TECHNICAL SUPPORT AND IT CAPACITY DEMAND: EVIDENCE FROM THE CLOUD

Completed Research Paper

German Retana
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree NW
Atlanta, GA 30308
german.retana@scheller.gatech.edu

Chris Forman
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree NW
Atlanta, GA 30308
chris.forman@scheller.gatech.edu

Sridhar Narasimhan
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree NW
Atlanta, GA 30308
sri.narasimhan@scheller.gatech.edu

Marius Florin Niculescu
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree NW
Atlanta, GA 30308
marius.niculescu@scheller.gatech.edu

D. J. Wu
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree NW
Atlanta, GA 30308
dj.wu@scheller.gatech.edu

Abstract

Using a unique data set on public cloud infrastructure services consumption by 15,076 firms over the period from March 2009 to October 2011, we address the question of how a provider's technical support influences cloud demand. The provider's customers can choose (and switch between) two levels of support, basic and managed, which differ in the extent to which the provider helps customers adapt the cloud infrastructure to their specific business needs. We find that customers who access managed support consume, on average, 110% more IT capacity than those who only access basic support. The former are also 15.5 percentage points more likely to deploy more complex infrastructure architectures that make better use of the cloud's features (e.g., its scalability). Customers who switch from managed to basic support continue consuming an average of 90% more IT capacity than customers who only access basic support.

Keywords: IS economics, IT infrastructure, IT services, cloud computing, technical support.

Introduction

The cloud computing model, in which IT capacity is offered on-demand, has been envisioned by some IS scholars as a general purpose technology (GPT) (Bresnahan and Trajtenberg 1995) that will serve as a catalyst for innovation and an engine for economic growth (e.g., Brynjolfsson et al. 2010; Varian 2010; Varian 2011). Nonetheless, current adoption rates of cloud services do not reflect such expectations. Surveys have suggested that only 29% of small and medium-sized businesses (SMBs) were paying for one or more cloud services in 2010 (Microsoft and Edge Strategies 2011) and that in 2011 only 4% of IT professionals had implemented cloud infrastructure services for production applications (SearchDataCenter.com 2011). A potential reason for the low adoption is that customers do not understand how to use the novel cloud infrastructure services. A 2011 survey found that only 25% of IT staff in global organizations had cloud experience with public infrastructure or platform-as-a-service, and 50% of the organizations claimed that their staff was "less than somewhat prepared to handle" these services (Symantec 2011).

This industry insight suggests that the fact the cloud providers have made technical progress and made their GPT available does not imply that customers can immediately leverage on the cloud's promised capabilities, much less meet the expected demand. For the full demand to be realized, it is necessary that customers invest in a process of *co-invention* (Bresnahan and Greenstein 1997; Bresnahan and Greenstein 2001) and adapt the service to their specific business needs. Nonetheless, performing such adaptations to make the most out of the cloud's capabilities (e.g., its scalability) is a daunting task that poses strong knowledge barriers to firms (Attewell 1992). It has been hypothesized that service providers can help firms overcome these knowledge barriers by filling the gaps between the new technologies and their uses (Attewell 1992; Bresnahan and Greenstein 1997), thus increasing the demand for their IT services. However, to the best of our knowledge, this hypothesis has not been systematically tested due to lack of sufficiently detailed data. This is an important gap in understanding.

We aim to take a first step in narrowing this gap by studying the role that the technical support offered by service providers may play in cloud services' demand. Our central research question is: *Does a provider's technical support influence IT capacity demand?* If so, *how?*

We have collected unique data on public cloud infrastructure services consumption from a major provider. The provider's customers can choose (and switch between) two levels of support, *basic* and *managed*. The levels of support differ in the extent to which the provider helps customers adapt the cloud infrastructure to their specific business needs. Specifically, when receiving managed support – the premium service, customers have the opportunity to learn from the provider's prior experience in deploying applications in cloud architectures; when receiving basic support customers must adapt their applications on their own and only rely on their internal knowledge stocks. Our data consists of 15,076 firms that used the provider's public cloud infrastructure service (also known as Infrastructure-as-a-Service, or IaaS) at some point between March 2009 and October 2011. We aggregate our data at the monthly level, so our unbalanced panel has 32 periods. The longitudinal nature of our data provides unique advantages, as it allows us to directly measure and identify the impact of the provider's service level on the demand for IT capacity.

