

TECHNOLOGY SUPPORT AND POST-ADOPTION IT SERVICE USE: EVIDENCE FROM THE CLOUD

German F. Retana

INCAE Business School, Alajuela, COSTA RICA {gf@incae.edu}

Chris Forman

Dyson School of Applied Economics and Management, S. C. Johnson College of Business, Cornell University,
Ithaca, NY 14850 U.S.A. {chris.forman@cornell.edu}

Sridhar Narasimhan, Marius Florin Niculescu, and D. J. Wu

Scheller College of Business, Georgia Institute of Technology, 800 West Peachtree NW, Atlanta, GA 30303 U.S.A.
{sri.narasimhan@scheller.gatech.edu} {marius.niculescu@scheller.gatech.edu} {dj.wu@scheller.gatech.edu}

Appendix A

Provider Cloud Infrastructure Service and Technology Support Offerings

This appendix offers additional details to those presented in the “Data and Sample Construction” section of the paper in relation to the research context and the provider’s service characteristics. In our particular setting, the cloud provider has recognized that the novelty of the service plus the complexities involved in deploying distributed architectures that best leverage the cloud’s scalability may pose significant knowledge barriers to buyers attempting to use the service. In response to this, the provider offers them the option to contract and receive full support. We discuss first the pricing and terms of the cloud infrastructure service offering, and then elaborate on what characterizes full support.

One of the essential characteristics of cloud infrastructure services is that they are offered on-demand (Mell and Grance 2011). Buyers only pay for what they use, and nothing else: there are no sign-up fees, no minimum spending requirements, no periodical subscription fees, and, since buyers can choose not to use their service as well, there are no contract termination penalties either. Moreover, in the particular case of our provider, the computing resources are offered to buyers at fixed hourly rates that increase in server size or capacity, generally in a linear fashion. Server capacity is defined in terms of memory (GB of RAM), processing power (number of virtual CPUs), and local storage (GB of space of local hard disk). During our observation period, the three capacity metrics tend to vary together as a bundle, meaning that more of one is generally associated with more of the other two, yet prices are set and buyers usually make infrastructure sizing decisions in terms of memory. Prices also vary depending on the operating system chosen for a server (e.g., Windows servers cost more than Linux servers), yet such heterogeneity does not alter our main findings. The results considering operating system heterogeneity are available upon request.

Buyers in our context can launch as many servers and of any size they want, when they want. However, as is discussed in the “IT Service Use” subsection, there are important technical challenges in deploying horizontally scalable configurations where several cloud servers work in parallel. These challenges may in turn limit buyers’ ability to use many servers at once. Finally, there are no usage caps, with the only exceptions to this being that the provider may have limited hardware installed at its data centers or may take security measures to prevent misuse of its service (e.g., spamming). In other words, for legitimate buyers, there is no predefined cap or limit to how much they can choose to use the service.

The provider complements its infrastructure offering with full support, which is offered for a fixed price premium per server-hour used plus an additional fixed monthly fee. For instance, instead of paying \$0.10 per hour for a 2GB RAM Linux server under basic support, a full support

buyer would pay \$0.12 more (i.e., \$0.22 per hour). Similarly, for the 4GB RAM server priced at \$0.20 per hour under basic support, the full support buyer would pay \$0.32 per hour. The monthly fee is paid as a monthly subscription, which is a fee high enough to deter buyers with very low willingness to pay (i.e., bloggers that use a single very small server). There are no sign-up or termination fees for the full support service. The only explicit switching cost from one support level to another is technical rather than monetary: when downgrading from full support to basic support, because of technical limitations in the service offering (during our observation period), buyers must redeploy their servers on their own under the new support regime. The redeployment will involve launching new servers with virgin operating systems (i.e., “out of the box”), and then installing and configuring their business applications on them.

