Spatial Competition and Missing Data: an Application to Cloud Computing

Shuang Wang*, Jacob LaRiviere†, Aadharsh Kannan‡

February 28, 2019

Abstract

We introduce a 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. Such a hybrid data structure is common to firm managers, economic consultants, and in merger analysis, but leveraging it for jointly estimating preferences for proximity and price sensitivity is not common. We use EM algorithm to tackle the customer level missing data problem of the second firm. Simulation shows that both the demand parameters and consumers' spatial distribution can be precisely recovered. We then estimate the model using a proprietary anonymized dataset from the fast growing cloud computing sector. Specifically, we use anonymized purchase level data from Microsoft's Azure purchase and aggregate market level Amazon Web Service revenue data. Estimation results show consumers' preference for geographic proximity (customer location and data center location) and imply a substantial variation in local market shares. Counterfactuals show that a new data center opened in a first best location can generate a market share gain 40% higher than a suboptimal location, and a price drop is most effective where the spatial competition is relatively intensive.

Keywords: hybrid data, EM algorithm, mixed logit

*Corresponding Author. Boston University, Department of Economics, shuangw@bu.edu. Thanks to Greg Lewis, Preston McAfee, Marc Rysman and Michael Schwarz for very helpful comments. All mistakes are ours.

†Microsoft, Office of the Chief Economist, Microsoft, jlariv@microsoft.com

‡Amazon, aadharshkannan@gmail.com

1 Introduction

Managers often have to make strategic decisions with incomplete information. For example, firms often have detailed data on their own customers' purchase history but no detailed sales data from their competitors. Missing competitors' purchase data makes understanding market dynamics challenging because firms can't tell whether increased sales are attributable to taking market share from competitors or increasing market size.

When there is a spatial element to competition and both customer locations and store locations matter, the missing data problem faced by firms is worse. In spatial competition, each store location interacts with customer location to determine customer-store level product characteristics. Thus, the missing data problem increases in dimensionality: each potential customer's distance to every store (both own firm and competitor firm) is now a product characteristic. Without knowing customer locations of competitor firms, it is challenging for managers to infer how the market responds to their own store openings and closings. Thus, there are actually two types of missing data for managers faced with the spatial competition: missing data on store level sales of competitors and missing data on customer locations of competitors.

In this study, we introduce a 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. Cloud computing is a nice example of spatial competition because compute latency is proportional to distance between application user and server location. Further, cloud computing is a rapidly growing sector important in its own right but understudied. 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.[1] Cloud computing growth is predicated upon the expansion of the data center (DC) footprints which house the servers that are the cloud. Microsoft's cloud, called Azure, introduced more than 15 data centers worldwide over the past year, with plans announced for 12 additional regions[2]. Cloud computing is 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.

We investigate consumers' preference over specific data centers in 2016 North American market. When a customer deploys a cloud computing workload, they pick a specific location to deploy the workload in. Location matters for cloud users: further distance between compute and users creates latency reducing experience quality, where national borders exist data sovereignty regulation can impact location decisions, and consumers might simply have preferences for different locations.[3]. To this end, cloud providers can compete spatially through their location choice of DCs. We focus on spatial competition between Amazon Web Services (AWS) and Microsoft Azure, the top two cloud providers in the US market.[4] So long as customers have preferences for proximity, any demand model of cloud which doesn't

[1]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

[2]See "Azure regions", accessed on 09/04/2018, https://azure.microsoft.com/en-us/global-infrastructure/regions/

[3]See "What is cloud computing? Everything you need to know about the cloud, explained", 01/24/2018, https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know-from-public-and-private-cloud-to-software-as-a/

[4]Spatial competition depends on two primary consumer preferences: consumers' willingness-to-pay (WTP) for proximity and consumers' geographic distribution. Cost heterogeneity is clearly a third important component which we largely do not model in this paper. In general though, a Hotelling line or Salop Circle model of competition highlights these consumer attributes.

account for customer and data center location will be mis-specified. While any cloud manager will have detailed data on purchasing behavior of their own customers, like any other industry it is uncommon to have detailed data on the behavior of their competitor's customers, especially their locations.

Our empirical technique uses the detailed data of one firm to estimate the spatial component of demand leveraging new point of sale openings (e.g., DC openings), projects those choice set changes onto the consumer market, then iteratively compares the model's predictions to observed aggregate sales with the expectation-maximization (EM) algorithm. After showing the approach recovers underlying demand parameters via simulation, we take the model to data with a proprietary dataset from the cloud computing sector. Specifically, we use purchase data anonymized to the zip code level for Microsoft's Azure, aggregate market level AWS revenue data and publicly available DC opening dates and locations.[5]

We use an iterative EM algorithm to address the missing data problem in our spatial demand problem. In each iteration, 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. Consumers' taste is assumed to be homogeneous conditional on product characteristics and consumer attributes. This is a reasonable first order assumption for a good like cloud compute resources which is a form of technological infrastructure. Given this assumption the individual level data of Microsoft alone could identify taste parameters. We leverage the rollout of new data centers by both AWS and Azure which are closer versus further to the set of possible cloud customers. By observing the rate at which new customers begin purchasing Azure when new AWS or Azure data centers go live and how those rates vary over space, we can identify preferences for proximity to data centers. Next, we project demand parameters identified from the Azure data to the observed AWS data center characteristics and sum across data centers. The gap between the projection and the observed AWS market share is attributed to different 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.

We show via simulation that the model successfully recovers the demand parameters and unobserved consumer spatial distribution then take the model to the data. Results show that cloud users have a preference for nearby data centers. Therefore the spatial layout of DCs induces a significant variation in local market power. Using the estimated parameters we perform counterfactual exercises to determine how market structure would change if a new data center is introduced in different locations or if market prices change. Among the six possible counterfactual Microsoft Azure data center locations, the most profitable one could generate a market share gain over 40% higher than the least. Thus, the costs of placing a data center sub-optimally are large. The model lets us decompose that across customers purchasing the outside good (no cloud) versus purchasing from a competitor. Consistent with economic theory, we find that the benefits of price competition are greatest where both Azure and AWS have a data center.