Our econometric approach uses fixed effects panel data models and a difference-in-difference identification strategy to compare customers' demand for IT capacity before they adopt managed support, during their continued access to managed support, and after they have switched back from managed to basic support. To address concerns about how omitted variable bias in our models might influence our results (e.g., how changes in unobserved business needs may drive both the choice of using managed support and the changes in our dependent variables), we employ lagged values of our variables as instruments (Arellano and Bover 1995; Blundell and Bond 1998) in dynamic panel models using generalized method of moments (GMM) estimation. The estimates from these models are qualitatively consistent with our main findings.

Our findings suggest that the provider's technical support has a positive and economically significant effect on customer demand for IT capacity. Customers who employ managed support consume, on average, 110% more IT capacity than those who only rely on basic support. Similarly, the former group is 15.5 percentage points more likely than the latter to employ complex infrastructure architectures that make use of the cloud's scalability. Moreover, we have found some initial evidence that the positive effects of managed support are stronger for customers who are just starting to use the cloud than for more experienced ones, which suggests that internal knowledge and external knowledge are weak substitutes for each other. We have additionally found that even after switching from managed support to basic support, former managed support customers continue using 90% more IT capacity than those who never had access to this type of support. Similarly, customers only decrease by 1.89 percentage points their likelihood of using complex and horizontally scalable architectures after they switch from managed to basic support. This suggests the reduction in the cost of co-invention due to access to the provider's managed support is durable, as knowledge transferred from the provider to the consumers does not depreciate quickly, if at all.

Our work contributes to several literature streams. First, we contribute to the research on co-invention by examining the role of service providers in facilitating customers' adaptation to a GPT. Prior research has shown that co-invention costs play a key role in determining the adoption rate of GPTs such as client/server computing (Bresnahan and Greenstein 1997). Nonetheless, much less is known regarding the mechanisms by which firms may overcome such co-invention costs, mainly because of the lack of sufficiently detailed data about firms' co-invention processes over time (e.g., Bresnahan and Greenstein 2001). Our work, which leverages on the detailed data available on how customers use cloud services, takes an important step in this direction. Second, our finding that internal and external knowledge sources may be substitutes adds to the ongoing debate that inquires if they are complements or substitutes for each other (see Argote and Miron-Spektor (2011) and Ceccagnoli et al. (2011) for reviews). Finally, the cloud context offers a unique opportunity to examine the IT investment of small startups who, given their low levels of IT spending traditionally have not as frequently been captured by data sources that make them amenable for study (e.g., Aral et al. 2006; Brynjolfsson and Hitt 1995; Brynjolfsson and Hitt 1996; Hitt et al. 2002). In our research, we observe the adoption and actual usage of an IT service by tens of thousands of very small firms (i.e., less than $1M in revenue and less than 100 employees) for whom co-invention costs are potentially prohibitive.

Framework and Hypotheses

Cloud computing has been defined by the National Institute of Standards and Technology (NIST) as a "model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services)" (Mell and Grance 2011). The pay-per-use model provides cloud infrastructure customers the freedom to consume and pay only for the computing resources they need, and correspondingly we conceptualize the demand for cloud infrastructure services as *IT capacity demand*. We say that one customer has a greater IT capacity demand than another if it uses greater server computing capacity (i.e., more CPUs and GB of RAM) over the same time span.

Additionally, the on-demand nature of the service along with its rapid elasticity provides firms the opportunity to reduce idle computing capacity waste and eliminate the necessity of an up-front capital commitment in overprovisioning resources (Armbrust et al. 2009; Harms and Yamartino 2010). Yet, to do this, firms must be able to scale their capacity in a cost-efficient manner. In particular, they should have a horizontally scalable architecture, which comes at the cost of a significant increase in *architecture complexity* relative to simpler, less easily scalable vertical architectures. Thus, we use the complexity of a customer's architecture as a proxy for how well they have adapted the cloud infrastructure service to their business needs.

