2)	0.167 (0.036)***	0.229 (0.036)***	0.509 (0.013)***
Cubic AR(1) term (γ_3)	-0.034 (0.009)***	-0.057 (0.009)***	-0.167 (0.004)***
Cloud adoption (γ_4)	0.005 (0.001)***	0.004 (0.001)***	0.002 (0.000)***
COVID-19 shock (γ_5)	-0.049 (0.001)***	-0.054 (0.001)***	-0.034 (0.001)***
Observations	14,893	37,948	153,230
Firms	4,877	12,791	70,321

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] wage control for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

firms that adopt cloud computing experience productivity gains of 0.2%-0.6%, depending on number of providers used. The equivalent gains are 0.4%-1.2% in low-tech manufacturing and 0.5%-1.5% in high-tech manufacturing. Figure 6 plots the steady state sector-level productivity distributions for firms that adopt a single cloud provider vs non-adopters. Both manufacturing sectors show a clear benefit of cloud computing. The benefit to firms in the services sector is less noticeable, both because the distribution of productivity varies less and because of the smaller coefficient.[50] Finally, Figure 7 displays the mean log productivity paths for a firm in each sector, comparing a non-adopter with a firm that has adopted one cloud provider.

Are these effects reasonable? Most studies of the impact of cloud computing tend to merge IaaS with PaaS and/or SaaS products (e.g. DeStefano et al. [2023]). The two exceptions are Jin and Bai [2022], who use the most similar definition to mine in Table A3.[51] They find a productivity increase of 0.47% per cloud-related job posting in the prior quarter. Using the mean value of this variable, 6.441, this comes to productivity gains of 3.03%. This effect may be due to the fact that their results focus on public US firms, which experience higher productivity gains from IT adoption than non-US firms due to better management practices

[50] Appendix D.4 displays the results for adopting all 3 cloud suppliers.

[51] This measure includes IaaS, PaaS, and SaaS computing and storage-related postings but ignores distributed computing and cloud server postings.

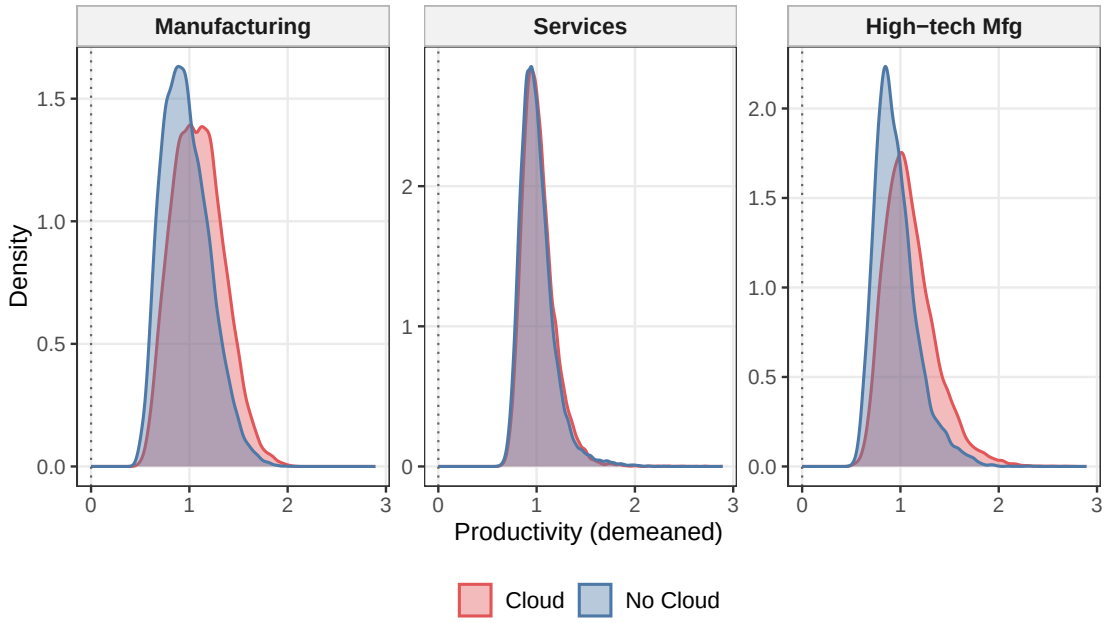


Figure 6: Steady state productivity distribution

Note: Simulated steady state productivity distributions for firms that adopt 1 cloud supplier vs those that have not adopted any suppliers.

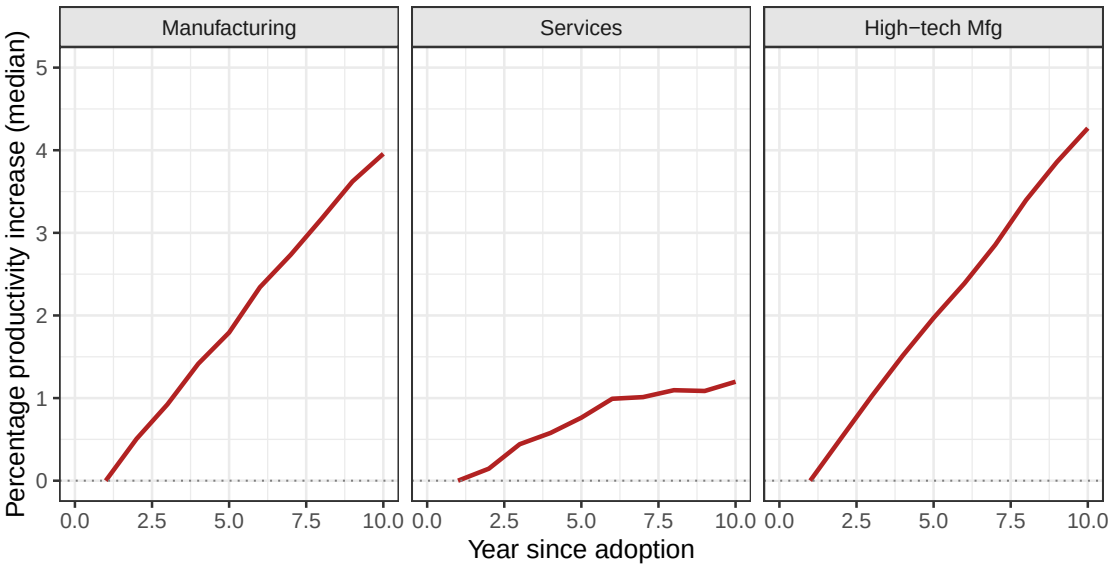


Figure 7: Median productivity difference between adopters and non-adopters.

Note: Difference in median productivity for adopters and non-adopters, from 10,000 simulations over 10 periods.

(Bloom et al. [2012]). As another exercise, I compute the gains from adopting Microsoft Azure IaaS in the case study reported in North et al. [2019]. Including cost savings, and ignoring equilibrium effects of non-cloud inputs, their results imply a productivity gain on average of 0.17% for the first 3 years when including cost savings and a gain of 0.05% when only including increases in revenue.[52] While empirical evidence remains scarce, it is heartening that my estimates fall between these two.[53]

6.2 Cost parameters

Using my production function estimates, I then estimate the remaining cost parameters via MSM. Results are reporting in Table 7, along with data moments and their robust standard errors. First, recall that costs are scaled up by firms' data stocks, D_{ft}. In order to allow my model to be computationally tractable, I assume that firm production of capital is Cobb-Douglas in D_{ft} and non-IT capital. Then, assuming homogeneous adjustment costs for each capital type, I can write:

$$\log(I^D_{ft}) = \zeta_{0i} + \zeta_{1i} \log(I_{ft}) + \Xi_l + \tau_i + \epsilon^D_{ft} \tag{28}$$

where I^D_{ft} is data investment in petabytes[54], Ξ_l are NUTS3 region l fixed effects, τ_i are 4-digit NACE code industry fixed effects, and $\epsilon^D_{ft} \sim \mathcal{N}(0,1)$.[55] In this regression, I focus on 2015-2016 before cloud computing was widely adopted, since Aberdeen only contains data on on-premise storage investments, while cloud computing investments are included in total software investments. I use investments rather than stocks because Aberdeen does not report detailed estimates of the data stock, and the pre-cloud time window is too short to use the perpetual inventory method. Aberdeen reports data stocks in euros. To convert to PB, I first convert to USD using the average annual exchange rate. I scrape HardDriveBenchmark.net [n.d.] for HDD prices and volumes back to 2018 and compute the annual average cost in

[52] For full details of the comparison, see Appendix D.3.

[53] Other issues such as endogeneity and measurement error remain. See Appendix D for robustness checks to deal with these issues.

[54] 1 PB = 1 million GB.

[55] I present the microfoundation for this equation in Appendix B.4.

Table 7: Elasticity of data investment with respect to total investment

Variable	High-tech manufacturing	Low-tech manufacturing	Services
ζ_0	-1.694 (0.583)**	-1.472 (0.167)***	-2.170 (0.120)***
ζ_1	0.388 (0.014)***	0.289 (0.009)***	0.146 (0.005)***
Observations	3,221	5,187	16,579
Firms	2,379	6,454	18,692
Region FE	✓	✓	✓
Industry FE	✓	✓	✓
R^2	0.4131	0.3630	0.2651
Adj. R^2	0.3571	0.3216	0.2428

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions estimated using OLS with NUTS3 and NACE4 fixed effects.

\$/TB.[56] Then, I use the US BLS price index for disk storage to obtain the price per TB back to 2015 (McCallum and of Labor Statistics [2024]).[57] Coefficients for this regression are presented in Table 7, showing a positive and significant relationship between data investments and total investments.

Next, Table 8 reports the estimates of the dynamic choice parameters.

Entry costs are falling rapidly; for a firm with a capital stock of €$70M$, the cost falls from €$9.05M$ in 2017 to €1.11 by 2021, consistent with the rapid rise in adoption observed in the data. Switching costs are relatively high, but a majority in the variation in supplier choice is coming from price and feature parameters. For example, a doubling of the price per core-hour increases maintenance costs by €0.52 per petabyte (PB) of data in the cloud. The complementarity of Microsoft Azure and on-premise Microsoft server products is especially large to rationalize the high covariance between Microsoft products observed in the data. Finally, the variation in entry costs is relatively small, while the variances of maintenance costs and investment are larger.

Model fit is reported in Figure 8 for entry rates. As shown, the model fits entry dynamics

[56] I use hard drives rather than solid state drives, as prior research has shown that even by 2016, most storage volume sold was on HDDs rather than SSDs (Igami and Uetake [2020]). Recent industry reports corroborate that this trend continued into more recent years (TRENDFOCUS, Inc. [2019]).

[57] The narrowest consumer price index in Eurostat that contains storage is HICP 09.1.3 ("Peripheral equipment"), which bundles goods with a rapid price decline such as storage with others like monitors and keyboards. Process innovation in storage media is the largest driver of annual price changes for storage, rather than cross-country currency fluctuations between the US and Eurozone. Therefore, I use the more precise BLS index.

Table 8: Cost Parameter Estimates

Variable	Estimate	SE
Entry Costs		
Constant (κ_0^E)	3.6449***	(0.1277)
Slope (κ_1^E)	-0.6799***	(0.0271)
Switching cost		
Switching cost (κ^S)	2.9778***	(0.0842)
Maintenance Costs		
Baseline (α_0)	2.4928***	(0.1948)
Price (α_1)	18.2175***	(0.0929)
Multihoming cost (α_2)	3.9289***	(0.4163)
VM log quality (α_3)	-0.0429***	(0.1585)
Azure-MS server complementarity (α_4)	-4.1624***	(0.4068)
Brand Effects		
Amazon brand effect (ξ_{AWS})	-0.6006***	(0.0704)
Microsoft brand effect (ξ_{MS})	-1.4773***	(0.1068)
Capital Costs		
Convex adjustment cost (C_1^K)	2.2356***	(0.3001)
Cost Heterogeneity		
Entry cost shock variance (σ^E)	1.1091***	(0.1429)
Maintenance cost shock variance (σ^b)	1.0135***	(0.0442)
Investment cost shock variance (σ^K)	2.2500***	(0.0128)
Supplier Conduct		
Supplier conduct (ζ^S)	0.1003***	(0.005)

Notes: Estimates are recovered using MSM. Standard errors computed using 80 block bootstrapped samples, which reflect sampling uncertainty from second-stage moments but not production function parameters.

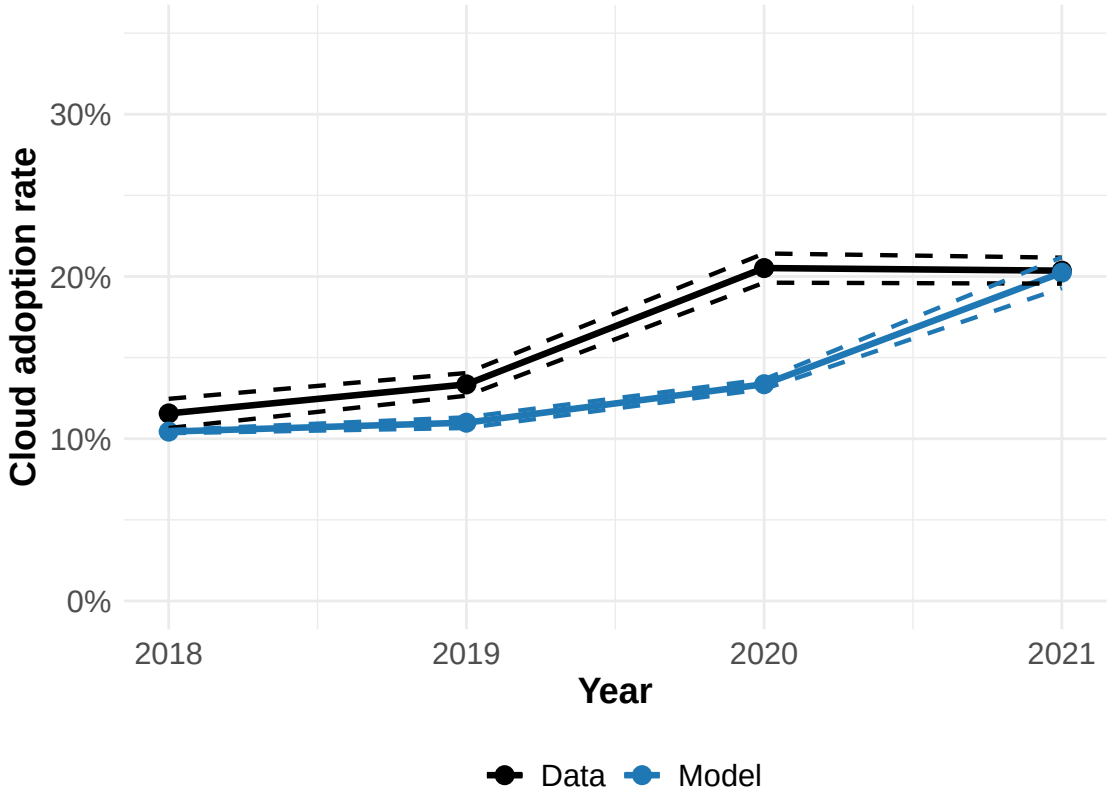


Figure 8: Actual vs simulated cloud adoption rates
Dashed lines represent 95% CI, which account for sampling and simulation uncertainty for the simulated moments and sampling uncertainty for the data moments.

relatively well, if a bit low. Next, Table 9 presents how well the model fits my chosen moments. Wit the exception of multihoming rates and switching rates, the model fits the data reasonably well.

