13

The Rise of Cloud Computing
Minding Your Ps, Qs, and Ks

David Byrne, Carol Corrado, and Daniel Sichel

13.1 Introduction

A transformation is underway that is revolutionizing the way computing services are provided to businesses, households, and the government. This new way of accessing computing services, typically referred to as "the cloud" or "cloud computing," represents the latest transition to a new computing platform—one in which computing is done on a network of off-site computing resources accessed through the internet.[1] As this chapter shows, the changes are extraordinary and likely will have important consequences for the structure of the economy, productivity growth, and economic measurement.

Yet because the advent of these services is relatively recent and because they largely are intermediate business inputs rather than final demand, their

David Byrne is a principal economist at the Board of Governors of the Federal Reserve System.

Carol Corrado is senior advisor and research director in economics at the Conference Board and a senior policy scholar at the Center for Business and Public Policy, McDonough School of Business, Georgetown University.

Daniel Sichel is professor of economics at Wellesley College and a research associate of the National Bureau of Economic Research.

We thank Victoria Angelova, Prianka Bhatia, and Liang Zhang for extraordinary research assistance. We extend a special acknowledgment to Liang Zhang for her excellent Wellesley College thesis that developed semiannual price indexes for Amazon Web Services' basic compute product. All views expressed in this chapter are those of the authors alone and should not be attributed to organizations with which they are affiliated. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/books-and-chapters/measuring-and-accounting-innovation-21st-century/rise-cloud-computing-minding-your-ps-qs-and-ks.

1. The notion of technological change in computing as a platform shift was introduced by Bresnahan and Greenstein (1999), who analyzed the disruptive effects of the introduction of PC/client-server platform on the computer industry.

imprint on the economy is difficult to identify in official statistics. Byrne and Corrado (2017b) assessed the macroeconomic impact of the shift to cloud computing and concluded that the productivity-enhancing impacts of the shift to cloud computing were not yet particularly evident in macroeconomic data—even after taking major steps to improve the measurement of information and communications technology (ICT) asset prices (Byrne and Corrado 2017a) whose prices should be indicative of cloud services prices.[2]

This chapter, building on Byrne and Corrado's work, develops measures to quantify the service prices and quantities and the capital investment relevant for tracking the US cloud services industry—the Ps, Qs, and Ks of the title. Our basic finding is that prices for cloud services have fallen rapidly and that the use of the cloud has grown tremendously, as has investment in the related infrastructure of IT equipment and software.[3]

For our analysis of prices, we assembled a unique dataset with quarterly data on prices and characteristics for cloud services offered by the largest provider, Amazon Web Services (AWS), since the first quarter of 2009, when AWS began posting prices on the internet. The data cover AWS's basic compute, database, and storage products.[4]

For AWS's compute product, prices fell at an average rate of about 7 percent during 2009–16. Price declines were slower before 2014 and more rapid starting in the beginning of 2014. Interestingly, 2014 is the year when Microsoft and Google began posting prices for their cloud offerings on the internet. We suspect that AWS's large price declines were, in part, a response to that change in the competitive environment. For AWS's database product, prices fell at an average rate of more than 11 percent during 2010–16. Here too, prices fell relatively modestly until the beginning of 2014, after which they fell at an average rate of more than 22 percent through the end of 2016. AWS's storage product followed a similar pattern, with prices falling at an average annual rate of about 17 percent during 2009–16 and even faster declines starting in 2014. These price declines are quite rapid and highlight how rapid advances in digital *products* are showing through to prices of digital *services*.

The extremely rapid growth of capital expenditures by large providers of cloud services that we document raises a measurement puzzle. Why has investment in IT equipment in the National Income and Product Accounts (NIPAs) been so weak if large and important firms are rapidly expand-

2. Other first-order macroeconomic impacts of the shift to cloud computing include (1) a weakening in the demand for IT equipment for a given volume of ICT services, (2) a lowering of the cost of supplying a given volume of ICT services (e.g., power consumption costs), and (3) an increase in the productivity of software development.

3. After this chapter was written, Coyle and Nguyen (2018) developed a price index for AWS's compute product for the United Kingdom. Their paper also documents the rapid growth of cloud computing.

4. We also collected data for Microsoft's and Google's basic compute, storage, and database services. We intend to develop price indexes for those in future work.

ing their capital expenditures for this equipment? In part, this tension could reflect, as noted in Byrne and Corrado (2017b), higher utilization of this equipment at cloud providers than at individual businesses that had deployed this equipment previously. That higher utilization would imply less demand for IT equipment for a given demand for computing services. But there is another possibility: cloud providers appear to be designing and assembling IT equipment (on an own-account basis) that is not fully counted as IT investment in the NIPAs. We believe that this own-account investment should be included in the figures for business investment in IT, and we present some back-of-the-envelope numbers suggesting that this own-account investment is large. Our calculation suggests that if this own-account investment were included in business IT investment, then the growth rate of nominal investment in IT equipment during 2007–15 would have averaged a little more than 2 percentage points higher, and real GDP average annual growth would have been a touch higher as well.[5]

The chapter is organized as follows. Section 13.2 defines cloud computing and provides nomenclature for describing different cloud service products. This section also discusses the key technologies underlying cloud infrastructure. Section 13.3 describes our new price indexes for cloud computing services, including the data, methodology, and results. Section 13.4 uses several different metrics to demonstrate the exceptionally rapid growth of cloud computing and the associated infrastructure. We also highlight the puzzle described above concerning IT capital investment. Section 13.5 concludes.

13.2 What Is Cloud Computing?

Because cloud computing is so new and has not been studied extensively by economists, we begin with some basic definitions and nomenclature. In particular, we start with the definition developed by the National Institute of Standards and Technology (NIST) and generally affirmed in the literature (e.g., Kushida, Murray, and Zysman 2011), then discuss the range of cloud services available, and finally turn to a brief review of key technologies underlying the development of cloud computing.

13.2.1 The NIST Definition of Cloud Computing

A definition of cloud computing was created by NIST in November 2009 and, after consultations with many industry and government experts and stakeholders, published in final form in September 2011 (Mell and Grance 2011). Their definition remains relevant and makes more concrete and complete the brief definition given above. After noting that cloud computing is an evolving paradigm, NIST states, "Cloud computing is a model for enabling

5. The level of nominal GDP in 2015 would have been $117 billion higher if our estimate of own-account investment in IT equipment were included.

ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort of service provider interaction."

NIST describes the following types of clouds:

- *private cloud* (a cloud infrastructure provisioned for a single organization or specific community of organizations; it may exist on or off premises)[6]
- *public cloud* (a cloud infrastructure provisioned for open use by the public; it exists on the premises of the cloud provider)
- *hybrid cloud* (a combination of the above bound together by standardized or proprietary technology that enables data and application portability)

Finally, NIST provides a concise description of the infrastructure that underlies the cloud as "the collection of hardware and software that enables the five essential characteristics of cloud computing. The cloud infrastructure can be viewed as containing both a physical layer and an abstraction layer. The physical layer consists of the hardware resources that are necessary to support the cloud services being provided, and typically includes server, storage and network components. *The abstraction layer consists of the software deployed across the physical layer, which manifests the essential cloud characteristics*. Conceptually the abstraction layer sits above the physical layer" (italics added; Mell and Grance 2011, 2).

13.2.2 Cloud Products

The NIST cloud computing definition also includes a description of service models, or service offerings. In measurement nomenclature, these services correspond to "product types" or product classes. These product classes include

- infrastructure as a service (IaaS),
- platform as a service (PaaS), and
- software as a service (SaaS),

with each described more fully in the box. As discussed below and in the box, we would add "serverless" or function as a service (FaaS) to NIST's list.

This collection of product types often is referred to as the cloud "stack," and the earlier point about a layer of abstraction lying across the physical layer becomes important for understanding the relationship among these products. As one moves up the stack from IaaS to PaaS and so on, the level of abstraction increases in the sense that the final user can abstract from (or ignore) more and more of the underlying infrastructure. As highlighted by the italicized sentences in the box, for IaaS, the user still needs to think

6. The NIST "community cloud" deployment model is grouped with the "private cloud" model for ease of exposition.

Definitions of Cloud Service Products

IaaS (infrastructure as a service): Provides computer processing, storage, networks, and other fundamental computing resources, where the consumer can deploy and run arbitrary software, including operating systems as well as applications. *The consumer manages or controls some aspects of the underlying cloud infrastructure (such as operating systems, storage, and select network components) and deployed applications.*

PaaS (platform as a service): Provides ability to deploy consumer-created applications created using programming languages, libraries, services, and tools. *The consumer neither manages nor controls the underlying cloud infrastructure but has control over the deployed applications.*

Serverless, also known as FaaS (function as a service): Provides the capability of deploying functions (code) on a cloud infrastructure on a metered basis—only charging the user when the function is operating. *The consumer (who would be a software developer) neither manages nor controls the underlying cloud infrastructure and, in contrast to PaaS, does not control the computing program.* An API (application program interface) gateway controls all aspects of execution.

SaaS (software as a service): Provides the capability of running providers' applications on a cloud infrastructure. The applications are accessible from various client devices through either a thin-client interface (e.g., web browser) or a program interface. *The consumer neither manages nor controls the underlying cloud infrastructure, including network, servers, operating systems, storage, or even individual application capabilities, apart from limited user-specific application configuration settings.*

Sources: Authors' update of NIST service models. See also Mell and Grance (2011), Cohen (2017), and Avram (2016).

about operating systems, storage, and other computing resources. For PaaS, the final user needs to think only about the deployed application and can abstract from (or largely ignore) other aspects of the infrastructure.

Since the NIST definition was published, the industry has introduced a new layer of abstraction, called "serverless" or FaaS. At this level of abstraction, the final user only needs to think about functions or code that are to be performed, and the cloud services provider manages all other aspects of the infrastructure. Serverless can be regarded as sitting above PaaS in the NIST stack (as in the box), although it may also be regarded as a refined PaaS service.

As a final point about nomenclature for cloud service products, we connect this discussion to the state of computing precloud by noting the role of traditional data centers. By using a data center, the final user could abstract from the physical hosting environment, a lower level of abstraction than in any of the cloud services described in the box. The growth of cloud computing thus has its roots, at least in part, in the competitive advantage the cloud offers customers in terms of cost, flexibility, and scalability. At the same time, the growth and popularity of the technology also reflects how the layers of abstraction in its products (especially the distinction between PaaS and SaaS) serve distinct classes of customers. With abundant computing resources, value in the stack moves up toward applications and platform, and the lower infrastructure layers become commoditized (Kushida, Murray, and Zysman 2015).

