Spatial Competition and Missing Data: an Application to Cloud Computing

November 13, 2020

Abstract

The internet was hypothesized to be the "death of distance". We investigate this hypothesis with a novel anonymized customer level dataset on demand for cloud computing accounting for both spatial and price competition among public cloud providers. We introduce a mixed logit demand model of spatial competition estimable with detailed data of a single firm but only aggregate sales data of a second. We leverage the EM algorithm to tackle the customer level missing data problem of the second firm. Estimation results and counterfactuals show that standard spatial competition economics hold even when distances for cloud latency is trivial.

Keywords: hybrid data, EM algorithm, mixed logit

1 Introduction

For a variety of reasons, firms historically care about physical location when making decisions about where to invest in physical capital.[1] Because the internet lowers costs of communication over space, it was reasonable to suspect internet adoption could mitigate the importance of physical location for investment decisions. According to this theory, the internet allows rural firms to have access to similar resources as urban firms without moving to urban locations, allows faster rural growth and ameliorates incidence issues of agglomeration economies (Forman et al. [2008] and Glaeser and Gottlieb [2009]).

Evidence of the internet leading to the "death of distance" is scant, however, for regional economic outcomes. At the regional level, Forman et al. [2012] shows that advanced internet is not sufficient, in and of itself, to enable wage growth in a city. Rather, they find advanced internet is a complement to a city's existing human capital and stock of firms. At the firm level, Giroud [2013] shows that between 1977-2005, decreased travel times to plants through new plane routes or roads led to increased investment and productivity. Hence, traditional investment decisions vary positively with proximity. A sharp test for the death of distance hypothesis, then, is how the internet impacts the affinity for proximity in firm investment decisions.

In this paper, we ask whether firm investment decisions enabled solely by the internet, cloud computing, systematically shows a predilection for proximity. Cloud computing providers, like Amazon Web Services (AWS) and Microsoft's Azure, rent compute resources called "virtual machines" to customers who connect to them via the internet. Once connected to a virtual machine (VM) via the internet, cloud users can perform functions previously required to be hosted by on premise servers like compute operations, read and write data operations or web application hosting.[2]

There is a large and growing demand for cloud computing: according to Gartner, the worldwide public cloud services market is projected to grow 17.3 percent in 2019 to total $206.2 billion, up from a $175.8 billion forecast in 2018.[3] Cloud computing is also important for general economic growth and productivity because renting cloud compute resources lowers fixed hardware capital costs for start-up firms and turns them into marginal operating expenses. Despite this fast growth and economic importance, there is very little empirical work on understanding the cloud computing market and welfare derived from it.

When a firm deploys a cloud computing instance, they must choose a physical location where to deploy it. Cloud providers have multiple data center locations and prices vary by location. Working with a far-away data center can impact latency. Latency is the time between a task request and task execution. As a rule of thumb, distance related latency is on the order of one millisecond per 100 miles.[4] To put that in context, 2017 research by Google finds that "average time to first byte" (a measure of web server responsiveness) was roughly 2,000 milliseconds when averaged across all websites in the U.S.[5] Thus, an increase of roughly 1,500 miles of distance between a VM user and the data center location

[1] Monitoring and information acquisition is one explanation (Giroud [2013]). Transportation costs, despite long run declines, are another (Glaeser and Kohlhase [2004]) Agglomeration economies are a third (Glaeser and Gottlieb [2009]).

[2] What was formally a fixed investment costs of server ownership into variable costs. With some configurations or "stacks" cloud users can outsource IT personnel to experts employed by the cloud providers.

[3] See "Gartner Forecasts Worldwide Public Cloud Revenue to Grow 17.3 Percent in 2019", accessed on 09/30/2018, https://www.gartner.com/en/newsroom/press-releases/2018-09-12-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2019

[4] See https://www.365datacenters.com/portfolio-items/beyond-bandwidth-distance-matters-choosing-data-center/. Data packets travel at the speed of light in a vacuum but in actual fiber which powers the internet, the speed is a bit less. Further, bends in fiber cables can slow down data speeds. Lastly, there are other factors not related to distance which can increase latency like congestion and reading data packets are not directly related to distance.

[5] See https://think.storage.googleapis.com/docs/mobile-page-speed-new-industry-benchmarks.pdf.

corresponds to a latency increase of ∼15 milliseconds (less than a 1% increase in time to first byte for the average U.S. web server). Hence, over relatively small distances like those faced by customers in our dataset (e.g., choosing between a data center 1,000 versus 1,500 miles away) latency might not always be the sole disutility of selecting far away data centers outside of niche use cases like high frequency trading.

We test two hypotheses for how distance impacts firm investment decisions in this paper. First, we investigate the strength of firm preferences for proximity when investing in the cloud. Whereas some firms, such as high frequency traders, care a great deal about 5 milliseconds of latency, many users do not. As a result, we view this as a strong test for the importance of distance and firm investment decisions: if firms are willing to pay more to use the nearest data center when only marginally closer than another data center, it is evidence against the "death of distance" hypothesis.

Second, we estimate how competition impacts the "death of distance" hypothesis. Public cloud providers like AWS, Azure and Google Cloud Platform (GCP) are fiercely competing in the rapidly growing cloud computing market. Market competition can impact both which cloud provider customers choose and, for their chosen provider, which specific data center location customers choose. We develop and estimate a structural model of cloud demand to investigate how strategic firm level decisions interact with preferences for proximity.

Having the type of data we have, detailed data for a single firm and aggregate data for another, is a common problem with both developing business strategy and in competition policy. We introduce a novel mixed logit demand model of spatial competition that is estimable with detailed data of a single firm but only aggregate sales data of a second and apply it to the cloud computing industry. The model lets us perform counterfactual analysis over 1) how spatial competition between cloud providers impacts optimal price setting behavior and 2) optimal data center locating decisions. We can thereby show how firms' cloud investment patterns change with competition upstream in the cloud computing provider industry.

We use a proprietary dataset with anonymized customer level zip codes linked to the location of data centers they choose. The dataset consists of all customers who deployed one popular type of VM on Microsoft's Azure in 2016. At the time, Azure was the second largest public cloud provider in the world behind AWS. We restrict the dataset to focus on location decisions of US and Canadian firms to locate in US and Canadian data centers. We leverage the rollout of new data centers in the U.S. and Canada over our time period to provide variation in the choice set of data centers and identify key demand parameters, allowing a subset of demand parameters to vary by a cloud user's industry. Due to large lead times in data center construction and 2016 being early in the public cloud sector, we argue data center location is plausibly exogenous.[6]

We have detailed data of a single firm, Azure, but only aggregate sales data of a second, AWS. Specifically, we use quarterly cloud revenue data from AWS available in their 10-Ks. We treat absent customer level data from the AWS as a missing data problem and leverage the structural model, detailed Azure data and Expectation Maximization (EM) algorithm to back out AWS customer location. The EM algorithm addresses the missing data problem iteratively: we first construct an expectation of the likelihood by integrating over the latent consumer locations based on their posterior distribution, and then maximize the likelihood function over demand parameters.

Identifying key demand parameters relies on the rollout of new data centers by both AWS and Azure and 2016 price changes. By observing the rate at which new customers begin purchasing Azure when

[6]We argue below that over the planning period for these data centers, reliability for servicing internal workloads was the primary reason for data center construction. For example, a two year lead time for data center construction implies that 2014 was the planning period. In 2014, share prices for Microsoft hadn't yet responded to increased cloud revenue reported in 10-Ks.

new AWS or Azure data centers open and how those rates vary over space, we can identify preferences for proximity to data centers. We argue that using data from early days of the cloud computing market and the long lead times to construct new data centers as a source of identification is adequate to identify preference parameters. We project demand parameters identified from the granular Azure data to the observed AWS data center characteristics and sum across AWS data centers. The gap between the projection and the observed AWS market share is attributed to fixed effects of cloud providers. Lastly, the population distribution of consumers can be inferred by the choice probabilities calculated from the identified demand model and the observed market shares of Azure, accounting for the presence of an outside good (on premise servers). We show via simulation that the model successfully recovers the demand parameters and unobserved consumer spatial distribution then take the model to the data given Azure's market share.

Our core empirical finding is that cloud users have a preference for nearby data centers. As a result, the spatial layout of DCs relative to customer location induces a significant variation in local market power. We have no identifying variation to test whether this preference for proximity is driven by latency concerns or other factors like a secular preference for proximity. It is hard to imagine latency is driving the magnitude of preference for proximity we find in the data. Our data covers North America and Canada only. Introduction of new data centers in our sample often change distance to nearest data center by only a few hundred miles or latency decreases of a few milliseconds. However, we find that cloud customers are willing to pay roughly 60% premiums for a reduction in distance of roughly 600 miles (i.e., 1000 kilometers).

We use estimated parameters to perform counterfactual exercises to determine how market structure would change if new data centers are introduced in different locations. Among the six possible counterfactual Microsoft Azure data center locations, the most profitable one could generate a market share gain roughly 25% higher than the least. Thus, the revenue reductions to cloud providers of placing a data center sub-optimally are large. The model lets us decompose increases in market share across customers purchasing the outside good (on premise servers) versus purchasing from a competitor and we find that much of the increase in market share is from the outside good although a meaningful share is from the competitor.

We also perform a counterfactual where we decrease price of all Azure data centers by 15% and assess changes on market shares. Consistent with economic theory of spatial competition, we find that the benefits of price competition are greatest where both Azure and AWS have a data center. Thus, our results provide evidence that spatial competition is important in the early stages of the cloud computing industry. Comparing the two counterfactuals, opening a new data center increases consumer surplus by roughly 75% of the consumer surplus from the price decrease. This is large since the new data center would impact only ∼10% of all Azure customers (e.g., surplus increases only for cloud users that deploy there) but it is plausible given the implied willingness to pay for proximity.[7]

There are three main lessons from this research. First, we find evidence that cloud customers display a material preference for proximity in deploying VMs that is hard to explain with latency issues. Indeed, we show that a large fraction of cloud users do not deploy in the nearest DC implying that latency is

[7] As a back of the envelope calculation if all customers receive a 15% price decrease their customer surplus increases by 15%. A new data center impacts those customers that deploy in it and there were 10 DCs at the end of our sample so roughly 10% of customers benefit from the new DC. Recall that aggregate consumer surplus from the new DC is 75% of the welfare increase from a 15% price decrease for all newly deploying customers. If N are the total number of cloud customers then $.1 * N * \Delta CS_{newDC} = .75 * N \Delta CS_{PriceChange} = .75 * N * .15$ and solving for ΔCS_{newDC} yields the increase in consumer surplus for customers deploying in the new DC in our counterfactual. Hence we must observe an increase in consumer surplus of (.15/.1)*.75 = 112.5%. This is plausible: a new proximate DC could be worth roughly twice as much to cloud users as distant DC based on our parameter estimates.

often not a major hurdle to cloud deployments. Data center age, for example, correlates with where cloud customers deploy. By focusing on the North American market we ignore data sovereignty issues but highlight that those are likely to also be important. We view a preference for proximity as being inconsistent with internet enabling the "death of distance" at least over our sample in the early stages of cloud adoption. We acknowledge that latency could be an important issue for some cloud applications, but the magnitude and scale over which preferences for proximity manifest and the observed distances in our dataset (we observe only North American customers) makes it difficult for latency to be plausibly responsible.

Second, our findings imply that market competition could help mitigate incidence issues from spatial allocation of capital. The "death of distance" narrative promised increase growth in rural areas attributable to better access to information and freer flow of goods and services. Our results imply that increased strategic spatial competition as the cloud market matures would reduce equilibrium prices and also increase incentives to invest in additional data centers. Although we don't endogenous DC location decisions in this paper to address it formally, intense competition among the major cloud providers (e.g., AWS, Azure and Google Cloud Platform, Alibaba, etc.) is likely increasing access and surplus to the cloud for all potential cloud customers across both the price and distance margins.

Third, more generally our results show that product managers for goods characterized by spatial competition can effectively estimate demand for their goods using detailed data of only a single firm so long as market data for the competitor firm exists and there is variation in the number of stores over time. In addition to benefits to managers, we highlight how this technique can also be used by economists to perform welfare analysis. While our use case is cloud computing, the technique could be useful for managers and researchers interested in questions regarding the impacts of opening and closing of brick and mortar stores faced with increasing online competition.

This paper contributes to three strands of literature. First, understanding the economic geography of the internet has important incidence implications. Despite higher adoption rates for early internet in rural areas (Forman et al. [2005]) it appears that the benefits enabled by the internet accrue in only a subset of cities (Forman et al. [2012] and Forman et al. [2008]). Further, recent research suggests that proximity to data centers could cause increased growth Jin and McElheran [2019]. Our work pushes these findings by investigating how spatial competition could impact the economic geography of internet enabled economic gains.

Second, in the field of discrete choice modeling, applications of EM algorithm date back at least to Bhat [1997], Train [2007] and Train [2008]. Many of these applications use EM to address missing data on consumer attributes. In what might be the most closely related EM based approach to ours, Conlon and Mortimer [2013] addresses missing data on product availability. At a high level, competitor sales are similar to missing data regarding any product generally. Unlike these previous papers, though, the data structure in our case has two problems: the aggregate level competitors' data makes both their consumer's attributes and disaggregated (e.g., store level) sales unobservable. Because this is a spatial model of competition, the consumer-store level attributes of AWS are of added importance.

While we view the EM algorithm as the most appropriate remedy for our missing data problem for both efficiency and computational feasibility, there are other related techniques in the literature. Other demand frameworks for a similar data structure include those in Berry et al. [2004], a Berry et al. [1995] inspired model leveraging micro moments of consumer characteristics. However, these "Micro-BLP" models are less efficient than maximum likelihood estimation (MLE) by attenuating the information on choices at individual level. The marketing literature often uses Bayesian techniques in the sense that demand parameters are also treated as latent variables. Examples includes but are not restricted to Chen and Yang [2007], Musalem et al. [2008], Jiang et al. [2009] , Musalem et al. [2010] and Zheng et al. [2012].