The provider has recognized that the novelty of the service plus the complexities involved in scaling IT capacity may pose significant knowledge barriers to its customers. As with other GPTs, the associated knowledge barriers increase the customers' co-invention (Bresnahan and Greenstein 1997) costs and may, in turn, inhibit the adoption and demand of the service. In response to this, the provider offers its customers the option to contract and access managed support for a price premium on the IT capacity hourly rates plus a fixed monthly fee. In addition to helping the customers adapt the service to their particular business needs, a prime goal of managed support is to educate customers on how to best use the cloud. Our hypotheses below aim to test if the provider's offering of managed support effectively reduces customers' co-invention costs. We do this by studying the implications of managed support adoption for cloud demand, described both in terms of the IT capacity consumed by customers and their needs for cloud specific features (e.g., scalability) as measured in the complexity of their infrastructure architecture deployments.

Technical Support, Learning and Reduction of Co-Invention Costs

Next, we explain the mechanism by which a provider's technical support may positively affect IT capacity demand. The mechanism is that customers who adopt managed cloud will learn from the provider via technical support. This learning will reduce their knowledge barriers and thus total co-invention costs, and in turn increase their demand for cloud services.

Prior work has suggested that the provider's support may play an important role in reducing knowledge barriers (Attewell 1992). Das (2003) mentions that "for high-technology vendors, technical support is not only a competitive necessity, but also a potential source of revenue in markets where profits from product sales are increasingly restricted by price competition", suggesting support can be used as a mechanism to influence demand for commoditized or weakly differentiated services such as cloud infrastructure services.

Second, managed support, such as that offered by the provider in our context, opens the door to knowledge transfer. We follow Darr and Kurtzberg (2000) in arguing that knowledge "transfer has occurred when a contributor shares knowledge that is *used* by an adopter", and define knowledge transfer as "the communication of knowledge from a source so that it is learned and applied by a recipient" (Ko et al. 2005).

When offering managed support, the provider takes a proactive approach in helping users configure their software applications so that they effectively scale in the cloud. Examples of these applications include marketing campaigns—where a large amount of IT capacity is needed for a short period of time—or e-commerce applications with uncertain customer-driven IT capacity demand. The relationship between the managed support customer and the cloud provider is similar to that of clients and consultants in enterprise software implementations, where initially the client has the business knowledge and the consultant the technical knowledge, and through their interactions knowledge is transferred between them (Ko et al. 2005). Thus, managed support is not pure outsourcing where the provider does everything for the customer and "takes the burden of learning off the back of a potential user" (Attewell 1992).

Finally, customers' knowledge is positively associated with their usage of IT systems. IS researchers have long known that "firms delay in-house adoption of complex technology until they obtain sufficient technical know-how to implement and operate it successfully" (Attewell 1992). Moreover, firms are known to delay not only the adoption (initial purchase) but also the actual assimilation of a technology because of knowledge barriers (Fichman and Kemerer 1999).

There are several reasons why overcoming knowledge barriers may play an important role in enabling the demand for cloud services. Many of the expected features of enterprise-grade servers, such as redundant components that ensure high availability and physical access to servers, are not present in the cloud. The cloud requires users to design for failure (Reese 2009) and consider how to keep an application running if any given server randomly disappears. Moreover, the cloud's scaling capabilities can only be truly exploited if the applications scale out horizontally (i.e., employ several servers performing functions in parallel). Neither of these nuances have been the norm in traditional application architectures.

Given this, customers with weak understanding of cloud infrastructure services may not feel confident enough to deploy production applications in the cloud. Thus, their usage will likely be limited to testing and development environments or to non-mission critical applications resulting in lower demands for IT capacity. On the other hand, if through managed support customers learn from the provider, they will be much more likely to feel confident enough to deploy higher grade production applications or even invent new services using the cloud's frontier technology (Bresnahan and Greenstein 2001). Either of these translates into a greater demand for IT capacity. We formalize this in our first hypothesis:

> ***H1a**: The adoption of and continued access to managed support by customers is associated with greater IT capacity demand compared to customers who only access basic support.*