There are two main lessons from this research. First, our results provide evidence that spatial competition is important in the cloud computing industry. By focusing on the North American market we ignore data sovereignty issues but highlight that those are likely to also be important. Second, 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

[5] Precise location data is not publicly available but we show that coarse location data (e.g., state-region) in our use case is sufficient to recover key parameters.

interested in questions regarding the impacts of opening and closing of brick and mortar stores faced with increasing online competition.

This paper contributes to two strands of literature. 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 disaggregate (e.g., store level) sales unobservable. Because this is a spatial model of competition, the consumer-store level attributes 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 of world cup, albeit in a context of a multivariate model and computationally more burdensome because of the inevitable Markov Chain Monte Carlo simulation.

Second, 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 easily 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 how EM algorithm can be employed to address the missing data problem. In Section 5, we discusses how parameters are identified by the available data. Section 6 shows the performance of a Monte Carlo experiment. In Section 7, we describe data resources and necessary assumptions to facilitate estimation. Section 8 gives the estimation results from the data. We do not report cross price elasticity estimates to protect trade secrets but rather perform two counterfactual exercises follow in Section 9 which show the costs of suboptimal management actions. We conclude this paper in Section 10.

2 Industrial Background

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

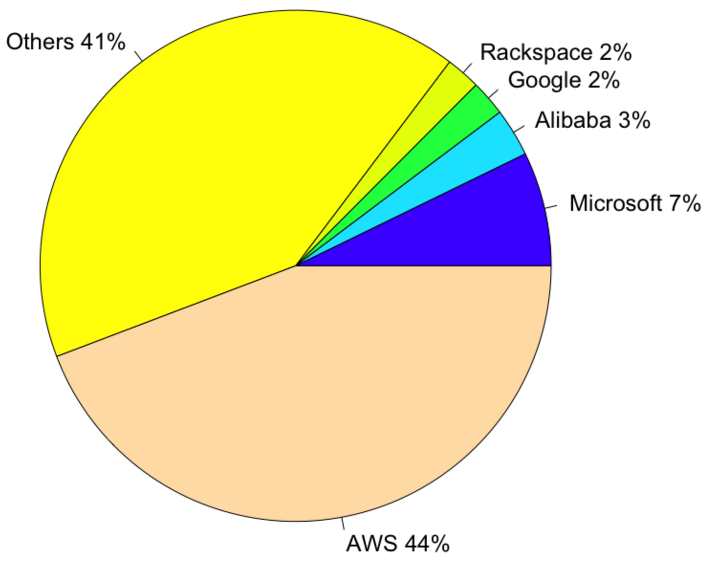


Figure 1: 2016 IaaS Public Cloud Computing Market Share

"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. 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 by the cloud providers' economies of scale and management of hardware and security.

Amazon Web Services (AWS) and Microsoft Azure are the two firms that have the largest market shares in global IaaS cloud market. In 2016, the total value of this market reached $22,160$ million U.S. dollars, of which AWS commands 44%, followed by Microsoft Azure at 7.1% [6], as shown in Figure 1.

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 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 in 2016. We allow consumers to have a demand shifter for being in a domestic data center in addition to explicitly controlling for distance between a customer's location and different data centers. 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.

[6]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.

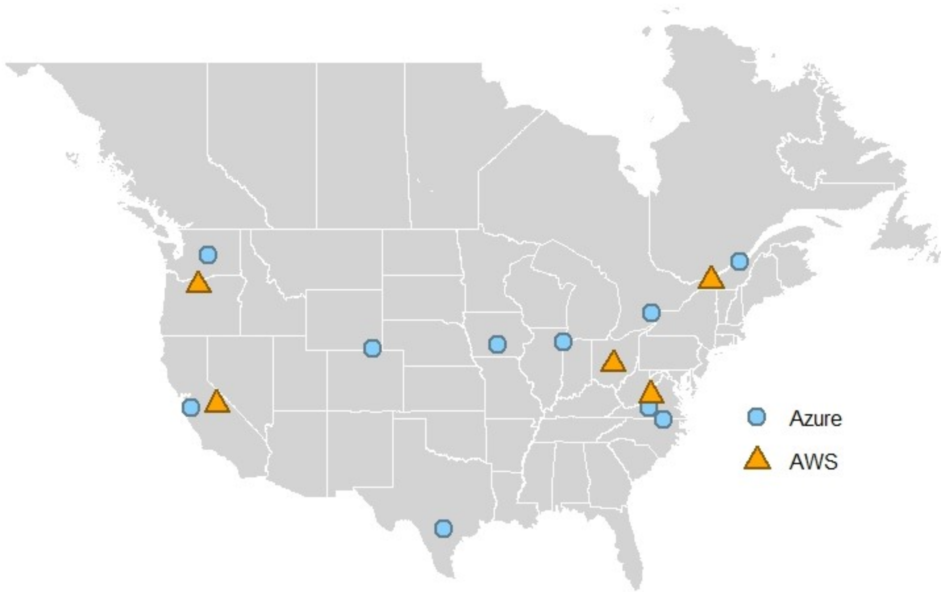


Figure 2: AWS and Azure Data Center Locations

Spatial proximity is likely an important aspect of DC differentiation, especially for those speed-sensitive users. Data transfer takes time, and the resulted latency could be further amplified due to security protocols. For example, for a consumer located in London, the download speed from a DC in Amsterdam is usually 40-50 Mbps, 3 times of that from New York City. Since spatial proximity is determined by both DC location and consumer location, consumer heterogeneity plays 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, EM algorithm enables a simultaneous estimation of both demand parameters and consumer spatial distribution, which is needed for any policy analysis respecting spatial demand preferences.