Table 9: Empirical vs. simulated moments

Moment	Data	Simulation
Adoption rate 2018 ($\frac{\sum_f \cdot N_{2018} \mathbb{1}\{b_{ft} \neq \varnothing\}}{\sum_f N_{2018}}$)	0.1156	0.1044
	(0.0009)	
Adoption rate 2019 ($\frac{\sum_f \cdot N_{2019} \mathbb{1}\{b_{ft} \neq \varnothing\}}{\sum_f N_{2018}}$)	0.1335	0.1099
	(0.0007)	
Adoption rate 2020 ($\frac{\sum_f \cdot N_{2020} \mathbb{1}\{b_{ft} \neq \varnothing\}}{\sum_f N_{2018}}$)	0.2052	0.1336

Continued on next page

Moment	Data	Simulation
	(0.0008)	
Adoption rate 2021 ($\frac{\sum_f \cdot N_{2020} \mathbb{1}\{b_{ft} \neq \varnothing\}}{\sum_f N_{2018}}$)	0.2036	0.2024
	(0.0008)	
Bundle autocorrelation ($\mathbb{E}[\mathbb{1}\{b_{ft} = b_{ft-1}\} \vert b_{ft} \neq \varnothing]$)	0.9371	0.8463
	(0.0012)	
Multi-cloud rate ($\mathbb{E}[\vert \mathbb{1}\{b_{ft}\vert > 1\} \vert b_{ft} \neq \varnothing]$)	0.3384	0.4166
	(0.0026)	
Exit rate ($\mathbb{E}[\mathbb{1}\{b_{ft} = \varnothing\} \vert b_{ft-1} \neq \varnothing]$)	0.0338	0.0547
	(0.0006)	
Size premium ($\mathbb{E}[K_{ft} \vert b_{ft} \neq \varnothing] / \mathbb{E}[K_{ft} \vert b_{ft} = \varnothing]$)	2.4961	1.6047
	(0.0401)	
Mean investment rate ($\mathbb{E}[\frac{I_{ft}}{K_{ft}}]$)	0.1576	0.2260
	(0.0006)	
Investment rate variance ($Var(\frac{I_{ft}}{K_{ft}})$)	0.0844	0.0526
	(0.0011)	
Own price correlation (new entrants) ($\mathbb{E}[p_{fbt} \cdot \mathbb{1}\{b_{ft}\} \vert b_{ft-1} = \varnothing]$)	0.0559	0.0696
	(0.0005)	
Own price corr. (no MS server) ($\mathbb{E}[p_{fbt} \cdot \mathbb{1}\{b_{ft}\} \vert b_{ft-1} = \varnothing, MS_{ft} = 0]$)	0.0557	0.0623
	(0.0010)	
Own quality correlation (new entrants) ($\mathbb{E}[\log(q_{bft}) \cdot \mathbb{1}\{b_{ft}\} \vert b_{ft-1} = \varnothing]$)	1.8008	1.8252
	(0.0020)	
Own quality corr. (no MS server) ($\mathbb{E}[\log(q_{bft}) \cdot \mathbb{1}\{b_{ft}\} \vert b_{ft-1} = \varnothing, MS_{ft} = 0]$)	1.8245	1.8309
	(0.0049)	
Azure-Microsoft correlation ($\mathbb{E}[\mathbb{1}\{Azure \in b_{ft}\} \cdot \mathbb{1}\{MS_{ft}\} \vert b_{ft-1} = \varnothing]]$)	0.5997	0.5934
	(0.0030)	

Continued on next page

Moment	Data	Simulation
AWS entrant share (MS server) ($\mathbb{E}[\mathbb{1}\{Amazon \in b_{ft}\}\|b_{ft-1} = \varnothing, MS_{ft} = 0]$)	0.1923	0.1870
	(0.0018)	
Azure entrant share (MS server) ($\mathbb{E}[\mathbb{1}\{Azure \in b_{ft}\}\|b_{ft-1} = \varnothing, MS_{ft} = 0]$)	0.6793	0.7129
	(0.0024)	
AWS entrant share (no MS server) ($\mathbb{E}[\mathbb{1}\{Amazon \in b_{ft}\}\|b_{ft-1} = \varnothing, MS_{ft} = 0]$)	0.3348	0.3461
	(0.0053)	
Azure entrant share (no MS server) ($\mathbb{E}[\mathbb{1}\{Azure \in b_{ft}\}\|b_{ft-1} = \varnothing, MS_{ft} = 0]$)	0.4099	0.4519
	(0.0050)	
Multi-cloud size premium ($\mathbb{E}[K_{ft}\|\|b_{ft}\| > 1]/\mathbb{E}[K_{ft}\|\|b_{ft}\| = 1]$)	1.4987	0.9423
	(0.0624)	
Amazon price ($\mathbb{E}[p_{Amazon,t}]$)	0.0516	0.0603
	(0.0000)	
Microsoft price ($\mathbb{E}[p_{Microsoft,t}]$)	0.0583	0.0848
	(0.0001)	
Google price ($\mathbb{E}[p_{Google,t}]$)	0.0454	0.0768
	(0.0000)	
Entry rate ($\mathbb{E}[\mathbb{1}\{b_{ft} \neq \varnothing\}\|b_{ft-1} = \varnothing]$)	0.0617	0.0117
	(0.0003)	

Note: Standard deviation of empirical moments calculated across 80 block bootstrapped samples, which reflect sampling uncertainty. Supplier price moments are computing using suppliers' optimization problem, while the remaining simulated moments come from consumer firms' optimization problem.

Table 10 presents estimates of the costs of using various cloud providers in the manufacturing sector, for firms at various sizes in the size distribution used in the simulation.[58] As

[58] In appendix **[In progress.]** I display the corresponding estimates for the services and high-tech manufacturing sectors.

expected, Microsoft and Amazon are lower priced than Google, largely due to high vertical quality terms that rationalize their higher adoption rates relative to Google. How do these numbers compare to real-world estimates? While representative statistics are difficult to obtain, an analysis by North et al. [2019] estimates that a firm with revenue of €3.4 billion saw entry and maintenance costs of using Microsoft Azure that sum to €3.9 million. Using the median K_{ft}/Y_{ft} value in the manufacturing sector in my sample, I find that the corresponding costs are €15.2 million in 2017, which falls to €4.2 million in 2021. While these numbers are slightly higher, it is encouraging that my simplified back of the envelope calculation is nevertheless quite similar to the cloud computing expenditures of a sample of real firms.[59]

Maintenance costs					
Bundle	Minimum	10th Percentile	Median	90th Percentile	Maximum
Azure	0.1751	0.2293	0.2984	0.4154	0.5838
Amazon	0.3801	0.4976	0.6478	0.9017	1.2672
Google	0.4643	0.6078	0.7912	1.1014	1.5478
Entry costs					
Year	Minimum	10th Percentile	Median	90th Percentile	Maximum
2017	2.5774	3.3741	4.3919	6.1137	8.5916
2021	0.3166	0.4145	0.5395	0.7511	1.0555

Table 10: Manufacturing sector cloud cost distribution

Note: Costs computed for firms used in the simulation, in millions of real euros. Bundle prices and qualities are computed using 2017 values. I assume firms are not using Microsoft server products.

7 Counterfactual policies

To study the welfare of competition policies for cloud computing, I use my model to conduct counterfactual scenarios. I first study the impact of a ban on data egress fees implemented in the EU Data Act (2024) and studied by other regulators (e.g. Korea Fair Trade Commission [2022]). Next, I consider the case in which firms using Microsoft Server products see gains

[59]For the full back of the envelope calculation, as well as results for other sectors, see Appendix **In proress.**.

from using these products on Google or Amazon cloud, as would be the case if a regulator mandated price parity or higher feature compatibility for using Microsoft products on rival cloud platforms. For each of these first two counterfactual policies, I consider the impact with and without a supplier price response.[60] Finally, to study the impacts of market power directly, I simulate a merger between Microsoft Azure and Google Cloud Platform and study the resulting implication for social welfare.

Recalling I make the simplifying assumption that prices in the steady state are fixed at time T prices, to avoid bias, I compute welfare along $t = 2017, ..., \tilde{T}$, where $\tilde{T} = T - 20 = 2031$. Per period welfare can be written as

$$\mathbb{W}_t = \underbrace{\int N_t(s_{ft})[\pi_{ft}(s_{ft}) - \kappa_t(D_{ft}) - C(I_{ft}, K_{ft})\epsilon^K_{ft}]dF(s_{ft})}_{\text{Consumer surplus}} + \underbrace{\sum_j \pi^S_{jt}(\mathbf{s}_t)}_{\text{Producer surplus}} \tag{29}$$

Then, social welfare can be written as

$$\mathcal{W} = \sum_{t=2017}^{\tilde{T}} \beta^t \mathbb{W}_t \tag{30}$$

To solve the counterfactual, I initialize the model using the observed 2017 distributions over states. Next, I assume that the policy arrives as an unexpected shock after firms have made their initial 2017 decisions, affecting all actions in period $t = 2018, ..., T$. For the counterfactual scenarios without price increases, I simply re-solve the inner value function fixed point holding prices fixed. For the counterfactual with endogenous price updates, I unfreeze the MME loop and re-solve for prices and downstream firms' beliefs. Finally, counterfactuals can change suppliers' invest-harvest incentives. In the appendix, I show that, under simplifying assumptions, the counterfactual conduct parameter can be written as

$$\zeta^{CF} \approx 1 + (\zeta - 1) \cdot \frac{\frac{1}{Q^{CF}_{jt}} \cdot \frac{\partial \mathbb{E}_t[W^{CF}_{jt+1}(s^{CF}_{t+1})]}{\partial P_{jt}}}{\frac{1}{Q_{jt}} \cdot \frac{\partial \mathbb{E}_t[W_{jt+1}(s_{t+1})]}{\partial P_{jt}}} \tag{31}$$

[60] For example, Competition and Markets Authority [2025g] finds that up to June 2025, cloud providers have not altered prices since eliminating some egress fees.

I then solve a quasi-Newton fixed point until ζ^{CF} converges, which I show is a contraction mapping as long as marginal costs are low enough. For robustness I solve for counterfactual pricing with and without ζ updates.

7.1 Ban on data egress fees

The first policy is a ban on data egress fees by cloud providers. The rationale behind such a ban is as follows. Egress fees act as a form of switching cost to consumers, locking them into long-run relationships with suppliers, even when better alternatives are available. Even with a rapidly-expanding consumer base, the high fixed costs of entry on the supply side means that the market remains concentrated. Theoretical analysis of an egress fee ban are mixed, with welfare results dependent on the demand elasticity (Do and Miklós-Thal [2025], Gans et al. [2023]).

To simulate the counterfactual scenario, I assume that, beginning in 2017, cloud computing suppliers are banned from charging egress fees, and simulate sectors forward until T, computing welfare until $\tilde{T}$. I consider two scenarios - one without supplier price changes and one with supplier price updates. The former mirrors the current reality: as of May 2025, none of the major cloud providers have raised IaaS prices in response to the removal of egress fees in January and March of 2024, respectively (Competition and Markets Authority [2025g], 44). In the second, I allow cloud suppliers to update their per-VM prices.

If a firm switches from a single supplier to another or exits the cloud computing market entirely, I assume they pay the full egress fee schedule listed under "egress to the public internet" or the equivalent. Otherwise, I do not observe the fraction of data a firm stores in each supplier when multihoming, which is endogenous to egress fees (Competition and Markets Authority [2025c]). I assume that when firms change bundles, they keep an equal fraction of data on all providers. Thus, if moving from $b_t = \{\text{Amazon}\}$ to $b_{t+1} = \{\text{Amazon, Microsoft}\}$, a firm will move 50% of its year t data stock. For each firm, I subtract this cost from κ_S. In addition, firms may move a fraction of their data between suppliers in each period when multihoming. For example, a firm may need to move customer data from its website, hosted on Microsoft Azure, to be analyzed using machine learning tools, hosted on Google Cloud Platform. To model this, I assume that firms move a fraction λ of data between providers in

each period t. In my simulations, I set $\lambda \in \{0.15\}$, following Piper and Cohen [2023], which finds that multicloud firms on average store 85% of data on a single supplier's servers.

7.2 Increased Microsoft server interoperability

The second policy I consider is a higher level of adoption benefits for firms using Microsoft server products on non-Microsoft cloud platforms. The UK Competition and Markets Authority has found that Microsoft charges more for Windows Server licenses for use on on rival cloud platforms than on Microsoft Azure, leading them to open an investigation into anticompetitive bundling. In addition, programs such as Azure Hybird Benefit, which allows consumers to host Azure services on excess local computing capacity for a reduced cost, further increase the synergy between the two products. I capture the impact of a hypothetical mandate to increase price or feature parity for Microsoft server products across products through allowing firms not using Microsoft Azure to capture a portion $\varepsilon \in \{0.1, 0.2\}$ of the benefit that Azure users do by reducing maintenance costs by $\varepsilon \cdot \alpha_4$ if they use Microsoft Server products.

7.3 Horizontal merger

Finally, to study the direct impacts of market power on cloud adoption, I simulate a merger between Microsoft Azure and Google Cloud Platform, such that the upstream market becomes a duopoly.[61] Given that this scenario has not occurred in the data, I make the conservative assumption that the merger only results in joint pricing decisions by the merged firm, without additional synergies that result either in reduced switching frictions or lower marginal costs.[62]

[61] The closest such example in the industry is IBM's 2019 acquisition of Red Hat, a firm that provides open-source cloud-adjacent software and hardware, such as containerization, hybrid cloud infrastructure, and the Red Hat Enterprise Linux OS.

[62] This can represent medium term merger dynamics, before the merged firm is able to optimize software interoperability or find ways to reduce marginal costs.

7.4 Results

The results are displayed in Table **??**. Both with and without a pricing response, the egress fee ban produces an overall increase in annual welfare - €47.9 million in the case without price increases and €157.2 million in the case in which suppliers are allowed to update prices. Surprisingly, the fall in producer surplus is similar in both scenarios - a loss of 3.55% (3.48%) in the case with (without) a price update, indicating that most of the loss net loss in revenue comes from the loss in egress fees. These results indicate that conditions are such that banning egress fees is a net gain to France, regardless of how suppliers respond with prices (Do and Miklós-Thal [2025]). However, results are overall quite modest, which is reasonable given that IT is relatively a small part of firms' input mix and cloud computing egress fees are even smaller.

Next, the increased interoperability of Microsoft server products also increases consumer welfare - when suppliers update their prices, consumer welfare rises from €132.6-296.0 million (0.05%-0.10%). Surprisingly, while Microsoft Azure's profit declines, most of the losses are picked up by firms switching to Google or Amazon, resulting in modest producer profit losses of only 0.09%-0.21%. While this result does not take into account the gains and losses of Microsoft's Windows division no long-run innovation incentives, it strongly suggests that, at least in the short run, enforcing price or feature parity across platforms can increase welfare.

Finally, the horizontal merger, surprisingly, produces a fall in welfare for both sides of the market. While the loss to suppliers is small (only 0.41% of profits), the loss to consumers is larger, an average of €837.5 million per year, or 0.28% of profits.

Microfounding the J-curve. Tracing out the path of downstream firm profits reveals that cloud computing adoption displays a similar J-curve to other general purpose technologies (cf. McElheran et al. [2025], Brynjolfsson et al. [2021]). In Figure 9 below, I plot the annual welfare gain from the egress fee ban, which resulted in additional multihoming and switching. Except for the COVID-19 shock, we see a classic J-curve, with firms paying fixed costs of adoption and switching in early years before experiencing productivity gains in the future.

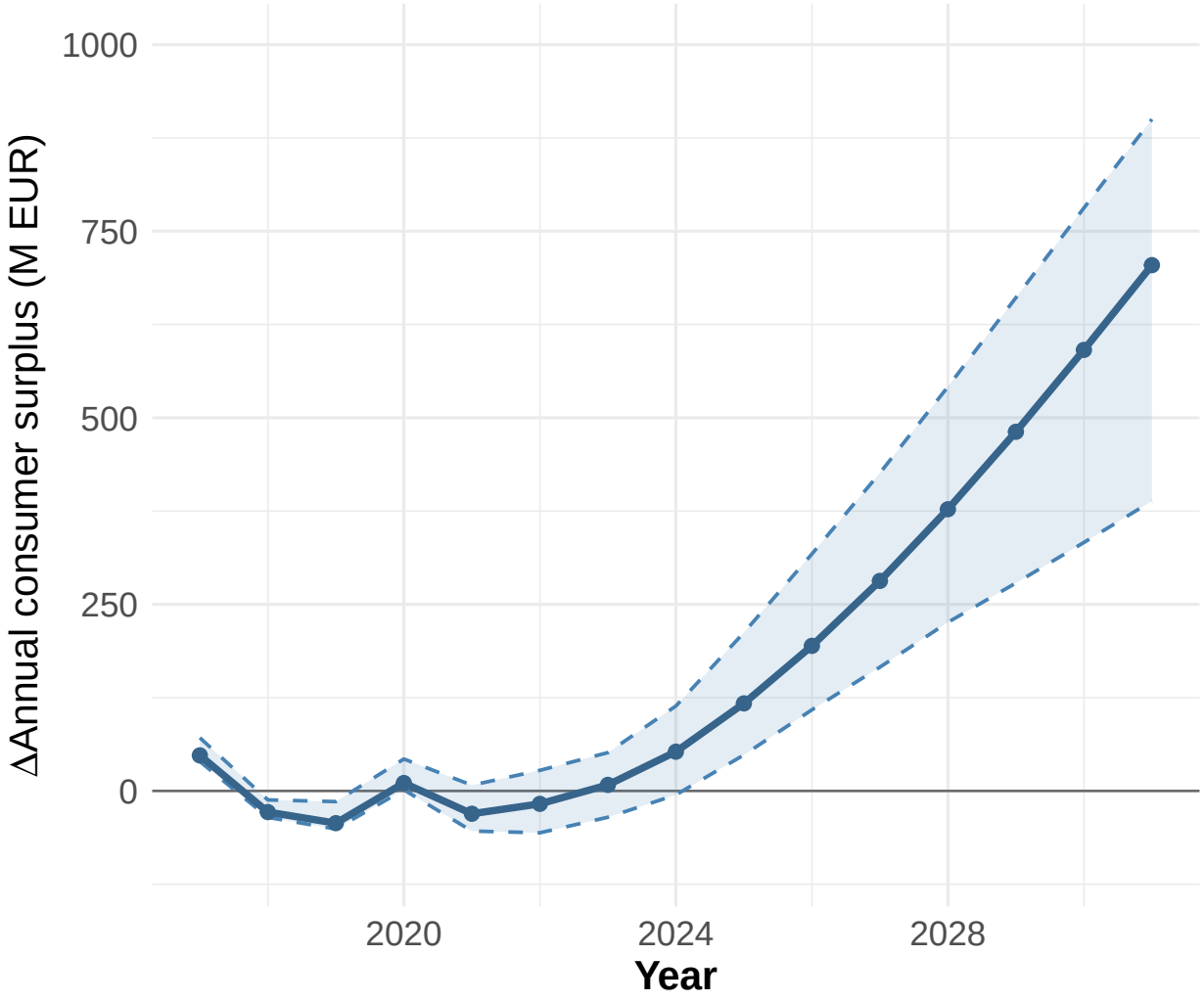


Figure 9: J-curve in downstream profit after adoption

Annual change in consumer surplus, 2017-2031, with 95% CI from 80 boostrap samples.