Specifically, Feit et al. [2013] is probably the most related work to ours. They use a mixture of individual level usage data for digital platforms and aggregate data on usage for traditional platforms to estimate the multi-platform media consumption, albeit in a context of a multivariate model and computationally more burdensome because of the inevitable Markov Chain Monte Carlo simulation.

Third, this paper expands the existing literature on spatial competition broadly in addition to our application regarding the cloud computing industry. In terms of data structure, previous works on spatial competition usually use either aggregate or disaggregate data only. For instance, Davis [2006] estimated a model of spatial competition in the movie theater industry with market share data. Davis [2006] aggregates consumer heterogeneity with an observed geographic consumer distribution from census data and then focuses on identifying the functional form of travel cost. Smith [2004] estimates a two-stage discrete-continuous model for the supermarket industry, and the complexity of unobserved consumer attributes is circumvented by consumer level data from a survey. While spatial competition and firm entry decisions are important economic questions (Seim [2006]), our novel demand estimation approach to combine micro and macro data can be applied to estimating demand elasticities as well.

The remainder of the article proceeds as follows. In Section 2 and 3, we give a brief introduction of the IaaS public cloud industry and describe the general framework of the model. Section 4 describes the model, describes how EM algorithm can be employed to address the missing data problem, and identification. Section 5 shows the performance of a Monte Carlo experiment. In Section 6, gives the estimation results from the data. Section 7 performs two counterfactual exercises highlighting the spatial competition aspects of cloud implied by estimated preference parameters. We conclude this paper in Section 8.

2 Industrial Background

According to the beginner's guide on the website of Microsoft Azure,

"Cloud computing is the delivery of computing services–servers, storage, databases, networking, software, analytics, and more–over the Internet ('the cloud'). Companies offering these computing services are called cloud providers and typically charge for cloud computing services based on usage, similar to how you are billed for water or electricity at home."

Most cloud computing services fall into one of three broad categories: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). In this paper, we focus on IaaS and model consumer's problem as a discrete choice among data centers. Focusing on IaaS over PaaS and SaaS is ideal in our setting because the user must specify a specific DC to deploy their cloud resources. Alternative PaaS and SaaS offerings often have a more curated experience in which the firm makes deployment decisions.

DCs are facilities that house computer systems and associated components, such as telecommunications and storage systems. Consumers rent virtual machines (VMs) at DCs as complements to local machines on a pay-as-you-go basis. The value proposition to customers is driven capacity management, cloud providers' economies of scale and management of hardware and security. Put another way, replacing lumpy capital expenditures on wholly-owned servers with smoother operating expenses in the cloud, being able to scale up and down demand for compute resources but not always provision for max demand as with own servers, and outsourcing hardware security concerns all are valuable. Some use cases include housing large datasets, serving website, web App or Application content and performing period machine learning model training.

Amazon Web Services (AWS) and Microsoft Azure are the two firms that have the largest market

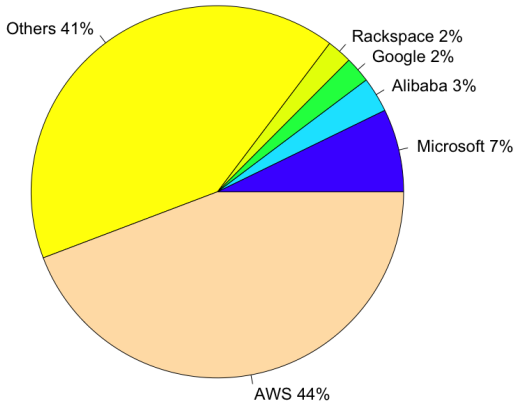
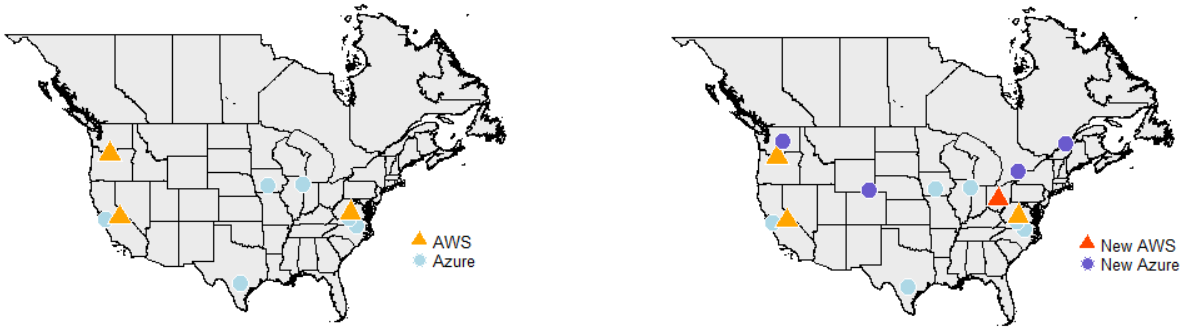


Figure 1: 2016 IaaS Public Cloud Computing Market Share from Gartner. Gartner data from survey results of firms. Market shares materially changed over the past five years so that Azure and Google's GCP now have much larger shares.

shares in global IaaS public cloud market. In 2016, the year our data spans, the total value of this market reached $22 billion U.S. dollars, of which AWS had 44%, followed by Microsoft Azure at 7.1% [8], as shown in Figure 1. Total cloud demand has increased significantly since 2016 to $44.4 Billion in 2019 with AWS's market share staying roughly constant but Azure's growing to 18%.[9]



(a) Jan 2016 (b) Dec 2016

Figure 2: North American Data Center (DC) layout in 2016. During the calendar year both AWS and Azure added DCs in different parts of the U.S. and Canada. We leverage how new DC introduction differentially impacts customers in different locations to estimate preferences fro DC proximity.

The core distinction between on-premise servers and cloud computing is that cloud customers rent compute resources from a public cloud provider. When purchasing an on premise server, a firm puts the server in their compute facility, normally in their office building. When a customer decides to rent compute resources and configure a VM, they select a physical location for that VM to be located. Figure 2 shows the location for all U.S. and Canadian data centers of both AWS and Azure. A shorter

[8] See "Gartner Says Worldwide IaaS Public Cloud Services Market Grew 31 Percent in 2016", accessed on 11/01/2018, https://www.gartner.com/en/newsroom/press-releases/2017-09-27-gartner-says-worldwide-iaas-public-cloud-services-market-grew-31-percent-in-2016.

[9] See https://www.gartner.com/en/newsroom/press-releases/2020-08-10-gartner-says-worldwide-iaas-public-cloud-services-market-grew-37-point-3-percent-in-2019.

Table 1: Comparison between Microsoft *basic A1* and AWS *t2.small*

Name	Brand	vCPUs	RAM(GiB)
basic A1	Mircosoft	1	1.75
t2.small	Amazon	1	2

Note: We compare demand for customers' first deployment of basic A1 for Microsoft and estimate first deployments of AWS' t2.small. These products are similar in terms of performance. Differences in product fixed effects will be covered by AWS fixed effects in the empirical model.

physical distance between a VM and its users is correlated with lower latency (e.g., shorter wait times for webpages to load). Each firm had a footprint in Canada by the end of 2016. We estimate demand of a single popular SKU for all U.S. and Canadian consumers with workloads in any DC in either the U.S. or Canada.

Spatial proximity is likely an important aspect of DC differentiation for speed-sensitive users. Data transfer takes time, and the resulted latency could be further amplified due to security protocols. For example, this is likely to be a real concern when considering leverage cloud servers across on the other side of the globe. Our dataset, though, includes only US and Canadian customer demand for VMs within US data center locations. It is plausibly less likely to be an issue for domestic data center location decisions where the U.S. is roughly 3000 miles across and as a rule of thumb, distance related latency is on the order of 100 miles per millisecond. Observing a preference for proximity could be related to latency preferences or non-performance related preferences to be physically close to data centers (sometimes called "server hugging").

Spatial proximity is determined by both DC location and consumer location and consumer heterogeneity along this margin could play a critical role in this demand system. Therefore, the estimation for demand parameters overlooking consumer heterogeneity in location could miss an important consumer preference. This motivates our mixed logit framework. Although consumer location is only observable for Microsoft customers, the EM algorithm we detail below enables a simultaneous estimation of both demand parameters and consumer spatial distribution, which is needed for any policy analysis respecting spatial demand preferences.

3 Data

We merge several datasets together for our analysis. These datasets include actual purchase data from Azure customers including customer locations for a single popular general purpose product or shop keeping unit (SKU), aggregate sales for AWS, data center locations, pricing data for AWS and Azure, census data on business locations for the U.S. and the analog for Canadian businesses.

The most novel attribute of our data is a random subset of Microsoft customer level choice data for the a general purpose cloud computing SKU: *basic A1* SKU. The analog of the *basic A1* SKU for AWS we consider is the *t2.small* SKU. A detailed technical comparison between *basic A1* and *t2.small* can be found in Table 1. The VMs are similar across CPU and RAM. One difference between the virtual machines SKUs is across product quality: whereas *basic A1* is a dedicated core, *t2.small* is a burstable VM. That means scaling up *t2.small* cores due to a "burst" in compute demand might not always be available if deployed whereas a dedicated core would be. This will be picked up in the brand/product fixed effects we estimate in the empirical model.[10]

[10]Because we only evaluate one product from both AWS and Azure, brand and product fixed effects are operationally identical in this paper.

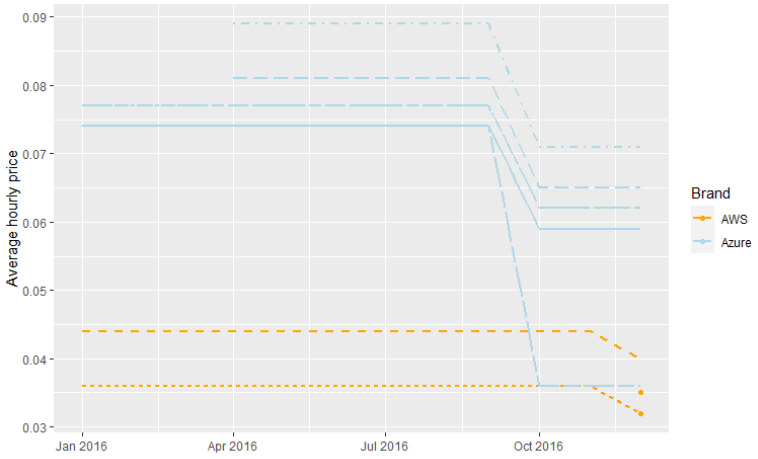


Figure 3: Prices of Microsoft *basic A1* and AWS *t2.small* Across Regions.
Note: Figure shows price dispersion for AWS and Azure over time and space. One large price drop for Azure and one small drop for AWS is responsible for identifying price coefficients. New region prices are shown as appearing midway through the year.

There was modest price variation in our sample period. Figure 3 shows region level prices across DCs for Azure's *basic A1* and AWS's *t2.small* in 2016. Prices are quoted in the hourly price of deploying a one core Azure *basic A1* or AWS *t2.small* VM. To put these prices into context, at \$0.08/hour a one core VM would cost \$700.80 if deployed for 24 hours a day for all 365 days in a year, which is more than a one core personal computer would have cost in 2016. The premium accounts for the ability to only pay for what is used (e.g., used the VM for 20 hours then shut it off), in addition to outsourcing security and IT.

Figure 3 shows when the first Canadian DCs of Microsoft and AWS were introduced in April, 2016 and Dec, 2016, those prices were slightly higher than prices in other regions for Azure. The price drop in Microsoft DCs in October 2016 and price decrease in December 2016 from AWS helps us identify price coefficients. Finally, prices were higher for Azures *basic A1* SKU relative to AWS's *t2.small* reflecting some time invariant differences in attributes which will be picked up by the brand/product fixed effects. Generally speaking, cloud providers tended to have a few large data centers at low prices then smaller data centers more geographically dispersed at slightly higher prices over our sample period.

Related to pricing, we focus on the location of initial VM deployment decision of customers in the paper. When a cloud user deploys a VM they pick a DC where the VM will be deployed. One advantage of the cloud relative to wholly owned servers is cloud user can turn off their VM at any time and stop paying in a "pay as you go" cloud computing business model. Usage decisions are second order for our question of willingness to pay for proximity for the cloud. In order to usage decisions to matter, there would need to be a substitution margin along which cloud users choose a location as a function of both their physical location and the expected duration of the deployment. We view this as unlikely although we discuss how adding a usage decision could impact the model and findings below. Also, by focusing on the initial deployment, we circumvent the complications of multiple DC users.

The Azure customer purchase data includes date of initial purchase, the specific DC location where the *basic A1* VM is deployed, the zip code for the customer and the industry of the customer when available. Table 2 shows a summary of observables for both the anonymized Azure data and the pricing data for both AWS and Azure. Table 2 also highlights the increase in data centers across both Azure (four to ten) and AWS (three to five) in 2016. The table also shows explicitly that we only observe the location of Azure customers. For this reason we leverage the EM algorithm to infer the location of AWS

customers.

Most customers do not have an industry associated with them so we classify them as "unknown". Roughly 25% of customers, however, do have an industry noted in our data. Observing industries is very likely non-random so the industry composition in Table 2 likely isn't representative of the overall customer industry of Azure. Based upon conversations with Microsoft employees, industry is often reported when a cloud user leverages an intermediary to deploy their cloud workloads (e.g., when the end customer uses a vendor cloud service provider to manage their cloud resources). Hence observing a reported industry could be a proxy for leveraging a vendor to operate cloud IT.