3 Model

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.

Because the model we propose is very much tied to the structure of the data in our problem so first explain the data structure before discussing the model. 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 then make two very strong assumptions on the product mix for these two firms which we detail further in the data section below. Our description here is not precisely correct, but it is to a first order approximation for the purposes of placing the model in context and support exposition. First, we assume that the cloud revenue for *basic A1* relative to all Azure revenue is the same as the revenue for *t2.small* relative to all AWS revenue. Put another way, after accounting for publicly available prices for these two products if there are \$1,000 in U.S. and Canada Azure sales for each *basic A1* purchase, we assume there is also \$1000 in sales for each *t2.small* AWS purchase. Second, we assume that both firms earn the same percentage of their total cloud revenue in the U.S. and Canada. Accounting for prices, product mixes and revenue mixes, this lets us back out quantities for AWS. We motivate and discuss the drawbacks to these assumptions in the data section at length below but we feel being up front about the structure of the data before introducing the model eases the model's introduction.

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

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

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.
- $d(\mathbf{l}_j, \mathbf{l}_i)$ 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.
- $price_{jt}$ is the price of DC j in period t
- $DCAge_{jt}$ 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 live programming stream 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. Furthermore, consumers' tolerance for latency could be different for domestic DCs and non-domestic DCs, i.e. $\dfrac{\partial u_{ijt}}{\partial d(\mathbf{l}_j, \mathbf{l}_i)} = \gamma + \rho \times \mathbb{1}_{ij}\{domestic\}$.

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. 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.

In the indirectly utility function specified above, there is no subscript i for all taste parameters so that there are no differences in marginal utility for latency, DC age nor domestic location of DCs across firms. In other words, if there are customers at exactly the same location, they will have identical preference over the DCs up to an idiosyncratic shock, ϵ_{ijt}. While the model allows for random taste parameters in principle, on average there must be no differences in marginal utility for latency, DC age nor domestic location of DCs for our inference to be valid. It enables us to identify taste parameters for cloud computing attributes with only a single firm's data. For cloud computing this is likely to be a reasonable assumption: latency degrades customer experience homogeneously across cloud providers. More details about identification are in Section 5.

3.1 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:

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

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 Azure, $\mathcal{F}_t^M$, and that of 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 including locations of all available DCs in period t
- $\mathbf{z}_{jt} = (price_{jt},\ DCAge_{jt},\ \mathbb{1}_j^{AWS})$ are product characteristics, and $\mathbf{z}_{\cdot t} = \{\mathbf{z}_{jt}, \forall j \in \mathcal{F}_t\}$ is simply a collection of them all.

- $\boldsymbol{\theta}_1 = (\gamma,\ \beta,\ \psi,\ \rho,\ \xi,\ \zeta,\ \alpha,\ \tau)$ is the set of utility parameters.

3.1.1 Missing DC level AWS demand

Although we can directly use Eq.(3) to denote the probability that a Microsoft customer chooses any specific Microsoft DC, the DC level choices of AWS customers are unobservable in our Microsoft Azure dataset . Therefore, 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, i.e.

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

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

Finally, the probability of not signing up for Microsoft Azure or AWS is

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

3.1.2 Missing consumer locations

In addition to DC level AWS demand, the locations of non-Microsoft customers are also unobservable, which makes the calculation of the choice probabilities conditional on consumer location infeasible. To circumvent this problem, we will take consumer location as a random variable, get the joint probability of location and choice, then integrate out its uncertainty in location for AWS and the outside option consumers. Particularly, the likelihood function in period t can be written as

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

where $C_t^f, f = M, A, O$ are the costumer sets for Microsoft, Amazon and non-cloud users respectively, and $f(\mathbf{l}_i|\boldsymbol{\theta}_2)$ is the spatial distribution of all consumers in the market, which is parameterized by $\boldsymbol{\theta}_2$. In practice we take the spatial distribution of consumers in the market to be the spatial distribution of medium and large firms across U.S. states and Canadian provinces.

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

$$\begin{aligned} LL_t(\boldsymbol{\theta}) \quad = \quad & \sum_{i \in C_t^M} \sum_{j \in \mathcal{F}_t^M} y_{ijt} log(P(y_{ijt} = 1|\mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) f_t(\mathbf{l}_i|\boldsymbol{\theta}_2)) \\ & + \sum_{i \in C_t^A} log \int_{l_i} P(y_{iAt} = 1|\mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) f_t(\mathbf{l}_i|\boldsymbol{\theta}_2) dl_i \\ & + \sum_{i \in C_t^O} log \int_{l_i} P(y_{iOt} = 1|\mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1) f_t(\mathbf{l}_i|\boldsymbol{\theta}_2) dl_i \end{aligned} \tag{7}$$

Then, summing up the log likelihood function over time, we can get

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

where $P_{it}^j(\boldsymbol{\theta}_1)$, $P_{bt}^A(\boldsymbol{\theta}_1)$ and $P_{bt}^O(\boldsymbol{\theta}_1)$ simplifies $P(y_{ijt} = 1|\mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1)$, $P(y_{iAt} = 1|\mathbf{l}_i \in C_b, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1)$ and $P(y_{ijt} = 1|\mathbf{l}_i, \mathbf{l}_t^{DC}, \mathbf{z}_{\cdot t}, \boldsymbol{\theta}_1)$ correspondingly. Once we take expectation over the unknown consumer's locations, the expected choice probability is same for every AWS consumer or any potential consumer. Therefore, we multiply them by the total quantities Q_t^A and Q_t^O.[7]

4 EM Algorithm

Maximizing the log likelihood function above with the usual Newton or quasi-Newton routines can be numerically difficult and computationally unstable, for which EM algorithm can be remedial. It 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.