	ΔPrice (avg.)	ΔCS (avg.)	ΔPS (avg.)	ΔWelfare (avg.)	ΔCS (NPV)	ΔPS (NPV)	ΔWelfare (NPV)
Panel A: Egress fee ban							
Fixed prices	0.000	74.3	-26.5	47.9	630.1	-226.4	403.7
	(0.00%)	(0.03%)	(-3.55%)	(0.02%)	(0.03%)	(-3.53%)	(0.02%)
Price response	-0.0005	183.1	-25.9	157.2	1292.9	-222.6	1070.3
	(-0.61%)	(0.06%)	(-3.48%)	(0.05%)	(0.06%)	(-3.47%)	(0.05%)
Panel B: MS-server spillover, $\varepsilon = 0.10$							
Fixed prices	0.000	26.1	-0.2	25.9	246.5	-1.3	245.2
	(0.00%)	(0.01%)	(-0.03%)	(0.01%)	(0.01%)	(-0.02%)	(0.01%)
Price response	-0.0004	132.6	0.7	133.2	1089.0	6.5	1095.4
	(-0.57%)	(0.05%)	(0.09%)	(0.05%)	(0.05%)	(0.10%)	(0.05%)
Panel C: MS-server spillover, $\varepsilon = 0.20$							
Fixed prices	0.000	53.6	-0.4	53.2	507.1	-2.9	504.2
	(0.00%)	(0.02%)	(-0.06%)	(0.02%)	(0.02%)	(-0.04%)	(0.02%)
Price response	-0.0010	296.0	1.6	297.5	2434.3	14.9	2449.2
	(-1.29%)	(0.10%)	(0.21%)	(0.10%)	(0.11%)	(0.23%)	(0.11%)
Panel D: Merger (GCP + Azure)							
Price response	0.0192	-837.5	-4.1	-841.7	-9003.2	-45.8	-9049.0
	(25.90%)	(-0.28%)	(-0.56%)	(-0.28%)	(-0.38%)	(-0.71%)	(-0.38%)

Table 11: Welfare effects of counterfactual policies and market structure

Notes: Welfare changes relative to the estimated baseline over 2017-2031, in millions of real 2015 euros, averaged over 80 bootstrap samples. "Avg." columns report average annual changes; "NPV" columns report 2017 present values discounted. ΔPrice is the unweighted mean change in supplier subscription prices over the same window. "Fixed prices" holds supplier prices at the baseline equilibrium path; "price response" re-solves the supplier pricing equilibrium.

8 Conclusion

In this paper, I examine the impact of adoption, switching, and integration frictions on firms' technology input choices and production decisions using the setting of cloud computing in France. I find that adoption is rising rapidly, and multihoming is common. Despite rapid evolution of the industry, I find that firms' decisions are highly persistent, and supplier concentration is high.

To rationalize these findings and explore the impact of recent competition policies, I develop a model of industry dynamics with endogenous firm IT adoption choices that may

increase productivity but incur entry, switching, and maintenance costs. Estimating the model, I find gains from adopting cloud computing of between 0.2% and 1.5%. Firms' supplier preferences are driven by complementarities with on-premise technologies as well as price; switching costs are moderate. Using the model, I estimate the impacts of an egress fee ban in the EU Data Act. I find that welfare falls for both consumers and producers in the short run, but, absent supplier price responses, increases overall welfare by 0.05%, or €157.2 million per year.

The characteristics in the cloud services market - widespread adoption, high supplier concentration, and multihoming - are common to many software IT services industries. The modeling approach and insights in the present work may be used to study dynamics in other such industries, such as that for large-language models (Demirer et al. [2025]).

While the present work studies the long-run impact of pro-competition policies for cloud computing, it focuses on price and cost changes only. In an industry with rapid product improvement, policies that alter incentives to lock in consumers or mandate a certain interoperability level may affect innovation incentives. A study of the effect of such policies on supplier innovation is left for future research.

A Data appendix

A.1 Dataset description

Our primary firm data comes from Bureau van Dijk's Orbis dataset, which covers unconsolidated firm financial data for a large fraction of public and private French firms. Prior research has found a high level of coverage of the dataset,[63] and I find that the dataset covers **In progress.**% of firms and **In progress.** % of output in the manufacturing and services sectors, respectively.

The data on cloud and other IT investments comes from Aberdeen (formerly Harte Hanks), a market research firm, which collects IT data to sell to computer software and hardware vendors. Aberdeen's data collection relies on a series of over 10,000 monthly surveys, from which they obtain data on firm IT stock and software vendor choice at the establishment level (Aberdeen Group [2017]). In addition, using these survey responses, they estimate the value of IT investments at each establishment, and rely on web scraping and interpolation to impute missing values, which can introduce errors (Brynjolfsson and Hitt [2003], Demirer et al. [2024]).

Historically, the firm has targeted for survey all firms with 100 or more employees, though recently it has expanded to include smaller firms. From 1998-2007, Aberdeen had a 45.1% response rate among European firms, increasing in firm size (Bloom et al. [2012]). By comparison, our sample has a 24.2% match rate among *all* French firms in Orbis.

A.2 Matching procedure

As the datasets do not have any common firm identifiers, we are forced to use other identifying information to match firms in both datasets. Our matching algorithm is as follows. First, to ensure that we are matching at the same level of ownership, we restrict our Orbis sample to firms with unconsolidated financial data.[64] For Aberdeen, we merge all establishments that have the same enterprise ID and name.[65] Second, following the approach in Hazell

[63] Asker et al. [2014]
[64] CONSCODE of "U1" or "U2."
[65] For example, Google and YouTube have the same enterprise ID but are separate firms in our sample.

et al. [2022], we clean firm names by converting them to lowercase, removing common words and abbreviations like "company" and "ltd." We then run the following matching algorithm, modified from Demirer et al. [2024]:

1. Match unique exact firm names
2. Match firms with the same URL
3. Match firms with the same phone number
4. Match firms in the same postal code, and the Jaro-Winkler name distance is less than 0.1
5. Match firms in the same city, and the name distance is less than 0.08
6. Match firms in the same province, and the name distance is less than 0.07
7. Match firms in the same country, and the name distance is less than 0.065

For city and province matches, since a subsidiary firm may have multiple headquarters listed, I am unable to determine which headquarters location is the corresponding one to that in Orbis. Therefore, I consider a match to be positive if the match conditions are met and any establishment is in the same city or province as in the Orbis data.[66] Using this procedure, I am able to match 69.5% of firms in the Aberdeen sample and 24.2% of firms in the Orbis sample. Manual spot checks ensure the accuracy of the merged dataset.

A.3 Products observed in data

Table 12 lists the products observed in Aberdeen's surveys. Note that the level of detail varies by cloud provider, with only Amazon listing both storage and computation, restricting the level of detail at which we are able to estimate egress fees. However, egress fees apply to both virtual machine storage and dedicated storage blobs, so my assumption that all firms exiting a specific IaaS provider are subject to egress fees seems reasonable.

[66] This may introduce bias if Aberdeen does not sample every establishment for a given firm. Until 2009, Aberdeen included data on total number of establishments at each firm as well as the fraction surveyed. For each year in the US sample until 2009, the mean firm had data on above 95% of establishments, and the median firm had data on 100% of establishments.

Table 12: Cloud products grouped by manufacturer

Cloud provider	**Products**
Amazon	Amazon S3, Amazon EC2, Amazon Web Services
Microsoft	Azure, Azure Blob Storage, Azure Virtual Machines
Google	Cloud Platform, Compute Engine

List of all product names that have appeared it at least one year of Aberdeen's data from 2017-2021. Names have been standardized across survey waves.

A.4 Summary statistics

I present summary statistics for the full Orbis sample in the following table.[67] Confirming prior research (e.g., Bloom et al. [2012]) the Aberdeen sample is weighted toward larger firms, with average revenue that is 3.28 times the overall sample and average employee count that is 1.44 times the overall sample. Further, firms in the matched sample are around 10 years older than firms in the full sample.

Table 13: Full Orbis sample summary statistics

Variable	Observations	Mean	25th Percentile	Median	75th Percentile
Revenue	4,462,383	6.23	0.06	0.29	1.34
Labor	1,780,539	46.92	3.00	8.00	20.00
Capital	5,040,883	9.57	0.01	0.10	0.48
Materials	4,119,885	4.71	0.03	0.13	0.69
Firm Age	5,041,128	13.78	4.00	10.00	19.00

Notes: The units for revenue, capital, and materials are in millions of euros. The units for labor are in employees. The unit for firm age is years.

A.5 Capital stock computation

Because I do not have enough observations in Orbis, I use the book value of the capital stock for each firm. In this appendix, I describe how I recover total investment and non-IT investment. First, I deflate all variables using capital-type specific deflators from Eurostat. Next, Orbis contains depreciation data for a subset of firms. For the remaining firms, I compute

[67] Revenue and labor figures are sourced from Aberdeen if the data is missing in Orbis.

$$\delta_{\iota t} = \mathbb{E}_{\iota t}\left[1 + \frac{I_{ft}}{K_{ft}} - \frac{p_{\iota t+1}K_{ft+1}}{p_{\iota t}K_{ft}}\right] \tag{32}$$

where $\delta_{\iota t}$ is then the average annual depreciation rate for industry ι at the 4-digit NACE code level. Using this, I compute the value of the remaining investment stock. Finally, Aberdeen has values for establishment level IT capital stocks (storage, hardware, and software), which I sum up to the firm level, deflate by their relevant Eurostat deflators, and subtract from I_{ft} for each f, t.

A.6 Computing VM prices

This section describes my price dataset in more detail and describes the process to merge the Azure usage and price data.

A.6.1 VM usage dataset description

To compute the average price paid at the firm-provider level, I utilize anonymized virtual machine usage data from Microsoft Azure (Cortez et al. [2017]). The first trace data is from 2017 and contains 30 consecutive days of VM usage for 5,958 subscriptions and 2,013,767 VMs. The second is from 2019 and contains 30 consecutive days of VM usage for 6,689 subscriptions and 2,695,548 VMs. This dataset has been used in prior economics research to understand firm learning-by-doing in virtual machine utilization (Brand et al. [2024]). The data contains the category (interactive vs delay-insensitive), categories for number of cores (0-2, 4-8, etc.), and memory size (0-2 GB, 2-4 GB, etc.).[68] To merge the data, for each region, I assume that firms pay the average price among all VMs in a given category.

Below, I present summary statistics for VM prices. Let w_{fpt} be the usage (measured in seconds) of VM instance type p used by firm f in year t, and let $W_{ft} = \sum_j w_{fpt}$. Then I can decompose the total variance of VM usage into the variance within firms and variance between firms as follows:

[68] The public data is from either internal Microsoft-owned firms' IaaS usage or external firms' usage of Microsoft PaaS. The sample has been selected to be representative of external Azure users in the month-region of study. For further details on the 2017 trace, see Cortez et al. [2017], and on the 2019 trace, see Microsoft Azure [2019].

$$\sigma^2_{t,total} = \underbrace{\frac{\sum_f W_{ft} \frac{\sum_p w_{fpt}(p_{fpt} - \bar{p}_{tf})^2}{\sum_p w_{fpt}}}{\sum_f W_{ft}}}_{\sigma^2_{t,within}} + \underbrace{\frac{\sum_f W_{ft}(\bar{p}_{ft} - \bar{p})^2}{\sum_f W_{ft}}}_{\sigma^2_{t,between}} \tag{33}$$

$$\text{where } \bar{p}_{ft} = \frac{\sum_p w_{fpt} p_{fpt}}{\sum_p w_{fpt}} \tag{34}$$

$$\bar{p}_{ft} = \frac{\sum_{f,p} w_{fpt} p_{fpt}}{\sum_{f,p} w_{fpt}}. \tag{35}$$

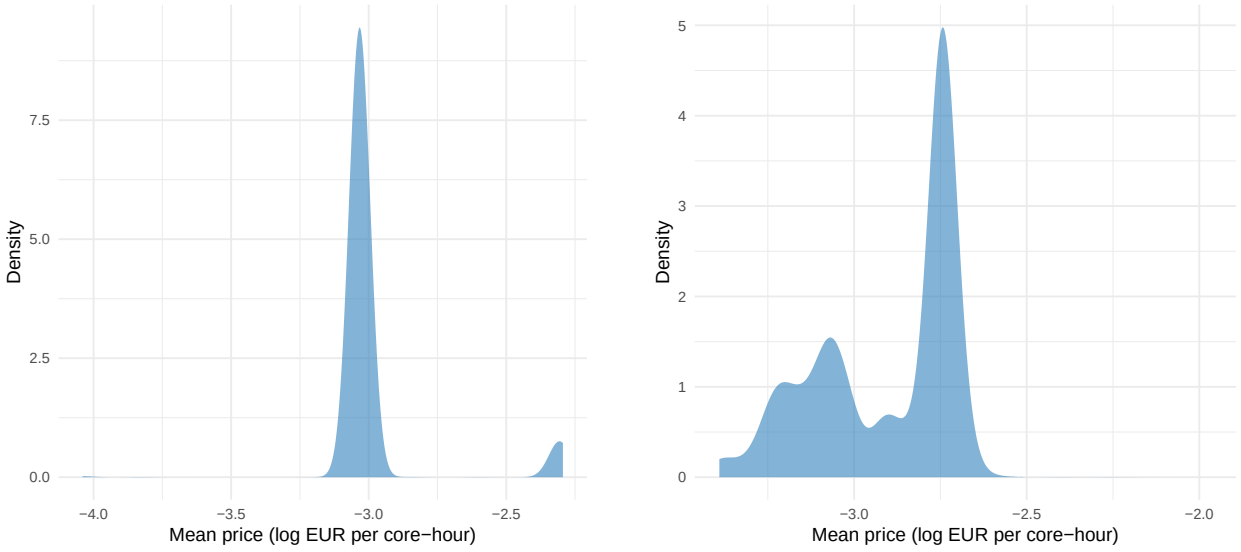
Results are presented for the region West Europe in Table 14. While variance between firms is similar in both time periods, the variance within firms increased, likely due to an expanding set of VM offerings in 2019 (139) compared to 2017 (65). In addition, Cortez et al. [2017] and Microsoft Azure [2019] show an increasing rate of per-VM CPU utilization, indicating firms in 2019 are better than firms in 2017 at matching VM compute to appropriate workloads (cf. Brand et al. [2024]).

Next, Figures 10 and 11 display the density of VM prices between and within firms, respectively. While there is a single peak of prices that most firms may in both years, there are several smaller modes, especially in 2017. Likewise, the within-firm distribution is either degenerate for smaller firms or approximately normally distributed. The mean price paid does not vary much by firm size.

Table 14: Variance decomposition of VM prices per core-hour

Variance	2017	2019
Between variance ($\sigma^2_{between}$)	1.84×10^{-4}	1.06×10^{-4}
Within variance (σ^2_{within})	2.30×10^{-6}	4.55×10^{-5}
Total variance ($\sigma^2_{between} + \sigma^2_{within}$)	1.86×10^{-4}	1.67×10^{-4}

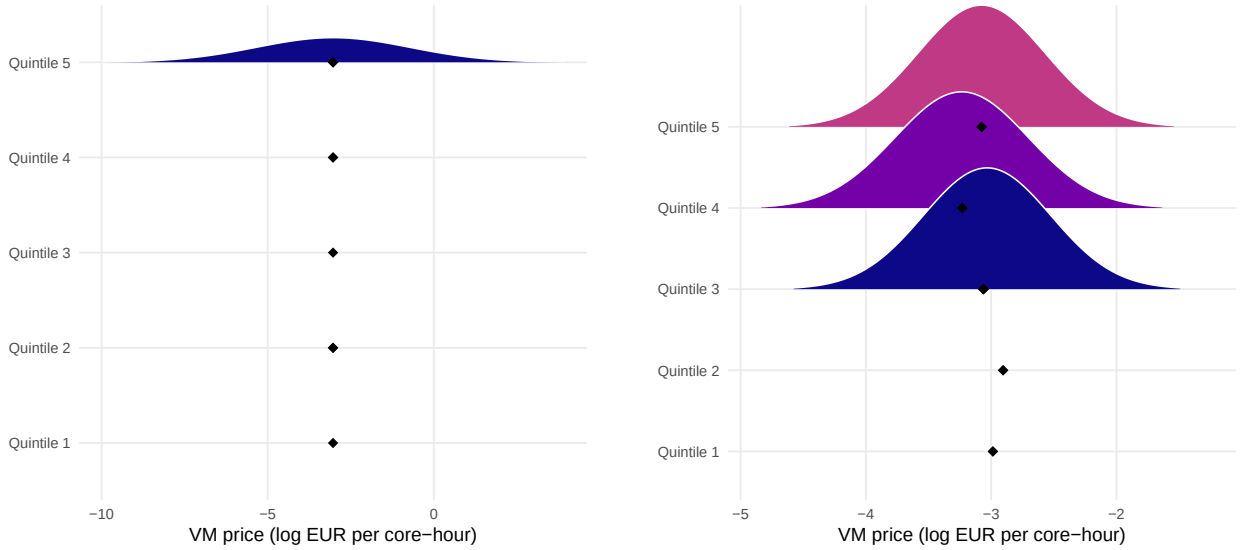
Notes: Figure displays the weighted variance in price of VM core-hour within and between firms for an anonymous Microsoft Azure region in the period November 16, 2016 to February 16, 2017 and some period ending before July 2019. For the former, price data from February 2017 is used. For the latter, price data from March 2019 is used. In 2019, 1.7% of firm-VM type observations are dropped due to the lack of a corresponding SKU in West Europe during that time.



(a) 2017 (b) 2019

Figure 10: Density of log VM prices between firms

Note: Weighted average of firm level VM prices in log euros.



(a) 2017 (b) 2019

Figure 11: Density of log VM prices within firm by quintile

Note: Weighted average of firm level VM prices in log euros. Firms are selected as the median firm in each quintile, measured by total VM usage in seconds.

A.6.2 Usage and price merge

Given that I only observe prices for a single Azure region in 2017 and in 2019, I assume that all firms use the same distribution of virtual machines for every j, l:

Assumption 1. *Let q_{plt} denote the quality of instance type p in region l at time t and let $\mathcal{F}_j^w(q_{plt})$ and $\mathcal{F}_j^b(q_{plt})$ denote the within and between consumer distribution of q_{plt}, respectively. Then*

$$\begin{aligned}
\mathcal{F}_j^w(q_{plt}) &= \mathcal{F}_{j'}^w(q_{plt}) \\
\mathcal{F}_j^w(q_{plt}) &= \mathcal{F}_j^w(q_{pl't}) \\
\mathcal{F}_j^b(q_{plt}) &= \mathcal{F}_{j'}^b(q_{plt}) \\
\mathcal{F}_j^b(q_{plt}) &= \mathcal{F}_j^b(q_{pl't}).
\end{aligned}$$

To merge Azure price data, I perform an exact match on VMs based on number of cores and memory, taking the mean price per core hour within each category. As the SKUs offered by Amazon and Google are not exact matches for Microsoft SKUs, I employ a fuzzy matching procedure in which I match exactly on CPU core count and match memory size in GB according to the size bins $(2^k, 2^{k+1}]$ for integers $k \in [0, 7]$.[69] I match the 2017 Azure trace dataset to 2016-2017 pricing data and the 2019 trace dataset to 2018-2021 pricing data.[70]

The following Table 15 displays, for each supplier, the annual weighted average price per core hour in nominal euros and fraction of VMs in the trace data that matched successfully to price data. Google's match rate is lowest, likely due to the fact that it offered fewer SKUs during this period than its competitors.