Recent developments in the cloud that facilitate the work of software developers could be particularly significant and could, in time, have important macroeconomic consequences. As cloud vendors adapt technologies that enable them to develop products "higher up the stack" and offer services with greater abstractions, the work of software development is simplified. Thus although all classes of customers benefit from the move to greater abstraction in the technologies deployed, the benefits enjoyed by software product developers are especially significant (Cohen 2017). As a specific example, the movement to serverless services with Amazon's 2014 release of the Lambda computing platform has enabled developers to focus only on code and its rapid deployment. This has lowered costs of new software product development among providers of software products for final sale (via SaaS or regular licensing) as well as for applications developed for use within a developer's own firm (or custom-developed for use within a given firm).[7]

Thus far, we have barely discussed SaaS. In the usual nomenclature, SaaS products sit on the top of the stack. However, we believe that SaaS is best understood as a category of software product services (albeit complex) rather than cloud services per se. SaaS products are usually supplied with transactional metering—that is, not as a collection of elastically provisioned services per the NIST definition. Thus SaaS products may thus be equally regarded as software products sold via an online subscription business model—a business model whose use has grown in the digital economy.[8,9]

7. Managed services featured at Amazon's 2017 developers conference, for example, included tools for business to leverage sophisticated deep-learning models and data without having to deal with complex infrastructure issues (Murray 2017).

8. For further discussion of the role of business models in services provision, see OECD (2014), chapter 4, "The Digital Economy, New Business Models and Key Features."

9. As reported by Rackspace, a leading IaaS provider, "In recent years there has been a move by traditional software vendors to market solutions as Cloud Computing which are generally accepted to not fall within the definition of true Cloud Computing." Rackspace goes on to describe SaaS as "software delivered over the web," which is precisely our point. Technically, some SaaS products satisfy the NIST definition of cloud—for example, the Salesforce Cus-

Accordingly, the prices and quantities we study as cloud computing in the remainder of this chapter exclude SaaS products.

13.2.3 Cloud Technologies

The cloud platform relies on a suite of technologies—mainly virtualization, grid computing, and microservices architectures—but also everything that makes high-speed broadband possible. Arguably, IT history is at the point where the tagline Sun Microsystems coined in the early 1990s, "The Network Is the Computer," is finally right.[10] The network is no longer a mere bridge between autonomous nodes on independent missions and prone to choke points (as in provision of transport). The continuous increase in network capacity and a near disappearance of limitations that could choke traffic in an earlier era (hardline security policies, storage performance issues, last-mile WAN hindrances) are the foundation of this latest platform shift in computing.

Behind a virtual machine host on a network of today, computing resources—storage, memory, networking, and CPUs—are physically distributed and managed via processing queues. Long before enterprises began moving onto the cloud, mainframes and servers were virtualized, and an essential element of computing focused on the function of processing queues. With cloud computing, some resource queue end-points are moved offsite, and more than ever, computing resource acquisition and allocation becomes the central task of cloud providers. One can be far more technical about the transformation of computing as it has undergone virtualization and moved to a cloud platform, but it is hard to be more prosaic than the old Sun tagline.

Cloud vendors have made increasing use of virtualization and grid computing to elastically supply information-processing services since the advent of the millennium, with the growth in capacity especially rapid since 2006, when Amazon Web Services opened its doors. The virtualization technology that is the primary enabler of cloud computing has been in commercial use since the 1970s via IBM mainframes. Modern IBM mainframes (circa the System/390 introduced in 1990 and renamed *zSeries* in 2000) are exceptionally adept at handling large, diverse, and varying workloads and remain in use today, though they have lost much force in the large datacenter market with the rise in cloud computing (Byrne and Corrado 2017b). Grid computing is applying the resources of many computers in a network to a single problem at the same time; the technology was first used in 1989 to link supercomputers and thereafter grew and evolved along with the internet (De Roure et al. 2003).

"Containers" are another new cloud technology. Containers—a scalable

tomer Relationship Management (CRM) product—but many others, including other CRM products, do not. See https://support.rackspace.com/white-paper/understanding-the-cloud-computing-stack-saas-paas-iaas/ accessed February 25, 2017.

10. The Sun Microsystems tagline is attributed to John Gage (Reiss 1996). The discussion in this paragraph draws from Hubbard (2014).

form of virtualization technology—allow users to run and deploy applications without launching a new virtual machine for each application, increasing the speed of software application development, deployment, and scalability. In terms of enterprise applications outside of Silicon Valley, it is still early in the application of containers. Indeed, the technology generally was not widely understood outside cloud vendors until the release of open source LINUX formats (Docker 1.0) in March 2013. Docker transformed container technology to a product for enterprise use. The consultancy IDC estimated that in 2014, only 1 percent of enterprise applications were running on containers that could readily be scaled, but reportedly growth in Docker adoption has been very rapid since then.[11]

One final point of history connects this discussion to the earlier use of commercial time-sharing services. These services were an important part of the computing environment in its earliest days. There was a period of frantic growth (1955–65), after which the industry flourished for another 20 years due to a competitive advantage that "arose from the nonlinear relationship between total operating costs and performance—the larger the time-sharing system, the lower the per-user cost" (Campbell-Kelly and Garcia-Swartz 2008, 27). Commercial time-sharing services underwent a complete industrial boom-to-bust cycle— like typewriters and punched-card machines—after the advent of the PC.[12]

13.3 Prices of Cloud Computing Services

Outside of sporadic media reports and research by some private consultants, relatively little is known about the prices of cloud computing services. This chapter develops new price indexes for three basic products provided by one of the leading providers of cloud services.

13.3.1 Data

We collected prices on a quarterly basis from AWS, the earliest and largest provider. We collected prices from when AWS began posting prices on the internet, with the earliest prices from 2009. To collect historical prices, we used the Internet Archive (also known as the Wayback Machine) to pull posted prices from web pages as they appeared in prior periods. We collected prices for a compute product (renting virtual machines), a selection of database products that offer SQL as well as other database software, and a range of disk storage products.

11. See DataDog (2015). See also Elliot and Perry (2018).

12. According to Campbell-Kelly and Garcia-Swartz (2008), the market for time-sharing existed because it was the only means at that time of providing a personal computing experience at a reasonable cost. They also present econometric evidence showing that the growth of time-sharing services in its heyday slowed down the growth of mainframe computer shipments; see their online appendix.

Of course, the services for which we gathered prices are just a subset of the wide array of services available, and they are at the lower end of the "stack" of cloud products described above. In particular, we place the compute and storage products in the IaaS category, and we place the database products in PaaS. That said, these compute, database, and storage services are key foundational elements on which many of the services that are higher in the stack are based. Accordingly, we believe that the compute, database, and storage products considered in this chapter provide a very useful and broadly representative sample of available cloud services.

AWS has been the market leader and has posted prices on the internet since 2009. Microsoft began posting prices in early 2014, and Google began posting prices in late 2014. We believe that AWS is broadly representative of the market, though future work on prices of other providers is needed to confirm that.

We note one important limitation of our data. We obtained data on prices and product characteristics but not on quantities because cloud service providers do not make product-level sales information readily available. We also were unable to obtain private data on quantities.

13.3.2 Amazon Web Services (AWS)

AWS offers an amazing array of products. One common feature across all products is that customers choose among regions—that is, where the servers are located on which they are running applications and storing data. Currently, AWS offers four regions in the United States, including Virginia, California, Oregon, and Ohio. (Amazon also offers many regions outside the United States.) For this chapter, we collected prices for Virginia, California, and Oregon. (The Ohio region was only introduced in October 2016.) For an AWS customer, choosing a region that is geographically closer reduces latency, and some customers will store data in multiple regions for redundancy. Prices differ across regions, with prices in California generally higher than those in Virginia or Oregon. In general, the differences in prices across regions are in levels, while changes in prices tend to be very similar across regions.

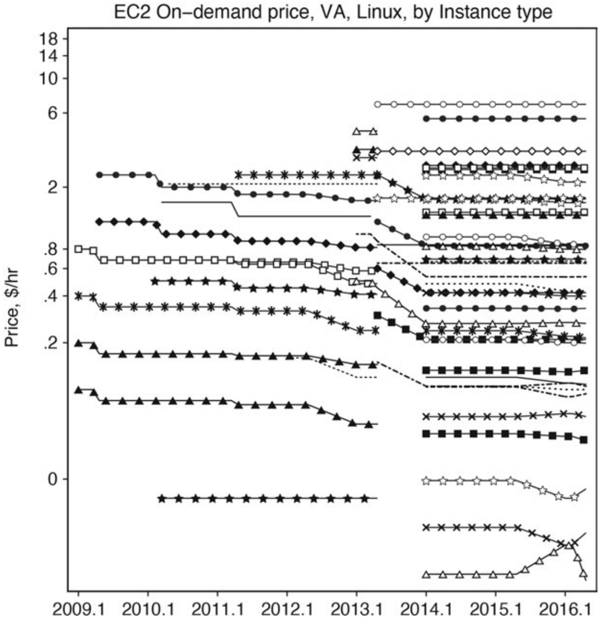
Compute Product (EC2—Elastic Compute Cloud). Using this product amounts to renting a virtual machine (PC or server) from AWS, and this product is priced in terms of dollars per hour. In cloud computing nomenclature, the use of a virtual machine is known as an "instance," and AWS offers instances in a wide range of configurations. During the span of our data from 2009 to 2016, AWS offered 55 different configurations of virtual machines. Each configuration has specified characteristics in terms of the power of the processor, the amount of RAM, and the amount of disk space allocated. In addition, customers can choose between Linux and Windows operating systems. For every available configuration, we collected prices as well as characteristics, and we have a total of 4,079 observations for EC2 prices. The characteristics are important, and we will use them to construct hedonic price indexes.

AWS offers several different pricing schemes for instances. For EC2, we collected data for only "on-demand" instances, which can be purchased at any time with no commitment. AWS also offers "reserved" instances, for which a customer pays in advance for a set volume of instances whether or not the instances are used. Prices of reserved instances are lower than those of on-demand instances. In addition, AWS runs a spot market for instances. Customers can bid for instances at a price of the customers' choosing. The customer will receive the instances if they are available but will not receive them if some other customer offered to pay a higher price for available instances. Prices of spot instances also tend to be below those of on-demand instances. Finally, AWS offers quantity discounts to heavier users.