4.1 Expectation

Following Bhat [1997], it can be shown that maximizing Eq.(8) 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.[8] 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.

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

[8] Take the second term in Eq.(8) as an example, the necessary first-order conditions for maximizing it is

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

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

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

4.2 Maximization

Equation (9) can be maximized iteratively: starting from some initial values $\boldsymbol{\theta}^s$, we update $\boldsymbol{\theta}$ with $\boldsymbol{\theta}^{s+1}$ which maximizes Eq.(9) conditional on $h^A_{\mathbf{l}_it}(\boldsymbol{\theta}^s)$ and $h^O_{\mathbf{l}_it}(\boldsymbol{\theta}^s)$. Formally,

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

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,

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

4.3 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 [9]

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

[9] For Microsoft customers,

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

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.

$$\begin{aligned}\boldsymbol{\varepsilon}_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}^s) &= \sum_t(\sum_{i\in C_t^M}\sum_{j\in\mathcal{F}_t^M} y_{ijt}logP_{it}^j(\boldsymbol{\theta}_1) + Q_t^A\sum_b h_{bt}^A(\boldsymbol{\theta}^s)logP_{bt}^A(\boldsymbol{\theta}_1) + Q_t^O\sum_b h_{bt}^O(\boldsymbol{\theta}^s)logP_{bt}^O(\boldsymbol{\theta}_1))\\ \boldsymbol{\varepsilon}_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}^s) &= \sum_t(Q_t^M\sum_b q_{bt}^M logq_{bt} + Q_t^A\sum_b h_{bt}^A(\boldsymbol{\theta}^s)logq_{bt} + Q_t^O\sum_b h_{bt}^O(\boldsymbol{\theta}^s)logq_{bt})\end{aligned}$$

where

$$h_{bt}^A(\boldsymbol{\theta}^s) = \frac{P_{bt}^A(\boldsymbol{\theta}_1^s)q_{bt}^s}{\sum_b P_{bt}^A(\boldsymbol{\theta}_1^s)q_{bt}^s} \tag{13}$$

$$h_{bt}^O(\boldsymbol{\theta}^s) = \frac{P_{bt}^O(\boldsymbol{\theta}_1^s)q_{bt}^s}{\sum_b P_{bt}^O(\boldsymbol{\theta}_1^s)q_{bt}^s} \tag{14}$$

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

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

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.

5 Identification

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

First, $(\gamma,\beta,\psi,\rho,\xi)$ are identified from the substitution pattern of Microsoft customers among Microsoft DCs. Since our Microsoft data is at individual level, including the locations of each customer, both $d(\mathbf{l}_i,\mathbf{l}_j)$'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*

Therefore $Q_t^M q_{bt}^M$ is the number of Microsoft customers in bin B_b, which is observable in our dataset .

(IIA) property of logit model makes it possible to focus on only a subset of products (Train [2009]).

Next, if we consider ζ and $\{v_{Ot}\}_t$ [10] 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 $\boldsymbol{\theta}_{1,2} = (\zeta, \alpha, \tau)$.

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

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

Note that μ_i is individual-specific due to consumers' heterogeneous locations.

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

$$\frac{Q^A_{bt}}{Q^M_{bt}} = \frac{\sum\limits_{j\in\mathcal{F}^A_t} exp(\zeta + \mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum\limits_{j\in\mathcal{F}^M_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} = exp(\zeta)\frac{\sum\limits_{j\in\mathcal{F}^A_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum\limits_{j\in\mathcal{F}^M_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}$$

Relate this to the observed market level AWS demand by $Q^A_t = \sum\limits_b Q^A_{bt}$, we have

$$Q^A_t = exp(\zeta)\sum_b \frac{\sum\limits_{j\in\mathcal{F}^A_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum\limits_{j\in\mathcal{F}^M_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{bt}$$

which gives

$$\zeta = log(Q^A_t / \sum_b \frac{\sum\limits_{j\in\mathcal{F}^A_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))}{\sum\limits_{j\in\mathcal{F}^M_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{bt})$$

Similarly,

$$v_{Ot} = log(Q^O_t / \sum_b \frac{1}{\sum\limits_{j\in\mathcal{F}^M_t} exp(\mu_b(l_j, \mathbf{z}^1_{jt}, \boldsymbol{\theta}_{1,1}))} Q^M_{bt})$$

[10] 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.

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

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

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

6 Monte Carlo Experiment

To test the model's identification, we did a Monte Carlo experiment. The data generation process is as follows,

1. We construct a hypothetic geographic rectangle, ranging from 20°N to 60°N, 70°E to 160°E. We then evenly divide it into 36 rectangular bins. 40°N is taken as the boarder between "country" C and "country" U.
2. Randomly generate the time-varying consumer distribution q_{bt}'s, the probability that a consumer is located in bin b in period t, where $b = 1, 2, ..., 36$, $t = 1, 2, 3$
3. There are 2 firms in the market, M and A, they each offer multiple DCs and an outside good is also available. Both firm M and firm A start from country U and sequentially rollout their DCs into country C. Specifically, in period 1, both M and A only have DCs in country U; in period 2, firm M introduces their first DC in country C and then firm A follows in period 3.
4. The market size grows over time: $M_1 = 8000$, $M_2 = 10000$, $M_3 = 12000$. Then consumer sets are generated based on consumer spatial distribution q_{bt} and market size M_t, and consumers spread evenly within each bin.
5. The utility function is specified as in Eq.(1) and Eq.(2), where
   - prices are from a uniform distribution $U[1, 6]$, $i.i.d$ across j, t;
   - $\mathbb{1}_{ij}\{domestic\} = 1$ if the DC and the consumer are on the same side of the "border"; 0 otherwise;
   - ϵ_{ijt} are $i.i.d$ type I extreme value errors
   - $\boldsymbol{\theta}_1 = (\gamma, \beta, \psi, \rho, \xi, \zeta, \alpha, \tau) = (-4.14 \times 10^{-4},\ 6.9 \times 10^{-1},\ 4.33 \times 10^{-5},\ 1.37,\ 1.38,\ 1.95,\ 1,\ -5 \times 10^{-1})$,

We then generate 100 different dataset s. For firm M, we keep the data at consumer level, while the data for firm A and the outside good is aggregated to market shares. Taste parameters and the distribution parameters are estimated using the model as described in Section 3 by EM algorithm. As shown in Table 1, all taste parameter estimates are concentrated at the true values with very small variations.