A.6.3 Price, quality, and region merge

Since it is far more complete than the Orbis location data, I use the headquarter latitude and longitude data from Aberdeen to match firms to datacenters. The matching algorithm is

[69] For example, Microsoft's basic A1 VM has 1.88 GB of RAM while the equivalent Amazon SKU, the t2.small, has 2.15 GB.

[70] The 2019 trace data has a higher match rate to 2018 Azure pricing data than the 2017 trace data. I adopt this assumption across all providers.

Table 15: Virtual machine price summary statistics

		Price per core hour			Usage match rate		
Provider	Year	Mean	Min	Max	Mean	Min	Max
Amazon	2017	0.0559	0.0492	0.0626	0.7510	0.5030	1.0000
Amazon	2018	0.0499	0.0380	0.0605	0.8040	0.7170	0.9790
Amazon	2019	0.0575	0.0551	0.0602	0.8040	0.7170	0.9790
Amazon	2020	0.0533	0.0438	0.0577	0.8480	0.7170	0.9790
Amazon	2021	0.0434	0.0365	0.0515	0.9790	0.9790	0.9790
Google	2017	0.0649	0.0608	0.0670	0.0691	0.0691	0.0691
Google	2018	0.0487	0.0449	0.0525	0.8560	0.8560	0.8560
Google	2019	0.0477	0.0379	0.0601	0.8560	0.8560	0.8560
Google	2020	0.0440	0.0394	0.0468	0.8560	0.8560	0.8560
Google	2021	0.0421	0.0368	0.0463	0.8560	0.8560	0.8560
Microsoft	2017	0.0531	0.0531	0.0531	1.0000	1.0000	1.0000
Microsoft	2018	0.0595	0.0552	0.0724	0.9790	0.9790	0.9790
Microsoft	2019	0.0643	0.0553	0.0810	0.9790	0.9790	0.9790
Microsoft	2020	0.0626	0.0476	0.0769	0.9790	0.9790	0.9790
Microsoft	2021	0.0572	0.0407	0.0712	0.9790	0.9790	0.9790

as follows. For each firm, I compute the Haversine distance between the firm's headquarters and the nearest datacenter for each provider used as follows, where headquarters are defined as follows. If the firm is single-establishment, I use that establishment. If the firm has an establishment designated as "Ultimate HQ," which is unique in the Aberdeen data, I use that. Otherwise, if there is at least one establishment designated as "Headquarters," I randomly choose one. Finally, if the firm has no in-country headquarters, I randomly assign an establishment labeled as "Branch." I drop firms that do not have location coordinate data.

As this process may introduce measurement error in prices and qualities, I compute summary statistics for 1) my full matched sample, 2) my matched sample excluding firms with random headquarter assignment, 3) my matched sample excluding multinational firms.[71] The results are presented in Table [**In progress.**]

[71] Even if multinational firms have a local headquarters in France, they still may store data at the datacenter nearest their global headquarters. I thank Daniela Muhaj for making this observation.

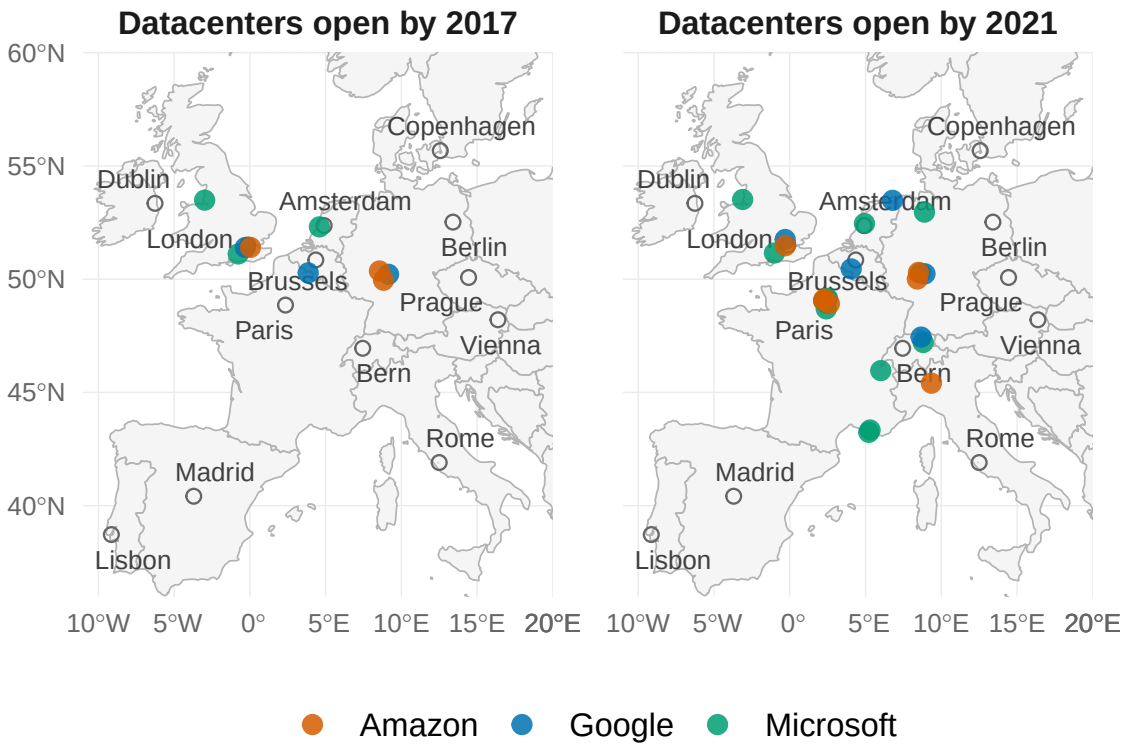


Figure 12: Datacenter locations in 2017 and 2021

A.7 Datacenter locations

The map in Figure 12 provides locations of datacenters for each of the cloud providers in my data at the beginning and end of each sample period.

A.8 Microsoft server adoption

To model the evolution of Microsoft server product adoption, I run the following AR(1) model for each sector i:

$$server_{ft} = \rho_0^s + \rho_i^s server_{ft-1} + \epsilon_{ft}^s$$

The results are displayed in Table 16. A full list of Microsoft server-related products is displayed in Table 17.

Table 16: AR(1) persistence of Microsoft server adoption (2017-2021)

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Intercept	0.0735 (0.0018)***	0.0453 (0.0008)***	0.0389 (0.0003)***
ρ_i^s	0.8917 (0.0033)***	0.9177 (0.0019)***	0.9238 (0.0007)***
Observations	25,328	82,962	582,236
Residual Std. Error	0.2397	0.2049	0.1927
R^2	0.7470	0.7311	0.7603
Adj. R^2	0.7470	0.7311	0.7603

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note: OLS regressions of Microsoft server adoption on its lag.

A.9 Computing virtual machine benchmarks

This appendix describes how I compute the average virtual machine quality using CPU benchmarks, which provide a unidimensional indicator of a CPU's performance (Goettler and Gordon [2011]). I use benchmarks from the PerformanceTEST software developed by PassMark Software [2026a]. There does not exist a CPU benchmark suite that allows for comparisons of VM quality over time.[72] An additional complication is that multiple VMs may share a single CPU, or even a single core. Some VM instances have a single vCPU, which corresponds to a single thread (Amazon Web Services [n.d.]). Thus, the total multithread performance on PassMark Software [2026a] may not adequately represent VM quality.

To compute the relative performance of a VM, therefore, I take the multithread score and divide it by the number of threads on a CPU to get a per-thread performance measure. For example, Microsoft Azure's A10 instance uses Intel Xeon E5-2660 v0, which has a multithread score of 8076 and a total of 16 threads, for a per-thread score of 504.75 (PassMark Software [2026b]). Wherever possible, I use official documentation; otherwise, I use third party cloud platform comparison websites, which run benchmarking software that is able to identify the underlying processor of a given instance type. When multiple models are used for the same instance type, I take the average score. Finally, I follow the same approach as in Appendix A.6 to compute the average benchmark score at the provider-datacenter-year level. Table 18 lists my sources for the CPUs used by each SKU. Summary statistics are listed in Table

[72] Geekbench comes closest, but newer versions of the benchmarking suite do not cover old VM instances and vice versa.

[In progress.]

Source Name	URL
Official Documentation	
Azure A-Series & Av2-Series Virtual Machines	docs.microsoft.com/.../av2-series
Azure Bsv2 / Bs-Series Virtual Machines	docs.microsoft.com/.../bsv2-series
Azure Dv2 and DSv2-Series Virtual Machines	docs.microsoft.com/.../dv2-dsv2-series
Azure Dv3 and Dsv3-Series Virtual Machines	docs.microsoft.com/.../dv3-dsv3-series
Azure Dv4 and Dsv4-Series Virtual Machines	docs.microsoft.com/.../dv4-dsv4-series
Azure Dv5 and Dsv5-Series Virtual Machines	docs.microsoft.com/.../dv5-dsv5-series
Azure Dav4 and Dasv4-Series Virtual Machines	docs.microsoft.com/.../dav4-dasv4-series
Azure Dasv5 and Dadsv5-Series Virtual Machines	docs.microsoft.com/.../dasv5-dadsv5-series
Azure DCsv2 / DC-Series (Confidential Computing)	docs.microsoft.com/.../dcv2-series
Azure Ev3 and Esv3-Series Virtual Machines	docs.microsoft.com/.../ev3-esv3-series
Azure Ev4 and Esv4-Series Virtual Machines	docs.microsoft.com/.../ev4-esv4-series
Azure Ev5 and Esv5-Series Virtual Machines	docs.microsoft.com/.../ev5-esv5-series

(continued on next page)

(continued from previous page)

Source Name	URL
Azure Eav4 and Easv4-Series Virtual Machines	docs.microsoft.com/.../eav4-easv4-series
Azure Easv5 and Eadsv5-Series Virtual Machines	docs.microsoft.com/.../easv5-eadsv5-series
Azure F-Series, Fs-Series, and Fsv2-Series VMs	docs.microsoft.com/.../fsv2-series
Azure G-Series and Gs-Series Virtual Machines	docs.microsoft.com/.../sizes-previous-gen
Azure H-Series Virtual Machines	docs.microsoft.com/.../h-series
Azure Ls-Series and Lsv2-Series Virtual Machines	docs.microsoft.com/.../lsv2-series
Azure NC-Series Virtual Machines	docs.microsoft.com/.../nc-series
Azure NCas T4 v3-Series Virtual Machines	docs.microsoft.com/.../nct4-v3-series
Azure NV-Series Virtual Machines	docs.microsoft.com/.../nv-series
Azure NVv4-Series Virtual Machines	docs.microsoft.com/.../nvv4-series
Azure: Sizes for Virtual Machines (Overview)	docs.microsoft.com/.../sizes
Azure: Previous Generation VM Sizes	docs.microsoft.com/.../sizes-previous-gen
GCP: N1 General-Purpose Machine Family	cloud.google.com/.../general-purpose-machines#n1
GCP: N2 General-Purpose Machine Family	cloud.google.com/.../general-purpose-machines#n2
GCP: N2D General-Purpose Machine Family	cloud.google.com/.../general-purpose-machines#n2d

(continued on next page)

(continued from previous page)

Source Name	**URL**
GCP: E2 General-Purpose Machine Family	`cloud.google.com/.../general-purpose-machines#e2`
GCP: C2 Compute-Optimized Machine Family	`cloud.google.com/.../compute-optimized-machines#c2`
GCP: T2D General-Purpose Machine Family	`cloud.google.com/.../general-purpose-machines#t2d`
GCP: Machine Families Resource and Comparison Guide	`cloud.google.com/.../machine-resource`
Azure Ebdsv5 and Ebsv5-Series Virtual Machines	`learn.microsoft.com/.../ebdsv5-ebsv5-series`
Third-Party Sources	
PassMark: ARM Cortex-A72 16 Core 2000 MHz	`cpubenchmark.net/.../id=6436`
PassMark: ARM Neoverse-N1 4 Core	`cpubenchmark.net/.../id=4397`
Geekbench 2: Intel Xeon E5506 @ 2.13 GHz (EC2)	`browser.geekbench.com/geekbench2/compare/1585176`
Geekbench 3: Intel Xeon E5507 @ 2.27 GHz (EC2)	`browser.geekbench.com/v3/cpu/546695`
Geekbench 5: Amazon EC2 t2.2xlarge	`browser.geekbench.com/v5/cpu/16843977`
Geekbench 6: AWS EC2 t2.large (baseline)	`browser.geekbench.com/v6/cpu/compare/2502090`
Geekbench 6: Amazon EC2 t3.2xlarge vs t2.large	`browser.geekbench.com/v6/cpu/compare/9526213`
Geekbench 6: Amazon EC2 t3a.large vs t2.large	`browser.geekbench.com/v6/cpu/compare/3064894`

(continued on next page)

(continued from previous page)

Source Name	URL
VPSBenchmarks: Google Compute Engine CPUs	vpsbenchmarks.com/.../google_compute_engine/cpus
gcloud-compute.com: f1-micro Specifications	gcloud-compute.com/f1-micro.html
emma: AWS EC2 g4dn.xlarge Spec & Pricing Guide	emma.ms/blog/aws-ec2-g4dn-xlarge
CloudOptimo: AWS EC2 m1.large Pricing & Specs	costcalc.cloudoptimo.com/.../m1.large
RunsOn: AWS EC2 m6g Instance Family	go.runs-on.com/instances/ec2/m6g
EC2Instances.info: Easy Amazon EC2 Instance Comparison	ec2instances.info

A.10 Other exogenous state transitions

I have four other time-varying state variables: mean industry wage, mean industry materials price, capital rental rate, and log VM quality. As they are not a focus of my analysis, I assume that the industry-level wage, materials price, and rental rate remain fixed at their 2021 values for all $t > 2021$. In addition, I assume that, beyond the data years, log VM quality follows an exogenous exponential growth rate, as a consequence of Moore's law.[73][74] The regression for log quality has $N = 61$ observations and is given by

$$d\log(q_{jt}) = \underbrace{\rho^q}_{0.0322(0.009)} + \epsilon^q_{jt}$$

[73] I also run a version with provider fixed effects, but these are not statistically significant.

[74] Quality does not perfectly follow the rate of improvements in semiconductors because firms continue to offer VMs based on old vintages.

A.11 Calibrating marginal costs

This appendix contains details for calibrating the marginal costs of supplying a vCPU for one hour and for the network costs of moving one GB of data off platform.

A.11.1 vCPU marginal cost

To model the marginal cost of running a vCPU for one hour, I rely on prior engineering estimates of the variable costs of running cloud computing servers (Mach and Schikuta [2011]). I can write the cost model as follows:

$$
\begin{aligned}
C_{jt}^{vm} &= C_{jt}^{depr} + C_{jt}^{power} \\
\text{where } C_{jt}^{depr} &= \frac{\delta^{VM}}{8760} \cdot \left[\frac{\bar{C}_{jt}^{server}}{\#vCPU} + \frac{\bar{C}_{t}^{SSD}}{\#GB/SDD} \right] \\
\text{and } C_{jt}^{power} &= (0.5 + 0.5\bar{u}_{jt})\bar{k}_{jt}^{max} \cdot P_{jt}^{elec}
\end{aligned}
$$

where δ^{VM} is the annual depreciation rate for computing hardware, calibrated to be 0.26 (Bloom et al. [2012]), which I divided by 8760 to convert to hourly depreciation. Then, $\bar{C}_{jt}^{server}, \bar{C}_{t}^{SSD}$ are the mean server and SSD prices, weighted by usage as described above. I do not observe oligopsony power in the input market, so I instead use the resale price of CPUs and SSDs from Passmark.com as an upper bound on the cost.[75] The expression for C_{jt}^{power} comes from Brand et al. [2024], where $\bar{u}_{jt}$ is the average annual usage rate; $\bar{k}_{jt}^{max}$ is the mean maximum energy consumption, which I proxy using manufacturers' thermal density power (TDP) data from Passmark.com; and P_{jt}^{elec} is the average annual electricity price in kWh, weighted by consumer-level datacenter choices. Electricity prices come from Eurostat (France, Germany, Netherlands, Belgium); DESNZ (United Kingdom); and ElCom (Switzerland). All prices are converted to euros; for each country, I use the largest band grid. In the model, calibrated marginal costs are weighted by firms' choice of datacenter. Results are displayed in Table 19.

[75] I do not observe SSD models used by cloud suppliers, though Igami and Uetake [2020] notes that computer storage is a homogenous good.