A prime goal of full support is to educate buyers on how to best use the cloud infrastructure service and adapt it to their idiosyncratic business needs. When receiving full support, buyers receive personalized guidance and training, and thus have the opportunity to learn directly from the provider’s prior experience in deploying applications in the cloud. Buyers not willing to pay the price premiums will only receive a basic level of support that has limited scope in the sense that it is intended to aid buyers with issues concerning account management or overall performance of the infrastructure service. For example, while a full support buyer may be personally guided step by step on how to deploy a web server through phone conversations, live chat sessions or support tickets, basic support buyers will be referred to a knowledge base. Similarly, if a cloud server failed, which happens much more frequently than in traditional data center settings given the commodity hardware employed and the multi-tenant architecture (i.e., multiple organizations’ virtual servers are hosted in the same and shared physical server), the provider would work together with full support buyers in solving the issues, while basic support users would only be notified about the failure, if anything. Thus, basic support buyers do not have fluid access to external knowledge from the provider and have to rely mostly on their internal capabilities to make use of the service.

Appendix B

Description of CEM Procedure and Sample Construction

Overview of CEM Procedure

We run our models on a subsample defined using a coarsened exact matching (CEM) procedure (Blackwell et al. 2010; Iacus et al. 2012). For matching purposes, we consider buyers who adopted full support at any point in their tenure as *treated* and those that relied exclusively on basic support as *controls*. Matching reduces endogeneity concerns (Ho et al. 2007), and CEM has been used extensively in recent work to improve the identification of appropriate control groups in difference-in-differences estimation (e.g., Azoulay et al. 2011; Azoulay et al. 2010; Furman et al. 2012).

CEM is particularly convenient for our setting because it is a nonparametric procedure that does not require the estimation of propensity scores. This is useful because, aside from the exogenous failures, we have limited data that would allow us to directly predict the likelihood of full support. Each unique vector formed by combinations of the coarsened covariates describes a stratum. Since the number of treated and control observations in each strata may be different, observations are weighted according to the size of their strata (Iacus et al. 2012). The differences in means between the treated and the controls across the various matching variables are almost all statistically significant. However, once we apply the CEM weights, the samples are perfectly balanced and any mean differences are eliminated. All our regressions with the CEM-based sample employ these weights. When exact matching is possible, such that for every treated observation there is a control observation identical to the first one across all possible covariates except for the treatment, a simple difference in means of the dependent variables would provide an estimate of the causal effect of interest. Nonetheless, since it is nearly impossible to use exact matching in observational data and thus there is always a concern about the influence of omitted variables, we continue using our fixed effects panel data model to control for them.

We match buyers based on six attributes: (1) pre-upgrade volume of IT use (i.e., memory use), (2) pre-upgrade efficiency of IT use (i.e., fraction of servers running in parallel), (3) pre-upgrade frequency of cloud infrastructure resizing (i.e., how often buyers launch a server, halt a server, or resize an existing one), (4) operating system of preference, (5) employment, and (6) intended use case for the cloud infrastructure service. The first four attributes are derived directly from buyers’ observed usage of the cloud service. The latter two attributes come from an optional sign-up survey.

The survey is optional and administered as part of the online sign-up web form; the response rate is 43.4 percent, and we have not found systematic differences between respondents and nonrespondents. The survey was first administered in June 2010, and we have all buyers’ responses until February 2012. Although there can only be one survey response per account, since buyers can have multiple accounts, we may also have multiple responses per buyer. In our data we have 6,152 survey responses from 5,565 different buyers in the baseline sample, 431

of which changed their response to at least one item across their surveys. However, for 42.3 percent of the buyers with varying responses the time gap between the survey responses is too short (i.e., less than 3 months) as to suggest that the variance is due to changes in firms' sizes or goals. Given this, we do not rely on variance across responses for our analysis and rather only consider the 5,134 buyers that either have a single survey response or that have consistent responses across all their submissions. Further, we have not considered firm attributes in the survey as controls in our models since they do not vary over time and thus would be absorbed by the firm fixed effect.

For the matching process, we only consider treated buyers who started using the cloud service with basic support and upgraded to full support later on. This allows us to match the upgraders to controls based on their usage behavior before they adopted full support, during what can be described as their pre-treatment periods. We elaborate on the exact process below. This approach, which is similar to the one implemented by Azoulay et al. (2010) and Singh and Agrawal (2011), ensures to the extent possible that treated firms do not exhibit differential usage behavior before they adopt full support relative to controls. Among the 5,134 buyers for which we have all this data (i.e., they answered the sign-up survey), 1,259 are treated and 3,875 are potential controls. Using the six criteria described above, we develop a weighted matched subsample. As part of our research, we ran our models with varying permutations of matching criteria which, in addition to the six already mentioned, included buyer industry. Our results were consistent across the various subsamples and are available upon request.