Table 1: Monte Carlo Experiment

Panel A: Taste Parameters			
	Mean Absolute Error	Median Absolute Error	Std. dev Estimator
γ	-1.10×10^{-6}	-4.91×10^{-8}	3.94×10^{-5}
β	6.88×10^{-4}	1.07×10^{-3}	1.13×10^{-2}
ψ	6.14×10^{-3}	-2.27×10^{-2}	2.21×10^{-1}
ρ	6.68×10^{-6}	6.60×10^{-6}	3.89×10^{-5}
ξ	4.16×10^{-2}	4.27×10^{-4}	3.37×10^{-1}
ζ	5.46×10^{-3}	-1.05×10^{-3}	7.05×10^{-2}
α	2.69×10^{-2}	-8.30×10^{-3}	3.19×10^{-1}
τ	1.70×10^{-2}	-4.30×10^{-2}	1.61×10^{-1}
Panel B: Consumer Spatial Distribution Parameters			
	Mean Absolute Error	Median Absolute Error	Std. dev Estimator
Mean q_b	2.07×10^{-11}	0	3.38×10^{-10}
Varaince q_b	2.08×10^{-5}	9.52×10^{-6}	3.92×10^{-5}

7 Data

We take this model to the data by focusing on the cloud computing industry in 2016 North American market, including the U.S. and Canada, for Microsoft Azure's popular *basic A1* offering and AWS's popular *t2.small* offering. In 2016, both Microsoft Azure and AWS introduced their first DCs in Canada. Therefore, there is reasonably large variation in DC locations during this period. We leverage this rollout in new DCs to identify the marginal utility of distance and domestic location.

There was also some modest price variation in our sample period. Figure 3 shows average prices across DCs for Azure's *basic A1* and AWS's *t2.small*. When the first Canadian DCs of Microsoft and AWS were introduced in April, 2016 and Dec, 2016, prices changed to reflect local DC level pricing decisions. The price drop of Microsoft DCs in Oct, 2016 was a strategic price decrease which we leverage to identify demand elasticities.

Our main data resources are as follows:

- **Microsoft Internal Data**: Three parts of our data are from internal Microsoft resources: (1) we have a subset of Microsoft internal monthly customer level choice data for the *basic A1* SKU, including their DC level choices and locations; (2) Microsoft quarterly revenues from cloud computing business; (3) product characteristics of Microsoft, including prices, locations and start dates.
- **Amazon Quarterly Revenue Report**: As a listed company, Amazon publishes their financial results in each quarter, where the net sales of AWS is specified[11]. Note that these net sales are worldwide, not restricted to North American market. To recover the revenue ratio of Microsoft Azure to AWS in North America, we need to assume that such a ratio is constant across the world. With such a ratio, the brand level demand for AWS will be recovered based on the observed Microsoft Azure demand. More details can be found later in this section. Also note that the DC level demand and the customer locations of AWS are unobservable, that is why the traditional demand estimation such as mixed logit is infeasible and motivates our EM algorithm design. We

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

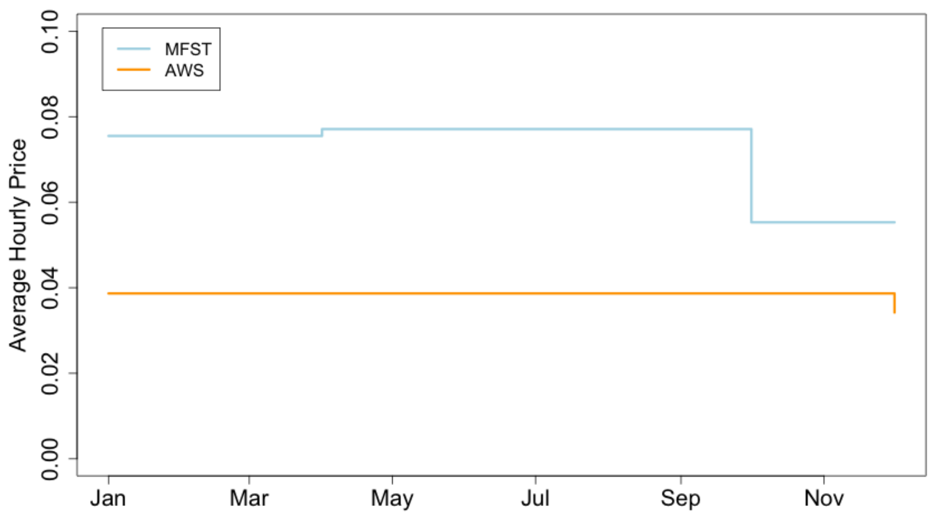


Figure 3: Average Hourly Price of Microsoft *basic A1* and AWS *t2.small*

also scraped historical price data for the *t2.small* offering and dates for when AWS data centers begin to operate.

- **Market Share Data**: Here, we further assume that the market for cloud computing is defined by all firms in the U.S. and Canada, so that the demand for the outside option can be recTable overed based on our knowledge of Microsoft demand and AWS demand. It is actually a mild assumption, since usually local machines are sufficient for the computational demand of most households and individuals. Specifically, for the U.S. market, the total number of businesses is from Henry J Kaiser Family Foundation (KFF), a non-profit organization focusing on national health issues. They have data on number of private sector firms by size in order to keep track of their health insurance coverage. While for the Canadian market, such a number is from Statistics Canada, a Canadian government agency which can be considered as the counterpart of the U.S. Census Bureau[12] . Specifically, 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.