Tracking prices for all of these different types of instances was beyond the scope of this chapter. For the purpose of constructing price indexes, a key question is whether the price trends for on-demand instances differ in systematic ways from those of other types of instances. Our sense is that prices within these different pricing schemes tend to move together, but that remains an open question. That said, we suspect that *individual* customers experience price declines that are more rapid for a time than are the trends we estimate. In particular, as customers gain experience with AWS and migrate more applications to the cloud, we suspect that they increasingly shift toward reserved instances and avail themselves of quantity discounts. This shift toward lower-priced instances generates faster price declines during the shift than we estimate from tracking prices of on-demand instances. Of course, once a customer has finished the shift toward lower-priced instance types, the trend in prices experienced by that customer likely would be in line with the price trends that we estimate.

Our raw data for EC2 prices are plotted in figure 13.1. This figure plots AWS's posted prices for each instance type for the full time it is in the market, with a different line style capturing each different instance type. In the figure, we show separate plots for each region and operating system pair, with each column of graphs covering a region and each row covering an operating system. The graphs, plotted with a log scale, indicate that prices tend to follow downward step functions, with longish periods of no price change. It also is evident that AWS revamped its offering of instance types around the beginning of 2014, dropping most extant instance types and introducing new ones. Of course the graphs reflect no controls for characteristics or quality of the instances, and as shown below, it turns out that this revamping was associated with a large drop in quality-adjusted prices.

Database Product (RDS—Relational Database Service). Using this product amounts to renting database software along with a virtual machine (called an instance class) to run the software. It is priced in terms of dollars per hour. AWS offers several different database engines, including MySQL,



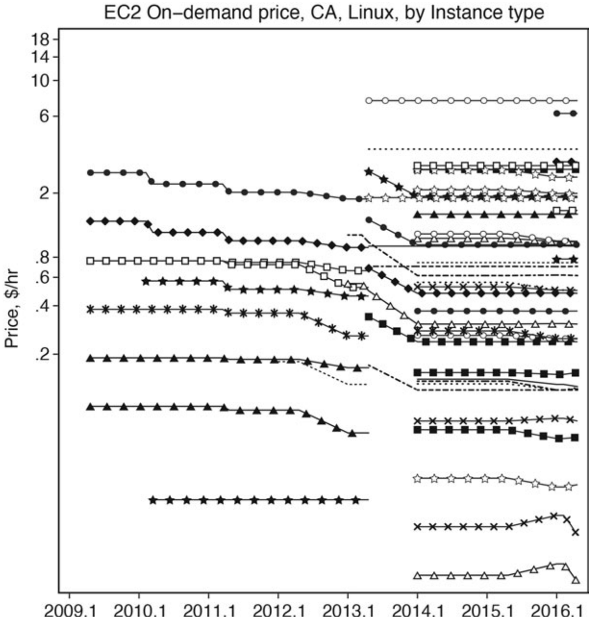
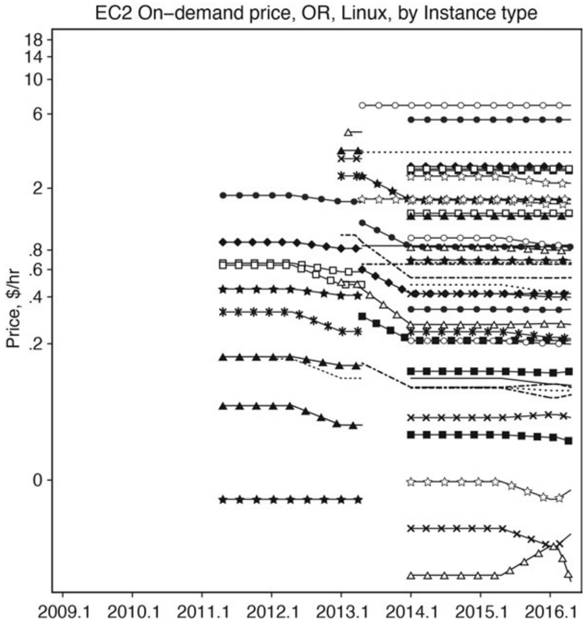


Fig. 13.1 Amazon EC2 posted prices by instance for each region and for Linux and Windows


EC2 On-demand price, OR, Linux, by Instance type
Price, $/hr
18
14
10
6
2
.8
.6
.4
.2
0
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1

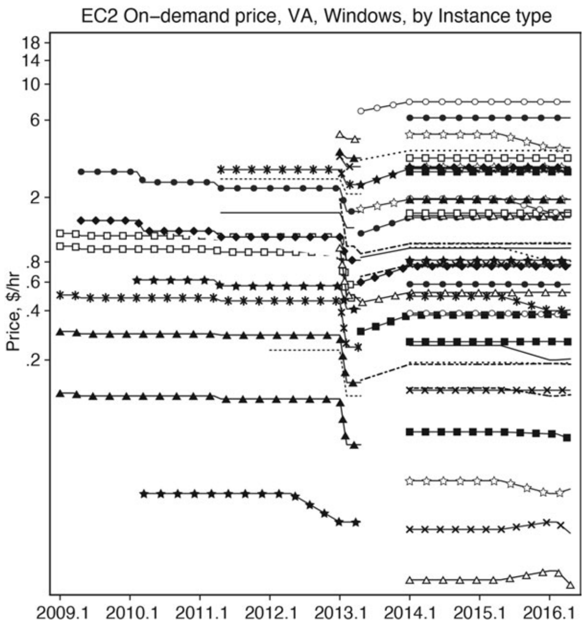
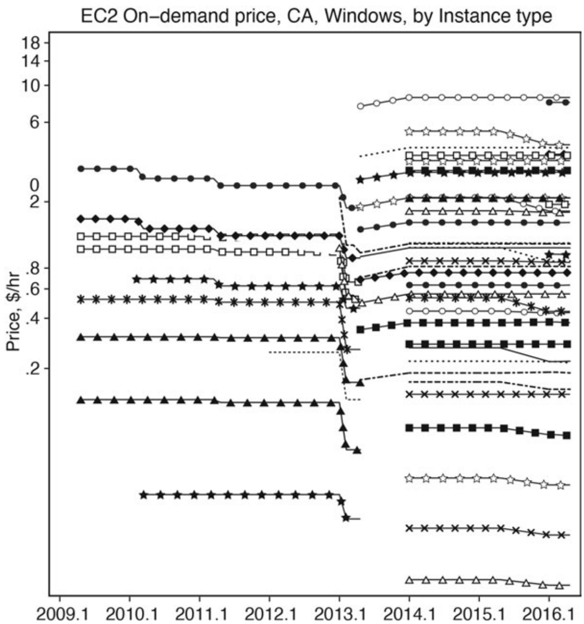

EC2 On-demand price, VA, Windows, by Instance type
Price, $/hr
18
14
10
6
2
.8
.6
.4
.2
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1

Fig. 13.1 (cont.)


EC2 On-demand price, CA, Windows, by Instance type
Price, $/hr
18
14
10
6
0
2
.8
.6
.4
.2
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1

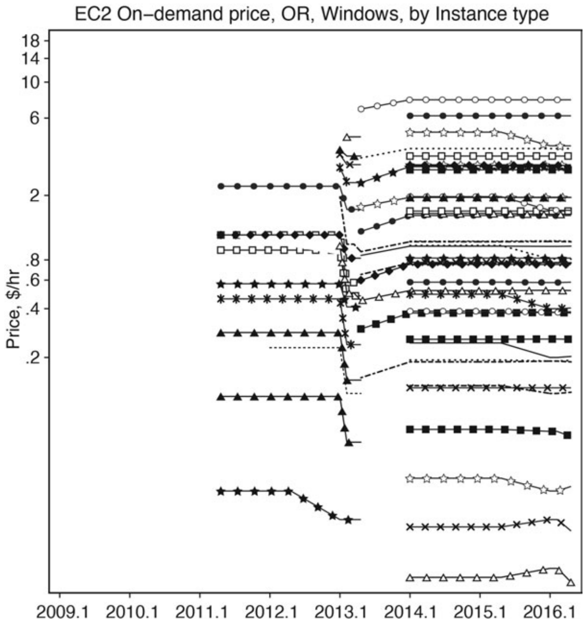

EC2 On-demand price, OR, Windows, by Instance type
Price, $/hr
18
14
10
6
2
.8
.6
.4
.2
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1

Fig. 13.1 (cont.)

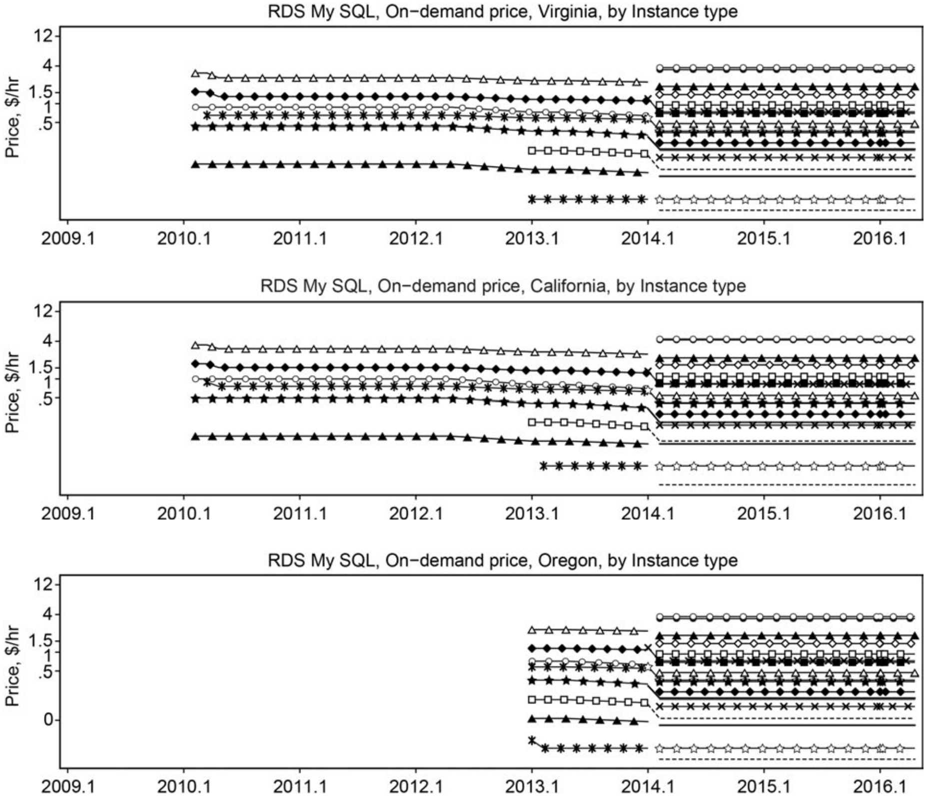


Fig. 13.2 Amazon RDS posted prices by instance for MySQL in the Virginia, California, and Oregon regions

SQL, SQL Standard, SQL Express, SQL Web, SQL Enterprise, PostgreSQL, Oracle, Aurora, and MariaDB. Some of these are open source, while others are proprietary and require a license. For those requiring a license, AWS offers instances for which customers use their own license as well as instances for which AWS provides the license (for a higher price). AWS also offers several different instance classes with differences in the CPU power of the virtual machine, the amount of RAM, network performance, and whether the instance class is optimized for input-output to storage. For every available configuration, we collected prices as well as characteristics for on-demand instances. (AWS also offers reserved instances for its database product.) In total, we have 5,340 observations on RDS prices.