Table 2: Summary Statistics

Panel A: Consumer Characteristics		
Consumer Characteristics	**Microsoft Azure**	**AWS**
Locations	observable	unobservable
Industry	observable	unobservbale
Discrete Manufacturing	4.5%	
Education	1.1%	
Health	1.3%	
Hospitality & Transportation	1.0%	
Insurance	0.6%	
Media / Telecome and Utilities	1.3%	
Nonprofit	0.5%	
Professional Services	13.2%	
Choice	DC level	Brand level
Panel B: DC Characteristics		
DC Characteristics	**Microsoft Azure**	**AWS**
Locations	observable	observbale
Changes in number of DCs in 2016	$4 \rightarrow 10$	$3 \rightarrow 5$
Start date of Canadian DC	Apr, 2016	Dec, 2016
Average hourly price	*basic A1*	*t2.small*
	\$0.0708 (0.0120)	\$0.0381 (0.0040)

Note: Table summarizes data used in the analysis. Industry only reported for roughly 30% of observations in our data and professional services and discrete manufacturing appear overly represented for those observations reporting industry. Both AWS and Azure saw and increase in the number of DCs over the time period.

We use variation in the number of DCs over time to identify taste parameters for proximity to data centers. The intuition is as follows: consider two sets of cloud users all from Ohio. Azure has no DC in Ohio over our sample but AWS opens a DC in mid-2016. The first set of customers need to deploy in January 2016 before AWS opened a DC in Ohio. Hence, we would observe some customers with Ohio zip codes as signing up in Azure DCs in January 2016. AWS then opens a DC in Ohio. Assume the second set of customers, also from Ohio, want to deploy in December 2016. If Ohio cloud customers value proximity then we would observe few Ohio customers deploying in Azure DCs in December. The same logic applies to opening new Azure DCs for the spatial distribution of customers in other Azure DCs.

Figure 4 shows average distance between the zip codes of customers making new deployments and the zip codes of the DCs they deploy to in the Azure data by month. Vertical lines indicate the dates would new DCs are available to customers. If the new DCs had no impact on deployment decisions the lines would be flat over time. The line shows some month on month variation in addition to a decreasing

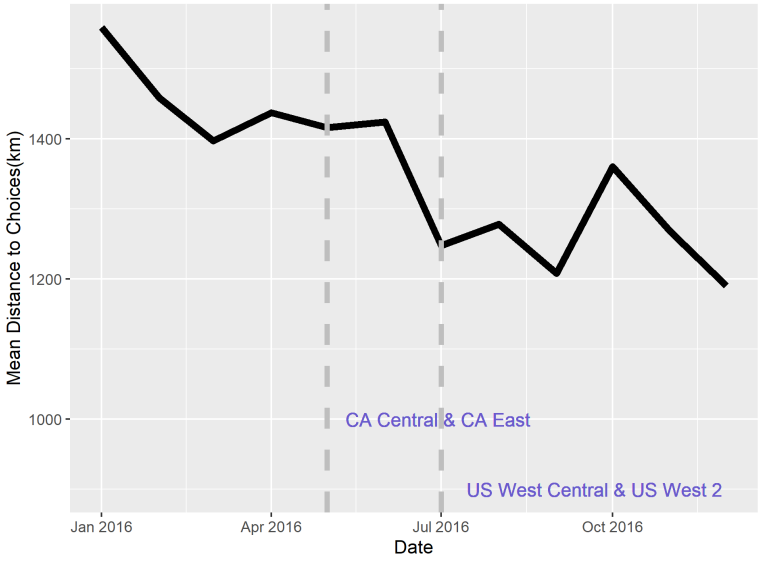


Figure 4: Average distance of deployments to customer zip code

Note: Figure shows drop in average distance between a deploying customer's billing zip code and their choice of data center over time. Sharpest one month drop occurs in same time period as new Azure DC locations.

trend over time. Hence, there is some reduced form evidence in the Azure data for a preference for proximity.

Figure 5 shows a density of customer location relative to deployed DC for customers that choose the closest DC relative to those that do not choose the closest DC at the time of deployment. The green shaded density shows the location in kilometers for customers that choose the nearest and the grey density customers who do not choose the nearest. The Figure is meant to highlight that when customers choose the nearest DC location they are often moving from something like 1000 to 4000 kilometers to being within 1000 kilometers. The average distance distance is roughly 1000 kilometers (vertical dashed lines). If 100 miles maps to one millisecond of latency then the gain in latency between the two densities is roughly six milliseconds. Aside from niche use cases like high frequency traders, six milliseconds is not likely to be material for most cloud VMs.

Our method relies on having detailed customer sales for a single firm (Azure) and aggregate customer sales for the second (AWS). While we have very good data on Azure customers, we have no customer level data for AWS customers. However, we observe aggregate global cloud sales from AWS from 10-K SEC filings in 2016.[11] SEC reported sales are worldwide, not restricted to North American market, but we observe U.S. sales as a percent of worldwide sales for Azure.

We make three strong but plausible assumptions to back out US AWS sales for *t2.small* in 2016 from their 10-K leveraging insights from Azure data. First, we apply the global revenue share of North America for Azure to AWS. While this is likely to be imperfect, it is hard to imagine that geographical revenue shares are significantly different across the providers. Second, we calculate the revenue share of the Azure SKU, *basic A1*, within North America relative to all other cloud products. We then apply that product revenue share to AWS. Third, we calculate the average sales of *basic A1* customers and apply it to the inferred AWS *t2.small* customers to get a customer count for *t2.small* for AWS customers.

This inferred approach is appealing because it permits us to get a plausible customer base for AWS customers. A simpler version in the same spirit would be to multiply the number of observed Azure

[11] More details can be found on http://phx.corporate-ir.net/phoenix.zhtml?c=97664&p=irol-reportsother

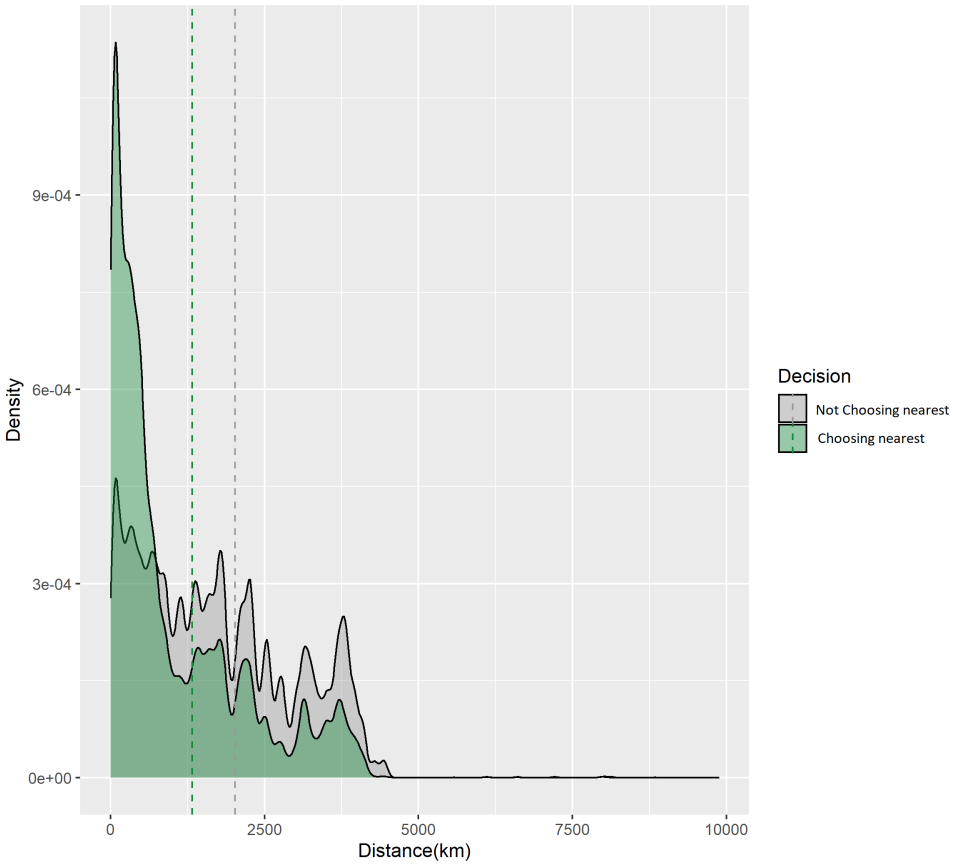


Figure 5: Average distance conditional on proximity decision

Note: Figure plots smoothed distance distribution for customers choosing the closest DC to them at time of deployment (shaded dark green) versus those not choosing the closest DC at time of deployment (shaded light grey). Average distance difference across each group is on the over of 600 kilometers and supports broadly overlap.

customers in our sample by the market share ratio of AWS to Azure show in Figure 1. In practice the two approaches give qualitatively similar customer count numbers, which we don't report due to confidentiality clauses in our data sharing agreements since it would provide customer count data for Azure customers. We discuss in our results section how results could be impacted by getting inferred AWS customer counts wrong.

Having both detailed data for Azure and aggregate data for AWS, the final piece of data is aggregate market size data for the cloud computing market in North America by state or province. The vast majority of cloud computing resources are used by firms as opposed to sold directly to consumers and the cloud is a substitute for on premise compute resources. We thus assume the market for cloud computing is defined by the total number of private sector firms in the U.S. and Canada. For the U.S. market we take the total number of businesses by state from Henry J Kaiser Family Foundation (KFF).[12] KFF tracks data on number of private sector firms by size.

For the Canadian market, we take data from Statistics Canada, a Canadian government agency which can be considered as the counterpart of the U.S. Census Bureau[13]. The data is from the Business Register (BR), a continuously-maintained central repository of baseline information on businesses and institutions operating in Canada. The variable is referred as "Canadian Business Counts" in the repository, including all active Canadian locations with employees. The number we use was collected in December, 2016.

We trim the market level data in two ways. First, because the data is at yearly level we take the numbers in 2015 and 2016 as they were collected at the end of each year, and then extrapolate them into each month in 2016 based on a constant growth rate assumption for both US and Canada data. Furthermore, we only consider firms with more than 50 employees as potential cloud users, they are 24.43% of all private firms in the U.S. in 2016 and 4.7% for Canada. We only look at larger firms since in 2016 cloud computing was more likely to be utilized by larger, tech savvy firms.[14]

4 Model

This section introduces our structural demand model for cloud deployments. We model all Canadian and U.S. consumers' utility to take the standard random utility model (RUM) form. In addition to allowing price and firm fixed effects to impact utility, we explicitly include distance between a consumer and data center and a shifter for if the data center is domestic. Finally, we leverage the EM algorithm to get around the missing data problem of unobserved AWS customer locations.

Cloud computing is a classic discrete-continuous good because consumers first decide to rent cloud computing resources, then decide how much to rent (Hanemann [1984]). For simplicity in what is already a non-trivial problem, we focus only on the initial purchase decision for two of the most popular general compute cloud products during this time: *t2.small* for AWS and *basic A1* for Azure. We do not model continued deployment decisions in this paper and focus on the location of new VM deployments.

We assume the utility of customer i choosing DC j in period t is

[12] See https://www.kff.org/other/state-indicator/number-of-firms-by-size.

[13] More detailed information can be found on the following website:https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3310003401.

[14] Of course, many smaller tech savvy start ups also leverage the cloud. We discuss robustness around this trimming decision in the results section. We make a final technical assumption to multiply the market size by *basic A1*'s demand share within Azure, so that the patterns in market shares are kept consistent with *basic A1* and *t2.small* demand. As we discuss below, this final resizing only changes the magnitude of the outside option relative to shares for AWS and Azure.

$$
\begin{aligned}
u_{ijt} \quad = \quad & \gamma_{m_i} \times d(\mathbf{l}_i, \mathbf{l}_j) + \beta \times price_{jt} + \psi \times \mathbb{1}_{ij}\{domestic\} \\
& + \xi \times DCAge_{jt} + \zeta \times \mathbb{1}_j^{AWS} + \epsilon_{ijt}, \forall j \in \mathcal{F}^t
\end{aligned} \tag{1}
$$

where

- $i = 1, 2, ..., I$ is the index for customers, $j = 1, 2, ..., J$ is the index for DCs and $t = 1, 2, ..., T$ is the time index.
- $\mathbf{l}$ is a 2-dimensional vector indicating locations, with the first component as longitude and the second as latitude.
- $d(\mathbf{l}_i, \mathbf{l}_j)$ is a function returning the distance between consumer i and DC j, i.e. $d(\mathbf{l}_j, \mathbf{l}_i) = ||\mathbf{l}_i - \mathbf{l}_j||$, where $||\cdot||$ is the great-circle distance. We allow preferences for proximity to vary based on consumer i's industry.
- m_i indictes consumer i's industry, we allow industry-specific distance coefficient to reflect the fact that different industries may have distinguished degrees of latency aversion.
- $price_{jt}$ is the price of DC j in period t.
- $\mathbb{1}_{ij}\{domestic\}$ is an indicator variable for if customer i is in the same country as the data center j.
- $DCAge_{jt}$ is the age of DC j in period t.
- $\mathbb{1}_j^{AWS}$ is an indicator for AWS DCs.
- ϵ_{ijt} is a type I extreme value that is $i.i.d$ across $\forall i,\ j,\ t$.

Equation (1) has a standard form but a couple of attributes merit discussion. First is the distance metric. We determine a customer's location based upon their observed billing address zip code and the approximate location of different data centers (nearest city). This introduces some measurement error: cloud customers care about latency between their deployment and the user of that deployment. For example, Netflix, a streaming video on demand provider, might prefer to put their cloud workloads close to their customers' locations rather than their corporate headquarters. While there is correlation between cloud customer's location and the location of their customers, that correlation is not perfect. This introduces measurement error and thus attenuation bias. As a result, the impacts of distance we estimate are likely a lower bound. Also, by including an indicator for domestic DCs, $\mathbb{1}_{ij}\{domestic\}$, we allow a general preference for domestic DCs due to concerns about information security or logistic convenience.

Second, since consumers' utility of different DCs vary with their locations, it is possible in principle to model utility function in a "random coefficient" fashion. Specifically, whereas we can calculate distance explicitly for Azure consumers, distances for AWS or non-cloud users are unknown. Consumers' heterogeneous tastes across DCs could be thought of as determined by their unobserved attributes, therefore similar to a "random coefficient" model. We put more structure on the problem by making assumptions about the spatial distribution of all possible cloud consumers because the counterfactual exercise we want to perform are the welfare implications of changing the location of DCs. Thus our modeling assumptions are driven by the nature of problem we seek to solve.