To facilitate estimation, we make a few more assumptions on DC level price, Amazon market share and market size. Both Microsoft and Amazon provide various VM's at each DC, which makes the definition of a DC level price unclear. Although the most straightforward strategy could be the average VM price weighted by their demands, it is infeasible since the VM level AWS demands are unobservable in our dataset . Therefore, for simplicity we take a popular Microsoft Azure VM, *basic A1* and choose AWS's *t2.small* as its counterpart, which has almost an identical technical specification as *basic A1* as benchmarks. We believe that the general compute targeting of *basic A1* and *t2.small* make them the best possible choice. While exact percentages are a trade secret, *basic A1* made up a non-trivial amount

[12] More detailed information can be found on the following website.
U.S. https://www.kff.org/other/state-indicator/number-of-firms-by-size;
Canada: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3310003401.

Table 2: Comparison between Microsoft *basic A1* and AWS *t2.small*

Name	Brand	vCPUs	RAM(GiB)	Current Price/Hour (California)
basic A1	Mircosoft	1	2	\$0.0650/hour
t2.small	Amazon	1	2	\$0.0368/hour

of Azure business. Therefore, we use the prices of *basic A1* and *t2.small* at each DC to approximate a DC level price. Furthermore, for both *basic A1* and *t2.small*, we use the prices for Windows operation system.

A more detailed technical comparison between *basic A1* and *t2.small* can be found in Table 2. One caveat to the table is that the virtual machines are different in dimension of product quality; whereas *basic A1* is a dedicated core, *t2.small* is a burstable VM. That means a *t2.small* core might not be available if deployed whereas a dedicated core would be. This is reflected in the price difference but also will be picked up in the brand fixed effects we estimate in the model.

Next, the demand for Amazon is recovered from ratio of Microsoft Azure revenue to AWS revenue, as in their quarterly financial reports. And we extrapolate the quarterly data to monthly level by assuming that the revenue grows at a constant rate within each quarter. Then, since the ratio of quantities equals that of revenue adjusted by average price across different DCs, the demand for Amazon should be the price-adjusted revenue ratio multiplied by the observed demand for Microsoft Azure.

Finally, the demand for the outside option is the market size excluding the demand for Microsoft Azure and AWS. Specifically, as discussed before, we have the total number of private sector firms in the U.S. from KFF. The data is at yearly level, so 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, again based on the constant growth rate assumption. We do similar things for Canadian market. 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; while for Canada it is 4.7%. Finally, the market size was multiplied by *basic A1*'s demand share within Microsoft, so that the patterns in market shares are kept consistent even when focus on *basic A1* and *t2.small* only.

A summary of observables and unobservables follows in Table 3.

Table 3: Summary Statistics

Panel A: Consumer Characteristics		
Consumer Characteristics	**Microsoft Azure**	**AWS**
Locations	observable	unobservable
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)