Our raw data for a selection of RDS prices are plotted in figure 13.2. This figure plots AWS's posted prices for each RDS instance type for the MySQL database software for the Virginia, California, and Oregon regions. Because of the multiplicity of types of database software, it is not feasible to plot all our data in a single figure. That said, the data in this figure are broadly representative of those for other regions and database software. The graphs are plotted with a log scale and show the same overall pattern as the EC2 price plots. Prices tend to follow downward step functions, with longish periods of no price change. As with EC2, AWS revamped its offerings around the beginning of 2014, dropping most extant instance types and introducing new ones.

Storage Product (S3—Simple Storage Solution). Using this product amounts to renting hard disk space. It is priced in terms of dollars per terabyte (TB) per month.[13] The pricing scheme for S3 builds in volume discounts directly with pricing tiers. For example, customers pay one price for the first TB used, a lower price for the next 49 TB used, a still lower price for the next 50 TB used, and so on.[14] AWS also offers three different types of storage: "standard" allows immediate access to stored data; "infrequent" access is for longer-term storage, and data can be retrieved only with a delay; and "glacier" storage has an even longer delay for retrieval. As with other AWS products, customers can choose among regions. We collected prices for all pricing tiers, all three types of storage, and the Virginia, California, and Oregon regions. In total, we have 445 observations on S3 prices.

Our raw data for S3 prices are plotted in figure 13.3. This figure plots AWS's posted prices for each price tier for the full time it is in the market for each region and type of storage pair. (Each different price tier is represented by a different line style.) In the figure, each column is for a region, and each row is for a different type of storage (standard, infrequent, and glacier).

13.3.3 Results

The new quality-adjusted price indexes presented here for EC2 (compute) and RDS (database) are based on adjacent-quarter regressions. For S3 (storage), quality does not change appreciably because the product is just a TB of storage, so we rely on matched-model indexes.

To explain our rationale for using adjacent-quarter regressions, we first describe a dummy-variable hedonic specification:[15]

$$\ln(P_{i,t}) = \alpha + \sum_{k} \beta_k X_{k,i,t} + \delta_t D_{i,t} + \varepsilon_{i,t}, \quad (1)$$

where $P_{i,t}$ is the price of product i in period t, $X_{k,i,t}$ is the value of characteristic k for that product in period t (measured in logs or levels, as appropriate), $D_{i,t}$ is a time dummy variable (fixed effect) that equals 1 if the price i is observed in period t and 0 otherwise, and $\varepsilon_{i,t}$ is an error term.

A potential shortcoming of equation (1), highlighted by Pakes (2003) and

13. A terabyte of data is 1,014 gigabytes. The prefix *tera* is from the Greek word for monster.

14. The pricing tiers have changed over time. For example, early on, prices dropped after the first TB of data, while now pricing does not drop until after the first 50 TB of data. This change reflects the ongoing decline in the price of storage.

15. The language used here to describe adjacent-quarter regressions draws heavily from Byrne, Oliner, and Sichel (2018).

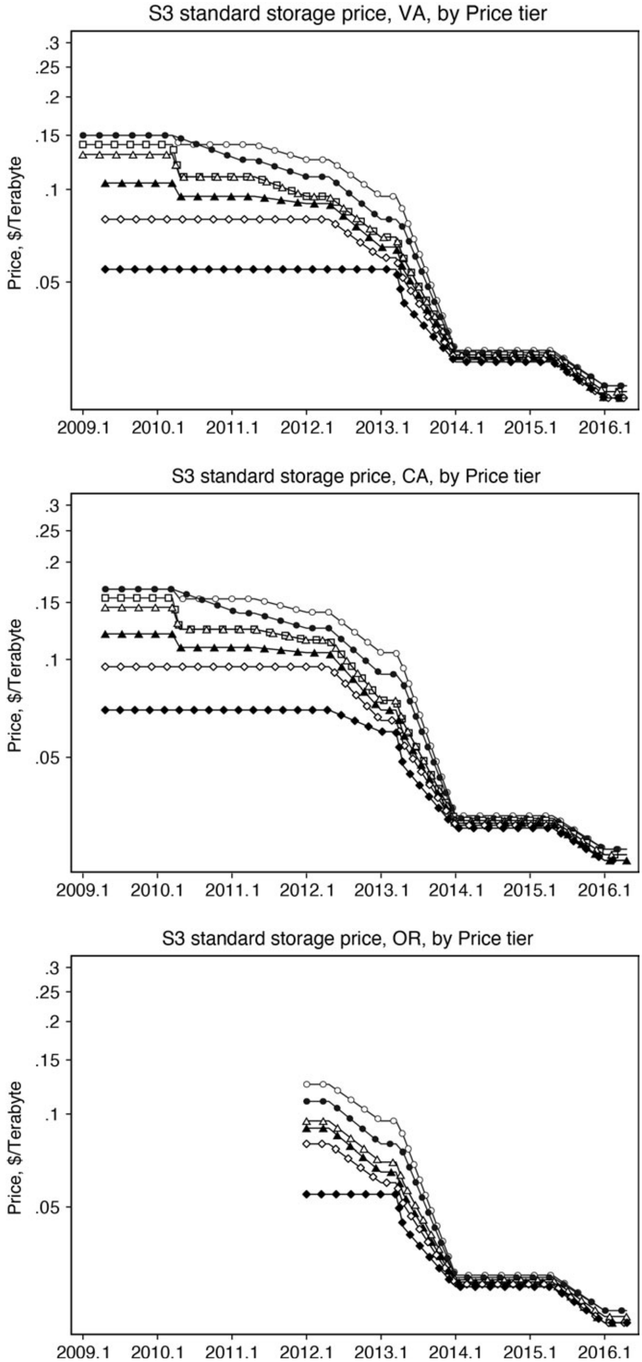


Fig. 13.3 Amazon S3 posted prices by price tier for each region and storage type

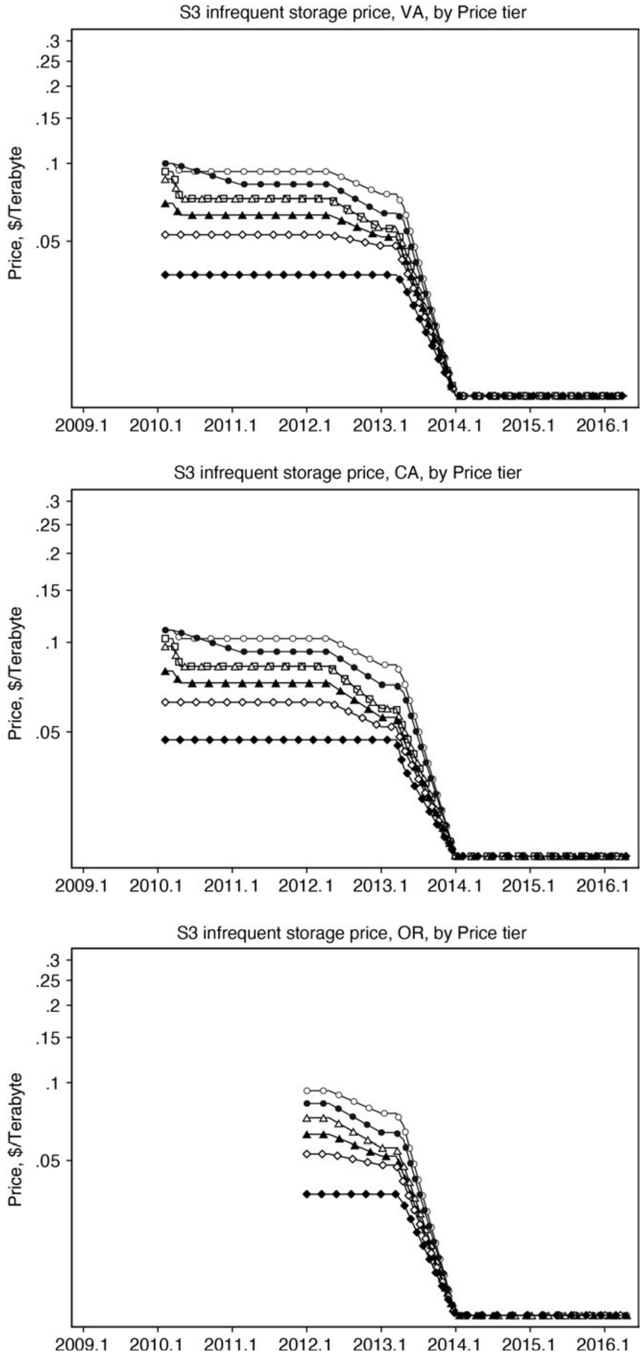

S3 infrequent storage price, VA, by Price tier
Price, $/Terabyte
.3
.25
.2
.15
.1
.05
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1
S3 infrequent storage price, CA, by Price tier
Price, $/Terabyte
.3
.25
.2
.15
.1
.05
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1
S3 infrequent storage price, OR, by Price tier
Price, $/Terabyte
.3
.25
.2
.15
.1
.05
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1

Fig. 13.3 (cont.)