Third, we allow for the utility of data centers to vary by the age of the data center measured in months. This allows for cloud customers to learn about new data centers over time. It also allows for growth in complementary services: our analysis examines only a single cloud computing product but

there are complementarities between products (e.g., VMs and data storage). Allowing for DC age to impact utility is a reduced form way of allowing complementarities to manifest.

We model the outside option as on-premise IT infrastructure. We assume that all consumers have one such option in their choice set, denoted as $j = O$ with characteristics $d(\mathbf{l}_i, \mathbf{l}_O) = 0, \forall i$, $price_{Ot} = 0, \forall t$, $\mathbb{1}_{ij}\{domestic\} = 1, \forall i$. Since all consumers had been using in-house infrastructure before cloud, it is unnecessary to model learning effects with a time-variant variable such as $DCAge_{jt}$, thus $\xi \times DCAge_O$ can be normalized up to a constant. Instead, we assume there is a time-variant fixed effect for the outside option, $\alpha + \tau ln(t)$, which can be interpreted as the general time trend of cloud computing. A negative coefficient on τ would reflect the general increase in market share of cloud computing relative to on-premise offerings. Therefore, the utility of the on-premise option available to all possible cloud customers is:

$$u_{iOt} = \alpha + \tau \times ln(t) + \epsilon_{iOt} \tag{2}$$

where α includes the domestic effect as well as the constant term in time trend.

4.1 The Likelihood Function

Since we assume ϵ_{ijt} are from type I extreme value distribution, the probability for customer i to choose DC j in period t takes the familiar logit form[15]:

$$P(y_{ijt} = 1 | m_i, \mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_t, \boldsymbol{\theta}_1) = \frac{exp(v_{ijt})}{exp(v_{iOt}) + \sum_{k \in \mathcal{F}_t} exp(v_{ikt})} \tag{3}$$

where

- v denotes the deterministic part of the utility function, i.e. $v_{ijt} = u_{ijt} - \epsilon_{ijt}$; $v_{iOt} = u_{iOt} - \epsilon_{iOt}$
- y_{ijt} is a 0-1 binary variable indicates whether consumer i signs up for DC j in period t.
- $\mathcal{F}_t$ is the product set in period t, including the product set of Microsoft's Azure, $\mathcal{F}_t^M$, and that of Amazon's AWS, $\mathcal{F}_t^A$, i.e. $\mathcal{F}_t = \mathcal{F}_t^M \cup \mathcal{F}_t^A$
- $\mathbf{l}_t^{DC} = \{\mathbf{l}_j, \forall j \in \mathcal{F}_t\}$ is the set collecting the locations of all available DCs in period t
- $\mathbf{z}_{jt} = (price_{jt},\ DCAge_{jt},\ \mathbb{1}_j^{AWS})$ is the product characteristics vector, and $\mathbf{z}_t = \{\mathbf{z}_{jt}, \forall j \in \mathcal{F}_t\}$ collects $\mathbf{z}_{jt}$ across all DC's.
- $\boldsymbol{\theta}_1 = (\gamma_m,\ \beta,\ \psi,\ \xi,\ \zeta,\ \alpha,\ \tau)$ is the set of utility parameters.

The probability of not signing up for Microsoft Azure or AWS is

$$P(y_{iOt} = 1 | m_i, \mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) = \frac{exp(v_{iOt})}{exp(v_{iOt}) + \sum_{k \in \mathcal{F}_t} exp(v_{ikt})} \tag{4}$$

[15] In our Azure data, consumers are from different purchase channels. In this estimation, we focus on two of them, web direct and volumn license. The reason is that consumers from other channels such as "Benefits" may have a different pricing scheme.

4.1.1 Unobserved AWS demand

Although we can directly use Eq.(3) to denote the probability that a Azure customer chooses any specific DC, the industries and locations AWS customers as well as their DC-level choices are unobservable in our dataset. We leverage our inferred AWS product revenue, conditional probabilities and the EM algorithm to get around this problem.

First, we write the likelihood as the probability of choosing AWS as a brand, which is the sum of probabilities of choosing any of their DCs:

$$P(y_{iAt} = 1|m_i, \mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) = \frac{\sum_{j \in \mathcal{F}_t^A} exp(v_{ijt})}{exp(v_{iOt}) + \sum_{k \in \mathcal{F}_t} exp(v_{ikt})} \tag{5}$$

where y_{iAt} indicates whether consumer i chooses AWS in period t.

4.1.2 Missing consumer locations

The industries and locations of non-Microsoft customers are unobservable in our data which makes the calculation of the conditional choice probabilities infeasible. To circumvent this problem, we will take consumer's industry and location as 2 random variables, get the joint probability of industry, location and choice, then integrate out its uncertainty in industry and location for AWS and the outside option consumers. Particularly, the likelihood function in period t can be written as

$$\begin{aligned} L_t(\boldsymbol{\theta}) \quad = \quad & \prod_{i \in C_t^M} \prod_{j \in \mathcal{F}_t^M} (P(y_{ijt} = 1|m_i, \mathbf{l}_i, \mathbf{l}^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) f_t(m_i, \mathbf{l}_i|\boldsymbol{\theta}_2))^{y_{ijt}} \\ & \times \prod_{i \in C_t^A} \int_{m_i, \mathbf{l}_i} P(y_{iAt} = 1|m_i, \mathbf{l}_i, \mathbf{l}^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) f_t(m_i, \mathbf{l}_i|\boldsymbol{\theta}_2) dm_i d\mathbf{l}_i \\ & \times \prod_{i \in C_t^O} \int_{m_i, \mathbf{l}_i} P(y_{iOt} = 1|m_i, \mathbf{l}_i, \mathbf{l}^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) f_t(m_i, \mathbf{l}_i|\boldsymbol{\theta}_2) dm_i d\mathbf{l}_i \end{aligned} \tag{6}$$

where $C_t^f, f = M, A, O$ are the sets of consumers for Microsoft's Azure, Amazon's AWS and non-cloud users respectively. The key attribute of equation (6) is the distribution of location for AWS and outside option purchasers on the second and third lines. The density $f(m_i, \mathbf{l}_i|\boldsymbol{\theta}_2)$ can be viewed as a industry-specific spatial distribution of consumers of all options in the market. Although we observe the industries and locations of Azure customers, we write their joint probabilities of industry, location and choice separately to keep the format consistent across brands. This will also enable us to infer $f_t(m_i, \mathbf{l}_i|\theta_2)$ based on the observed Azure customer locations. More details can be found in Section 6. In practice we take the industry-specific spatial distribution of consumers in the market to be the that of medium and large firms across U.S. states and Canadian provinces as described in the Data section above.

Taking logs for the function above and compacting $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ as $\boldsymbol{\theta}$ then summing over time gives:

$$
\begin{aligned}
LL(\boldsymbol{\theta}) &= \sum_t LL_t(\boldsymbol{\theta}) \\
&= \sum_t \Big(\sum_{i \in C_t^M} \sum_{j \in \mathcal{F}_t^M} y_{ijt} log(P_{it}^j(\boldsymbol{\theta}_1) f(m_i, \mathbf{l}_i | \boldsymbol{\theta}_2)) \\
&\qquad + Q_t^A log(\int_{m_i, \mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(m_i, \mathbf{l}_i | \boldsymbol{\theta}_2) dm_i d\mathbf{l}_i) \\
&\qquad + Q_t^O log(\int_{m_i, \mathbf{l}_i} P_{it}^O(\boldsymbol{\theta}_1) f(m_i, \mathbf{l}_i | \boldsymbol{\theta}_2) dm_i d\mathbf{l}_i)) \qquad (7)
\end{aligned}
$$

where $P_{it}^j(\boldsymbol{\theta}_1)$, $P_{it}^A(\boldsymbol{\theta}_1)$ and $P_{it}^O(\boldsymbol{\theta}_1)$ simplifies $P(y_{ijt} = 1 | m_i, \mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1)$, $P(y_{iAt} = 1 | m_i, \mathbf{l}_i \in C_b, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1)$ and $P(y_{ijt} = 1 | m_i, \mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1)$ correspondingly. Since we take expectation over the unknown consumer's industry and location, the expected choice probability is same for every AWS consumer or any potential cloud consumer. Therefore, we multiply them by the total quantities Q_t^A and Q_t^O.[16]

4.2 EM Algorithm

Maximizing the log likelihood function above with the usual Newton or quasi-Newton routines can be numerically difficult and computationally unstable. This is a key motivation for leveraging the EM algorithm. The EM algorithm is a two-stage iterative method which involves calculating an expectation of the log likelihood function weighted by the Bayes' probabilities at some initial values and then updating the parameters by maximization.

Following Bhat [1997], it can be shown that with a given distribution of firms in North American and a set of preference parameters (θ_1), maximizing Eq.(7) is mathematically equivalent to maximizing the alternative log likelihood function in Eq.(8), where $f(m_i, \mathbf{l}_i | \boldsymbol{\theta}_2)$ are replaced by its Bayesian posterior counterparts, i.e. the probabilities that an AWS customer or a non-cloud user is from industry m_i and located at $\mathbf{l}_i$ which we denote as $h_{m_i, \mathbf{l}_i, t}^A$ and $h_{m_i, \mathbf{l}_i, t}^O$ respectively.

$$
\begin{aligned}
&\sum_t \Big(\sum_{i \in C_t^M} \sum_{j \in \mathcal{F}_t^M} y_{ijt} log(P_{it}^j(\boldsymbol{\theta}_1) f(m_i, \mathbf{l}_i | \boldsymbol{\theta}_2)) \\
&\quad + Q_t^A \int_{m_i, \mathbf{l}_i} h_{m_i, \mathbf{l}_i, t}^A(\boldsymbol{\theta}) log(P_{it}^A(\boldsymbol{\theta}_1) f(m_i, \mathbf{l}_i | \boldsymbol{\theta}_2)) dm_i d\mathbf{l}_i \\
&\quad + Q_t^O \int_{m_i, \mathbf{l}_i} h_{m_i, \mathbf{l}_i, t}^O(\boldsymbol{\theta}) log(P_{bt}^O(\boldsymbol{\theta}_1) f(m_i, \mathbf{l}_i | \boldsymbol{\theta}_2)) dm_i d\mathbf{l}_i \qquad (8)
\end{aligned}
$$

Then this maximization problem can be solved iteratively: starting from some initial values $\boldsymbol{\theta}^s$, we first update the Bayesion posterior probabilites, and then maximize Eq.(8) for $\boldsymbol{\theta}^{s+1}$ conditional on the Bayesian posteriors. Details of the approach are carefully described in the Appendix.

Lastly, due to the property of log operation, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can be separately updated. Specifically , we iteratively maximize the following two objective functions,

[16] Recall we make some assumptions on AWS revenue composition to get Q_t^A and the market size broadly to get Q_t^O.

$$
\begin{aligned}
\varepsilon_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}^s) &= \sum_t \Big(\sum_{i\in C_t^M} \sum_{j\in \mathcal{F}_t^M} y_{ijt} log P_{it}^j(\boldsymbol{\theta}_1) \\
&\quad + Q_t^A \int_{m_i,\mathbf{l}_i} h_{m_i,\mathbf{l}_i,t}^A(\boldsymbol{\theta}^s) log P_{it}^A(\boldsymbol{\theta}_1) dm_i d\mathbf{l}_i \\
&\quad + Q_t^O \int_{m_i,\mathbf{l}_i} h_{m_i,\mathbf{l}_i,t}^O(\boldsymbol{\theta}^s) log P_{bt}^O(\boldsymbol{\theta}_1) dm_i d\mathbf{l}_i \Big) \\
\varepsilon_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}^s) &= \sum_t \Big(\sum_{i\in C_t^M} \sum_{j\in \mathcal{F}_t^M} y_{ijt} log f(m_i,\mathbf{l}_i|\boldsymbol{\theta}_2) \\
&\quad + Q_t^A \int_{m_i,\mathbf{l}_i} h_{m_i,\mathbf{l}_i,t}^A(\boldsymbol{\theta}^s) log f(m_i,\mathbf{l}_i|\boldsymbol{\theta}_2) dm_i d\mathbf{l}_i \\
&\quad + Q_t^O \int_{m_i,\mathbf{l}_i} h_{m_i,\mathbf{l}_i,t}^O(\boldsymbol{\theta}^s) log f(m_i,\mathbf{l}_i|\boldsymbol{\theta}_2) dm_i d\mathbf{l}_i \Big)
\end{aligned}
$$

In practice, instead of assuming a parametric distribution for $f(m_i, \mathbf{l}_i|\boldsymbol{\theta}_2)$, we assume a discrete distribution of consumer's industry and location, or say a discrete industry-specific spatial distribution. The discrete distribution can approximate any arbitrary distribution when discretization is fine enough. Specifically, we take each U.S. state and Canadian province as a bin b, and then the probability that a consumer (including non-cloud users) from industry m belongs to a certain bin b in period t is q_{mbt}, and these q_{mbt}'s are treated as parameters to estimate. Using states and provinces is both convenient and appropriate since data on market size (medium and large firms) is available at the state level and provides good variation in distance from newly introduced DCs. Details of this approach are again in the Appendix.

In sum there are a few departures from normal log likelihood maximization we make in our approach. First, we replace location probabilities with Bayesian posteriors. Second, we iteratively solve for parameters governing the distribution of consumers for each product and preference for the product. Third, we discretize the spatial distribution of North America. This final step is an advantage for us since variation in Azure demand in geographical bins over time in response to new Azure and AWS DCs help us identify the model's parameters. The iterative maximization process across geographical and preference parameters continues until convergence as we describe in detail in the next section.

4.3 Identification

In this section, we show that the parameters can be identified in the following order: (1) $\boldsymbol{\theta}_{1,1} = (\gamma_m, \beta, \psi, \rho, \xi)$; (2) $\boldsymbol{\theta}_{1,2} = (\zeta, \alpha, \tau)$; (3) $\boldsymbol{\theta}_2 = \{q_{mbt}\}_{mb}, t = 1, 2, ..., T$.