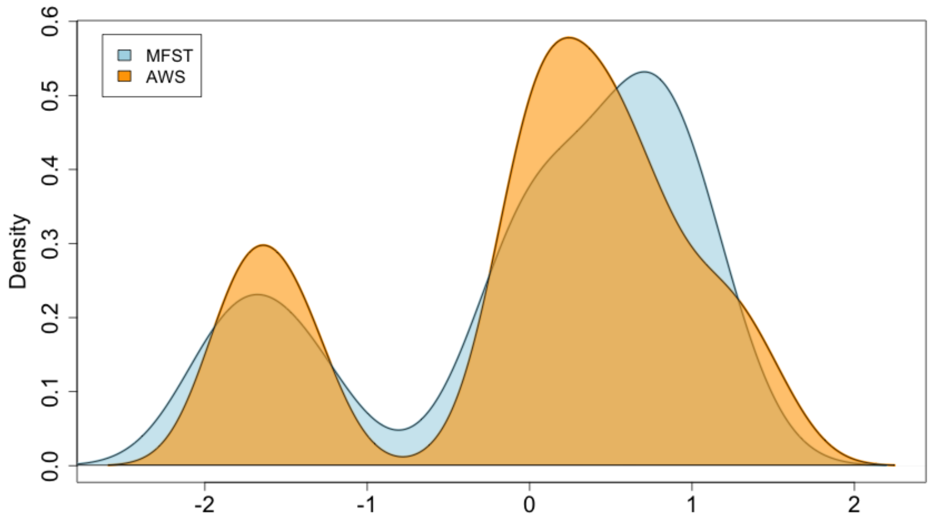


Figure 4: Microsoft vs AWS Market Share Distribution

8 Estimation Results

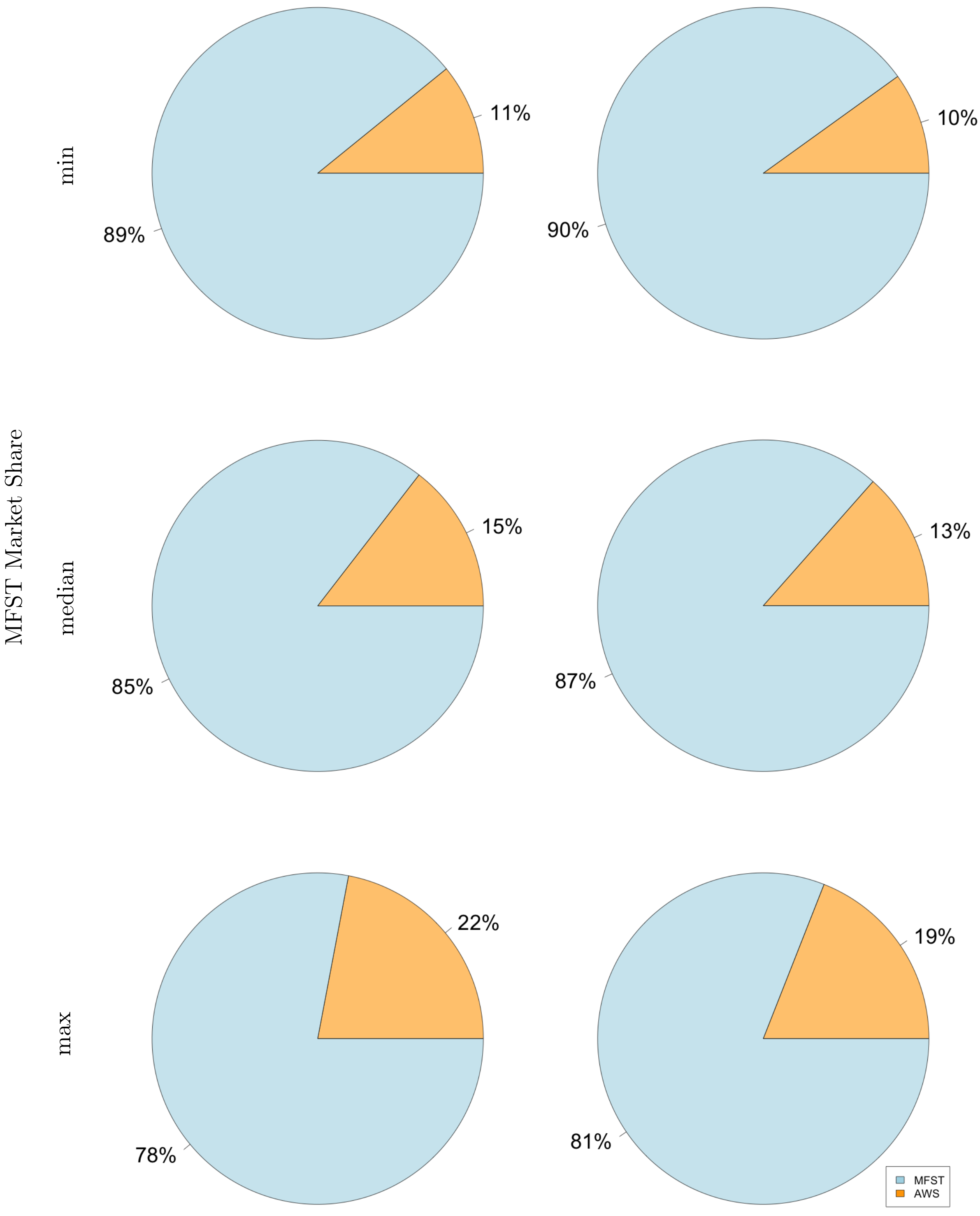
As discussed in Section 5, we estimate the IaaS cloud demand in North America during Jan, 2016 to Dec, 2016 and recover the local market size for each U.S. state and Canadian province. We do not report preference parameters in order to protect trade secrets.[13] We can discuss parameters generally, though. All parameters are sharply estimated and have expected sign. For example, customers dislike distances and high prices but generally prefer domestic DCs. AWS has a positive product fixed effect rationalizing their market size over this time period. Thus, the model recovers a very reasonable characterization of the market.

As for consumer spatial distribution q_{bt}'s, the mean estimate is 0.0156, with a standard deviation of 0.0294. Recall that q_{bt} is the percent of total market volume (both AWS, Microsoft's Azure and the outside option) which comes from a particular geographical bin b. Of note is that the comparatively large standard deviation; it implies a considerable variation in local market size, which one would expect to be crucial to cloud providers' decision on DC location.

In addition to aggregate market size by region, the relative market shares between Azure and AWS also vary across the U.S. and Canada. Figure 4 gives the densities of each firm's estimated market shares normalized with respect to each firm's mean and standard deviation respectively. Both firms exhibit bimodal market shares over space: both firms have some regions above the firm-specific mean market share and some regions below, but the distribution is not single-peaked. Even though AWS was the market leader during 2016, in some regions Azure had higher market share than it did on average. Recall that 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.

While Figure 4 shows variation in market share within firms over space, Figure 5 shows variation in market shares across firms within space. We obsfuscate precise locations to protect trade secrets and instead show variation in the minimum, median and maximum Azure market share across markets sized

[13]Indeed, it is well known that Amazon has a large number of quality economists working on similar issues!



(a) Bottom-sized markets (b) Top-sized markets

Figure 5: Microsoft vs AWS Market Share in Top-sized & Bottom-sized Markets

below the median (bottom-sized markets) and those sized above (top-sized markets). Note that Figure 5 represents a relatively small level of market penetration highlighting that the cloud computing industry is still very young.

We estimate changes in market share across regions of roughly 100% for both relatively small and relatively large markets. 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. Recall that these market shares are from 2016 data and since then Azure has grown in market share. Thus these numbers do not reflect current shares.

9 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. Thus our main counterfactual focuses on using the model to measure the gains to leveraging the model to optimize data center location. We also condition on current DC layout in North America, and predict the market share responses to a counterfactual 15% price change for all Microsoft DCs.

First, we propose 6 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 4 new DCs in North America in 2016 which increased its total number to 10, twice that of AWS. Therefore, it is reasonable to quantify the impact of a denser product space.

All the data used in counterfactual analysis is the December, 2016 data. Lastly, note that equilibrium effects are beyond the scope of these exercises. Neither AWS nor Microsoft itself is allowed to alter to the current DC layout or adjust the prices of existing DCs according to the counterfactual changes in the market.