CEM Matching Criteria

Six different attributes of firms were used to match treated and controls. In this section we describe each of them. They are summarized in Table B1.

Name	Description	# of Categories	Categories
Memory Use	Average memory usage in GB of RAM in months before upgrade	9	< 0.5, 0.5–1, 1–2, 2–4, 4–8, 8–16, 16–32, 32–64, > 64
Architecture Complexity	Average fraction of servers running in parallel in months before upgrade	5	0.00, 0.00–.25, .25–.50, .50–.75, .> 75
Adjustments	Frequency of infrastructure resizing in months before upgrade	5	0, 1–2, 3–9, 10–43, > 43
OS Preference	Primary OS before upgrade	6	Linux, Windows, RedHat, SQL, Mix
Employment	Employment	5	0–10, 11–50, 51–100, 101–250, > 250
Use Case	General use cases (can have more than 1)	5	High variance, low variance, back office, hosting, test, and development

IT Use, Architecture Complexity and Frequency of Infrastructure Sizing Adjustments

With regard to overall use (i.e., memory use) and frequency of infrastructure resizing, when creating our baseline sample we had already discarded basic support users with very small and/or rather static deployments over the early periods of their tenure. We excluded buyers who averaged 512 MB RAM/hour or less during their first 6 months (excluding first month) or made no adjustments to size of their infrastructure during their first 6 months (excluding first month). Nonetheless, even among the remaining buyers there is considerable variation in these two variables.

The average memory usage, fraction of servers running in parallel (as a proxy for the architecture complexity of the deployment), and the frequency of infrastructure resizing actions used to match treated and control groups were computed as follows. Assume that a given treated buyer adopted the service with basic support in some period t_0 and switched from basic to full support in a later time period, t_{fs} , $t_{fs} > t_0$. Then, we consider the set of controls (i.e., buyers who exclusively used basic support) who also adopted the service in month t_0 and used the service (i.e., did not churn) at least up to t_{fs} . This ensures all buyers were using the service during the same calendar time frame and have very similar tenure by period t_{fs} . For the treated group and all these controls, we compute the average memory usage, fraction of servers running in parallel, and frequency of scaling actions in the periods during which all buyers were using basic support: from t_0 up to t_{fs-1} . Finally, we use this metric, which represents their pre-upgrade behavior, to match buyers.

For average memory usage, we set our cutoff points at standard server sizes: 512MB, 1GB, 2GB, 4GB, 8GB, 16GB, 32GB, and 64GB of RAM. For the fraction of servers running in parallel, we opted to create five bins. Its distribution has a strong mass at zero, which justified the first

bin with all observations with zero value. Afterward we built 0.25-width bins which are close to the 25th, 50th, and 75th percentiles of the non-zero values. For frequency of infrastructure resizing, we base our cutoff points on percentiles of the distribution: the 25th percentile is a single change to the size of the deployment, the 50th percentile is 3 changes, the 75th percentile is 9 changes, and the 95th percentile is 43 changes. In total, as shown in Table B1, we have nine categories for memory use, five categories for fraction of servers running in parallel, and five categories for the frequency of infrastructure resizing to match on.

Operating System (OS) Preference

During the time span of our data, the provider offered its servers running four different OS and we observe which OS each individual server used:

1. Linux: Several distributions, although we do not observe which.
2. Windows: Several versions of Windows Server, although we do not observe which.
3. Red Hat Enterprise Linux.
4. SQL Server: This is really a Windows Server running SQL Server, yet it was offered under its own price scheme and hence is considered another operating system for this exercise.