First, $(\gamma_m, \beta, \psi, \xi)$ are identified from the substitution pattern of Microsoft customers among Microsoft DCs. Since our Microsoft data is at individual level, including the industry and location of each customer, m_i, $d(\mathbf{l}_i, \mathbf{l}_j)$'s and $\mathbb{1}_{ij}\{domestic\}$'s are deterministic, i.e. there is no unknown interaction between individual attributes and product characteristics. Therefore, the *Independence of Irrelevant Alternatives* (IIA) property of logit model makes it possible to focus on only a subset of products (Train [2009]).

Next, if we consider ζ and $\{v_{Ot}\}_t$ [17] as the general preference for all AWS DCs and the outside option over Microsoft, with the product characteristics and $\boldsymbol{\theta}_{1,1}$ as given, the unexplained part of market share ratios should be attributed to that "general preference", which gives the identification of $\boldsymbol{\theta}_{1,2} = (\zeta, \alpha, \tau)$.

[17] With a slight abuse of notation, we suppress the subscription i in v_{iOt} since the deterministic utility from the outside option is individual-invariant.

Specifically, for $\forall j \in \mathcal{F}_t$, we write the AWS fixed effect separately from other components in the utility index, i.e.

$$v_{ijt} = \mu_i(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}) + \zeta \mathbb{1}^{AWS}_j,$$

where

$$\begin{aligned} \mathbf{z}^1_{jt} &= (price_{jt},\ DCAge_{jt}) \\ \mu_i(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}) &= \beta price_{jt} + \gamma_{m_i} d(\mathbf{l}_i, \mathbf{l}_j) + \psi \mathbb{1}_{ij}\{domestic\} + \xi DCAge_{jt} \end{aligned}$$

Note that μ_i is individual-specific due to the consumer's heterogeneous industry and location.

Then, within each combinataion of m and b, the model gives the market share ratio of AWS to Microsoft as the fraction of the exponentials of their inclusive values,

$$\frac{Q^A_{mbt}}{Q^M_{mbt}} = \frac{\sum_{j \in \mathcal{F}^A_t} exp(\zeta + \mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum_{j \in \mathcal{F}^M_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} = exp(\zeta) \frac{\sum_{j \in \mathcal{F}^A_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum_{j \in \mathcal{F}^M_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}$$

Here with a little abuse of notation, we use subscript m, b to emphasize that function $\mu_i(\cdot)$ is the same for consumers from the same industry m and located in bin b. Also, for Azure consumers, even though $\mathbf{l}_i$ is observed, we lower the granularity to bin b level in this section just to illustrate the concept.

Relate this to the observed market level AWS demand by $Q^A_t = \sum_{m,b} Q^A_{mbt}$, we have

$$Q^A_t = exp(\zeta) \sum_{m,b} \frac{\sum_{j \in \mathcal{F}^A_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum_{j \in \mathcal{F}^M_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{mbt}$$

which gives

$$\zeta = log(Q^A_t / \sum_{m,b} \frac{\sum_{j \in \mathcal{F}^A_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum_{j \in \mathcal{F}^M_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{mbt})$$

Similarly,

$$v_{Ot} = log(Q^O_t / \sum_{m,b} \frac{1}{\sum_{j \in \mathcal{F}^M_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{mbt})$$

Then α and τ are identified by the linear relation $v_{Ot} = \alpha + \tau ln(t)$.

Finally, given $\boldsymbol{\theta}_1$, the model could infer the local market size based on the observed local demand of

Microsoft, i.e.

$$
\begin{aligned}
q_{mbt} &= \frac{Q^M_{mbt} + Q^A_{mbt} + Q^O_{mbt}}{M_t} \\
&= \frac{Q^M_{mbt} + exp(\zeta)\dfrac{\sum_{j\in\mathcal{F}^A_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum_{j\in\mathcal{F}^M_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{mbt} + exp(v_{Ot})\dfrac{1}{\sum_{j\in\mathcal{F}^M_t} exp(\mu_{mb}(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{mbt}}{M_t}
\end{aligned}
$$

where M_t is the market size in period t.

5 Monte Carlo Experiment

To test model's identification, we performed a Monte Carlo experiment. Itss important to assess whether the model's parameters are recoverable with only detailed Azure data because Azure had only a 7% market share over our sample. Accordingly, the basic structure of the simulated data sets used in the Monte Carlo borrows from the true data in two ways.[18] First, the number of consumers in each industry-state/province is generated based on the distribution (e.g., $\{q_{mbt}\}_{m,b,t}$) that we recover from estimation. Second, the taste parameters that we use to generate each consumer's DC choices are the same as the estimates from the actual data.

The main variation across these simulated data sets are the indiosyncratic random utility shocks ϵ_{ijt}. We simulated 100 data sets. For each data set, we let each consumer chooses the DC that gives the highest uility. We then keep the individual choices of Azure customers while aggregating AWS customers and those who choose the outside option up to market shares at period level. With the spatial distribution of consumers masked so that they must be estimated as when we estimate the model with our actual data, we estimate each simulated data set with the EM-algorithm described above.

The results of the Monte Carlo are shown in Appendix Table 5. The model performs reasonably well. For all 18 parameters except one the true simulated parameter is within the 95% confidence interval of the parameters estimated from the simulated data. The one parameter that is outside of the 95% confidence interval is the indicator variable for a DC being domestic. The domestic indicator is only marginally outside the confidence interval (CI): true value 1.58 and 95% CI of [1.415,1.516]. Thus, there is some evidence we estimate a domestic indicator that makes cloud customers look slightly less interested (less than 10%) in deploying their VM in country.

That the coefficient on the domestic indicator variable is somewhat imprecise is not surprising given the nature of our data. It is identified almost entirely off Canadian customers choosing to deploy VMs in the Canadian Azure DCs after they open halfway through 2016. However, the number of unique Canadian firms in our sample is an order of magnitude lower than the number of U.S. firms in the sample. As a result, the indicator variable is likely to be measured imprecisely and, perhaps, modestly downward biased. The downward bias could be due in part to the domestic preference loading onto the estimated distance preference for Canadian customers. That said, the Monte Carlo shows the true value of the distance coefficient (γ for unknown industries) is exactly in the center of the 95% CI.

[18] Of course, the DC layout as well as their prices are also consistent with our observed data in each period.

6 Estimation Results

Table 3 shows results from estimating the model with the data. We performed estimation in R and convergence times on a single PC were on the order of 10 hours. We do not report the number of observations so as to not reveal information on the number of unique customers for this Azure SKU over our sample, per the confidentiality agreement with Microsoft.

Table 3 shows that all parameters are precisely estimated and have the expected sign. The price coefficient is negative. The coefficients on DC age, the domestic indicator, AWS fixed effect, outside option (OO) fixed effects are all positive. The positive AWS and OO fixed effects reflect market share sizes over the sample (e.g., Azure $<$ AWS $<$ OO based upon our assumption of outside good market size).[19]

Table 3: Estimates

	Estimates	Std. Err.
Distance (in km)		
Discrete Manufacturing	-1.439×10^{-3} ***	1.223×10^{-5}
Education	-1.439×10^{-3} ***	1.348×10^{-5}
Health	-1.430×10^{-3} ***	1.510×10^{-5}
Hospitality & Transportation	-1.439×10^{-3} ***	1.269×10^{-5}
Insurance	-1.439×10^{-3} ***	1.447×10^{-5}
Media / Telecom and Utilities	-1.750×10^{-3} ***	1.516×10^{-5}
Nonprofit	-1.442×10^{-3} ***	1.594×10^{-5}
Professional Services	-1.334×10^{-3} ***	1.117×10^{-5}
Unknown	-5.168×10^{-4} ***	7.436×10^{-6}
Price	$-1.212 \times 10^{+1}$ ***	1.080×10^{-3}
Domestic	1.872***	1.853×10^{-2}
DC Age	9.047×10^{-1} ***	2.586×10^{-2}
AWS FE	2.162***	1.702×10^{-12}
OO FE	1.461***	8.867×10^{-3}
OO trend	2.295***	1.027×10^{-3}

Note: All parameters statistically significant. All coefficients have expected sign with distance and price both negative and highly significant. The model includes AWS and OO fixed effects in the first and seventh month of our data where there was some backfilled reporting from previous months due to a Microsoft reporting delays. Those coefficients are statistically significant and an order of magnitude lower than AWS and OO FEs; we don't report them as consider them nuisance parameters. We normalized DC age so that DC age is measured with respect to the oldest DC in the sample.

The key coefficient of interest is the coefficient on distance where we take the baseline to be preference for distance in unknown industries. The coefficient is negative and highly significant. Coefficients for other industries don't exhibit much variation and are roughly twice the magnitude of the estimated coefficient of distance for observations without a recorded industry. Thus, there is first order correlation between observing industry and preference for proximity. As mentioned above, based upon internal conversations it could be that observing industry is correlated with using a vendor to operate cloud resources. Because we don't fully observe the data generating process for that field in the data, we instead focus on cloud users with unknown industries, which make up the vast majority of the sample.

It is more informative to evaluate the ratio of the coefficient on distance to the coefficient on price rather than each coefficient in isolation. The ratio of distance to price is the willingness to pay for one kilometer. Hence, the willingness to pay to be 1,000 kilometers closer to a data center in our sample is 4.2

[19] The model includes AWS and OO fixed effects in the first and seventh month of our data where there was some backfilled reporting from previous months due to a Microsoft reporting delays. Those coefficients are statistically significant and an order of magnitude lower than AWS and OO FEs; we don't report them as consider them nuisance parameters. Due to this reporting issue in the timing of some of the Azure data, we don't put much stock in the sign of the coefficient on the logarithmic time trend of the outside good.

cents (per hour) for unknown industries (the majority of the sample). Recall that the average price over the sample for Azure's basic A1 product is 7.1 cents (per hour). Hence we estimate a price premium of roughly 60% of the average Azure hourly price. As a point of comparison DC level prices often varied by 20-50% within a public cloud provider over our sample. Hence, the implied point estimate for unknown industries seems modestly large but not out of the question where are the point estimate for cloud users that reported their industry seems larger than we expect. For customers in known industries, the estimated disutility of distance is stronger but, as mentioned above, we don't view those point estimates as reliable due to how industry data is recorded in our sample: many of those customers go through a third party to deploy their workloads.

There are two important caveats worth noting relating the negative and significant impact of distance on utility. First, the positive coefficient on DC age reflects that old DCs tend attract more deployments than newer DCs all else being equal. This could reflect some amount of inertia: as a customer deploys a new type of VM they are likely to put it in the same DC as where they might have older deployments. If so, this would reduce the likelihood of finding a strong negative utility for distance: identification of the distance parameter is driven by new DCs opening and evaluating how many customers proximate to its location start deploying workloads there. If older DCs have a stronger attraction, the likelihood of deploying in a new, proximate DC would be lower.

Second, the coefficient on domestic is strong and positive. We believe this could cause some modest downward bias (e.g., more negative) in the distance parameter so that we estimate a stronger dis-utility of distance than the true effect. The reason is that Azure opened two data centers in Canada which are both more proximate to many Canadians and also domestic. Hence, some preference for deploying in a domestic DC could be loaded on to the distance parameter. That said, there is significant variation in proximity for nearby but not domestic data centers for Canadians since the Canadian DCs are located in or east of Ontario. Thus, when Canadian DCs are opened halfway through our sample, cloud users in Vancouver, British Columbia can choose between a nearby DC in Washington state that is not domestic and a far away Canadian DC that is domestic. Therefore, any downward bias on the distance parameter is likely modest.

It is somewhat surprising that the data doesn't show any modest variation at the industry level. 2016 was still the early days of cloud usage. Some industries like health and education could have been later adopters and are now displaying similar distance preferences as discrete manufacturing and professional services did over our sample. For example, it might have been that more sophisticated cloud users display weaker preferences for proximity meaning that we estimate a "short run" effect in this paper. As noted above, though, it is possible that observing industry is really a proxy for using a third party to deploy and manage virtual machines. As a result, any underlying differences in preference for proximity could be second order to using third parties for IT management.

Figure 6 takes the estimates and aggregates consumers from all industries in a "bin" (e.g., U.S. state or Canadian province) to display heterogeneity in market shares over space for Azure and AWS. Figure 6 gives the densities of each firm's estimated market shares. The scale of the market share distribution on the left (AWS) is higher than that of the distribution on the right (Azure) but both are market shares accounting for the outside good. Hence a "bin" with 10% share for AWS and 2% share for Azure implies an 88% share for the outside good where the outside good is the number of firms with more than 50 employees in the geography.

The important aspect of Figure 6 is the non-trivial heterogeneity in market share across locations. Both firms exhibit bimodal market shares over space: they have some regions above the firm-specific mean market share and some regions below, and the distribution is not single-peaked. Even though AWS was the market leader during 2016, there are some areas where Azure has a market share in the

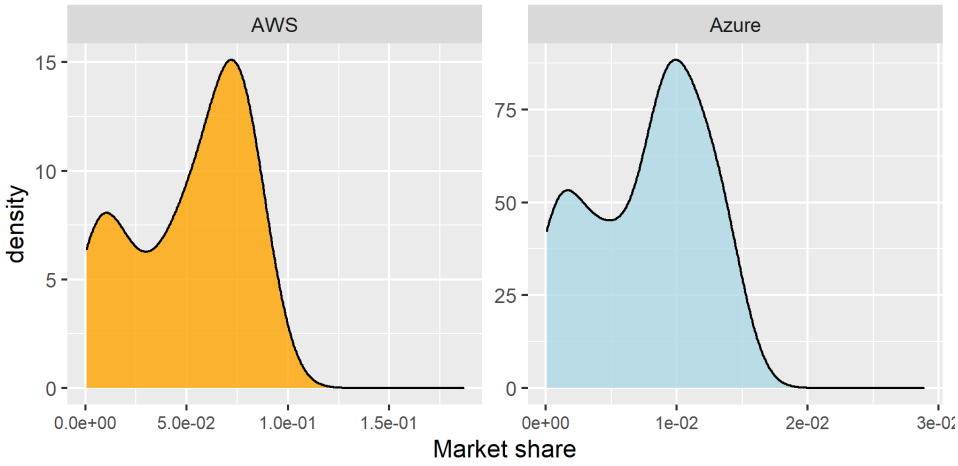


Figure 6: AWS vs Azure Market Share Distribution

Note: Market share distributions show material variation over states and provinces. Although not clear from this figure, it should be unsurprising that market shares also compliment each other: where AWS has a larger share Azure tends to have smaller share and vice versa.

low single digit percents and some nearly 20% of AWS's share.