9.1 New DC Location

The 6 proposed states spread evenly in southern U.S. from the west coast to the east coast. We anonymize the names of those states by referring to them as Region 1-6. Furthermore, we assume that the price is set at the average Microsoft price level so that the different demand responses could be attributed to the differences in local market size and DC layout.

Results are shown in Figure 6. The market share changes are normalized to that of our numeraire region, Region 1. It shows that introducing a new DC in Region 3 generates highest market share gains for Microsoft Azure, which is around 40% higher than the numeraire region. This result is consistent with both observed facts and estimation results: first, the DC density in Region 3 is comparatively lower than the others. Therefore, a newly-introduced DC would overturn the relative utilities from the outside good and cloud computing by reducing distance; second, according the estimated local market sizes, q_{bt}'s, Region 3 is ranked among the top in the country. Therefore, the size of consumers who previously choose the outside option and then switch to Microsoft because of the reduction in distance is relatively larger.

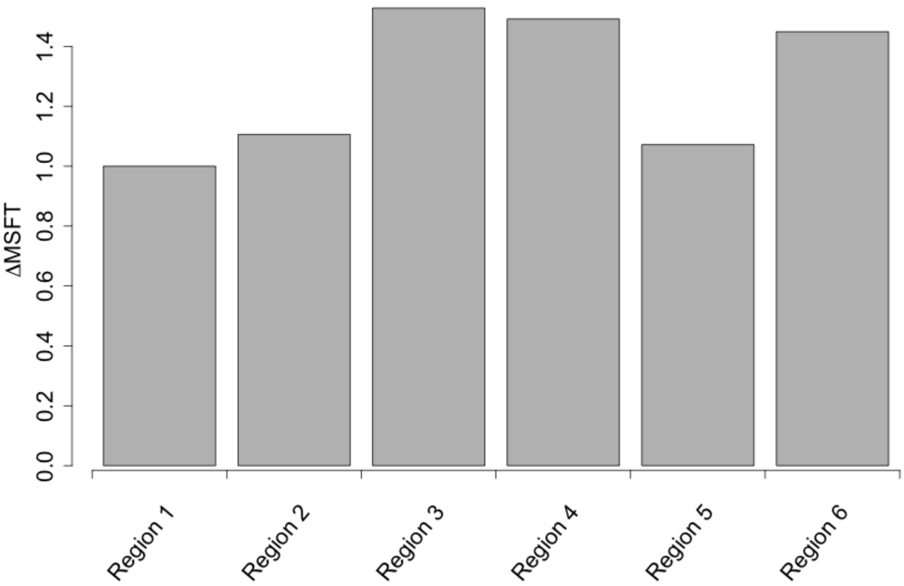


Figure 6: Introducing a New DC

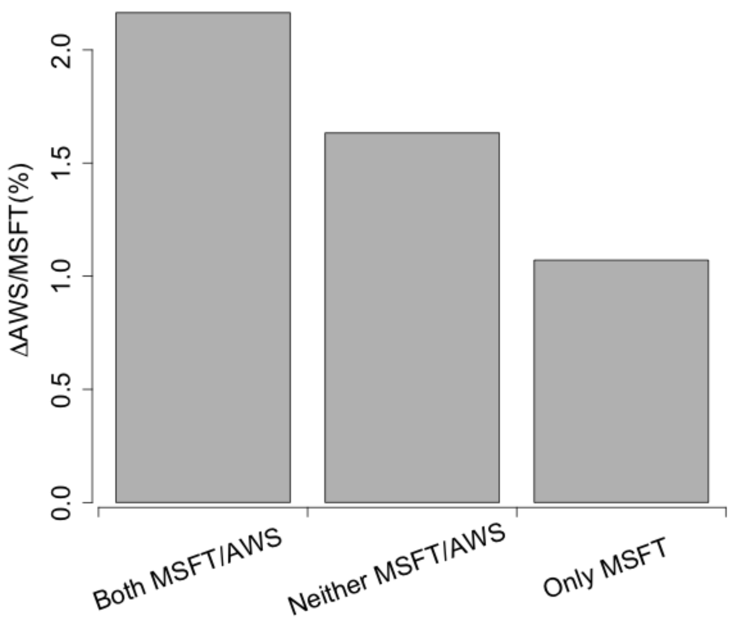


Figure 7: Price Competiton

9.2 Price Drop

Figure 7 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, neither a AWS nor a Azure DC or only an Azure DC, names of the states are anonymized to protect trade secret. Each bar shows the ratio of switchers from AWS to Azure relative to the original Azure market share in that state.

Figure 7 shows that the incremental change in market share varies by local market structure. Intuitively, in areas where both AWS and Microsoft DCs are available, price plays a relatively more important role in their competition, that is why price cut would have the most significant impact. On the contrary, the potential gain is quite restrictive in states 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.

Second, the amount of switchers are generally small in all three scenarios. 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 Betrand, winner take all market. This is consistent with AWS's early leadership in the cloud market.

10 Conclusion

In this paper, we explore the strategy of estimating demand system when the dataset is a fusion of disaggregate consumer level choice data and aggregate market share data. 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. A necessary assumption here is the homogeneous consumer taste given product chai.e. racteristics and consumer attributes. It enables the identification of demand parameters up to a product 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.

We also apply this estimation strategy to the spatial competition in IaaS public cloud industry in 2016 North American market. Results are consistent with intuition, saying that cloud users have significant preference for proximity. Together with estimated local market size, the estimated demand system can be used to do various counterfactual analysis, such as the two exercises on new DC locations and price drop discussed in this paper.

At a high level, there are two key learnings from this research. First we find evidence that spatial competition is important in the cloud computing industry. Policy makers and firms should account for this in developing policies relating to cloud computing. Second, the methods developed here are broadly applicable to goods characterized by spatial competition. Further research in this area can help firm managers and merger officials improve tools and thereby increase both profits and consumer welfare.

References

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

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.

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.

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.