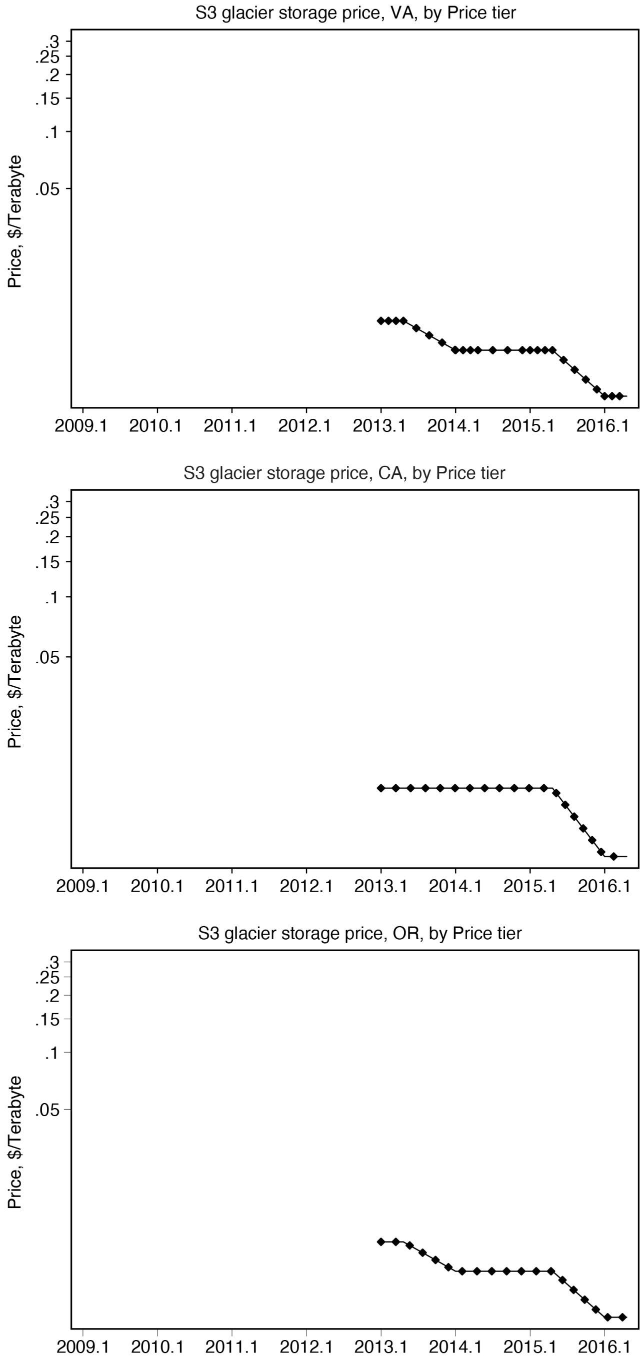

S3 glacier storage price, VA, by Price tier
Price, $/Terabyte
.3
.25
.2
.15
.1
.05
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1
S3 glacier storage price, CA, by Price tier
Price, $/Terabyte
.3
.25
.2
.15
.1
.05
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1
S3 glacier storage price, OR, by Price tier
Price, $/Terabyte
.3
.25
.2
.15
.1
.05
2009.1
2010.1
2011.1
2012.1
2013.1
2014.1
2015.1
2016.1

Fig. 13.3 (cont.)

Erickson and Pakes (2011), is that the coefficients on the characteristic are constrained to remain constant over the full sample period. Byrne, Oliner, and Sichel's (2018) study on microprocessors used adjacent-year regression; here, we follow their setup but use adjacent-quarter regressions.

To make things precise, we describe our adjacent-quarter procedure for EC2; the procedure for RDS is parallel. For EC2, we estimate the following regression for each two-quarter overlapping period:

$$\ln(P_{i,t}) = \alpha + \sum_{k} \beta_k X_{k,i,t} + \delta D_{2t} + \varepsilon_{i,t}, \quad (2)$$

where $P_{i,t}$ is the price of EC2 instance of type i in quarter t and $X_{k,i,t}$ is the k^{th} characteristic of instance i in quarter t. The dummy variable D_{2t} equals 1 if the price observation is for the second quarter of the two-quarter overlapping period and 0 otherwise.

To construct a price index from these sequences of regressions, we spliced together the percent changes implied by the estimated coefficients on the D_{2t} variables. All the reported results are bias adjusted to account for the transformation from log prices to a nonlog price index.[16]

Because we do not have quantity data, the adjacent-quarter regressions are unweighted so that each observation receives an equal weight in the regression. This approach is an unfortunate limitation of not having quantity data.

EC2. For the adjacent-quarter regressions for EC2, the following characteristics entered as natural logs: *ECU* (AWS's designation of the power of the processor), *Mem* (the amount of memory in GB), and *Storage* (the amount of disk storage in GB).[17] The regressions also include the following fixed effects: *storSSD* (= 1 if the disk storage is solid state), *pltfrm* (= 1 if the processor is 64 bit; = 0 if the processor is 32 bit), *System* (= 1 if the system is Linux; = 0 for Windows), *inO* (= 1 if the price is for the Oregon region), and *inC* (= 1 if the price is for the California region).

Results of these regressions are summarized in table 13.1. Because of the number of adjacent-quarter regressions, the table summarizes the regression results, showing the minimum, maximum, and median values of coefficient estimates across the regressions.[18] In addition, of the 31 adjacent-quarter

16. Because the exponential function is nonlinear, the translation from the natural log of prices to price levels requires an adjustment in order to be unbiased. We apply the standard adjustment based on the estimated variance of the coefficient δ, as described in van Dalen and Bode (2004).

17. In later periods, AWS began charging separately for disk storage for some instances. For these observations, *Storage* is set equal to zero.

18. As is evident in the table (as well as in our adjacent-quarter estimates for other cloud services), some parameters exhibit considerable variation across the adjacent-quarter regressions. Running adjacent-year regressions likely would damp this variation. We chose not to consider adjacent-year regressions for two reasons. First, because prices of these services change infrequently and by large amounts and because new products are introduced infrequently, we wanted

Table 13.1 **Amazon EC2 adjacent-quarter regressions, 2009:Q2–2016:Q4 (summary of coefficient estimates across all adjacent-quarter regressions)**

	Minimum	Maximum	Median	Fraction significant at 5%	Fraction significant at 10%
D2	−0.329	0.031	0.0	2/31	2/31
ECU	−0.114	0.604	0.212	28/31	29/31
Mem	−0.739	0.85	0.630	31/31	31/31
Storage	−0.66	0.199	−0.067	30/31	31/31
StorSSD	−0.049	0.017	0.0	0/31	10/31
System	−0.444	0	−0.341	29/31	29/31
pltfrm	−0.477	2.103	0.0	10/31	10/31
inO	−0.025	0.038	0.0	0/31	0/31
inC	0.0	0.146	0.127	23/31	24/31
Constant	−5.939	−0.926	−4.616	31/31	31/31

Notes: D2 is the dummy variable for the second quarter of the adjacent-quarter regression. *ECU*, *Mem*, *and Storage* are in natural logs. *ECU* measures processor power, *Mem* is the amount of RAM, and *Storage* is the amount of disk storage. Other variables enter as fixed effects. *StorSSD* = 1 if solid state storage, *System* = 1 if operating system is Linux, *pltfrm* = 1 if the processor is 64 bit, *inO* = 1 if the region is Oregon, and *inC* = 1 if the region is California. The omitted categories are the Windows operating system in the Virginia region with magnetic hard drive disk storage and a 32-bit processor.

regressions, the table shows the fraction of the estimates for each coefficient that are significant at the 5 and 10 percent significance levels.

The coefficient on the dummy variable capturing quality-adjusted price change, *D2*, has a median value of zero, reflecting that prices are not changing in most quarters. The coefficient for the variable for processor power, *ECU*, generally is positive and highly significant, as prices are higher for instances providing more processor power. The same pattern holds for the memory variable, *Mem*. The variable for disk storage is almost always significant, though its sign often is negative. Among the fixed effects, solid-state disk storage, *StorSSD*, has relatively little effect on prices, while instances running with Linux, the *System* variable, are priced at a hefty discount to instances running with Windows (for which AWS would be paying a license fee). The coefficient on the fixed effect distinguishing between 32- and 64-bit processors (*pltfrm*) is quite variable across regressions and significant in about a third of the regressions. Prices in the Oregon region, captured by the *inO* variable, are little different from those in Virginia, while prices in the California region, the *inC* variable, typically are more than 10 percent higher than prices in Virginia.

Table 13.2 reports the price indexes generated by these regressions as well as the number of observations and adjusted-R^2 for each adjacent-quarter

to be able to isolate these periods of change. Second, our quarterly frequency coincides with that in the National Accounts.

Table 13.2 **Amazon EC2 (compute product) price index**

	Price index	Percent change (quarterly rate)	Number of observations	Adjusted R^2
2009: 1	100.00			
2009: 2	100.00	0.0	20	0.996
2009: 3	100.00	0.0	20	0.996
2009: 4	95.29	−4.7	38	0.91
2010: 1	95.29	0.0	56	0.927
2010: 2	95.29	0.0	60	0.926
2010: 3	91.27	−4.2	69	0.955
2010: 4	91.77	0.5	75	0.968
2011: 1	91.77	0.0	76	0.969
2011: 2	91.77	0.0	76	0.969
2011: 3	91.77	0.0	76	0.969
2011: 4	84.92	−7.5	98	0.967
2012: 1	87.71	3.3	126	0.961
2012: 2	88.68	1.1	132	0.96
2012: 3	88.68	0.0	132	0.963
2012: 4	88.68	0.0	132	0.963
2013: 1	82.37	−7.1	156	0.953
2013: 2	77.98	−5.3	182	0.949
2013: 3	77.98	0.0	184	0.974
2013: 4	77.95	0.0	242	0.974
2014: 1	56.15	−28.0	370	0.944
2014: 2	56.15	0.0	440	0.956
2014: 3	56.15	0.0	440	0.956
2014: 4	56.15	0.0	440	0.956
2015: 1	56.15	0.0	440	0.956
2015: 2	56.15	0.0	440	0.956
2015: 3	56.15	0.0	440	0.956
2015: 4	56.15	0.0	440	0.956
2016: 1	48.84	−13.0	518	0.938
2016: 2	48.84	0.0	596	0.936
2016: 3	48.72	−0.2	596	0.936
2016: 4	48.72	0.0	298	0.936
Memo: Avg. at annual rate				
2009:1−2016:4		−6.9		
2009:1−2013:4		−5.1		
2014:1−2016:4		−10.5		

Notes: Based on adjacent-quarter hedonic regression as described in the text. All estimates are bias adjusted to account for the translation from log price to a price index. The last two columns show the number of observations and adjusted R^2s from each of the adjacent-quarter regressions.

regression.[19] The adjusted-R^2s are quite high, indicating that the right-hand-side variables are capturing most of the sources of variation in prices. The price index is shown in the first column, and percent changes at *quarterly*

19. The price trends for EC2 are similar to those reported by Zhang (2016).

rates are reported in the second column. These figures highlight that prices do not change in most quarters. Price declines are large in some quarters, with the biggest drop in the first quarter of 2014, when AWS revamped its offerings of EC2 instances. Although not evident in the plots of posted prices in figure 13.1, the newly offered instances provided much higher quality at prices that were, on their face, roughly comparable to the posted prices of the old offerings of instances. Accordingly, the hedonic regressions identify a very large quality-adjusted price decline in that period.