Microsoft had more data centers than AWS during this time period, possibly earning higher market share in the regions where AWS did not have a data center. This finding is consistent with cloud customers having preferences for proximity in our sample. It is also consistent with competition being important for welfare in this market insofar as competition leads to more DCs being built in different locations.

While Figure 6 shows estimated aggregate variation in market share within firms over space, Figure 7 shows estimated variation in market share across Azure (blue) and AWS (orange) at the state/province level. We only select six regions for clarity and don't report precise locations associated with each state/province per our data sharing agreement. We instead show variation in the minimum, median and maximum Azure market share across markets sized below the median (bottom-sized markets) and those sized above (top-sized markets). Note that Figure 7 represents a relatively small level of aggregate market penetration relative to the outside option for both AWS and Azure which highlights that the cloud computing industry is still young and rapidly growing.

Figure 7 shows that we estimate changes in market share across regions of more than 100% for relatively small markets (8% to 18%) and roughly 100% for relatively large markets (10% to 19%). The model estimates a right tail as well: median Azure market share was slight less than half of the difference between the minimum and maximum market share. Qualitatively, we do estimate relatively larger market shares in some states where Azure has a DC but AWS does not, and vice versa. Finally, these market shares are from 2016 data and since then Azure has grown in market share. Thus these numbers do not reflect current market shares nor do they necessarily represent what would happen if new DCs were built today since more DCs have been constructed between 2016 and 2020.

7 Counterfactuals

With the estimated taste parameters, we move on to counterfactual analysis. The strength of this modeling approach is the ability to estimate heterogeneous market shares over space using disaggregate data for one firm but aggregate data for another. Our counterfactuals focus on using the model to

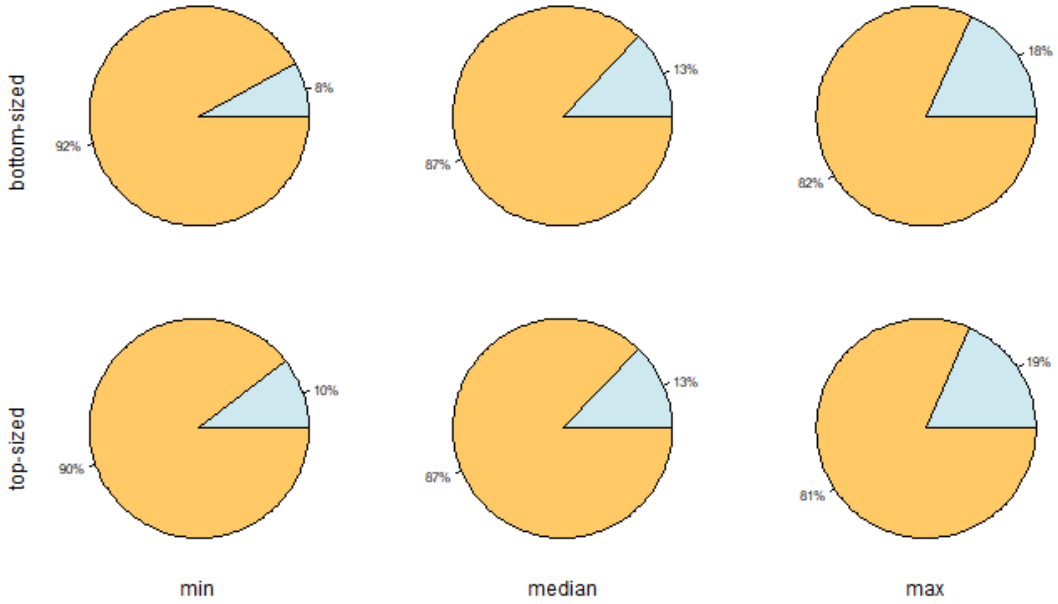


Figure 7: Microsoft vs AWS Market Share by Market Size

Note: Consistent with Figure 6, market shares vary across states. This is true for both large and small states where Azure market share can range from up to 19% of AWS market share down to less than 10%. Recall this figure does not report market share of the outside option nor other cloud providers so it is not directly comparable to Figure 1.

optimize data center location and examine the interplay of price competition and spatial competition in the cloud industry. All the data used in counterfactual analysis is the December 2016 data so that the counterfactuals reflect the most recent view of the data we observe.

First, we propose six states in southern U.S. where Microsoft currently has no DC, and ask which one would bring the most market share increase if Microsoft put one more DC there. These examples are chosen for their relevance. Microsoft Azure introduced four new DCs in North America in 2016 which increased its total number to ten, twice that of AWS. Therefore, it is reasonable to quantify the impact of a denser product space.

Second, we condition on the current DC layout in North America, and predict the market share responses to a counterfactual 15% price change for all Azure DCs. We then investigate how counterfactual changes in market shares vary based upon how vigorous spatial competition is. Put another way, we simulate a price decrease and evaluate how it impacts market shares in locations where both Azure and AWS have a DC, where only Azure has a DC and where neither Azure nor AWS have a DC.

Lastly, while we calculate changes in consumer surplus, fully capturing strategic supply side equilibrium responses is beyond the scope of these exercises. Neither AWS nor Azure alters DC layout or adjust prices of existing DCs in response to our counterfactual exercises. Accounting for equilibrium competition best responses is beyond the scope of our paper as our contribution highlights spatial competition for cloud computing rather than equilibrium competitive behavior.

7.1 New DC Location

The six proposed states for which we simulate Azure building a DC are spread evenly in southern U.S. where there was no DC in 2016 from Arizona to Florida. We assume the price for the newly constructed

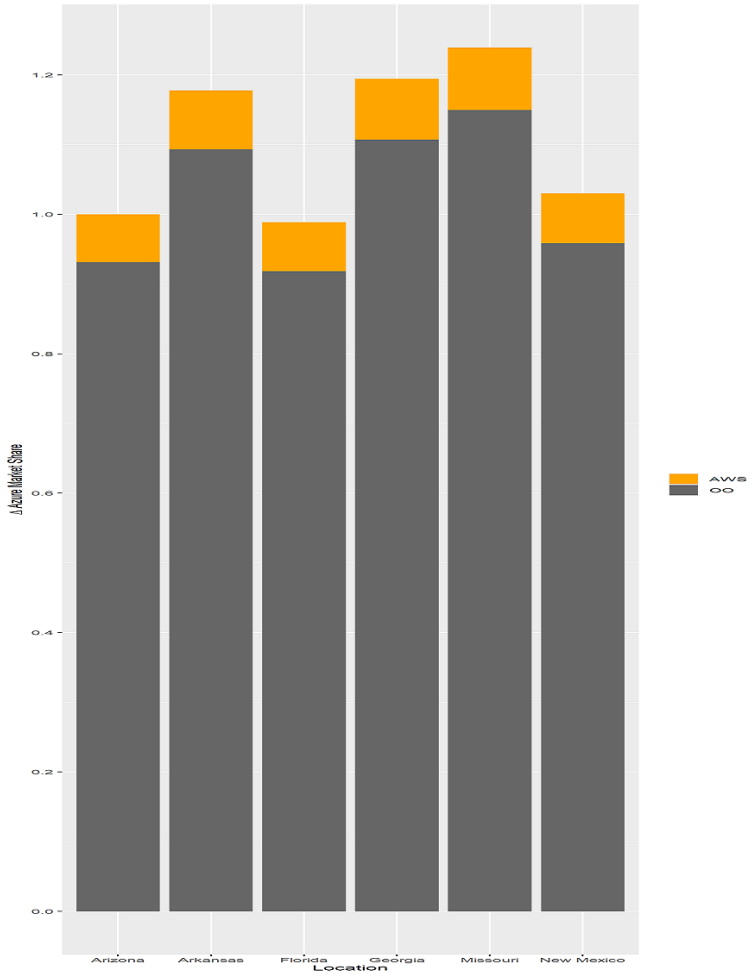


Figure 8: Introducing a New DC

Note: Figure reports the change in Azure market share from new customersin the counterfactual where Azure builds a new DC in one of six states. The increase in market share is reported relative to the percentage change in market share of a new DC introduced in Arizona. The Figure also reports where the increase in Azure market share comes from: the outside option over AWS. The southeast U.S. seem to indicate the largest percentage increase in share for AWS due in part due to a relatively large increase in share from acquired from AWS.

DC is set at the average Azure price level in December 2016 so that the different demand responses could be attributed to the differences in local market size and DC layout. All changes in Azure market share are normalized to changes from introducing a DC in Arizona, which Azure actually did enter in 2018. We decompose increases in market share by the "market stealing" effect of taking share from AWS and the cloud "market growing" effect of moving customers off their own premise and onto the cloud.

Figure 8 shows the results from the new DC counterfactual measured in percentage increase in market share relative to Arizona. There are a couple of important findings. First, it shows that introducing a new DC in Missouri generates highest market share gains for Microsoft Azure, which is around 25% higher than the "numeraire" state of Arizona. This result is consistent with estimation results since during this time period the DC density in Missouri is comparatively lower than the others. Therefore, a newly-introduced DC would provide greater utility increase relative to the outside good and AWS by reducing distance.[20]

Its useful to put the 25% number into perspective of the overall costs of running a DC to assess whether something like differences in wholesale electricity costs could drive location decisions. According to a report from the U.S. Chamber of Commerce roughly annual operating expenses are less than 10% of the capital costs of a data center and roughly 50-75% of operating expenses are electricity.[21] Hence electricity is on the order of 5-7.5% of annualized amortized DC costs. According to the U.S. Energy Information Administration, in 2016 average wholesale electricity prices in low cost Texas were \$27.16/MWh versus \$34.54 in high cost PJM for a 25% difference[22] The implication is that electricity cost differences on the order of 2% of annualized costs could explain DC location decisions that results in a 25% difference in market share changes. This seems unlikely.

Second, there is some modest variation in the size of the "market stealing" versus the "market growing". Building a DC in Georgia, Missouri and Arkansas leads to a larger market share increase than Arizona, Florida and New Mexico. In Georgia, Missouri and Arkansas there is a larger proportional increase in the "market stealing" versus the "market growing" effect driving the increase in market share. The implication is that appropriately siting DCs can lead to increased local market shares driven disproportionately by the market stealing effect. It is perhaps for this reason that all public cloud providers have dramatically increased their geographical footprint in the last five years, all roughly doubling the number of unique DC locations globally. This clearly is beneficial to consumers who, based on our estimates, appear to non-trivially value proximity. However, this strategic effect seems second order to the market growing effect based upon our sample.

Finally, we calculate the consumer surplus gain generated by a new Azure DCs. We define consumer surplus as the expected maximum money metric utility for new customers registered in Dec 2016, i.e. $t = T$. Put another way, we don't account for gains to existing customers since we only model the initial deployment decision. Because we don't account for the differences in usage intensity among consumers, consumer surplus estimates measured in dollars should be thought of as gains in the first hour of a single deployment of a one core VM. Since the lifespan of a VM is often many cores and many hours, the level of the surplus gains reported here are extreme lower bounds and as such we focus on percentage changes across counterfactuals. Specifically, for cloud users in industry m at location b at time T,

$$E(CS_{mbT}) = -\frac{1}{\beta} E[\max_{j \in \mathcal{F}_T} u_{mbjT}] = -\frac{1}{\beta} log(1 + \sum_{j \in \mathcal{F}_T} (exp(u_{mbjT}))) + C$$

[20] Of course, because we measure changes in within state market shares this says nothing of the aggregate increase in revenue for putting a new DC in Missouri relative to Arizona.

[21] See https://www.uschamber.com/sites/default/files/ctec_datacenterrpt_lowres.pdf.

[22] See https://www.eia.gov/electricity/wholesale/#history.

The subscripts m and b emphasize that utility depends on the hetergeneous cloud user industry m and location b. C is a constant term which is negligible when calculating the surplus differences.

The expectation value is defined relative the set of available data centers ($\mathcal{F}_T$) plus the outside option whose uitlity is normalized to 0. Thus when a new Azure DC j' opens, there will be one more element in the choice set thus makes the joint set $\mathcal{F}_T \cup j'$, and the execpation is supposed to increase since maximization function weakly increases with the number of choices. The strength of this approach is comparing by how much consumer surplus increases when DCs are placed in better versus worse locations. Finally, we aggregate this individual level expectation to the North American market level by summing over all industries and locations, i.e.

$$E(CS_T) = \sum_{sb} E(CS_{mbT}) q_{mbT} \times M_T,$$

where M_T is the market size, i.e. the total number of firms with 50 or more employees in North America. As in the exercise above we compare the percentage increase in consumer surplus to a single baseline state, Arizona.

Table 4: Consumer surplus effects of new DC locations (AZ baseline)

Location	Arizona	Arkansas	Florida	Georgia	Missouri	New Mexico
% $\Delta E(CS_T)$	100%	117.9%	98.9%	119.6%	124.1%	103.1%

The consumer surplus effect of each new DC location are summarized in the Table 4. Table 4 shows that percentage changes in consumer surplus by state are almost identical to changes in market share for Azure. This is not surprising: increases in market share indicate increases in consumer surplus as more cloud users begin to consume Azure.

7.2 Price Drop

Figure 9 provides demand responses to an overall 15% price drop of Microsoft Azure across all regions. We show results of the price impact in three representative U.S. states with different market structures: both a AWS and Azure DC (Virginia), neither a AWS nor a Azure DC (Georgia) or only an Azure DC (Texas). Each bar shows the percentage of switchers from AWS to Azure. All market share changes are pegged to Georgia in this counterfactual.