A.11.2 Egress marginal cost

The main component of egress fees are the network costs of transferring data. I use the lowest competitive rate in 2022 from Boudreau [2022] of \$0.05/Mbps-month, or approximately 0.00015 euros per GB.

$$C^{egress} = \frac{0.05\text{USD}}{\text{Mbps-month}} \cdot \frac{1}{324} \approx 0.00015 \frac{EUR}{GB}$$

As a robustness check, I also use within-supplier same-region data regress fees, which tend to be on the order of $0.01 \frac{EUR}{GB}$. Results are displayed in Table **[In progress.]**

Table 17: List of Microsoft server products

Category	Product name
Windows Server	Windows NT Server; Windows Server; Microsoft Virtual Server; Hyper-V Server
SQL Server	Microsoft SQL Server ; SQL Server Reporting Services; SQL Server Analysis Services; SQL Server Management Studio; SQL Server Business Intelligence; Transact-SQL
Exchange Server	Microsoft Exchange Server; Exchange Archiving; Exchange Online Protection
SharePoint	Microsoft SharePoint; Microsoft Enterprise Content Management
Unified Communications	Microsoft Lync; Lync Server; Lync Enterprise Voice; Skype for Business
System Center	Microsoft System Center; System Center Operations Manager; Configuration Manager; Data Protection Manager; Endpoint Protection; Virtual Machine Manager; Server Management Tools
Identity and Directory Services	Active Directory Federation Services; Kerberos Authentication; Group Policy Management; Microsoft DirectAccess
Virtualization and Application Virtualization	Microsoft Hyper-V; Microsoft Application Virtualization (App-V)
Enterprise Integration and Middleware	Microsoft BizTalk Server; Microsoft Commerce Server; Enterprise Project Management; Customer Care Framework; Enterprise Application Integration
Hosting, Storage, and Mobility Infrastructure	Microsoft Hosting Services; Web Hosting; Microsoft StorSimple; Storage Management; Microsoft Enterprise Mobility Suite
Security and Threat Management	Microsoft Forefront; Forefront Unified Access Gateway

List of all product names that have appeared it at least one year of Aberdeen's data from 2015-2021. Names have been standardized across survey waves.

Table 19: Marginal cost summary statistics

	Amazon	Google	Microsoft
Marginal cost	0.000963 (0.000603)	0.00148 (0.000241)	0.000979 (0.000255)

Mean costs in euros across datacenter-years are displayed, with standard deviations in parentheses.

B Modeling details appendix

This appendix derives the details of the closed form expressions for the dynamic choice probabilities for cloud computing bundles b_{ft} and non-IT capital K_{ft}.

B.1 Bundle choice details

The functional form of the entry and maintenance cost shocks means I can rewrite firms' value functions as follows. Throughout, I suppress dependence on industry i.

First, denote

$$\sigma^E(K_{ft}) \equiv \sigma^E \cdot D_{ft}(K_{ft}) \tag{36}$$

$$\sigma^b(K_{ft}) \equiv \sigma^b \cdot D_{ft}(K_{ft}) \tag{37}$$

as the effective cost shock incurred by a firm of size K_{ft}. Let the choice specific value functions be defined as follows

$$V_t^{Inc}(s_{ft}) = \max_{I_t} \left\{ \pi_t(s_{ft}) - C^K(I_{ft}, K_{ft}) - D_{ft} \cdot \kappa^M(s_{ft}) - D_{ft} \cdot \kappa^S \mathbb{1}\{b' \neq b_{ft}\} + \beta \mathbb{E}_t[\bar{V}_{t+1}(s_{ft+1})|s_{ft}] \right\} \tag{38}$$

$$V_t^{Ent}(s_{ft}) = \max_{I_{ft}} \left\{ \pi_t(s_{ft}) - C^K(I_{ft}, K_{ft}) - D_{ft} \cdot \kappa_t^E + \beta \mathbb{E}_t[\bar{V}_{t+1}(s_{ft+1})|s_{ft}] \right\} \tag{39}$$

$$V_t^0(s_{ft}) = \max_{I_{ft}} \left\{ \pi_t(s_{ft}) - C^K(I_{ft}, K_{ft}) + \beta \mathbb{E}_t[V_{t+1}(s_{ft+1})|s_{ft}, b_{ft+1} = \varnothing] \right\} \tag{40}$$

where $V_t^{Inc}(s_{ft}), V_t^{Ent}(s_{ft}), V_t^0(s_{ft})$ are the value functions for cloud incumbents, cloud potential entrants, and cloud non-takers in period t. Using these expressions and my assumption about the distribution of (σ^E, σ^b), I can integrate out the shocks to write the firm-specific value functions as follows:

$$V_t^{Inc}(s_{ft}) = \sigma^b(K_{ft}) \log \sum_{b' \in \mathcal{B}} \exp\left(V_t^{Inc}(s_{ft})/\sigma^b\right) + \sigma^b(K_{ft})\tilde{\gamma} \tag{41}$$

$$V_t^{Ent}(s_{ft}) = \sigma^E(K_{ft}) \log \left[\exp(V_t^0(s_{ft})/\sigma^E) + \sum_{b' \in \mathcal{B}/\varnothing} \exp(V_t^{Ent}(s_{ft}, b')/\sigma^E)\right] + \sigma^E(K_{ft})\tilde{\gamma} \tag{42}$$

where $\tilde{\gamma}$ is the Euler-Mascheroni constant. Then the probability of entry in state s_{ft} is given by

$$\mathbb{P}_t[b_{ft+1} \neq \varnothing | b_{ft} = \varnothing, s_{ft}] = \frac{\sum_{b' \in \mathcal{B}/\varnothing} \exp(V_t^{Ent}(s_{ft}, b_{ft+1})/\sigma^E(K_{ft}))}{\exp(V_t^0(s_{ft})/\sigma^E(K_{ft})) + \sum_{b_{ft+1} \in \mathcal{B}/\varnothing} \exp(V_t^{Ent}(s_{ft}, b_{ft+1})/\sigma^E(K_{ft}))} \tag{43}$$

while the conditional probability of choosing a cloud bundle conditional on entry is $b' \in \mathcal{B}$ is given by

$$\mathbb{P}_t[b_{ft+1}|b_{ft}, \tilde{s}_{ft}] = \frac{\exp(\frac{1}{\sigma^B(K_{ft})}\beta\mathbb{E}_{t+1}[V_{t+1}^{Inc}(s_{ft+1})|s_{ft}] - C^K(K_{ft}, I_{ft}) - D_{ft}\cdot\kappa_t^E \mathbb{1}\{b_{ft}=\varnothing\} - D_{ft}\cdot\kappa^S \mathbb{1}\{b_{ft+1}\neq b_{ft}\})}{\exp(\frac{1}{\sigma^B(K_{ft})}\beta\mathbb{E}_{t+1}[V_{t+1}^{0}(s_{ft+1})|s_{ft}] - C^K(K_{ft}, I_{ft})) + \sum_{b_{ft+1}\in\mathcal{B}/\varnothing}\exp(\frac{1}{\sigma^B(K_{ft})}\beta\mathbb{E}_{t+1}[V_{t+1}^{Inc}(s_{ft+1})|s_{ft}] - C^K(K_{ft}, I_{ft}) - D_{ft}\cdot\kappa_t^E \mathbb{1}\{b_{ft}=\varnothing\} - D_{ft}\cdot\kappa^S \mathbb{1}\{b_{ft+1}\neq b_{ft}\})}$$

B.2 Capital investment details

Following Gowrisankaran and Schmidt-Dengler [2024] and suppressing dependence on sector i, let the choice-specific value for selecting some capital stock K^j be

$$v_{ft}^j = -C^K(K^j, K_{ft}) + \beta\mathbb{E}_t[V_{t+1}(s_{ft+1}, \psi_{t+1})|s_{ft}, \psi_t] \tag{44}$$

for the j-th grid of the set of capital values, $\mathcal{K}$. Given a bundle choice b_{ft}, a firm chooses K^j if and only if

$$v_{ft}^j - I_{ft}^j \epsilon_{ft}^K \geqslant v_{ft}^k - I_{ft}^k \epsilon_{ft}^K \tag{45}$$

where

$$I^j_{ft} = K^j - (1-\delta)K_{ft} \tag{46}$$

Monotonicity of $I(\cdot, K_{ft})$ implies that the firm prefers K^j over K^k if and only if

$$\epsilon^K_{ft} \geqslant \frac{v^k_{ft} - v^j_{ft}}{I^k_{ft} - I^j_{ft}} \equiv \epsilon^K_{ft}(j,k) \tag{47}$$

for any $j > k$. The firm chooses K^j if and only if $\underline{\epsilon}_{ft}(j) \leqslant \epsilon^K_{ft} < \bar{\epsilon}_{ft}(j)$, where

$$\underline{\epsilon}_{ft}(j) = \begin{cases} \epsilon^K_{ft}(j, j+1), & j < J \\ -\infty, & j = J \end{cases} \tag{48}$$

$$\bar{\epsilon}_{ft}(j) = \begin{cases} \infty, & j = 1 \\ \epsilon^K_{ft}(j-1, j), & j > 1 \end{cases} \tag{49}$$

Then the conditional probability of choosing capacity K^j in state s_{ft} is given by

$$\mathbb{P}[K_{ft} = K^j | s_{ft}, \psi_t] = \Phi\left(\frac{\bar{\epsilon}_{ft}(j)}{\sigma^K}\right) - \Phi\left(\frac{\underline{\epsilon}_{ft}(j)}{\sigma^K}\right) \tag{50}$$

Finally, I can follow Gowrisankaran and Schmidt-Dengler [2024] and use the properties of the truncated normal distribution to derive the conditional mean of the shock to be

$$\mathbb{E}[\epsilon_{ft} | K_{ft} = K^j, s_{ft}, \psi_t] = \frac{\phi\left(\frac{\underline{\epsilon}_{ft}(j)}{\sigma^K}\right) - \phi\left(\frac{\bar{\epsilon}_{ft}(j)}{\sigma^K}\right)}{\Phi\left(\frac{\bar{\epsilon}_{ft}(j)}{\sigma^K}\right) - \Phi\left(\frac{\underline{\epsilon}_{ft}(j)}{\sigma^K}\right)} \tag{51}$$

B.3 Supplier pricing details

In this section, I derive suppliers' value functions as well as their backwards induction problem. To make the problem computationally tractable, I take analytic derivatives, fitting distributions over agents' states.

First, let a supplier's Bertrand pricing strategy be $(\{P_{jt+1}\}_t, \{E_{jt+1}\}_t) \in \mathbb{R}^t \times \mathbb{R}^t$ and denote E^{obs}_{jt+1} as the observed egress fee schedule. Recall supplier j's vale function is given

by

$$V^s_{ft+1}(\mathbf{s}_{jt}) = \arg \max_{P_{jt+1}, E_{jt+1}} \mathbb{E}_t[(P_{jt+1} - MC^p_{jt+1})Q_{jt+1} + (E_{jt+1} - MC^e_{jt+1})Q^{exit}_{jt+1}|\mathbf{s}_{jt}]$$

Denoting firm f's own state excluding bundle choice as as $\tilde{x}_{ft} = (K_{ft}, \Omega_{ft}, MS_{ft}, server_{ft})$, at any period t, total demand for product j is given by:

$$\begin{aligned} Q_{jt+1} = & \sum_{i \in 1,2,3} N^{\varnothing}_{it} \int_{K_{ft} \times \Omega_{ft}} P^e_{it}(b_{ft+1} \neq \varnothing | K_{ft} \Omega_{ft}, b_{ft} = \varnothing) \\ & \times \mathbb{E}_t \left[\frac{P^b_{it+1}(j \in b_{ft+1} | K_{ft+1}, \Omega_{ft+1}, b_{ft})}{|b_{ft+1}|} \right] dF_i(K_{ft} \times \Omega_{ft} | b_{ft} = \varnothing) \\ & + \sum_{i \in 1,2,3} \sum_{b \in \mathcal{B}/\varnothing} N^b_{it} \int_{K_{ft} \times \Omega_{ft}} \mathbb{E}_t \left[\frac{P^b_{ft+1}(j \in b_{ft+1} | K_{ft+1}, \Omega_{ft+1}, b_{ft})}{|b_{ft+1}|} \right] dF_i(K_{ft} \times \Omega_{ft} | b_{ft}) \end{aligned}$$

where N^b_{it} is the number of firms that are using bundle b in industry i at time t, and we divide shares by the bundle size $|b_{ft+1}|$ because we assume downstream firms use each provider in the bundle equally. Taking first order conditions gives

$$\frac{\partial Q_{jt+1}}{\partial p_{jt+1}} = \sum_{i\in 1,2,3} N_{it}^{\varnothing} \int_{K_{ft}\times\Omega_{ft}} \frac{\partial P_{it}^{e}(b_{ft+1}\neq\varnothing|K_{ft}\Omega_{ft}, b_{ft}=\varnothing)}{\partial p_{jt+1}}$$
$$\times \mathbb{E}_t\left[\frac{P_{it+1}^{b}(j\in b_{ft+1}|K_{ft+1},\Omega_{ft+1},b_{ft})}{|b_{ft+1}|}\right] dF_i(K_{ft}\times\Omega_{ft}|b_{ft}=\varnothing)$$
$$+ \sum_{i\in 1,2,3} N_{it}^{\varnothing} \int_{K_{ft}\times\Omega_{ft}} P_{it}^{e}(b_{ft+1}\neq\varnothing|K_{ft}\Omega_{ft}, b_{ft}=\varnothing)$$
$$\times \mathbb{E}_t\left[\frac{\partial}{\partial p_{ft+1}}\cdot\frac{P_{it+1}^{b}(j\in b_{ft+1}|K_{ft+1},\Omega_{ft+1},b_{ft})}{|b_{ft+1}|}\right] dF_i(K_{ft}\times\Omega_{ft}|b_{ft}=\varnothing)$$
$$+ \sum_{i\in 1,2,3}\sum_{b\in\mathcal{B}/\varnothing} N_{it}^{b} \int_{K_{ft}\times\Omega_{ft}} \mathbb{E}_t\left[\frac{\partial}{\partial p_{ft+1}}\cdot\frac{P_{it+1}^{b}(j\in b_{ft+1}|K_{ft+1},\Omega_{ft+1},b_{ft})}{|b_{ft+1}|}\right] dF_i(K_{ft}\times\Omega_{ft}|b_{ft})$$
$$+ \sum_{i\in 1,2,3} N_{it}^{\varnothing} \int_{K_{ft}\times\Omega_{ft}} P_{it}^{e}(b_{ft+1}\neq\varnothing|K_{ft}\Omega_{ft}, b_{ft}=\varnothing)$$
$$\times \mathbb{E}_t\left[\frac{P_{it+1}^{b}(j\in b_{ft+1}|K_{ft+1},\Omega_{ft+1},b_{ft})}{|b_{ft+1}|}\right] dF_i^{K}(K_{ft}\times\Omega_{ft}|b_{ft}=\varnothing)$$
$$+ \sum_{i\in 1,2,3}\sum_{b\in\mathcal{B}/\varnothing} N_{it}^{b} \int_{K_{ft}\times\Omega_{ft}} \mathbb{E}_t\left[\frac{P_{ft+1}^{b}(j\in b_{ft+1}|K_{ft+1},\Omega_{ft+1},b_{ft})}{|b_{ft+1}|}\right] dF_i^{K}(K_{ft}\times\Omega_{ft}|b_{ft}) = 0$$

where $dF_i^K(K_{ft}\times\Omega_{ft})$ is the derivative of $dF_i(K_{ft}\times\Omega_{ft}$ with respect to K_{ft}.

I first compute the pointwise derivative for each Ω_{ft}, K_{ft}. Observe that because $\epsilon_e \sim GEV1(\sigma_e)$ and $\epsilon_m \sim GEV1(\sigma_m)$, the conditional choice probabilities can be written in logit form as follows

$$P_{it}^{e}(b_{ft+1}\neq\varnothing|\tilde{x}_{ft}, b_{ft}=\varnothing) \equiv$$
$$\frac{\exp(\frac{1}{\sigma_e}(-D_{ft}\kappa_{Et} + \beta\mathbb{E}_t[V_{t+1}(\tilde{x}_{ft+1}, b_{ft+1}\neq\varnothing|\tilde{x}_{ft}, b_{ft})]))}{\exp(\frac{1}{\sigma_e}(\beta\mathbb{E}_t[V_{t+1}(\tilde{x}_{ft+1}, b_{ft+1}=\varnothing|\tilde{x}_{ft}, b_{ft}) + \exp(\frac{1}{\sigma_e}(-D_{ft}\kappa_{Et} + \beta\mathbb{E}_t[V_{t+1}(\tilde{x}_{ft+1}, b_{ft+1}\neq\varnothing|\tilde{x}_{ft}, b_{ft})}$$

$$P_{it}^{b}(b_{ft+1}|\tilde{x}_{ft}, b_{ft}\neq\varnothing) =$$
$$\frac{\exp(\frac{1}{\sigma_m}(-D_{ft}\kappa_{M}(b_{ft}) + \beta\mathbb{E}_t[V_{t+1}(\tilde{x}_{ft+1}, b_{ft+1}|\tilde{x}_{ft}, b_{ft})]))}{\sum_{b\in\mathcal{B}/\varnothing}\exp(\frac{1}{\sigma_m}(-D_{ft}\kappa_{M}(b_{ft}) + \beta\mathbb{E}_t[V_{t+1}(\tilde{x}_{ft+1}, b_{ft+1}|\tilde{x}_{ft}, b_{ft})}$$

Taking the derivatives of each of these, we get

$$\frac{\partial P^e_{it}(b_{ft+1} \neq \varnothing | \tilde{x}_{ft}, b_{ft} = \varnothing)}{\partial p_{jt+1}} =$$
$$- \frac{\alpha\beta}{\sigma_e} \mathbb{E}_t \left[\frac{D_{ft+1}}{|b_{ft+1}|} P^b_{it}(b_{ft+1} | \tilde{x}_{ft}, b_{ft} = \varnothing) \right] \cdot P^e_{it}(b_{ft+1} \neq \varnothing | \tilde{x}_{ft}, b_{ft} = \varnothing)(1 - P^e_{it}(b_{ft+1} \neq \varnothing | \tilde{x}_{ft}, b_{ft} = \varnothing))$$
$$\frac{\partial P^b_{it}(b_{ft} | \tilde{x}_{ft}, b_{ft} \neq \varnothing)}{\partial p_{ft}} =$$
$$- \frac{\alpha D_{ft}}{\sigma_m} P^b_{it}(b_{ft+1} | \tilde{x}_{ft}, b_{ft} \neq \varnothing)(1 - P^b_{ft}(b_{ft+1} | \tilde{x}_{ft}, b_{ft} \neq \varnothing))$$

Next, using estimates K_{ft}, Ω_{ft}, I fit $dF(K_{ft} \times \Omega_{ft})$ for each industry i.