Even though there were multiple OS available, as we show in Table B2, most buyers either exclusively or at least primarily used a single OS. To determine if a buyer is a user of a particular OS, we computed the proportion of the total amount of GB RAM-hours consumed by each buyer over its observed tenure that were consumed of each of the four different OSs. Then, using arbitrary yet high thresholds (e.g., from 85 percent up to 100 percent), we flag a buyer as user of a certain OS if the proportion of service use with that OS is greater or equal than the defined threshold. Using these proportions of workloads under each OS and varying thresholds, we populated each column of Table B2 as follows:

- **Threshold:** Indicates the percentage of total usage using a specific OS used to flag a buyer as a user of that OS.
- **Linux, Windows, Red Hat and SQL:** Indicate the proportion of buyers who used at least as much as the threshold of their total usage under each corresponding OS. For example, 57.35 percent of buyers in the baseline sample used at least 99 percent of all their GB RAM-hours on Linux servers.
- **Only 1:** Given a certain threshold, it shows the proportion of total buyers that used only a single OS. The column is the sum of the four different OS columns to the left.
- **Mixed:** Given a certain threshold, it shows the proportion of total buyers that used a mix of more than a single OS. The “Only 1” column and this column add up to 100 percent.

The main takeaway from Table B2, and in particular from the “Only 1” column, is that most customers primarily use a single OS. For instance, 66.96 percent of buyers in the baseline sample ran all their servers using a single OS, and 80.13 percent ran at least 95 percent of their workloads using a single OS.

Threshold	Proportion of Buyers Using OS					
	Linux	Windows	Red Hat	SQL	Only 1	Mixed
100%	52.38%	10.47%	2.31%	2.13%	66.96%	33.04%
99%	57.35%	13.29%	2.81%	2.36%	75.49%	24.51%
95%	60.07%	14.78%	3.09%	2.52%	80.13%	19.87%
90%	61.60%	15.89%	3.24%	2.65%	83.05%	16.95%
85%	62.79%	16.78%	3.42%	2.83%	85.49%	14.51%

To put this proportion into perspective, recall the median buyer in the sample consumes an average of 0.5 GB RAM (or 512 MB RAM) per hour over its tenure. Thus, over a month, a median buyer consumes $0.5 \text{ GB RAM/h} \times 24 \text{ h/day} \times 30 \text{ days/month} = 360 \text{ GB RAM/month}$. If a buyer uses the same OS for at least 95 percent of its workload, then it is using some other OS for at most 18 GB RAM during a month. This level of usage is equivalent to running a very small, 256 MB RAM (0.25 GB RAM) server for 3 days of the month (i.e., 72h). Even for a threshold of 85 percent, the remaining 15 percent is 54 GB RAM during the month, or 9 days of a very small 256 MB RAM server.

We feel that such levels of usage (e.g., a very small server during 9 days per month) are inconsistent with running production applications, even if they are only used for short time spans. Even a small blog would run in that 256 MB RAM server but for an entire month (i.e., 30 days), not 9 days, and any standard application will traditionally need at least a 512 MB RAM server, twice as large as this one. In other words, we are confident that customers who use at least 85 percent or more of their workloads on a single OS can be characterized as users of that OS. We find this is the case for 85.49 percent of the buyers in the baseline sample and 89.44 percent of the buyers in the CEM subsample described below. For our matching process, we employ the 85 percent threshold to flag buyers as users of each of the four different OS or a mix of any of the four, resulting in five categories.

Employment

The employment and intended use case data are collected from the sign-up survey. The proportions of buyers falling into the relevant categories within these two attributes are shown in Table B3. For the employment cutoff points, we broadly rely on the ranges used in the survey. Among the buyers with consistent survey responses across all their accounts, 65.60 percent indicated they have 10 or fewer employees. Another 19.75 percent indicated they have between 11 and 50 employees. We subdivide the remaining 15 percent of buyers in three bins each accounting for roughly 5 percent of our sample: from 51 to 100, from 101 to 250, and more than 250.

Buyer Role	All Buyers	Controls	Treated
Number of Buyers	5,134	3,875	1,259
Employment			
10 or less	65.60%	69.19%	54.57%
11 to 50	19.75%	18.68%	23.03%
51 to 100	5.03%	4.36%	7.07%
101 to 250	3.66%	3.02%	5.64%
More than 250	5.96%	4.75%	9.69%
Use Case			
High Use Uncertainty	46.34%	46.86%	44.72%
Low Use Uncertainty	59.14%	57.34%	64.65%
Back Office Applications	18.85%	19.48%	16.92%
Hosting	9.17%	9.29%	8.82%
Test & Development	29.31%	32.26%	20.25%