Figure 9 shows that the incremental change in market share varies by local market structure. Intuitively, in areas where both AWS and Microsoft DCs are available like Virginia, we estimate that price plays a relatively more important competitive role and that price cuts have a significant impact on market shares. On the contrary, the potential gain is less pronounced in states like Texas where Microsoft is the only cloud provider. In other words, the loss from a price rise would also be limited, a straightforward implication of local market power.

Another implication from this counterfactual is the amount of switchers across AWS and Azure are generally small in all three scenarios of price competition. Recall that during this time period AWS was both the early market leader and much larger in publicly reported revenue numbers. Put another way, this doesn't appear to be a Bertrand, winner take all market. This is consistent with AWS's early leadership in the cloud market but also spatial competition as being a material driver of increase Azure market share over this sample. It contradicts the idea of fully location agnostic demand in the cloud computing industry and the internet being the "death of distance".

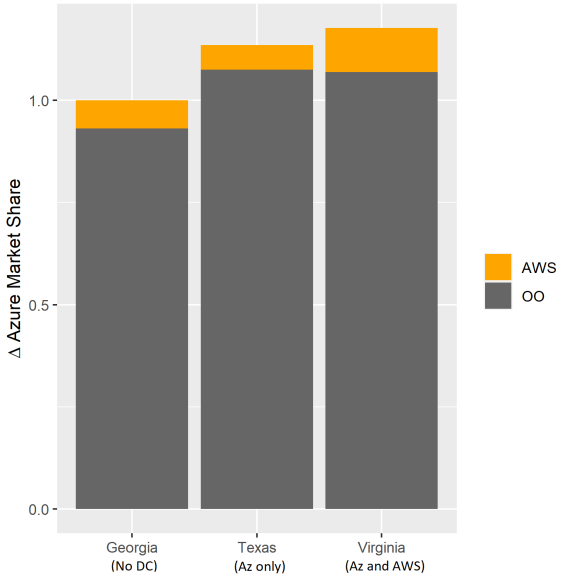


Figure 9: Price Competiton

Note: Figure reports the change in Azure market share from new customers in the counterfactual where Azure decreases the price of all Azure DCs by 15%. The increase in market share is reported relative to the percentage change in market share in Georgia. The Figure highlights that increases in market share from the price drop will vary based upon how much spatial competition there is in region. For example, both AWS and Azure have DCs in Virginia and we observe larger market share changes there due to a price decrease.

7.3 Counterfactual Comparison

We calculate changes in consumer surplus effect in the same way for the price drop counterfactual as for the new DC location counterfactual. Thus we can compare the change in consumer surplus from building a single new DC and compare it to the change in consumer surplus from a 15% price decrease from all Azure DCs. This serves as a sanity check for our estimates and the counterfactuals built upon them.

We find that the increase in consumer surplus from building a new DC in the six states from our coutnerfactual was 77% of the increase in consumer surplus from a 15% across the board price decrease. Recall that the 15% price decrease impacts all new customers in North America in a single month (the entire set of new customers in a month) whereas the new DC will cause a change in behavior of only a fraction of the monthly extensive margin (e.g., just the customers induced to move to Azure based upon the new DC). Assessing orders of magnitude, this makes sense at a high level: assume roughly 10% of total new customers deploy in the new DC in any given month (i.e., there were 10 Azure DCs at the end of the sample) those customers have a large reduction in distance between the previously closest DC and the new proximate one. Recalling that the implied willingness to pay for 1000kms (∼600 miles) in proximity was 60% of the average price of Azure, those 10% of new cloud customers' benefit implies an average decrease in distance of roughly 750 miles, which seems plausible since not every Azure customer deploys in the closest DC (see Figures 4 and 5 above).

There is another intuitive way to perform a back of the envelope calculation: all customers receiving a 15% price decrease in the price counterfactual have a 15% increase in their customer surplus to a first order approximation. Alternatively, in the new DC counterfactual, only those customers deploying in the new DC have their consumer surplus are impacted by it. There were 10 DCs at the end of our sample so roughly 10% of customers benefit from a new DC. Recall that aggregate consumer surplus from the new DC is 75% of the welfare increase from a 15% price decrease for all newly deploying customers. Hence, for customers deploying in the new DC in our counterfactual, we must observe an increase in consumer

surplus of (.15/.1)*.75 = 112.5%.[23] This is again plausible: a new proximate DC could be worth roughly twice as much to cloud users as distant DC based on our parameter estimates.

Given the staggering growth in cloud adoption in the last five years by firms, it is hard to imagine latency concerns being the sole driver of this barrier. For example, the distance decrease of a customer in Atlanta, Georgia to the nearest Azure DC at the time of our study was around 600 miles, or about 6 milliseconds of latency. While we cannot rule out latency as a driver with our data, these results indicate the presence of a secular preference for proximity consistent with "server hugging". If preference based server hugging does explain this result, our evidence suggests an alternative preference based rationale for why the internet may not lead to the "death of distance" in the case of cloud computing.

8 Conclusion

We find that cloud compute customers care about proximity to a surprising degree even within the US where latency difference across data centers are often separated in the single digit milliseconds. Our result is consistent with a growing body of work that finds that the internet has not in fact been the "death of distance" although we can't fully rule out strong preferences for reduced latency with our data. Because customers do care about distance, vigorous spatial competition of public cloud providers like AWS, Azure, GCP and Alibaba in the quickly maturing cloud market are likely to benefit cloud users a great deal and more quickly move firms from wholly owned on premise servers to remote rented cloud based compute resources. The number of data centers of each cloud provider has roughly doubled in the past five years.

While we do not model an equilibrium entry decision in this paper, there is clear room to expand this line of research in that dimension. Such work could have particular importance given that cloud computing resources lower barriers to entry of new firms and therein enable more productivity from the global labor force. Such models could be used by policy makers to encourage more competition in industries where spatial competition is important and enables aggregate productivity of the economy.

Our methodology could also be useful to other economists. We estimate our demand system when the dataset contains disaggregate consumer level choice data of one firm and aggregate market share data of another. We show that both the taste parameters and a discrete distribution of unobserved consumer attributes can be recovered with EM algorithm under the framework of mixed logit. It enables the identification of demand parameters up to brand level fixed effects which could be further pinned down by the observed market shares. Given demand parameters, the consumer spatial distribution, i.e. the local market sizes, is identified by the inverse of model-predicted local Microsoft market share. A Monte Carlo exercise supports identification.

Finally, our data is from 2016 which are the early days of the cloud computing industry. In 2016 and even in 2020 cloud revenue is growing rapidly. Cloud computing is not a product in long run equilibrium and preferences for cloud attributes are likely to change as cloud users learn and experiment with cloud resources. Hence, our results might not be externally valid in a fully mature cloud computing market.

References

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

[23] If N are the total number of cloud customers then $.1 * N * \Delta CS_{newDC} = .75 * N \Delta CS_{PriceChange} = .75 * N * .15$ and solve for ΔCS_{newDC}.

Steven Berry, James Levinsohn, and Ariel Pakes. Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy*, 112(1):68–105, 2004.

Chandra R Bhat. An endogenous segmentation mode choice model with an application to intercity travel. *Transportation science*, 31(1):34–48, 1997.

Yuxin Chen and Sha Yang. Estimating disaggregate models using aggregate data through augmentation of individual choice. *Journal of Marketing Research*, 44(4):613–621, 2007.

Christopher T Conlon and Julie Holland Mortimer. Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30, 2013.

Peter Davis. Spatial competition in retail markets: movie theaters. *The RAND Journal of Economics*, 37(4):964–982, 2006.

Eleanor McDonnell Feit, Pengyuan Wang, Eric T Bradlow, and Peter S Fader. Fusing aggregate and disaggregate data with an application to multiplatform media consumption. *Journal of Marketing Research*, 50(3):348–364, 2013.

Chris Forman, Avi Goldfarb, and Shane Greenstein. How did location affect adoption of the commercial internet? global village vs. urban leadership. *Journal of urban Economics*, 58(3):389–420, 2005.

Chris Forman, Avi Goldfarb, and Shane Greenstein. Understanding the inputs into innovation: Do cities substitute for internal firm resources? *Journal of Economics & Management Strategy*, 17(2):295–316, 2008.

Chris Forman, Avi Goldfarb, and Shane Greenstein. The internet and local wages: A puzzle. *American Economic Review*, 102(1):556–75, 2012.

Xavier Giroud. Proximity and investment: Evidence from plant-level data. *The Quarterly Journal of Economics*, 128(2):861–915, 2013.

Edward L Glaeser and Joshua D Gottlieb. The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of economic literature*, 47(4):983–1028, 2009.

Edward L Glaeser and Janet E Kohlhase. Cities, regions and the decline of transport costs. In *Fifty Years of Regional Science*, pages 197–228. Springer, 2004.

W Michael Hanemann. Discrete/continuous models of consumer demand. *Econometrica: Journal of the Econometric Society*, pages 541–561, 1984.

Renna Jiang, Puneet Manchanda, and Peter E Rossi. Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics*, 149(2):136–148, 2009.

Wang Jin and Kristina McElheran. Economies before scale: Learning, survival, and performance of young plants in the age of cloud computing. *University of Toronto Working Paper*, 2019.

Andres Musalem, Eric T Bradlow, and Jagmohan S Raju. Who's got the coupon? estimating consumer preferences and coupon usage from aggregate information. *Journal of Marketing Research*, 45(6): 715–730, 2008.

Andrés Musalem, Marcelo Olivares, Eric T Bradlow, Christian Terwiesch, and Daniel Corsten. Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197, 2010.

Katja Seim. An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics*, 37(3):619–640, 2006.

Howard Smith. Supermarket choice and supermarket competition in market equilibrium. *The Review of Economic Studies*, 71(1):235–263, 2004.

Kenneth Train. A recursive estimator for random coefficient models. *University of California, Berkeley*, 2007.

Kenneth E Train. Em algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, 1(1):40–69, 2008.

Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

Zhiqiang Zheng, Peter Fader, and Balaji Padmanabhan. From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data. *Information Systems Research*, 23(3-part-1):698–720, 2012.

9 Detailed treatment of the EM algorithm

Following Bhat [1997], it can be shown that maximizing Eq.(7) is mathematically equivalent to maximizing

$$\sum_t \Big(\sum_{i \in C_t^M} \sum_{j \in \mathcal{F}_t^M} y_{ijt} log(P_{it}^j(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)) + Q_t^A \int_{\mathbf{l}_i} h_{\mathbf{l}_i,t}^A(\boldsymbol{\theta}) log(P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)) d\mathbf{l}_i + Q_t^O \int_{\mathbf{l}_i} h_{\mathbf{l}_i,t}^O(\boldsymbol{\theta}) log(P_{bt}^O(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)) d\mathbf{l}_i \tag{9}$$

if $h_{\mathbf{l}_i,t}^A(\boldsymbol{\theta})$ and $h_{\mathbf{l}_i,t}^O(\boldsymbol{\theta})$ are taken as given.[24] Here, $h_{\mathbf{l}_i,t}^A(\boldsymbol{\theta})$ and $h_{\mathbf{l}_i,t}^O(\boldsymbol{\theta})$ are the Bayesian posterior probabilities that an AWS customer or a non-cloud user is located at $\mathbf{l}_i$, i.e.

$$h_{\mathbf{l}_i,t}^A(\boldsymbol{\theta}) = \frac{P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\int_{\mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i} \tag{10}$$

$$h_{\mathbf{l}_i,t}^O(\boldsymbol{\theta}) = \frac{P_{it}^O(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\int_{\mathbf{l}_i} P_{it}^O(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i} \tag{11}$$

[24] Take the second term in Eq.(7) as an example, the necessary first-order conditions for maximizing it is

$$\begin{aligned}
\frac{\partial log(\int_{\mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i)}{\partial \boldsymbol{\theta}} &= \frac{1}{\int_{\mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i} \frac{\partial \int_{\mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i}{\partial \boldsymbol{\theta}} \\
&= \int_{\mathbf{l}_i} \frac{1}{\int_{\mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i} \frac{\partial P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}} d\mathbf{l}_i \\
&= \int_{\mathbf{l}_i} \frac{P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\int_{\mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i} \frac{\partial P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}} / P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i \\
&= \int_{\mathbf{l}_i} \frac{P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\int_{\mathbf{l}_i} P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i} \frac{\partial log P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}} d\mathbf{l}_i \\
&= \int_{\mathbf{l}_i} h_{\mathbf{l}_i,t}^A(\boldsymbol{\theta}) \frac{\partial log P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}} d\mathbf{l}_i
\end{aligned}$$

It is equivalent to maximizing $\int_{\mathbf{l}_i} h_{\mathbf{l}_i,t}^A(\boldsymbol{\theta}) log(P_{it}^A(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)) d\mathbf{l}_i$ with $h_{\mathbf{l}_i,t}^A(\boldsymbol{\theta})$ as given.