[In progress.]

B.4 A model of capital and data

In this appendix, I show that the log-linear model that represents the relationship between capital investment and storage investment can be recovered up to a constant multiplier.

Suppose, as in Farboodi and Veldkamp [2021], that data is produced as a byproduct of firm f's production according to the following equation:

$$I^D_{ft} = z_i F_i(K_{ft}, L_{ft}, M_{ft})$$

where z_i is an industry level measure of how well a firm can convert production into data.[76] Next, suppose that data investment also follows a convex adjustment cost structure:

$$D_{ft} = p_{Dt} I^D_{ft} + C_D \left(\frac{I^D_{ft}}{D_{ft}} \right)^2 \cdot D_{ft}$$

Given my brief window before cloud computing was widely adopted (2015-2016), I make the following assumption to ensure stationarity:

Assumption 2. *For $t < 2017$, firms assume $\kappa^E_{ft} = \kappa^E_0$ and are thus in a steady state.*

That is, firms do not anticipate the fall in cloud adoption costs until 2017 and thus perceive themselves to be in a steady state (cf. the steady state assumptions in Hornbeck

[76] I am implicitly assuming that a firm purchases exactly enough storage to cover each period's additional data production.

et al. [2024]). Using these assumptions, then, I can derive D_{ft} as a function of known parameters, K_{ft}, I_{ft}, I_{ft}^D and input prices.

[In progress.]

C Estimation appendix

In this appendix I provide further details on how I estimate parameters. First, I provide details of the moment conditions identifying my industry-level production functions. Next, I describe estimation of my remaining parameters, detailing the MME algorithm and MSM procedure. Finally, I expand on how I computed the estimates.

C.1 Production function estimation details

[In progress.]

C.2 Dynamic parameter estimation details

In this appendix, I provide details on estimating the dynamic choice parameters. I use the Markov moment equilibrium (MME) (Ifrach and Weintraub [2017]) approximation to an MPE for supplier prices, and estimate the parameters via the method of simulated moments (MSM) (McFadden [1989], Pakes and Pollard [1989]).

Algorithm. The algorithm for computing an MME is similar to standard value function iteration for consumers and Bertrand pricing for suppliers, but with an added outer loop. First, I guess a candidate transition path of prices $\hat{M}^\phi$. In the inner loop, I solve the steady state value functions for consumers at some period T in the future, using prices at p_T as an approximation. I then compute values along the transition path using backwards induction. Finally, I simulate prices forward and solve for consumers' best responses and suppliers' Bertrand pricing strategies. I then fit an AR(1) model over the path of each supplier j's price path, $\hat{M}^\phi$. The algorithm repeats until convergence of $\hat{M}_\phi$.

Formally, denote decision rules for downstream firms and suppliers as follows:

$$a_{ft} = \psi_t^d(s_{ft}, \sigma_{ft}^K, \sigma_{ft}^E, \sigma_{ft}^b)$$
$$p_{jt} = \psi_t^s(s_t)$$

such that the year-t strategy profile can be written as

$$\psi^t = \{\psi_t^d(s_{ft}, \sigma_{ft}^K, \sigma_{ft}^E, \sigma_{ft}^b)\}_{f \in \mathcal{F}} \cup \{\psi_t^s(s_t)\}_{j \in \{1,2,3\}}$$

Then the perceived transition kernel for the is given by

$$\hat{M}_t^\psi = \Phi M_t^\psi \tag{52}$$

where $\hat{M}_t^\psi$ is the transition kernel of the moment-based state when agents use strategy ψ_t, M_t^ψ is the true transition kernel, and Φ is an operator such that Equation **?** holds. I approximate M_t^ψ as

$$\begin{bmatrix} P_{Amazon,t+1} \\ P_{Google,t+1} \\ P_{Microsoft,t+1} \end{bmatrix} = \begin{bmatrix} \rho_0^{Amazon} \\ \rho_0^{Google} \\ \rho_0^{Microsoft} \end{bmatrix} + \begin{bmatrix} \rho_0^{Amazon} P_{Amazon,t} & 0 & 0 \\ 0 & \rho_0^{Google} P_{Google,t} & 0 \\ 0 & 0 & \rho_0^{Microsoft} P_{Microsoft,t} \end{bmatrix}$$

The MME algorithm is detailed below. For each candidate parameter vector $\hat{\boldsymbol{\theta}}$, run:

1. Initialize $V_{ft}(s_{ft})$ for all $s_{ft} \in \mathcal{S}$.
2. Run inner loop:
   (a) Solve for terminal period value functions $V_T(s_{fT})$ of downstream firms at the terminal T for all states s_{fT}.
   (b) Compute value functions along the transition path $t = 0, .., T-1$ by backwards induction.
   (c) Initialize the distribution of states in $t = 0$, s_0, using data.

(d) Simulate the industry forward and compute supplier and downstream firm decision rules at each state to obtain ψ^t, $\{\hat{P}_{jt}\}_{j\in\{1,2,3\},t\in\mathcal{T}}$, and therefore $\hat{M}_\psi$.

3. Calculate the updated value functions $\tilde{V}_{ft}(s_{ft})$ and new equilibrium decision rules ψ^t.

4. Re-run the inner loop and obtain an updated $\hat{M}_\psi$.

5. Compute the norm $||\{\frac{\hat{M}_\psi - M_\psi}{M_\psi}\}||$, where $||\cdot||$ is the sup norm.

6. If the norm is greater than $\epsilon = 0.001$, repeat steps 2-5. Else terminate.

Optimization. Let $\boldsymbol{\theta} = [\kappa_0^E, \kappa_1^E, \kappa^S, \boldsymbol{\alpha}_{0:4}, \xi_{Amazon}, \xi_{Microsoft}, \sigma^E, \sigma^b, C_1^K, \sigma^K, \zeta]'$. I minimize a minimum distance objective function:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left((\boldsymbol{\Psi}(\boldsymbol{\theta}) - \hat{\boldsymbol{\Psi}})' \mathbf{W} (\boldsymbol{\Psi}(\boldsymbol{\theta}) - \hat{\boldsymbol{\Psi}}) \right) \tag{53}$$

The weighting matrix $\mathbf{W}$ is a diagonal weighting matrix in which the weights are chosen such that moments of different sizes are given roughly equal weight in the optimizer. In particular, moments above one receive a weight of 0.1, while all other moments receive a weight of 1. Observed statistics are averages over firm-years, in which sector level weights are given by the number of firms in each sector-year in the matched data, with the exception of entry costs, which are averaged across firms within each year.

C.3 Computational details

In this appendix I provide details on the simulation for both the structural parameter estimation and counterfactual scenarios.[77] The estimation is complex, requiring thousands of nested fixed point iterations over firm value functions as well as guesses of the price paths. Even supplying analytical derivatives for suppliers and modeling simple parametric price forecasts, the optimizer often fails to converge far from the true parameter values. To remedy this, I follow Gortmaker [2024] and begin my point estimates using Bayesian optimization, a derivative-free global approach that can handle well noisy objectives. I use the `bayesopt` function in MATLAB, which constructs a surrogate Gaussian Process model from

[77] The author benefited from Claude Opus 4.0 through 4.8 and Claude Fable 5 to optimize code.

the evaluated objective and uses an acquisition function to select the next best parameter values, balancing exploitation and exploration. I use the expected improvement plus acquisition function. I find that an active set size of 100 and exploration ratio of 0.08 work well for my optimizer. I use the settings from Gortmaker [2024]: 10 sets of 50 random initial evaluations within wide parameter bounds, followed by 150 Bayesian iterations for each sets. Failure rates reach up to 40% across initial guesses, though are far lower near the global optimum.[78]

After global optimization, I fine-tune with a round of local optimization starting from the parameters with the smallest objective across Bayesian evaluations, using the trust-based local refinement algorithm `fmincon` and the SQP algorithm. As an additional guard I apply a feasibility constraint that analyzes whether the MME loop has failed to converge:

$$\left| \frac{||\hat{M}_\psi - M_\psi||}{\epsilon^{tol}} \right| - 1 \leqslant 10^{-5}$$

Near the global optimum the constraint is never binding. For each boostrap iteration, I begin with local optimization, warm starting each iteration at the point estimate, switching from central to forward differences in the SQP algorithm in the interest of tractability.

D Results appendix

D.1 Difference-in-differences regression on cloud spot prices

To provide evidence about whether or not egress costs matter for firm switching in the short run, I leverage the random timing of the major firms' decisions to drop egress costs on the spot market price of virtual machines. In particular, Google announced it would drop fees for exiting consumers worldwide on January 11, 2024, while Amazon and Microsoft announced their fee eliminations on March 5 and 13, 2024, respectively, though their fee eliminations were restricted to exiting consumers. To estimate the short run effects of an egress fee elimination on demand for virtual machines, I run a difference-in-differences specification for

[78] If any parameter guess produces at least 10 supplier-years at which any supplier j's markup is $> €50$, I terminate the guess at outer iteration 40 and return an objective of `NaN`.

Variable	Model 1 (54 days)	Model 2 (54 days)	Model 3 (30 days)	Model 4 (30 days)
Treatment	-0.0255 (0.0243)	-0.0520 (0.0266)†	0.0262 (0.0235)	0.0048 (0.0280)
Treatment × EU		0.0895 (0.0319)**		0.0726 (0.0449)
N. obs	4,639,571	4,639,571	2,574,119	2,574,119
Adj. R^2	0.313944	0.313944	0.624046	0.624046

Table 20: Effect of egress fee elimination on VM spot prices

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run with product, provider, time, and region fixed effects and use heteroskedasticity-robust standard errors.

the 54 days surrounding Google's announcement. In the short run, if egress fees are a barrier to switching, we should expect to see a fall in the spot market price relative to the list price as consumers exit. I compute the discount offered by the spot market compared to the list price for virtual machine instance p on day d as:

$$\text{Savings}_{pd} = \frac{P_{pd}^{spot}}{P_{pd}^{list}} \times 100\% \tag{54}$$

I use Google's announcement as a treatment and run the following difference-in-differences specification:

$$\text{Savings}_{pd} = \sum_{d \neq -1} \beta \cdot \mathbb{1}\{j = \text{Google}\} \cdot \mathbb{1}\{t \geqslant \text{Jan 11, 2024}\} + \alpha_p + \beta_j + \tau_d + \gamma_r + \epsilon_{pd} \tag{55}$$

where $\mathbb{1}\{j = \text{Google}\}$ is an indicator; and $\alpha_p, \beta_j, \tau_d$, and γ_r are product, provider, region r, and time fixed effects, respectively. I use the companion dataset from Lee et al. [2022] and average spot price discounts at the daily level.[79] I run a specification for the full data ($\pm$ 54 days around treatment) as well as a shorter window ($\pm$ 30 days around treatment). Results are presented in Table 20 below. I run the basic regression as well as a regression in which I interact the treatment with an indicator that takes the value 1 if the datacenter is located in the European Union. All regressions produce null results, with the exception of model 2, which shows that savings rise after egress fees are eliminated, but only in the EU.

[79]The data for Amazon contains data at the region and availability zone (AZ) level. Since AZ names are anonymized for each customer and I only observe data for a single customer, I also average across AZ.

D.2 Additional production function estimates

This appendix presents a number of robustness checks to my baseline production function. The first estimates a translog production function, the second allows the productivity impact of IaaS to vary over time, the third re-defines IaaS-using firms as those in which Aberdeen has relatively high confidence of use, and the fourth uses total capital investment (non-IT + IT) as the proxy variable in the first stage.

D.2.1 Translog production function

My baseline model uses a Cobb-Douglas production function in order to allow me to partial out labor and materials from the firm's period profit function, dramatically simplifying the state space. To test the impact of a more flexible production function on my estimates, I estimate the following translog specification using the control function approach in Ackerberg et al. [2015]:

$$y_{ft} = \beta_0 + \beta_l l_{ft} + \beta_k k_{ft} + \beta_m m_{ft} + \tag{56}$$

$$\beta_{ll} l_{ft}^2 + \beta_{kk} k_{ft}^2 + \beta_{mm} m_{ft}^2 + \tag{57}$$

$$\beta_{lk} l_{ft} k_{ft} + \beta_{km} k_{ft} m_{ft} + \beta_{lm} l_{ft} m_{ft} + \epsilon_{ft} \tag{58}$$

[In progress.]

D.2.2 Learning-by-doing and productivity

Next, I allow for the impact of cloud adoption to vary over time, to model learning by doing effects (Brand et al. [2024], cf. Benkard [2000]). My modified productivity process is as follows:

$$\omega_{ft} = \gamma_0 + \gamma_1 \omega_{ft-1} + + \gamma_2 \omega_{ft-1}^2 + + \gamma_3 \omega_{ft-1}^3 \tag{59}$$

$$+ \gamma_\tau \tau_{IaaS} + \gamma_5 \cdot \mathbb{1}\{t = 2020\} + v_{ft} \tag{60}$$

where τ_{IaaS} is the number of years a firm has used cloud computing.[80] Results are presented in Tables 21 and 22. In both manufacturing and services, the variable is positive and significant, indicating that the productivity effects of cloud computing are rising with additional use, perhaps due to additional adjustment frictions (North et al. [2019], Asker et al. [2014]) or learning-by-doing effects (Brand et al. [2024], Humlum [2022]).

Table 21: Production function: Learning by doing

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Labor (β_l)	0.129 (0.081)	0.307 (0.048)***	0.231 (0.036)***
Materials (β_m)	0.583 (0.031)***	0.477 (0.054)***	0.720 (0.184)***
Capital (β_k)	0.092 (0.028)**	0.079 (0.010)***	0.043 (0.008)***
Wage (ϑ_w)	0.148 (0.064)*	0.310 (0.046)***	0.224 (0.084)**
Observations	14,893	37,948	153,230
Firms	4,877	14,491	70,321

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] controls for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

Table 22: Productivity evolution: Learning by doing

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Productivity intercept (γ_0)	0.163 (0.021)***	0.157 (0.020)***	0.243 (0.007)***
Linear AR(1) term (γ_1)	0.722 (0.047)***	0.680 (0.048)***	0.424 (0.017)***
Quadratic AR(1) term (γ_2)	0.159 (0.033)***	0.234 (0.038)***	0.432 (0.014)***
Cubic AR(1) term (γ_3)	-0.031 (0.008)***	-0.059 (0.010)***	-0.110 (0.003)***
Cloud years (τ_{IaaS})	0.002 (0.0003)***	0.001 (0.0002)***	0.0009 (0.0001)***
COVID-19 shock (γ_5)	-0.052 (0.001)***	-0.053 (0.001)***	-0.028 (0.004)***
Observations	14,893	37,948	153,230
Firms	4,877	14,491	70,321

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] controls for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

[80] Compare to the exporting variable in de Loecker [2013].

D.2.3 Sample restriction to high-confidence IaaS firms

In this appendix, to ensure robustness to error in the data collection process, I reclassify IaaS-suing firms as only those in which Aberdeen has the highest confidence. From 2017-2019, Aberdeen includes a "Cloud computing purchase likelihood score," which is an ordinal variable from 1-5 indicating the confidence level at which Aberdeen believes an establishment uses cloud computing, with 5 being most likely. From 2020-2021, Aberdeen includes a "product confidence ranking" that serves the same function, but is an ordinal variable from 1-3 at the product level (e.g. Amazon AWS).

I re-run my baseline ACF model in which I categorize as IaaS users only firms with at least one establishment for which the confidence score is $\geqslant 4$ for 2017-2019 or 3 for 2020-2021.

[In progress.]

D.2.4 Total investment proxy variable

Aberdeen estimates the value of IT investments in euros based on its computation of the value of PCs, servers, and storage devices reported by firms in interviews. To account for potential measurement error (Tambe and Hitt [2012], Brynjolfsson and Hitt [2003]), I re-run my baseline regressions but use total capital investment as my proxy variable for the first stage, as is standard in many IO applications (Ackerberg et al. [2015], Olley and Pakes [1996]). Results are presented in Tables 23 and 24. Even though median IT investment is 19.1%, 16.4%, and 18.9% of total investment in manufacturing, services, and high-tech manufacturing, respectively, differences from the baseline model are minimal.

D.2.5 Including exporting and market share quality sifters

Existing work in industrial organization and trade finds significant productivity effects from firms choosing to export (de Loecker [2013]). In addition, De Loecker et al. [2016] find that firm-level market share to be correlated with output quality. To model this I re-estimate my model and modify the first stage of Ackerberg et al. [2015] as follows:

$$\tilde{\omega}_{ft}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = \hat{\phi}_{ft} - f(\mathbf{x}_{ft}; \boldsymbol{\beta}) - H(p^l_{ft}, ms_{\tilde{i}ft}, EXP_{ft}; \boldsymbol{\vartheta}) \tag{61}$$

Table 23: Production function: Total investment proxy

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Labor (β_l)	0.191 (0.083)*	0.290 (0.047)***	0.343 (0.044)***
Materials (β_m)	0.608 (0.031)***	0.494 (0.052)***	0.686 (0.263)**
Capital (β_k)	0.064 (0.020)**	0.080 (0.010)***	0.012 (0.036)
Wage (ϑ_w)	0.214 (0.056)***	0.300 (0.044)***	0.048 (0.122)
Observations	15,903	46,631	164,500
Firms	4,974	14,995	73,127

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] controls for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

Table 24: Productivity evolution: Total investment proxy

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Productivity intercept (γ_0)	0.177 (0.028)***	0.191 (0.018)***	-0.013 (0.001)***
Linear AR(1) term (γ_1)	0.661 (0.007)***	0.589 (0.045)***	-0.877 (0.002)***
Quadratic AR(1) term (γ_2)	0.223 (0.052)***	0.308 (0.035)***	0.160 (0.004)***
Cubic AR(1) term (γ_3)	-0.052 (0.013)***	-0.079 (0.001)***	-0.137 (0.004)***
Cloud adoption (γ_4)	0.004 (0.001)***	0.004 (0.001)***	0.002 (0.001)***
COVID-19 shock (γ_5)	-0.045 (0.001)***	-0.049 (0.001)***	-0.039 (0.001)***
Observations	15,903	46,631	164,500
Firms	4,974	14,995	73,127

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] controls for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

where $ms_{\tilde{i}ft} \in [0,1]$ is firm f's market share in period t in industry $\tilde{i}$ measured at the 4-digit NACE code level and $EXP_{ft} \in \{0,1\}$ is an indicator for whether a firm f has positive export revenue in year t. Market shares are defined on the full Orbis sample[81], and export data comes from Orbis. Results for the production function are presented in Table 25 and for the productivity process in Table 26.