All told, quality-adjusted prices for EC2 instances fall at an average *annual* rate of about 7 percent over the full sample. Interestingly, prices fell at an annual average rate of about 5 percent from the beginning of 2009 to the end of 2013. Then, in early 2014, just as Microsoft had entered the market to a sufficient degree that they were posting their cloud prices on the internet (and shortly before Google started doing the same), AWS began cutting prices more rapidly. That started with the big price drop in early 2014, and over the period from the start of 2014 to the end of 2016, EC2 prices fell at an average annual rate of 10.5 percent.

RDS. For the adjacent-quarter regressions for RDS, the following characteristics entered as natural logs: *Vcpu* (AWS's designation of the power of the processor) and *Memory* (the amount of memory in GB). The regressions also include the variable *IOPerformance*, which is a qualitative variable indicating whether the network performance is low, moderate, high, or very high. In addition, the regressions include the following fixed effects: *Provisioned IOPS optimized* (= 1 if instance is optimized for input to and output from storage), *inO* (= 1 if the price is for the Oregon region), and *inC* (= 1 if the price is for the California region), a set of fixed effects for each type of database software offered (the omitted category is SQL Standard).

Results of these regressions are summarized in table 13.3. As for the EC2 results, the table summarizes the regression results, showing the minimum, maximum, and median values of coefficient estimates across the regressions. In addition, of the 25 adjacent-quarter regressions, the table shows the fraction of the estimates for each coefficient that are significant at the 5 percent and 10 percent significance levels.

The coefficient on the dummy variable capturing quality-adjusted price change, *D2*, has a median value of zero, reflecting that prices are not changing in most quarters. The coefficient for the variable for processor power, *Vcpu*, generally is positive and relatively significant, as prices are higher for instances providing more processor power. The same pattern holds for the memory variable, *Memory*. The variable *IOPerformance* also is always positive and almost always significant. Among the fixed effects, the variable *Provisioned IOPS optimized* (indicating optimization of storage input/output) is always positive and significant. Just as for EC2, prices in the Oregon region, captured by the *inO* variable, are little different from those in Virginia, while prices in the California region, the *inC* variable, typically are more than 10

Table 13.3 **Amazon RDS adjacent-quarter regressions, 2010:Q3–2016:Q4 (summary of coefficient estimates across all adjacent-quarter regressions)**

	Minimum	Maximum	Median	Fraction significant at 5%	Fraction significant at 10%
D2	−0.53	0.01	0.00	5/25	5/25
Vcpu	−0.15	0.22	0.03	16/25	16/25
Memory	0.57	0.74	0.69	25/25	25/25
IOPerformance	0.04	0.35	0.25	24/25	24/25
Provisioned IOPS optimized	0.07	0.22	0.13	25/25	25/25
inC	0.09	0.12	0.11	25/25	25/25
inO	−0.01	0.01	0.00	0/25	0/25
Aurora	−1.31	0.00	0.00	5/25	5/25
MySQL	−1.44	0.00	−1.00	18/25	18/25
Oracle (own license)	−1.43	0.00	−1.00	17/25	17/25
Oracle (AWS provided license)	0.00	0.76	0.37	21/25	21/25
PostgreSQL	−1.38	0.00	0.00	12/25	12/25
SQL (own license)	−1.02	0.00	−0.67	18/25	18/25
SQL express	−1.37	0.00	−0.96	18/25	18/25
SQL web	−0.66	0.00	−0.60	18/25	18/25
MariaDB	−1.44	0.00	0.00	4/25	4/25
Constant	−3.10	−1.99	−2.87	25/25	25/25

Notes: No observations for 2015:Q4 were available in the web archive.

percent higher than prices in Virginia. Among the fixed effects for different database software, most are priced at significant discounts relative to SQL Standard. Oracle is the big exception; if AWS provides the license, Oracle is priced significantly above SQL Standard.

Table 13.4 reports the price indexes generated by these regressions. The adjusted R^2s are quite high, indicating again that the right-hand-side variables are capturing most of the sources of variation in prices. The price index is shown in the first column, and percent changes at quarterly rates are reported in the second column. As for EC2, these figures highlight that prices do not change in most quarters. Price declines are large in some quarters, with the biggest drop at the beginning of 2014, when AWS revamped its offerings.

All told, quality-adjusted prices for RDS instances fall at an average *annual* rate of more than 11 percent over the full sample. Over subperiods, the pattern is the same as that for EC2 prices. Prices fell at an annual average rate of about 3 percent from the beginning of 2009 to the end of 2013. Then, in early 2014, just as Microsoft had entered the market to a sufficient degree that they were posting their cloud prices on the internet, AWS began cutting prices more rapidly. That started with the big price drop in early 2014, and over the period from the start of 2014 to the end of 2016, RDS prices fell at an average annual rate of more than 22 percent.

Table 13.4 **Amazon RDS (database product) price index, 2010:Q2–2016:Q4**

	Price index	Percent change (quarterly rate)	Number of observations	Adjusted R^2
2010: 2	100.00			
2010: 3	100.00	0.0%	22	0.999
2010: 4	93.73	−6.3%	24	0.997
2011: 1	93.73	0.0%	24	1
2011: 2	93.73	0.0%	44	0.999
2011: 3	93.73	0.0%	64	0.999
2011: 4	93.73	0.0%	64	0.999
2012: 1	93.73	0.0%	64	0.999
2012: 2	93.73	0.0%	133	0.971
2012: 3	93.73	0.0%	202	0.967
2012: 4	93.73	0.0%	202	0.967
2013: 1	87.25	−6.9%	242	0.971
2013: 2	87.19	−0.1%	282	0.976
2013: 3	87.19	0.0%	282	0.976
2013: 4	87.19	0.0%	308	0.978
2014: 1	82.29	−5.6%	420	0.977
2014: 2	48.30	−41.3%	601	0.975
2014: 3	48.30	0.0%	696	0.981
2014: 4	48.30	0.0%	696	0.981
2015: 1	48.30	0.0%	696	0.981
2015: 2	48.30	0.0%	696	0.981
2015: 3	48.55	0.5%	712	0.981
2015: 4	48.55	0.0%		
2016: 1	38.38	−20.9%	1,183	0.983
2016: 2	38.20	−0.5%	1,218	0.985
2016: 3	38.20	0.0%	702	0.984
2016: 4	38.20	0.0%	606	0.983
Memo: Avg. at annual rate				
2010:2–2016:4		−11.6		
2010:2–2013:4		−3.3		
2014:1–2016:4		−22.6		

Notes: Based on adjacent-quarter hedonic regression as described in the text. All estimates are bias adjusted to account for the translation from log price to a price index. The last two columns show the number of observations and adjusted R^2s from the adjacent-quarter regressions. No observations are available for 2015:Q4; we assumed no price change in that quarter.

S3. As noted, quality does not change appreciably over time for S3, the AWS storage product. Accordingly, we construct matched-model indexes by tracking price changes over time for each price tier. Table 13.5 reports the resulting price indexes for each price tier. As for EC2 and RDS, these figures indicate that prices do not change in most quarters. Price declines are large in some quarters, with the biggest drop at the beginning of 2014, as AWS appeared to be responding to a competitive threat from Microsoft (and Google later in the year).

The bottom three lines of the table provide summary figures that are an

Table 13.5 **Amazon S3 (storage product) price indexes, standard storage, Virginia (percent change, quarterly rate)**

	Terabyte (TB) range						
	s ≤ 1	1 < s ≤ 50	50 < s ≤ 100	100 < s ≤ 500	500 < s ≤ 1K	1K < x ≤ 5K	≥ 5K
2009: 2	0.0	0.0	0.0	0.0			
2009: 3	0.0	0.0	0.0	0.0			
2009: 4	0.0	0.0	0.0	0.0			
2010: 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2010: 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2010: 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2010: 4	−6.7	0.0	−21.4	−15.4	−9.5	0.0	0.0
2011: 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2011: 2	0.0	−16.7	0.0	0.0	0.0	0.0	0.0
2011: 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2011: 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2012: 1	−10.7	−12.0	−13.6	−13.6	−5.3	0.0	0.0
2012: 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2012: 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2012: 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2013: 1	−24.0	−27.3	−26.3	−26.3	−27.8	−25.0	0.0
2013: 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2013: 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2013: 4	−10.5	−6.2	−14.3	−14.3	−15.4	−15.0	−21.8
2014: 1	−64.7	−6.7	−51.7	−51.7	−48.2	−45.1	−36.0
2014: 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2014: 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2014: 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2015: 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2015: 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2015: 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2015: 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2016: 1	−23.3	−22.0	−24.1	−24.1	−26.3	−25.0	−23.6
2016: 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2016: 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2016: 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Memo: Avg. at annual rate							
2009:1−2016:4	−18.1	−18.7	−19.5	−18.8	−18.9	−15.7	−11.6
2009:1−2013:4	−10.9	−13.1	−15.9	−14.7	−14.5	−10.0	−5.5
2014:1−2016:4	−29.3	−27.6	−25.3	−25.3	−24.8	−23.4	−19.9
2009:1−2016:4 Average across all price tiers	−17.3						
2009:1−2013:4 Average across all price tiers	−12.1						
2014:1−2016:4 Average across all price tiers	−25.1						

Notes: Based on matched-model indexes for each price tier. AWS offered different sets of price tiers in different periods, so not all tiers have entries for every period.

unweighted average of price change across all the price tiers. All told, prices for S3 storage fall at an average *annual* rate of more than 17 percent over the full sample. Over subperiods, the pattern is the same as that for EC2 prices. Prices fell at an annual average rate of about 12 percent from the beginning of 2009 to the end of 2013. Then, in early 2014, just as Microsoft had entered the market to a sufficient degree that they were posting their cloud prices on the internet, AWS began cutting prices more rapidly. That started with the big price drop in early 2014, and over the period from the start of 2014 to the end of 2016, S3 prices fell at an average annual rate of about 25 percent.