9.1 Maximization

Equation (8) can be maximized iteratively: starting from some initial values $\boldsymbol{\theta}^s$, we update $\boldsymbol{\theta}$ with $\boldsymbol{\theta}^{s+1}$ which maximizes Eq.(8) conditional on $h^A_{\mathbf{l}_i t}(\boldsymbol{\theta}^s)$ and $h^O_{\mathbf{l}_i t}(\boldsymbol{\theta}^s)$. Formally,

$$\begin{aligned}
\varepsilon(\boldsymbol{\theta}|\boldsymbol{\theta}^s) &= \sum_t \Big(\sum_{i\in C_t^M} \sum_{j\in\mathcal{F}_t^M} y_{ijt} log(P^j_{it}(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)) \\
&\quad + Q^A_t \int_{\mathbf{l}_i} h^A_{\mathbf{l}_i,t}(\boldsymbol{\theta}^s) log(P^A_{it}(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)) d\mathbf{l}_i \\
&\quad + Q^O_t \int_{\mathbf{l}_i} h^O_{\mathbf{l}_i,t}(\boldsymbol{\theta}^s) log(P^O_{bt}(\boldsymbol{\theta}_1) f(\mathbf{l}_i|\boldsymbol{\theta}_2)) d\mathbf{l}_i \Big) \\
\boldsymbol{\theta}^{s+1} &= argmax_{\boldsymbol{\theta}} \quad \varepsilon(\boldsymbol{\theta}|\boldsymbol{\theta}^s)
\end{aligned} \tag{12}$$

Furthermore, due to the property of log operation, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can be separately updated by maximizing the following two objective functions,

$$\begin{aligned}
\varepsilon_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}^s) &= \sum_t \Big(\sum_{i\in C_t^M} \sum_{j\in\mathcal{F}_t^M} y_{ijt} log P^j_{it}(\boldsymbol{\theta}_1) + Q^A_t \int_{\mathbf{l}_i} h^A_{\mathbf{l}_i,t}(\boldsymbol{\theta}^s) log P^A_{it}(\boldsymbol{\theta}_1) d\mathbf{l}_i + Q^O_t \int_{\mathbf{l}_i} h^O_{\mathbf{l}_i,t}(\boldsymbol{\theta}^s) log P^O_{bt}(\boldsymbol{\theta}_1) d\mathbf{l}_i \Big) \\
\varepsilon_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}^s) &= \sum_t \Big(\sum_{i\in C_t^M} \sum_{j\in\mathcal{F}_t^M} y_{ijt} log f(\mathbf{l}_i|\boldsymbol{\theta}_2) + Q^A_t \int_{\mathbf{l}_i} h^A_{\mathbf{l}_i,t}(\boldsymbol{\theta}^s) log f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i + Q^O_t \int_{\mathbf{l}_i} h^O_{\mathbf{l}_i,t}(\boldsymbol{\theta}^s) log f(\mathbf{l}_i|\boldsymbol{\theta}_2) d\mathbf{l}_i \Big)
\end{aligned}$$

9.2 A Discrete Spatial Distribution

Instead of assuming a parametric distribution for $f(\mathbf{l}_i|\boldsymbol{\theta}_2)$, we assume a discrete spatial distribution of consumer locations, so in theory it can approximate any arbitrary distribution when the discretization is fine enough. Specifically, we take each U.S. state and Canadian province as a bin B, and then the probability that a consumer (including non-cloud users) belongs to a certain bin b in period t is q_{bt}, and these q_{bt}'s are treated as parameters to estimate. So the objective functions is given as [25]

$$\begin{aligned}
\varepsilon_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}^s) &= \sum_t \Big(\sum_{i\in C_t^M} \sum_{j\in\mathcal{F}_t^M} y_{ijt} log P^j_{it}(\boldsymbol{\theta}_1) + Q^A_t \sum_b h^A_{bt}(\boldsymbol{\theta}^s) log P^A_{bt}(\boldsymbol{\theta}_1) + Q^O_t \sum_b h^O_{bt}(\boldsymbol{\theta}^s) log P^O_{bt}(\boldsymbol{\theta}_1) \Big) \\
\varepsilon_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}^s) &= \sum_t (Q^M_t \sum_b q^M_{bt} log q_{bt} + Q^A_t \sum_b h^A_{bt}(\boldsymbol{\theta}^s) log q_{bt} + Q^O_t \sum_b h^O_{bt}(\boldsymbol{\theta}^s) log q_{bt})
\end{aligned}$$

where

[25] For Microsoft customers,

$$\begin{aligned}
\sum_{i\in C_t^M} \sum_{j\in\mathcal{F}_t^M} y_{ijt} log f(\mathbf{l}_i|\boldsymbol{\theta}_2) &= \sum_b \sum_{i\in B_b} \sum_{j\in\mathcal{F}_t^M} y_{ijt} log q_{bt} \\
&= \sum_b \sum_{i\in B_b} log q_{bt} \sum_{j\in\mathcal{F}_t^M} y_{ijt} \\
&= \sum_b \sum_{i\in B_b} log q_{bt} \\
&= \sum_b Q^M_t q^M_{bt} log q_{bt}
\end{aligned}$$

The third equation holds because these Microsoft customers must choose one of the Microsoft DCs. And Q^M_t is the demand for Microsoft in period t, and q^M_{bt} is the spatial distribution specific for Microsoft consumers. Therefore $Q^M_t q^M_{bt}$ is the number of Microsoft customers in bin B_b, which is observable in our dataset .

$$h_{bt}^{A}(\boldsymbol{\theta}^{s}) = \frac{P_{bt}^{A}(\boldsymbol{\theta}_{1}^{s})q_{bt}^{s}}{\sum_{b} P_{bt}^{A}(\boldsymbol{\theta}_{1}^{s})q_{bt}^{s}} \tag{13}$$

$$h_{bt}^{O}(\boldsymbol{\theta}^{s}) = \frac{P_{bt}^{O}(\boldsymbol{\theta}_{1}^{s})q_{bt}^{s}}{\sum_{b} P_{bt}^{O}(\boldsymbol{\theta}_{1}^{s})q_{bt}^{s}} \tag{14}$$

Intuitively, $\boldsymbol{\varepsilon}_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}^s)$ can be considered as a variant of an ordinary multinomial logit model: since AWS customers in bin b share the same log likelihood $logP_{bt}^{A}(\boldsymbol{\theta}_1)$, it is multiplied by $Q_t^A h_{bt}^A(\boldsymbol{\theta}^s)$, the "posterior" number of AWS customers in bin b. Parallely, $logP_{bt}^{O}(\boldsymbol{\theta}_1)$ is multiplied by the "posterior" number of people who choose the outside option, $Q_t^O h_{bt}^O(\boldsymbol{\theta}^s)$. Therefore, we are essentially matching the predicted choice probabilities, or say market shares, with the "observed" ones given by $\boldsymbol{\theta}^s$.

For $\boldsymbol{\varepsilon}_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}^s)$, if we rewrite it as

$$\boldsymbol{\varepsilon}_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}^s) = \sum_t \sum_b (Q_t^M q_{bt}^M + Q_t^A h_{bt}^A(\boldsymbol{\theta}^s) + Q_t^O h_{bt}^O(\boldsymbol{\theta}^s)) log q_{bt},$$

it can be interpreted as pairing each q_{bt} with the "observed" total probability that a consumer belongs bin b. Moreover, it has a closed-form optimizer, i.e.

$$q_{bt}^{s+1} = \frac{Q_t^M q_{bt}^M + Q_t^A h_{bt}^A(\boldsymbol{\theta}^s) + Q_t^O h_{bt}^O(\boldsymbol{\theta}^s)}{M_t},$$

where $M_t = Q_t^M + Q_t^A + Q_t^O$ is used to denote the market size in period t. This closed-form solution would significantly ease the computation.

Henceforth, we repeat the procedure in Eq.(12) until parameters converge.

10 Monte Carlo Results

Table 5: Monte Carlo Experiment

	True value	Mean absolute error	Median absolute error	95% confidence interval
Panel A: Taste Parameters				
γ				
Discrete Manufacturing	-9.197×10^{-4}	2.7882×10^{-5}	2.3078×10^{-5}	$[-9.852 \times 10^{-4}, -8.717 \times 10^{-4}]$
Education	-1.920×10^{-3}	6.4716×10^{-5}	5.3641×10^{-5}	$[-2.082 \times 10^{-3}, -1.783 \times 10^{-3}]$
Health	-1.593×10^{-3}	4.8878×10^{-5}	3.5984×10^{-5}	$[-1.752 \times 10^{-3}, -1.523 \times 10^{-3}]$
Hospitality & Transportation	-2.265×10^{-3}	7.3319×10^{-5}	5.9226×10^{-5}	$[-2.480 \times 10^{-3}, -2.148 \times 10^{-3}]$
Insurance	-4.550×10^{-3}	1.2034×10^{-4}	8.2875×10^{-5}	$[-4.758 \times 10^{-3}, -4.184 \times 10^{-3}]$
Media / Telecome and Utilities	-1.749×10^{-3}	5.2614×10^{-5}	3.7984×10^{-5}	$[-1.887 \times 10^{-3}, -1.621 \times 10^{-3}]$
Nonprofit	-4.438×10^{-3}	1.3285×10^{-4}	1.0621×10^{-4}	$[-4.475 \times 10^{-3}, -4.084 \times 10^{-3}]$
Professional Services	-7.098×10^{-4}	1.6197×10^{-5}	1.3858×10^{-5}	$[-7.509 \times 10^{-4}, -6.751 \times 10^{-4}]$
Unknwon	-4.897×10^{-4}	1.0663×10^{-5}	8.4040×10^{-6}	$[-5.216 \times 10^{-4}, -4.664 \times 10^{-4}]$
β	-1.809×10^{-2}	3.9278×10^{-2}	3.9295×10^{-2}	$[-2.160 \times 10^{-1}, -8.954 \times 10^{-2}]$
ψ	1.5836	8.5543×10^{-2}	8.5721×10^{-2}	[1.415, 1.561]
ξ	1.0679	4.4033×10^{-2}	3.5976×10^{-2}	[0.950, 1.123]
ζ	2.229	3.7300×10^{-2}	3.5322×10^{-2}	[2.194, 2.328]
ζ_7	-1.117×10^{-3}	1.2989×10^{-3}	2.6814×10^{-3}	$[-1.1823 \times 10^{-3}, -1.023 \times 10^{-3}]$
α	6.327	7.4440×10^{-2}	6.3820×10^{-2}	[6.127, 6.458]
α_7	-8.965×10^{-1}	1.1828×10^{-2}	9.6289×10^{-3}	$[-9.122 \times 10^{-1}, -8.655 \times 10^{-1}]$
τ	2.213×10^{-1}	5.475×10^{-3}	4.998×10^{-3}	$[1.986 \times 10^{-1}, 2.224 \times 10^{-1}]$
Panel B: Consumer Spatial Distribution Parameters				
Std Err. q_{mbt}	4.342×10^{-3}	1.4870×10^{-5}	1.5021×10^{-5}	$[4.120 \times 10^{-3}, 4.315 \times 10^{-3}]$

This Figure shows implied cloud market size by industry for different regions in the US with dark colors being large market sizes. Put another way, this Figure shows the total demand by industry of AWS plus Azure for different regions. It recovers sensible patterns such as discrete manufacturing is prominent in the upper midwest and west coast and professional services being largest on the east coast and and west coast. We take this as evidence the model is recovering sensible market level patterns.

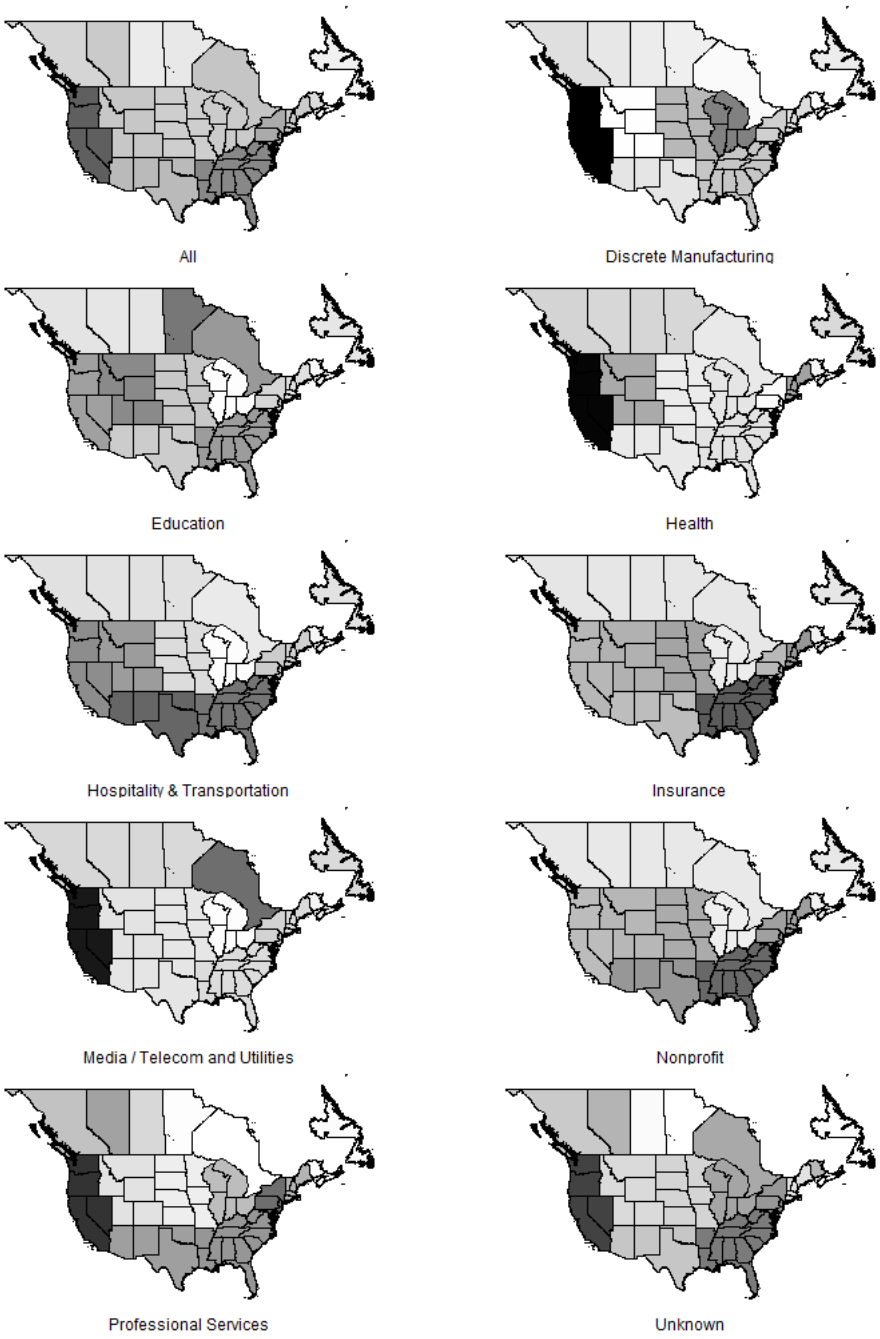


Figure 10: Market size map