[81] Missing revenue in Orbis is supplemented with Aberdeen data.

Table 25: Production function: Additional quality controls

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Labor (β_l)	0.246 (0.039)***	0.205 (0.031)***	0.162 (0.070)*
Materials (β_m)	0.629 (0.033)***	0.671 (0.051)***	0.559 (0.052)***
Capital (β_k)	0.039 (0.019)*	0.024 (0.018)	0.087 (0.028)**
Wage (ϑ_w)	0.212 (0.019)***	-0.090 (0.134)	0.247 (0.152)
Market share (ϑ_{ms})	0.268 (0.182)	0.471 (0.145)**	0.391 (0.152)*
Exporting (ϑ_{EXP})	1.367 (0.423)**	0.782 (0.389)*	1.470 (0.495)**
Observations	14,883	37,943	153,195
Firms	4,876	14,489	70,319

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] controls for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

Table 26: Productivity evolution: Additional quality controls

Variable	High-tech manufacturing	Low-tech manufacturing	Services
Productivity intercept (γ_0)	-0.011 (0.013)	-0.130 (0.003)***	0.215 (0.003)***
Linear AR(1) term (γ_1)	0.736 (0.020)***	0.733 (0.015)***	0.712 (0.009)***
Quadratic AR(1) term (γ_2)	-0.069 (0.029)*	-0.021 (0.035)	0.017 (0.012)
Cubic AR(1) term (γ_3)	-0.020 (0.036)***	-0.024 (0.024)	0.016 (0.004)
Cloud adoption (γ_4)	0.005 (0.006)	0.005 (0.002)†	0.003 (0.002)
COVID-19 shock (γ_5)	-0.046 (0.010)***	-0.042 (0.004)***	-0.038 (0.004)***
Observations	14,883	37,943	153,195
Firms	4,876	14,489	70,319

$^{***}p < 0$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Note: Regressions are run using the method from Ackerberg et al. [2015] with the de Loecker [2013] correction for cloud adoption in productivity and De Loecker et al. [2016] controls for input quality. Standard errors are from 200 block bootstrap samples at the firm level.

D.2.6 Cost distribution for remaining sectors

In this appendix, I present the estimates corresponding to Table 10 for the services and high-tech manufacturing sectors.

D.3 Comparison to observed productivity gains

In this appendix, I use the estimates from Forrester reports in North et al. [2019] to perform a back of the envelope calculation of the real-world adoption costs and productivity gains

Maintenance costs					
Bundle	Minimum	10th Percentile	Median	90th Percentile	Maximum
Azure	0.0913	0.1228	0.2514	0.4984	1.1493
Amazon	0.1981	0.2666	0.5456	1.0818	2.4946
Google	0.2420	0.3256	0.6664	1.3213	3.0471
Entry costs					
Year	Minimum	10th Percentile	Median	90th Percentile	Maximum
2017	1.3433	1.8074	3.6992	7.3345	16.9141
2021	0.1650	0.2220	0.4544	0.9010	2.0779

Table 27: Services sector cloud cost distribution

Note: Costs computed for firms used in the simulation, in millions of real euros. Bundle prices and qualities are computed using 2017 values.

Maintenance costs					
Bundle	Minimum	10th Percentile	Median	90th Percentile	Maximum
Azure	0.0699	0.1602	0.3811	1.2578	3.3447
Amazon	0.1516	0.3476	0.8272	2.7300	7.2595
Google	0.1852	0.4246	1.0104	3.3346	8.8672
Entry costs					
Year	Minimum	10th Percentile	Median	90th Percentile	Maximum
2017	1.0281	2.3570	5.6088	18.5101	49.2211
2021	0.1263	0.2896	0.6890	2.2739	6.0468

Table 28: High-tech manufacturing sector cloud cost distribution

Note: Costs computed for firms used in the simulation, in millions of real euros. Bundle prices and qualities are computed using 2017 values.

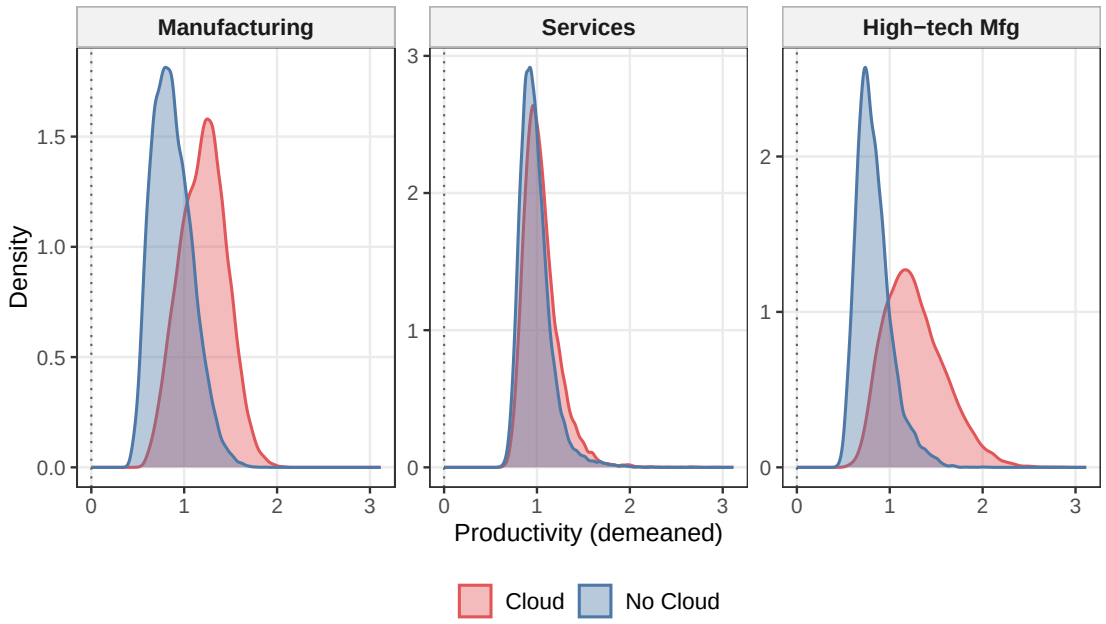


Figure 13: Steady state productivity distribution (0 vs. 3 suppliers)

from adopting IaaS and compare them to my estimates.

[**In progress.**]

D.4 Steady state productivity distribution for multihoming firms

Figure 13 compares the steady state productivity distribution for each sector, comparing non-adopting firms with those that have adopted all 3 cloud products.

E Further institutional details

This appendix provides further description of the cloud computing services market in France. I first provide details on the IaaS market. Then, I provide further details on the relevant provisions of the Data Act.

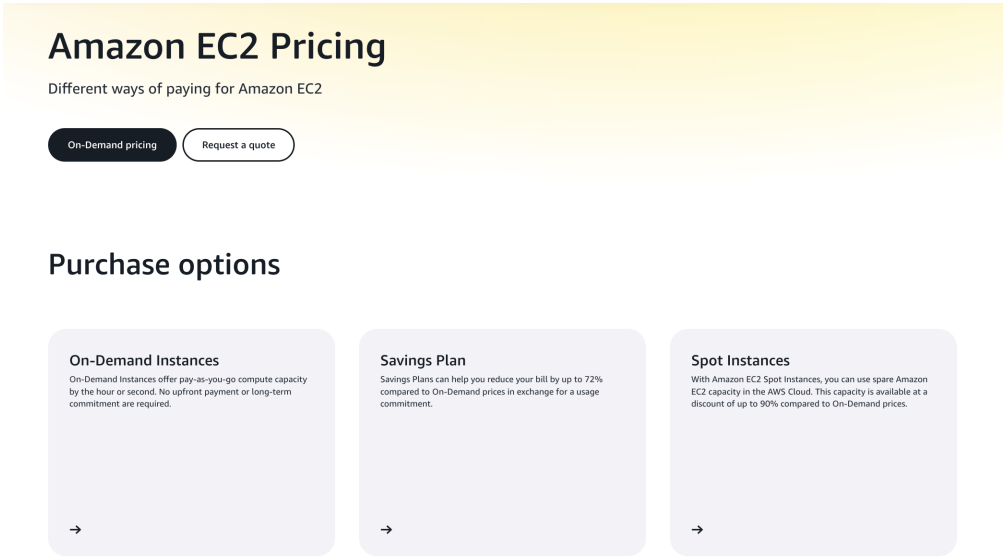
E.1 The Infrastructure-as-a-Service market

Cloud providers offer three types of cloud services: software-as-a-service (SaaS), infrastructure-as-a-service (IaaS), and platform-as-a-service (PaaS). SaaS is the broadest term and encom-

passes various software leasing, including including Google Docs, Microsoft Office 365, and Dropbox. While many of these services are IaaS bundled with a software interface, their contracts are generally highly varied and encompass bundling rental of IT hardware with software licenses.[82] PaaS also abstracts away from the underlying hardware by focusing on an application layer such as serverless computing or containers. I ignore this segment in this study for two reasons. First, I do not observe the specific bundling of software and hardware that firms in our sample purchase, limiting our ability to study pricing. In addition, evidence suggests that IaaS is more common among consumers.[83]

Finally, IaaS is the leasing of storage (usually as "blobs") or computation in the form of virtual machines (VMs). Users have a wide array of options for these - for the former, this includes data speeds, and for the latter this includes such factors as CPU clock speed, number of cores, and memory size. We focus on IaaS for two reasons. First, the economic mechanisms are clear - IaaS has the potential to lower the cost of computation through scale economies and allows firms to dynamically respond to productivity shocks, lowering IT capital adjustment frictions (Asker et al. [2014]). Second, the set of pricing options are fairly simple. For computation, firms can purchase on-demand instances, which are charged by the minute, or reserve computation for a period of 1 or 3 years for a reduced per unit price. Finally, firms can also purchase additional compute on the spot market, with the stipulation that the VM will be shut down if the firm is outbid for it.[84] In the data, most firm computation is done via on-demand prices, with a substantial minority of firms using reserved instances. A 2019 Flexera report found that 47% of AWS, 23% of Azure users, and 10% of Google Cloud users utilized reserved instances (Flexera [2019]). For storage, firms can purchase standard storage, with marginal price declining by volume, or inexpensive long-term storage that costs an additional amount for retrieval. The relative simplicity of IaaS pricing allows me to estimate firms' price elasticity of demand for cloud inputs, even though I do not observe firm-level prices in my data.

[82]Indeed, Dropbox used Amazon's services until 2020, when it switched to purchasing its own data centers.
[83]See Brand et al. [2024] Appendix A. According to the CMA, in 2022, Azure's top product was compute, or IaaS virtual machines (Competition and Markets Authority [2025a], 17).
[84]Though some larger firms such as Netflix sign contracts with individual cloud providers, discussions with a social scientist working with a large cloud provider suggest such contracts are rare.



(a) Pricing options for VMs

Compute Savings Plans for Amazon EC2

Compute Savings Plans apply to EC2 Instance usage regardless of instance family, size, AZ, AWS Region, OS or tenancy.

Select a location type and region

Location Type: AWS Region — Region: Europe (Paris)

Select terms for your Compute Savings Plans

Term length: 1 year — Payment options: All Upfront

Select an operating system and tenancy to view rates

Operating system: Linux — Tenancy: Shared

Viewing 448 available instances

1 2 **3** 4 5 6 7 ... 23

Instance name	Savings Plans rate	Savings over On-Demand	On-Demand rate	Region	Operating system	Tenancy
m7gd.2xlarge	$0.3755	25%	$0.499	Europe (Paris)	Linux	Shared
m7gd.4xlarge	$0.7509	25%	$0.9979	Europe (Paris)	Linux	Shared
m7gd.8xlarge	$1.5018	25%	$1.9958	Europe (Paris)	Linux	Shared
m7gd.12xlarge	$2.2527	25%	$2.9938	Europe (Paris)	Linux	Shared
m7gd.16xlarge	$3.0036	25%	$3.9917	Europe (Paris)	Linux	Shared
m7gd.metal	$3.0036	25%	$3.9917	Europe (Paris)	Linux	Shared
m7i.large	$0.08324	29%	$0.1176	Europe (Paris)	Linux	Shared

(b) A sample of VM prices

Figure 14: Pricing options for Amazon Web Services

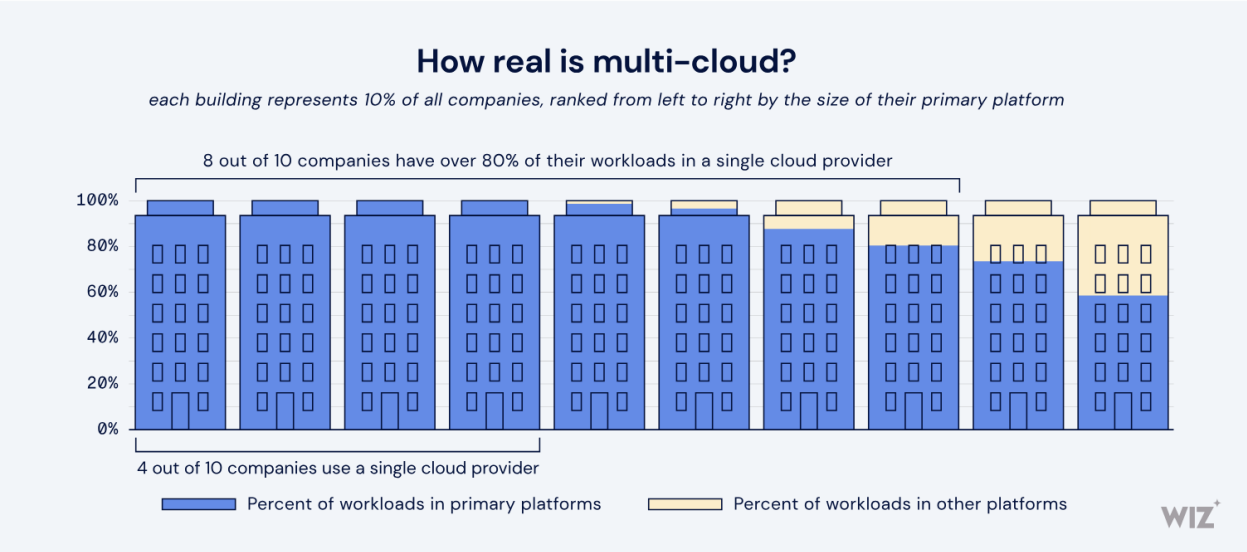


Figure 15: Distribution of data stored in the cloud among Wiz's customers in 2023.

E.2 The EU Data Act

In progress.

Figure 15 provides external evidence that cloud computing users tend to use a primary provider by showing the deciles of the data distribution for the customers of Wiz, a large cloud security firm.

References

Aberdeen Group. *Aberdeen Technology Data Cloud: Data Dictionary*. San Diego, CA and Les Ulis, France, revision 3.14 edition, 2017. URL `https://www.aberdeenservices.com`. © 2017 Aberdeen Group. All rights reserved.

Daniel A Ackerberg, Kevin Caves, and Garth Frazer. Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451, 2015.

Victor Aguirregabiria and Junichi Suzuki. Identification and counterfactuals in dynamic models of market entry and exit. *Quantitative Marketing and Economics*, 12(3):267–304, 2014.

Amazon Web Services. Price reduction — aws news blog. Web page, blog category,

2025. URL https://aws.amazon.com/blogs/aws/category/price-reduction/. Accessed: 2025-09-25.

Amazon Web Services. Rules for specifying cpu options for an amazon ec2 instance. https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-cpu-options-rules.html, n.d. Accessed: 2026-04-17.

Peter Arcidiacono and Robert A Miller. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867, 2011.

John Asker, Allan Collard-Wexler, and Jan De Loecker. Dynamic inputs and resource (mis) allocation. *Journal of Political Economy*, 122(5):1013–1063, 2014.

Bee Yan Aw, Mark J Roberts, and Daniel Yi Xu. R&d investment, exporting, and productivity dynamics. *American Economic Review*, 101(4):1312–1344, 2011.

Pablo D Azar. Moore's law and economic growth. *FRB of New York Staff Report*, (970), 2021.

Kevin Barefoot, Dave Curtis, William Jolliff, Jessica R Nicholson, Robert Omohundro, et al. Defining and measuring the digital economy. *US Department of Commerce Bureau of Economic Analysis, Washington, DC*, 15:210, 2018.

C Lanier Benkard. Learning and forgetting: The dynamics of aircraft production. *American Economic Review*, 90(4):1034–1054, 2000.

Tamay Besiroglu, Nicholas Emery-Xu, and Neil Thompson. Economic impacts of ai-augmented r&d. *Research Policy*, 53(7):105037, 2024.

Nicholas Bloom. The impact of uncertainty shocks. *econometrica*, 77(3):623–685, 2009.

Nicholas Bloom and Nicola Pierri. Cloud computing is helping smaller, newer firms compete. *Harvard Business Review*, 94(4), 2018.

Nicholas Bloom, Raffaella Sadun, and John Van Reenen. Americans do it better: Us multinationals and the productivity miracle. *American Economic Review*, 102(1):167–201, 2012.

Nicholas Bloom, Mirko Draca, and John Van Reenen. Trade induced technical change? the impact of chinese imports on innovation, it and productivity. *The review of economic studies*, 83(1):87–117, 2016.

Brianna Boudreau. Price erosion remains the universal norm, October 2022. URL `https://resources.telegeography.com/price-erosion-remains-the-universal-norm`.