13.4 How Big Is the Cloud?

Official revenue data for the cloud services industry and its main products according to nomenclature used in this chapter are not available. Nonetheless, a natural starting point is the Bureau of Economic Analysis (BEA) data on the closest intermediate-use category in the input-output account, 514 (Data Processing, Internet Publishing, and Other Information Services). This category includes data for North American Industry Classification System (NAICS) industry 518200 (Data Processing, Hosting, and Related Services), which subsumes much of the relevant core cloud services activity but includes other information services as well.[20] These data suggest the intensity of business use of purchased cloud services has been rising steadily (figure 13.4a). Because this category of spending is very coarse, it does not highlight the dynamism and explosive growth of cloud services, however. For example, the latest Census revenue data for Data Processing, Hosting, and Related Services (NAICS 518200) grew 8 percent and 10 percent in 2015 and 2016, respectively. While these rates of change are rapid relative to the overall economy, according to Amazon's company reports, AWS revenues grew 70 percent and 55 percent, respectively, in these calendar years.[21]

Using a broader definition of the cloud, Cisco Systems estimates that since emerging in the mid-2000s, the cloud model has rapidly dominated the data center market. Cloud data centers currently account for 90 percent of data center traffic and have accounted for essentially all growth since 2010 (figure 13.5). Indeed, traffic at cloud data centers rose at a 62 percent average annual rate between 2010 and 2016. This concept of cloud data centers, however, also does not correspond directly to the *purchased* services discussed in the previous paragraph for at least three reasons. First, it includes traffic

20. The structure of NAPCS (North American Product Classification System), introduced in 2017, usefully distinguishes among website hosting, data storage services, and so forth but does not distinguish between services provided by traditional data centers and those provided by cloud vendors. See the industry description at "North American Industry Classification System," US Census Bureau, https://www.census.gov/eos/www/naics/index.html, and the NAPCS structure at "North American Product Classification System," US Census Bureau, https://www.census.gov/eos/www/napcs/, both accessed March 5, 2017.

21. Data referred to in this paragraph were accessed September 10, 2018.

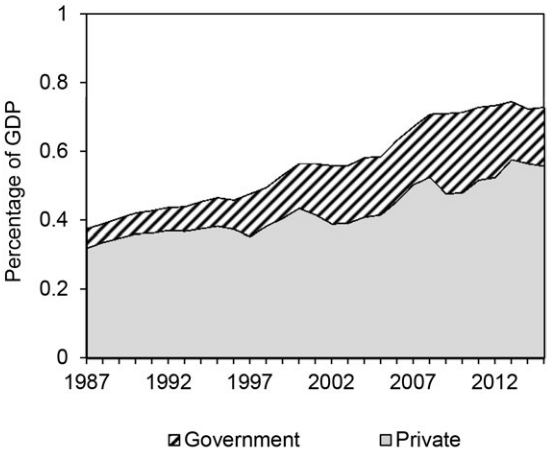


Fig. 13.4a Intermediate uses of information services, 1987 to 2015
Note: Data processing, hosting, and other information services products, wherever produced (BEA IO product code 514, covering 2002 NAICS 5182, 51913).

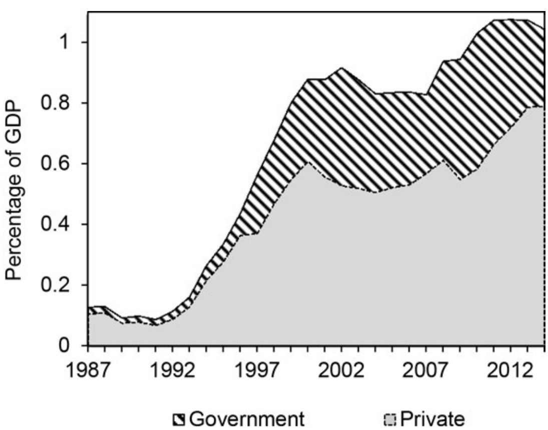


Fig. 13.4b Intermediate uses of computer and network design services, 1987 to 2015

related to the massive core centers used for "free services"—for example, Google's centers for its Gmail service. Second, Cisco's measure of cloud activity includes traffic at dedicated centers designed but not owned by IT services companies (e.g., IBM Cloud Services). Payments for these services likely are included in the NAICS 541512 (Computer and Network Design Services) industry. Revenues in this industry have grown especially rapidly relative to GDP (figure 13.4b).

Third, the Cisco measures reflect the rise of the "edge" cloud, which has

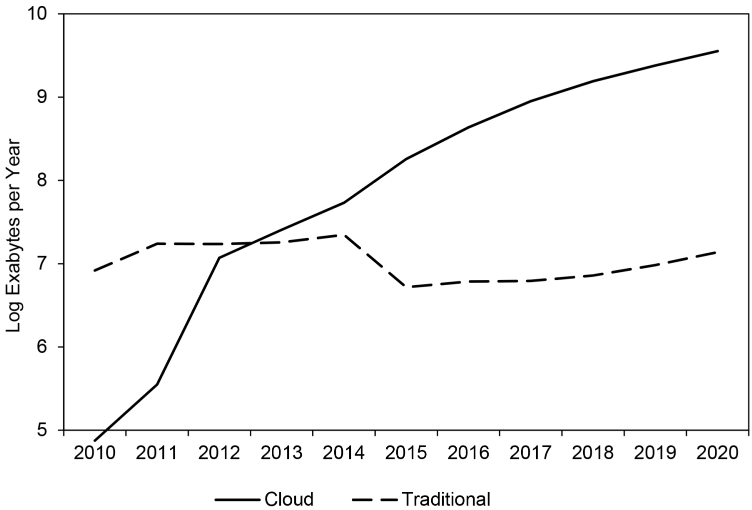


Fig. 13.5 Global data traffic by datacenter type, historical and projected, ratio scale
Source: Cisco Global Cloud Index, Forecast and Methodology, 2015–20 and earlier editions.

a restraining effect on both traffic and underlying business IT costs. A host of new technologies—including the Internet of Things (IoT), augmented and virtual reality, autonomous cars, drones, and smart cities—has led to an explosion in the volume of data that, given current bandwidth, cannot feasibly be transmitted to and from the cloud for processing in real time. Accordingly, this development has led businesses and governments to locate the processing and storage of their massive data collections locally or near the perimeter (i.e., near the "edge") of internet providers networks. Without going into details (but see AT&T 2017), edge computing streamlines the flow of data, transmitting only higher-value data (e.g., data from multiple IoT sources) to a shared central cloud center for further processing and analytic use.

Concurrently, capital expenditures at hyperscale cloud service providers have surged in recent years, rising at an annual rate of 21 percent during 2010 to 2015. Moreover, these expenditures now have reached roughly $50 billion per year, similar in magnitude to capital expenditures at telecom service providers (figure 13.6).[22]

Figure 13.7 shows the importance and rapid growth of the cloud from a different perspective: the share of the world's most powerful computers

22. Cisco classifies a data center operator as hyperscale if they have revenue of $1 billion in Iaas/Paas, $2 billion in SaaS, $4B from internet/search/social networking, or $8 billion from e-commerce / payment processing. Figure 13.6 includes the companies meeting this definition that provide cloud services.

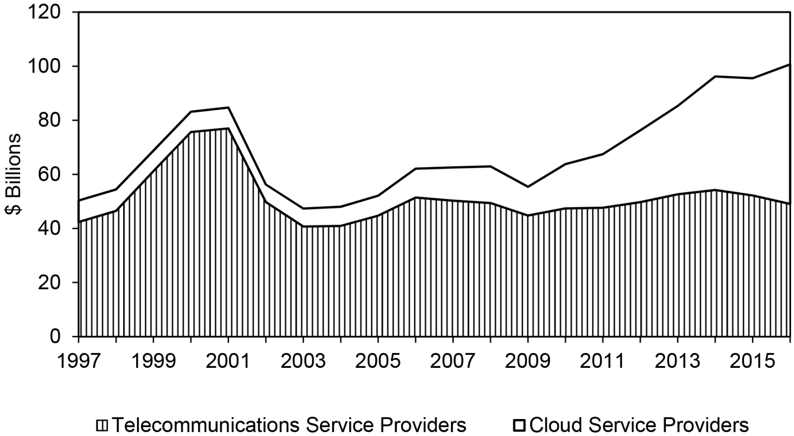


Fig. 13.6 US company capital expenditure Selected IT service providers
Source: Authors' tabulation of company financial filings.
Note: Included cloud service providers meet Cisco definition of hyperscale. Included telecommunications service providers are AT&T, Verizon, Sprint, T-Mobile US, Century Link, and related companies.

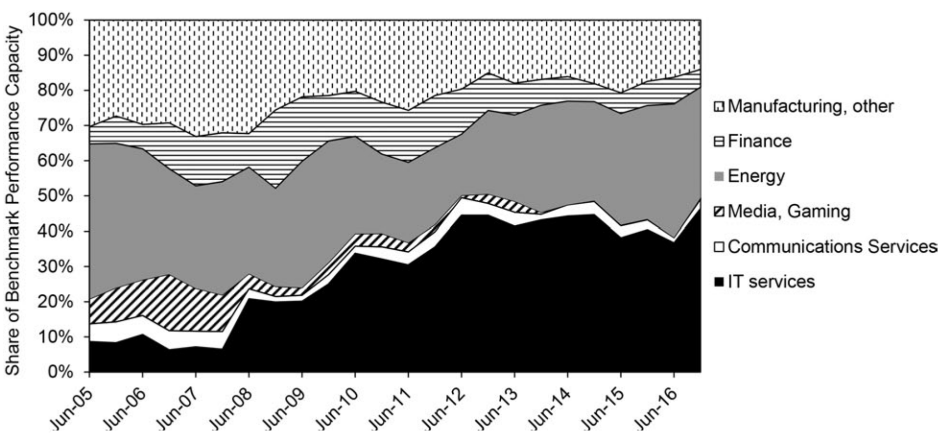


Fig. 13.7 Industrial supercomputer capacity by sector
Source: Top500.com, authors' calculations.

operated by IT service firms leapt from under 10 percent in 2006 to more than 40 percent in 2009 and has persisted at that level since.[23]

And tying back to the discussion of virtualization, IT consultancies com-

23. The IT services category is necessarily broader than cloud services because the descriptions of individual supercomputing sites vary in specificity. That being said, some sites are identified as Microsoft Azure and AWS.

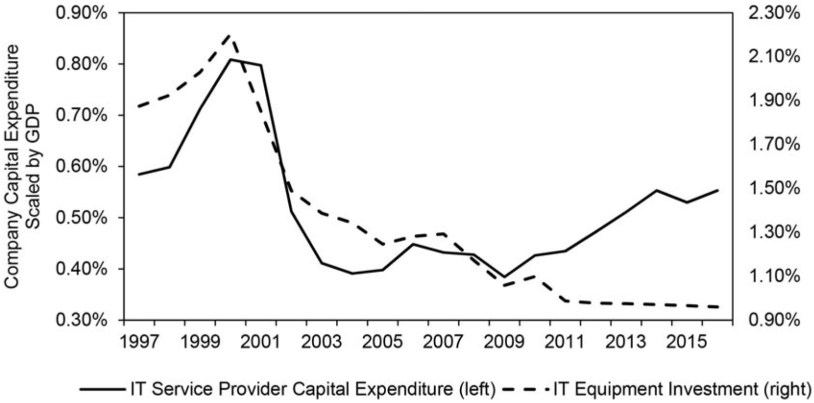


Fig. 13.8 Capital expenditure, selected US IT service providers and NIPA nominal IT equipment investment

Source: Bureau of Economic Analysis. Authors' tabulation of company financial reports.