Intended Use Case

The intended use case is collected by a multiple choice question (i.e., “Mark all that apply”) that asked buyers to “Please indicate what solution(s) you intend to use [the cloud infrastructure service] for.” The 20 options available to buyers are very specific, and finding matches across such specific use cases would be extremely hard. Instead, we group the specific use cases into three more general use cases based on two dimensions: if the use case is related to back office or front office applications, and, in the latter case, if it is likely that the volume of usage for the use case is predictable or not. Our first general use case, which we call “High Use Uncertainty,” includes customer-facing websites that are prone to unpredictable variance in their volume of usage. Examples of such use cases are social media sites, online gaming sites, online publishing sites, rich media sites (e.g., audio or video), and other Software-as-a-Service (SaaS) offerings. Our second general use case, “Low Use Uncertainty,” includes customer-facing websites used for regular operation of the firm that have steady or at least predictable use levels. Examples are corporate websites, collaboration platforms, online portals, and e-commerce sites. We chose to include e-commerce sites in this general use case since, although it may have a high variance, seasonality makes the peaks and valleys of the demand fairly predictable. Finally,

our “Back Office Applications” general use case includes applications or systems used internally for business operations. Examples are a company’s intranet and systems used for accounting, customer relationship management, human resources, supply chain management, or backup. We additionally consider web hosting services and running test and development environments as additional general use cases. Altogether, we have five general use cases.

CEM Sample

To construct our CEM sample we started off with our entire dataset which has a total 79,619 different buyers. However, as noted in the main text we do not employ all buyers in our baseline sample. For the baseline sample we have excluded buyers who (1) only received basic support and (2) averaged 512 MB RAM/hour or less during their first 6 months (excluding first month) or (3) made no adjustments to size of their infrastructure during their first 6 months (excluding first month). We do not consider their behavior during their first month in our threshold because most buyers are setting up their infrastructure during this time. The baseline sample has 22,179 buyers.

From the baseline sample, we can only include in our CEM sample those buyers for which we have a survey response, which are 5,134, and for whom we observe at least two months of tenure. This is so that we can match buyers based on their behavior in the period before upgrading from basic to full support. This leaves us 4,200 buyers for the matching process.

The CEM procedure leaves in our sample 1,525 buyers, of which 1,303 are controls who exclusively used basic support and 222 are treated buyers who started with basic support and upgraded to full support. There are on average 5.86 control buyers per treated buyer.

Appendix C

Support Interactions and Construction of Instruments

The content of the support interactions between the provider and its buyers was used to identify three types of exogenous failures experienced by buyers. The descriptions of the failure events can be found in Table 3 in the paper. The following are the keywords and phrases used to identify each of these types of support interactions. All support interactions that matched some keyword or phrase were visually examined to rule out false positives.

Failure Type	Variable Name	List of Keywords or Phrases Used for Identification of Failure
Service outage	<i>FailOutage</i>	Providers’ service status URL, cloud status, outage, scheduled maintenance, undergoing maintenance
Network-related failure	<i>FailNetwork</i>	Server does not respond to ARP requests, faulty switch, network issue in our data center, lb in error state, load-balancer hardware nodes, DDoS
Hardware-related failure	<i>FailHardware</i>	Hardware failure, degraded hardware, drive failing, drives failing, server outage, host failure, server is down, server down, is hosted on has become unresponsive, problem with our server, host server, physical host, physical hardware, physical machine, host machine, failing hardware, hardware failure, imminent hardware issues, migrate your cloud server to another host, queued for move, issue on the migrations, host server of your cloud servers

Once we identified the occurrence of the failures through the coding process, we calculated the accumulated number of them occurring over time for each buyer and each failure type. Letting $F \in \{FailOutage, FailNetwork, FailHardware\}$ represent a type of support interaction, whenever the count reached N incidents we turn the corresponding FN indicator on. For example, variable $FailOutage2_{i,t}$ will be equal to one if buyer i has accumulated at least two support interactions that have been coded as type *FailOutage* by month t .