James M Brand, Mert Demirer, Connor Finucane, and Avner A Kreps. Firm productivity and learning in the digital economy: Evidence from cloud computing. Technical report, National Bureau of Economic Research, 2024.

Timothy F Bresnahan. The oligopoly solution concept is identified. *Economics letters*, 10 (1-2):87–92, 1982.

Timothy F Bresnahan and Manuel Trajtenberg. General purpose technologies 'engines of growth'? *Journal of econometrics*, 65(1):83–108, 1995.

Erik Brynjolfsson and Lorin Hitt. Paradox lost? firm-level evidence on the returns to information systems spending. *Management science*, 42(4):541–558, 1996.

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

Erik Brynjolfsson, Daniel Rock, and Chad Syverson. The productivity j-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1):333–372, 2021.

David Byrne, Carol Corrado, and Daniel E Sichel. The rise of cloud computing: minding your p's, q's and k's. Technical report, National Bureau of Economic Research, 2018.

Varadarajan V Chari and Hugo Hopenhayn. Vintage human capital, growth, and the diffusion of new technology. *Journal of political Economy*, 99(6):1142–1165, 1991.

Allan Collard-Wexler and Jan De Loecker. Reallocation and technology: Evidence from the us steel industry. *American Economic Review*, 105(1):131–171, 2015.

Competition and Markets Authority. Appendix a: Demand for cloud services and how customers purchase cloud services. Appendix to final report, cloud services market investigation, UK Competition and Markets Authority, January 2025a. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1143160/Appendix_A__Demand_for_cloud_services_and_how_customers_purchase_cloud_services.pdf. Appendix A to the CMA's final decision on the market investigation into public cloud infrastructure services.

Competition and Markets Authority. Appendix i: Prevalence of multi-cloud. Appendix to final decision report, cloud services market investigation, UK Competition and Markets Authority, jul 2025b. URL https://assets.publishing.service.gov.uk/media/688b80af5612925da6a94da9/Appendix_I_-_Prevalence_of_multi-cloud.pdf. Appendix I — Analysis of providers' and customers' data on how common multi-cloud usage is.

Competition and Markets Authority. Appendix m: Egress fees – analysis of customers' cost scenarios. Appendix to final decision report, cloud services market investigation, UK Competition and Markets Authority, August 2025c. URL https://assets.publishing.service.gov.uk/media/688b80fdd8da16bcb8469599/Appendix_M_-_Analysis_of_customers%27_cost_scenarios.pdf. Appendix M — Egress fees: analysis of customers' cost scenarios.

Competition and Markets Authority. Appendix o: Customer views on egress fees. Appendix to final decision report, cloud services market investigation, UK Competition and Markets Authority, August 2025d. URL https://assets.publishing.service.gov.uk/media/688b817cfc784fa12a089072/Appendix_O_-_Customer_views_on_egress_fees.pdf. Appendix O — Customer views on egress fees.

Competition and Markets Authority. Appendix q: Egress fees — analysis of cloud providers' cost of egress. Appendix to final decision report, cloud services market investigation, UK Competition and Markets Authority, August 2025e. URL https://assets.publishing.service.gov.uk/media/688b81a7d8da16bcb8469525/

Appendix_Q_-_Analysis_of_cloud_providers__cost_of_egress.pdf. Appendix Q — Egress fees: analysis of cloud providers' cost of egress.

Competition and Markets Authority. Appendix t: Licensing analysis. Appendix to final decision report, cloud services market investigation, UK Competition and Markets Authority, August 2025f. URL https://assets.publishing.service.gov.uk/media/688b81a0d8da16bcb8469522/Appendix_T_-_Licensing_analysis.pdf. Appendix T to the final decision report on the Cloud Services Market Investigation.

Competition and Markets Authority. Appendix w: Remedies not taken forward in this market investigation. Appendix to final decision report, cloud services market investigation, UK Competition and Markets Authority, jul 2025g. URL https://assets.publishing.service.gov.uk/media/67976cc8cfd3deafa04fde60/Appendix_W_3.pdf. Remedies Appendix — potential remedies considered but not pursued under CMA powers.

Competition and Markets Authority. Chapter 3: Competitive landscape. In *Cloud services market investigation: Final decision report*, report chapter 3. January 2025h. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1143159/final-decision-report.pdf. Final Decision Report of the CMA.

Russell W Cooper and John C Haltiwanger. On the nature of capital adjustment costs. *The Review of Economic Studies*, 73(3):611–633, 2006.

Eli Cortez, Anand Bonde, Alexandre Muzio, Mark Russinovich, Marcus Fontoura, and Ricardo Bianchini. Resource central: Understanding and predicting workloads for improved resource management in large cloud platforms. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 153–167, 2017.

Datacenters.com. Search data centers: Colocation, bare metal, cloud. Online database / marketplace, 2025. URL https://www.datacenters.com/. Accessed: 2025-11-05.

Jan de Loecker. Detecting learning by exporting. *American Economic Journal: Microeconomics*, 5(3):1–21, 2013.

Jan De Loecker, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik. Prices, markups, and trade reform. *Econometrica*, 84(2):445–510, 2016.

Maarten De Ridder. Market power and innovation in the intangible economy. *American Economic Review*, 114(1):199–251, 2024.

Mert Demirer, Diego J Jiménez Hernández, Dean Li, and Sida Peng. Data, privacy laws and firm production: Evidence from the gdpr. Technical report, National Bureau of Economic Research, 2024.

Mert Demirer, Andrey Fradkin, Nadav Tadelis, and Sida Peng. The emerging market for intelligence: Pricing, supply, and demand for llms. Technical report, National Bureau of Economic Research, 2025.

Timothy DeStefano, Richard Kneller, and Jonathan Timmis. Cloud computing and firm growth. *Review of Economics and Statistics*, pages 1–47, 2023.

Jihwan Do and Jeanine Miklós-Thal. Price discrimination against multi-clouders. *Available at SSRN*, 2025.

Magali Dos Santos and Tatiana Ovsyannikova. Solutions saas : quelle comptabilisation dans les comptes établis en normes françaises? Web article, April 2023. URL `https://www.efl.fr/actualite/solutions-saa-comptabilisation-comptes-2022-etablis-regles-francaises_f3fa0f427-db4f-4b11-bbff-548c340996ca`. Accessed: 2025-09-19.

Jean-Pierre Dubé, Günter J Hitsch, and Peter E Rossi. Do switching costs make markets less competitive? *Journal of Marketing research*, 46(4):435–445, 2009.

EC2Instances.info Contributors. Ec2instances.info: Easy amazon ec2 instance comparison, 2025. URL `https://ec2instances.info`. Accessed: 2025-09-15.

Joseph Emeras, Sebastien Varrette, Valentin Plugaru, and Pascal Bouvry. Amazon elastic compute cloud (ec2) versus in-house hpc platform: A cost analysis. *IEEE Transactions on Cloud Computing*, 7(2):456–468, 2016.

Richard Ericson and Ariel Pakes. Markov-perfect industry dynamics: A framework for empirical work. *The Review of economic studies*, 62(1):53–82, 1995.

Eurostat / European Commission. 45 News release, Eurostat, December 2023. URL https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20231208-1. Accessed: 2025-09-25.

Maryam Farboodi and Laura Veldkamp. A model of the data economy. Technical report, National Bureau of Economic Research Cambridge, MA, USA, 2021.

Joseph Farrell and Paul Klemperer. Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, 3:1967–2072, 2007.

Flexera. Rightscale 2019 state of the cloud report from flexera. Technical report, Flexera, January 2019. URL https://resources.flexera.com/web/pdf/report-state-of-the-cloud-2019.pdf?elqTrackId=3886b5fc3baa42dea4110e39a84a9e1e&elqaid=7675&elqat=2. Survey of 786 technical professionals; licensed under Creative Commons Attribution4.0 International License.

Flexera. Rightscale 2020 state of the cloud report from flexera. Technical report, Flexera, January 2020. URL https://resources.flexera.com/web/pdf/report-state-of-the-cloud-2020.pdf.

Lucia Foster, John Haltiwanger, and Chad Syverson. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1): 394–425, 2008.

Fernando Galindo-Rueda and Fabien Verger. Oecd taxonomy of economic activities based on r&d intensity. Technical Report 2016/04, OECD Publishing, Paris, 2016. URL https://www.oecd.org/content/dam/oecd/en/publications/reports/2016/07/oecd-taxonomy-of-economic-activities-based-on-r-d-intensity_g17a283b/5jlv73sqqp8r-en.pdf.

Joshua Gans, Mikaël Hervé, and Muath Masri. Economic analysis of proposed regulations of cloud services in europe. *European Competition Journal*, 19(3):522–568, 2023.

Maria Garcia-Osipenko, Nicholas Vreugdenhil, and Nahim Bin Zahur. Endogenous rigidities and capital misallocation: Evidence from containerships. *Available at SSRN 5217587*, 2025.

Alessandro Gavazza. Leasing and secondary markets: Theory and evidence from commercial aircraft. *Journal of Political Economy*, 119(2):325–377, 2011.

Ronald L Goettler and Brett R Gordon. Does amd spur intel to innovate more? *Journal of Political Economy*, 119(6):1141–1200, 2011.

Jeff Gortmaker. Open source software policy in industry equilibrium. Technical report, Working paper, Tech. Rep, 2024.

Gautam Gowrisankaran and Marc Rysman. Dynamics of consumer demand for new durable goods. *Journal of political Economy*, 120(6):1173–1219, 2012.

Gautam Gowrisankaran and Philipp Schmidt-Dengler. A computable dynamic oligopoly model of capacity investment. Technical report, National Bureau of Economic Research, 2024.

HardDriveBenchmark.net. Hard drive benchmark. Online database, n.d. URL `https://www.harddrivebenchmark.net/`. Accessed: 2025-09-19.

Jonathon Hazell, Christina Patterson, Heather Sarsons, and Bledi Taska. National wage setting. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2022-150), 2022.

James J Heckman et al. Statistical models for discrete panel data. *Structural analysis of discrete data with econometric applications*, 114:178, 1981.

Elisabeth Honka. Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics*, 45(4):847–884, 2014.

Richard Hornbeck, Shanon Hsuan-Ming Hsu, Anders Humlum, and Martin Rotemberg. Gaining steam: Incumbent lock-in and entrant leapfrogging. Technical report, National Bureau of Economic Research, 2024.

Anders Humlum. *Robot adoption and labor market dynamics*. Rockwool Foundation Research Unit Berlin, Germany, 2022.

Bar Ifrach and Gabriel Y Weintraub. A framework for dynamic oligopoly in concentrated industries. *The Review of Economic Studies*, 84(3):1106–1150, 2017.

Mitsuru Igami and Kosuke Uetake. Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996–2016. *The Review of Economic Studies*, 87 (6):2672–2702, 2020.

Wang Jin and John Jianqiu Bai. Cloud adoption and firm performance: Evidence from labor demand. *Available at SSRN 4082436*, 2022.

Wang Jin and Kristina McElheran. Economies before scale: It strategy and performance dynamics of young us businesses. *Management Science*, 2024.

Andrew J. Kearns. Broadband deployment in equilibrium: Is competition the solution to the digital divide? Working paper, Federal Trade Commission, January 2024. URL `https://andrewkearns.github.io/andrewkearns/Broadband_Seattle_Andrew_J_Kearns.pdf`.

Ali Khajeh-Hosseini, David Greenwood, and Ian Sommerville. Cloud migration: A case study of migrating an enterprise it system to iaas. In *2010 IEEE 3rd International Conference on cloud computing*, pages 450–457. IEEE, 2010.

Cinar Kilcioglu and Justin M Rao. Competition on price and quality in cloud computing. In *Proceedings of the 25th international conference on world wide web*, pages 1123–1132, 2016.

Korea Fair Trade Commission. Kftc announces results of cloud service market study. `https://www.ftc.go.kr/eng/downloadBbsFile.do?atchmnflNo=17414`, December 2022. Press release, December 28, 2022 (hereinafter "S. Kor. FTC Cloud Study").

Sungjae Lee, Jaeil Hwang, and Kyungyong Lee. Spotlake: Diverse spot instance dataset archive service. In *2022 IEEE International Symposium on Workload Characterization (IISWC)*, pages 242–255. IEEE, 2022.

James Levinsohn and Amil Petrin. Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341, 2003.

Werner Mach and Erich Schikuta. A consumer-provider cloud cost model considering variable cost. In *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pages 628–635. IEEE, 2011.

Jacques Mairesse, Pierre Mohnen, Elisabeth Kremp, and Elizabeth KREMP. The importance of r&d and innovation for productivity: A reexamination in light of the french innovation survey. *Annales d'Economie et de Statistique*, pages 487–527, 2005.

John C. McCallum and U.S. Bureau of Labor Statistics. Disk, 2024. URL `https://ourworldindata.org/`. Based on: John C. McCallum (2023), *Price and Performance Changes of Computer Technology with Time*; U.S. Bureau of Labor Statistics (2024), *US Consumer Prices*; with minor processing by Our World in Data.

Kristina McElheran, Mu-Jeung Yang, Zachary Kroff, and Erik Brynjolfsson. *The Rise of Industrial AI in America: Microfoundations of the Productivity J-curve (s)*. US Census Bureau, Center for Economic Studies Suitland, MD, USA, 2025.

Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, pages 995–1026, 1989.

Microsoft Azure. Azure 2019 public dataset v2 – trace analysis. `https://github.com/Azure/AzurePublicDataset/blob/master/analysis/Azure%202019%20Public%20Dataset%20V2%20-%20Trace%20Analysis.ipynb`, 2019. Jupyter Notebook.

MultiCloud.Tips. Case studies of successful multi-cloud deployments. Web Article, n.d. URL `https://multicloud.tips/article/Case_studies_of_successful_multicloud_deployments.html`. Accessed: 2025-09-16.

Mary Anne North, Richard Cavallaro, and Dean Davison. The total economic impact™ of microsoft azure iaas: Cost savings and business benefits enabled by azure infrastructure-as-a-service (iaas). Technical report, Forrester Consulting, commissioned by Microsoft, August 2019. URL https://sunrise.co/wp-content/uploads/2022/08/TEI-Azure.pdf. Forrester Research, Inc.

Ofcom. Cloud services market study: Interim report. Technical report, Office of Communications (Ofcom), November 2023. URL https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/technology-research/cloud-services-market-study-interim-report.pdf. Accessed: 25 February 2026.

G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, November 1996. doi: 10.2307/2171831.

Ariel Pakes and David Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, pages 1027–1057, 1989.

Tommy Pan Fang and Shane Greenstein. Where the cloud rests: The economic geography of data centers. *Strategy Science*, 2025.

PassMark Software. Cpu benchmarks. https://www.cpubenchmark.net/, 2026a. Accessed: 2026-04-17.

PassMark Software. Intel xeon e5-2660 @ 2.20ghz cpu benchmark. https://www.cpubenchmark.net/cpu.php?cpu=Intel+Xeon+E5-2660+%40+2.20GHz&id=1219, 2026b. Accessed: 2026-04-17.

Bettina Peters, Mark J Roberts, Van Anh Vuong, and Helmut Fryges. Estimating dynamic r&d choice: an analysis of costs and long-run benefits. *The RAND Journal of Economics*, 48(2):409–437, 2017.

Greg Phillips and Sean Owens. The total economic impact™ of aws cloud operations: Cost savings and business benefits enabled by operating on aws. Technical report, Forrester Consulting, commissioned by AWS, May 2022. URL https://pages.awscloud.

com/rs/112-TZM-766/images/GEN_forrester-tei-cloud-ops_May-2022.pdf?trk=f7a0693e-d5d0-4f82-8bb3-278393da31a4&sc_channel=displayads. Forrester Research, Inc.

Scott Piper and Amitai Cohen. The state of the cloud 2023: The top cloud security threats to be aware of in 2023. Blog post, Wiz, February 2023. URL https://www.wiz.io/blog/the-top-cloud-security-threats-to-be-aware-of-in-2023. Accessed: 2025-09-18.

Maria Polyakova. Regulation of insurance with adverse selection and switching costs: Evidence from medicare part d. *American Economic Journal: Applied Economics*, 8(3): 165–195, 2016.

Michael Rubens. Labor-market power, deadweight loss, and technology adoption. 2025.

Oleksandr Shcherbakov. Measuring consumer switching costs in the television industry. *The RAND Journal of Economics*, 47(2):366–393, 2016.

Statista. Public cloud – infrastructure as a service (iaas), france. Online market forecast, 2025. URL https://www.statista.com/outlook/tmo/public-cloud/infrastructure-as-a-service/france. Accessed: 2025-09-19.

Prasanna Tambe and Lorin M Hitt. The productivity of information technology investments: New evidence from it labor data. *Information systems research*, 23(3-part-1):599–617, 2012.

Neil Thompson. The economic impact of moore's law: Evidence from when it faltered. *Available at SSRN 2899115*, 2017.

TRENDFOCUS, Inc. Cloud, hyperscale, & enterprise storage service: Cq1 '19 mid-quarter update. Industry report, TRENDFOCUS, Incorporated, Cupertino, CA, April 2019. URL https://www.trendfocus.com/. Published April 4, 2019. Authors: Mark Geenen, John Kim, John Chen, Don Jeanette, Jason Lervold.

Shuang Wang, Jacob LaRiviere, and Aadharsh Kannan. Spatial competition and missing data: an application to cloud computing. *Work. Pap*, pages 1–23, 2019.

Jungwon Yeo and Daniel P Miller. Estimating switching costs with market share data: an application to medicare part d. *International Journal of Industrial Organization*, 61: 459–501, 2018.

Ran Zhuo and Audrey Tiew. Industrial policy, location choice, and firm performance in high-tech manufacturing. *Location Choice, and Firm Performance in High-Tech Manufacturing (August 01, 2025)*, 2025.

Nikolas Zolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David N Beede, Cathy Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz. Advanced technologies adoption and use by us firms: Evidence from the annual business survey. Technical report, National Bureau of Economic Research, 2021.