In a similar vein, applications designed for cloud infrastructures are meant to be horizontally scalable and immune to the sudden failure of any given node in the architecture. Only by deploying horizontally scalable architectures can customers leverage on the cloud's scalability, so their usage is a good indicator of how well do customers use the cloud. Nonetheless, horizontally scalable architectures add significant complexity to the traditional deployment of applications and may impede a customer from doing so on her own. If the provider can guide customers in adapting their applications to the cloud, then it is more likely that they will use complex architectures that make the most out of the cloud's features:

> ***H1b****: The adoption of and continued access to managed support by customers is associated with greater architecture complexity compared to customers who only access basic support.*

Regarding these hypotheses, it is worth recalling that one of the key benefits of using complex and horizontally scalable architectures is the reduction of idle capacity waste (Armbrust et al. 2009; Harms and Yamartino 2010). Thus, given a software application, customers who benefit from managed support and use such complex architectures (per H1b) may be able to run the application more efficiently and demanding less IT capacity than customers who only access basic support. Even without employing complex architectures, it is reasonable to assume that more knowledgeable customers are able to better forecast their IT capacity demands such that they overprovision fewer computing resources than less knowledgeable ones. This suggests that managed support, by reducing customers' knowledge barriers and thus their co-invention costs, may have the opposite effect as that suggested by H1a: managed support reduces rather than increases IT capacity demand. Nonetheless, our interview with the provider reveals that, while this is plausible such as in the very rare case of capacity overbooking before adoption of managed support, there is a much stronger effect whereby knowledgeable customers are willing to deploy more applications and are more confident in exploiting the elasticity features in the cloud than inexperienced ones. This insight is consistent with the overall effects predicted in H1a and H1b.

Learning, Forgetting and Durability of Co-Invention Cost Reduction

In addition to deciding to adopt the provider's managed support, customers can also choose to cease its use and switch from managed to basic support. Prior research has already shown that once firms internalize knowledge transferred from external sources, their valuation of that knowledge decreases relative to their valuation of internal knowledge (e.g., Menon and Pfeffer 2003). Therefore, a potential reason why customers switch back to basic support is that they have learned from the provider and reduced their co-invention cost enough such that their marginal gains from having access to managed support are less than the price premium they must pay for it. We conjecture that if customers learn how to deploy cloud infrastructures from the service provider, then they will not exhibit any significant change in their demand for cloud services after they switch to basic support.

We argue that the deployment of a cloud project can be characterized as a process innovation customized to the context and needs of the customer, in which not forgetting is vital for continued success. Our context relates closely to the professional services industry in which subsequent projects, which are few rather than plentiful, cannot only use the expertise generated through prior projects but their success actually depends on the access to such knowledge. Extant research in this industry has found that organizational forgetting rates are near zero (Boone et al. 2008).

There are at least two reasons why we expect switching firms to retain knowledge. First, the nature of projects within a small firm, at least in the short-run, is likely to be similar from one project to another, and thus a new project can use knowledge gained in a prior project. For example, once a customer learns how to deploy several web servers behind a load balancer for one web-based software application, it is reasonable to assume she will be capable of doing so again for a similar application in the near future. Second, a very important nuance of how the service is offered makes knowledge retention a requirement to switch: if customers desire to continue running the same set of applications under the new support regime, they must redeploy on their own their entire infrastructure.

In conclusion, if customers do not forget quickly what they have learned, then we have no reason to expect a dramatic change in their behavior after they switch from managed support to basic support. At a minimum, their IT capacity demand and the complexity of their architectures should be greater than that of customers who never had access to managed support. This result would provide evidence of the durability of the effects of managed support. We formalize this in the following hypotheses:

***H2a**: Customers who have accessed managed support in the past will consume more IT capacity than customers who have only accessed basic support.*

***H2b**: Customers who have accessed managed support in the past will employ more complex architectures than customers who have only accessed basic support.*

The Role of Experience

In addition to learning from an external source such as the provider, cloud infrastructure users have a more commonly studied source of knowledge: their own experimentation and experience. The literature focusing on how experience impacts firms' behavior has had a very long tradition (e.g., Dutton and Thomas 1984) and its review is beyond the scope of this manuscript. Nonetheless, although we do not formally hypothesize it, in accordance with prior findings we expect that customer's experience, by having the effect of reducing co-invention costs, is positively associated with their demand for cloud services. Specifically, we expect a positive and diminishing marginal impact of experience on cloud demand.