Note: IT equipment investment includes communications equipment, computers, and peripherals. Included cloud service providers meet Cisco definition of hyperscale. Included telecommunications service providers include AT&T, Verizon, Sprint, T-Mobile US, Century Link, and related companies.

mented in 2008 that server virtualization had become the "killer app" for the business datacenter. Subsequently, IDC estimated that the number of virtual machines (VMs) per server in the United States—an indicator of the application workload of an enterprise server—advanced nearly 12 percent per year from 2007 to 2013 (Byrne and Corrado 2017b).

13.4.1 Where Has All This Investment Gone?

How well does this financial data align with official measures? Mapping company reports to official industry statistics is challenging. Companies providing cloud services provide a host of other IT services as well. Consequently, their establishments undoubtedly are classified to a variety of industries, most notably the industries in NAICS subsectors 511 (Publishing Industries, except Internet [includes Software]), 513 (Broadcasting and Telecommunications), and 519 (Other Information Services).

In light of this wave of investment by cloud service providers, the continuing shift away from IT equipment in business fixed investment in equipment and intangibles may be seen as puzzling. Figure 13.8 plots NIPA nominal IT investment and the capital expenditures figure for cloud service providers from figure 13.6 as shares of GDP. As shown, these two series tracked fairly closely from the mid-1990s through about 2009 as IT investment tailed off as a share of GDP. But after 2009, these series diverged sharply as capital expenditures surged while the series for NIPA IT investment remained sluggish. One possible explanation is the higher utilization that follows as firms outsource IT functions to the cloud. Such an increase in utilization could translate into weaker investment in the short run. Indeed, IDC Inc. reports that the nominal value of sales of servers to US firms fell at an annual average rate of 11 percent from 2004 to 2016, and the decline has accelerated since 2008.

That being said, we also consider another possibility: that cloud services firms have been building their own IT equipment, at least in part.[24] If so, then a portion of the capital expenditures reported above may be for components that have gone into IT equipment built on an own-account basis rather than for already-assembled IT equipment. Google, for example, is reported to have built both computing and network equipment from purchased components.[25] Consistent with this possibility, the "use tables" published by the US Bureau of Economic Analysis indicate that the output of the Computer and Electronics Manufacturing sector (NAICS 334) used by IT services sectors is substantial—$58.6 billion in 2015.[26] At the same time, the "make tables" indicate that these electronic intermediates are not made into final electronics sold by the IT services sector. This suggests that these components are used for own-account production of IT equipment used within the firms.

If this story is correct, this own-account investment should be (but we believe likely is not) counted in the NIPAs as business investment in IT equipment, albeit own-account investment. How much might this own-account investment add up to? For the sake of argument, we assume that the omitted investment value of the own-account production of final electronics is equal to the value of the electronic intermediates used.[27] With this valuation, the story for business investment in IT equipment changes markedly. As seen in figure 13.9, nominal IT equipment and software investment, including our estimate of own-account investment, would be $58 billion higher in 2015 than in the official estimates, amounting to 0.32 percent of GDP. For nominal investment in IT equipment, adding this own-account investment would boost the average annual growth rate during 2007–15 by roughly 2 percentage points compared with official estimates. For nominal GDP growth, including this own-account investment would add three basis points per year to the growth rate during this period.

13.5 Conclusion

We find that cloud computing has exploded. By available measures, the quantity of cloud activity has grown extremely rapidly, as has associated

24. A parallel presentation of own-account investment by cloud service providers appears in Byrne, Corrado, and Sichel (2017).

25. See *Wired* (2015).

26. We treat BEA categories 511, 512, 514, and 5415 as IT services. This group includes industry 518210 mentioned above (in category 514) as well as software publishing, telecom services, and computer design services.

27. We believe this assumption is conservative; although the details of data center server inputs are not available, Gartner Inc. reports that the market value of personal computers is roughly four times the value of electronic inputs.

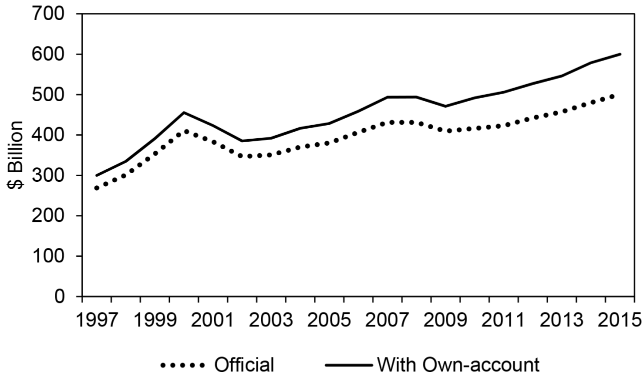


Fig. 13.9 IT equipment and software investment
Source: US Bureau of Economic Analysis, authors' calculations.

capital investment. At the same time, prices of basic cloud services have fallen rapidly since 2009, based on a unique dataset we assembled. However, because cloud is so new and so much of it is intermediate input, it is challenging to track in the statistical system, and the available data do not distinguish between cloud-based and traditional services, whether services are purchased or produced internally or generated at the "edge." We highlight one area where real GDP may be understated by a noticeable amount as a result of changes in the economy related to the rise of cloud computing.

References

AT&T Edge Cloud. 2017. AT&T Labs and AT&T Foundry, white paper. https://about.att.com/content/dam/innovationdocs/Edge_Compute_White_Paper%20FINAL2.pdf.

Avram, Abel. 2016. "FaaS, PaaS, and the Benefits of the Serverless Architecture." *InfoQ*, June 25, https://www.infoq.com/news/2016/06/faas-serverless-architecture.

Bresnahan, Timothy F., and Shane Greenstein. 1999. "Technological Competition and the Structure of the Computer Industry." *Journal of Industrial Economics* 47 (1): 1–40.

Byrne, David M., and Carol A. Corrado. 2017a. "Accounting for Innovation in Consumer Digital Services: Implications for Economic Growth and Consumer Welfare." Paper presented at the 5th IMF Statistical Forum, "Measuring the Digital Economy." IMF Headquarters, Washington, DC, November 16, 2015. http://www.imf.org/~/media/Files/Conferences/2017-stats-forum/carrado.ashx?la=en.

Byrne, David M., and Carol A. Corrado. 2017b. "ICT Prices and ICT Services: What Do They Tell Us about Productivity and Technology?" *International Productivity Monitor* 33 (Fall): 150–81.

Byrne, David M., Carol A. Corrado, and Daniel E. Sichel. 2017. "Own-Account IT Equipment Investment." FEDS Notes, October 4. Washington, DC: Federal Reserve Board. https://www.federalreserve.gov/econres/notes/feds-notes/own-account-it-equipment-investment-20171004.htm.

Byrne, David M., Stephen D. Oliner, and Daniel E. Sichel. 2018. "How Fast Are Semiconductor Prices Falling?" *Review of Income and Wealth* 64 (3): 679–702.
Campbell-Kelly, Martin, and Daniel D. Garcia-Swartz. 2008. "Economic Perspectives on the History of the Computer Time-Sharing Industry, 1965–1985." *IEEE Annals of the History of Computing* 30 (1): 16–36.
Cohen, Robert B. 2017. "Understanding the Next Production Revolution and its Relationship to Software Innovations." Mimeo, Economic Strategy Institute, January 22.
Coyle, Diane, and David Nguyen. 2018. "Cloud Computing and National Accounting." ESCoE Discussion Paper 2018-19, December.
DataDog. 2015. "8 Surprising Facts about Real Docker Adoption." October 2015, last updated June 2016, accessed February 25, 2017. https://www.datadoghq.com/docker-adoption/.
De Roure, David, Mark A. Baker, Nicholas R. Jennings, and Nigel R. Shadbolt. 2003. "The Evolution of the Grid." In *Grid Computing: Making the Global Infrastructure a Reality*, edited by F. Berman, G. C. Fox, and T. Hey, 65–100. Wiley Series in Communications Networking and Distributed Systems. Chichester: John Wiley.
Elliot, Stephen, and Randy Perry. 2018. "Adopting Multicloud: A Fact-Based Blueprint for Reducing Enterprise Business Risks." International Data Corporation White Paper, June.
Erickson, Tim, and Ariel Pakes. 2011. "An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods." *American Economic Review* 101 (August): 1707–38.
Hubbard, Patrick. 2014. "The Network in the Computer, Again." *Network Computing*, May 6. http://www.networkcomputing.com/cloud-infrastructure/network-computer-again/1827958867.
Kushida, Kenji E., Jonathan Murray, and John Zysman. 2011. "Diffusing the Cloud: Cloud Computing and Implications for Public Policy." *Journal of Industry, Competition and Trade* 11 (3): 209–37.
Kushida, Kenji E., Jonathan Murray, and John Zysman. 2015. "Cloud Computing: From Scarcity to Abundance." *Journal of Industry, Competition and Trade* 15 (1): 5–19.
Mell, Peter, and Timothy Grance. 2011. *The NIST Definition of Cloud Computing*. NIST Special Publication 800-145. http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.
Murray, Jonathan. 2017. "The Third Wave." *Medium*, December 19. https://medium.com/@adamalthus/the-third-wave-b4ec5380079a.
OECD. 2014. *Addressing the Tax Challenges of the Digital Economy*. OECD/G20 Base Erosion and Profit Shifting Project. Paris: OECD.
Pakes, Ariel. 2003. "A Reconsideration of Hedonic Price Indexes with an Application to PCs." *American Economic Review* 93 (5): 1578–96.
Reiss, Spencer. 1996. "Power to the People." *Wired Magazine*, December 1. https://www.wired.com/1996/12/esgage/.
van Dalen, Jan, and Ben Bode. 2004. "Estimation Bias in Quality-Adjusted Hedonic Price Indexes." Mimeo, Rotterdam School of Management. http://www.ipeer.ca/papers/vanDalenBodeOct.1,2004,SSHRC%20Paper17.pdf.
Wired. 2015. "Like Google and Facebook, Twitter Designs Its Own Servers." July 9, 2015. https://www.wired.com/2015/07/like-google-facebook-twitter-designs-computer-servers/.
Zhang, Liang. 2016. "Price Trends for Computing Services." Senior Thesis, Wellesley College, May.