Appendix D

Impulse Response Functions

An impulse response function represents the response of a dependent variable to a (one-time) unit change in some covariate while all other variables dated t or earlier are held constant (Hamilton 1994, pp. 318-323). In our case, we compute and plot difference quotients

$$\frac{\Delta \hat{y}_{i,t}}{\Delta x_{i,t-j}} \left(y_{i,t} \in \{InMemory_{i,t}, FractionParallel_{i,t}\}, x_{i,t} \in \{AdoptFull_{i,t}, CeaseFull_{i,t}\} \right)$$

over time to show how current memory usage or current fraction of servers running in parallel is influenced by adoption or dropping of full support j periods ago.

These difference quotients are the coefficients of the associated rational lag model (Greene 2008, pp. 683-686). The rational lags identify the effect that each lag of the covariate, on its own, has on the dependent variable. Since we used two lags ($p = 2$) of the dependent variables in our estimations of Model (2), the early lags (i.e., $j \leq 1$) of $AdoptFull_{i,t-j}$ or $DropFull_{i,t-j}$ have a specific formulation, while the later lags (i.e., $j \geq 2$) follow a recursive form. The approach is very similar to that of example 20.4 in Greene (2008, pp. 685-686). The rational lag coefficients, which we denote $\hat{\delta}_j$, are computed as follows (we show $\hat{\beta}_j$ coefficients as those multiplying the $AdoptFull_{i,t}$ lags, but algebra is identical for the $\hat{\gamma}_k$ coefficients multiplying lags of $DropFull_{i,t}$):

$$\begin{aligned} \hat{\delta}_0 &= \frac{\Delta \hat{y}_t}{\Delta x_t} = \hat{\beta}_0 \\ \hat{\delta}_1 &= \frac{\Delta \hat{y}_t}{\Delta x_{t-1}} = \hat{\beta}_1 + \hat{\lambda}_1 \hat{\delta}_0 \\ \hat{\delta}_2 &= \frac{\Delta \hat{y}_t}{\Delta x_{t-2}} = \hat{\beta}_2 + \hat{\lambda}_1 \hat{\delta}_1 + \hat{\lambda}_2 \hat{\delta}_0 \\ \hat{\delta}_j &= \frac{\Delta \hat{y}_t}{\Delta x_{t-j}} = \hat{\beta}_j + \hat{\lambda}_1 \hat{\delta}_{j-1} + \hat{\lambda}_2 \hat{\delta}_{j-2}, \quad 2 \leq j \leq r \end{aligned}$$

As an additional step, we use Monte Carlo simulation and draw 100,000 random samples of the vectors of the fitted $\hat{\beta}_j$ and $\hat{\lambda}_s$ coefficients using their estimates and their variance-covariance matrix. For each draw we compute and record the fitted rational lags $\hat{\delta}_j$, and use their distributions to estimate their 90 percent confidence interval. In Figure 1, the dashed lines represent the 5th and 95th percentiles of the distribution of each $\hat{\delta}_j$.

References

- Azoulay, P., Graff Zivin, J. S., and Sampat, B. N. 2011. "The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine," National Bureau of Economic Research.
- Azoulay, P., Graff Zivin, J. S., and Wang, J. 2010. "Superstar Extinction," *Quarterly Journal of Economics* (25), pp. 549-589.
- Blackwell, M., Iacus, S., King, G., and Porro, G. 2010. "cem: Coarsened Exact Matching in Stata," *Stata Journal* (9:4), pp. 524-546.
- Furman, J. L., Jensen, K., and Murray, F. 2012. "Governing Knowledge in the Scientific Community: Exploring the Role of Retractions in Biomedicine," *Research Policy* (41), pp. 276-290.
- Greene, W. H. 2008. *Econometric Analysis* (6th ed.), Upper Saddle River, NJ: Prentice Hall.
- Hamilton, J. D. 1994. *Time Series Analysis*, Princeton, NJ: Princeton University Press.
- Ho, D., Imai, K., King, G., and Stuart, E. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis* (15:3), pp. 199-236.
- Iacus, S. M., King, G., and Porro, G. 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching," *Political Analysis* (20:1), pp. 1-24.
- Mell, P., and Grance, T. 2011. "The NIST Definition of Cloud Computing," National Institute of Standards and Technology Information Technology Laboratory, Gaithersburg, MD.
- Singh, J., and Agrawal, A. 2011. "Recruiting for Ideas: How Firms Exploit the Prior Inventions of New Hires," *Management Science* (57:1), pp. 129-150.