An open question is whether the provider's technical support enhances or mitigates the impact of the internal knowledge generated by customers through their own experience or, in other words, if internal and external knowledge sources are complements or substitutes for each other. While prior research in the context of R&D efforts has suggested a complementarity relationship between these (e.g., Arora and Gambardella 1994; Cassiman and Veugelers 2006; Cohen and Levinthal 1989; Cohen and Levinthal 1990), recent work has also found evidence of a substitution effect in the context of process innovations (e.g., Forman et al. 2008; Vega-Jurado et al. 2009).

Given that the deployment of applications in the cloud is more akin to a process innovation than to an R&D endeavor, we suspect that in our context a customer's own experience mitigates (substitutes for) the effects of the external knowledge transferred by the provider in helping reduce co-invention costs. Consequently, a customer's experience will reduce the increases in IT capacity demand and architecture complexity associated with the adoption and continued access to managed support. Our intuition is further enriched by interviews with the provider in which we have learned that they believe managed support is most beneficial for customers still trying to understand the intricacies of the service. We test these claims through the following hypotheses:

***H3a**: The greater a customer's own experience using cloud infrastructure services, the lower its increase in IT capacity demand associated with the adoption of and continued access to managed support.*

***H3b**: The greater a customer's own experience using cloud infrastructure services, the lower its increase in architecture complexity associated with the adoption of and continued access to managed support.*

Empirical Model

We employ linear fixed effects panel data models along with a difference-in-difference identification strategy to tease out the effects of adoption of managed support and switching back from managed to basic support on cloud demand. The customer fixed effect allows us to control for unobserved time-invariant customer characteristics that are likely correlated with the decision to adopt and/or switch from managed support. Furthermore, the time fixed effect allows us to control for any shocks (e.g., holidays) that may cause changes in behavior across all customers.

The complexity of a customer's infrastructure architecture can be associated with their usage of horizontal scaling methods. There are essentially two ways of growing an IT infrastructure: vertically, or up, and horizontally, or out (Garcia et al. 2008; Michael et al. 2007; Reese 2009, p. 176). Scaling vertically implies increasing capacity of a server or spreading out the IT stack across several servers, in either case having at most one server per function with growth capped by the maximum server capacity available. Differently, scaling horizontally involves having several servers performing functions in parallel and offers virtually unlimited growth potential. Nonetheless, horizontal scaling's unlimited growth comes with additional complex challenges associated with load balancing and session management across servers, among others

(Casalicchio and Colajanni 2000; Cherkasova 2000). Given these increased complexities in horizontal scaling, we use an indicator of its usage to identify the architecture's complexity.

Our main regressors of interest will be dummy variables that indicate if managed support has been adopted by a customer or if she has switched from managed to basic support by a given period. We are aware that the customers' decisions to adopt and switch from managed support may be endogenous and that this would in turn bias our estimates. We consider this and perform a falsification test in which we examine the timing of the effects of managed support and also employ a dynamic panel data model as a robustness check. We first test the main effects of the adoption and switching events and then interact these dummies with the customers' tenure to identify the role played by a customer's experience in relation to these events.

Effects of Adoption of and Switching from Managed Support

Our first model tests if the usage or the prior usage of managed support is associated with greater IT capacity demand and complexity:

$$y_{it} = \alpha + \beta\ ManagedSupport_{it} + \gamma\ SwitchToBasic_{it} + \mu_i + \tau_t + l_{it} + \varepsilon_{it}. \qquad (1)$$

Subscripts i and t index individual customers and time periods respectively. Parameter μ_i is the customer fixed effect that will be mean-differenced out, and τ_t is the calendar time fixed effect. We also include a dummy variable, l_{it}, indicating in what period of her lifetime a customer is when period t starts. This allows us to control for the possibility that customers' IT capacity demand and the complexity of their deployments may increase in a non-linear fashion as they grow older and learn more about the cloud infrastructure service. Parameter ε_{it} is our error term which we assume is correlated only within individuals, but not across them.

Our first regressor of interest is $ManagedSupport_{it}$, which is a binary variable that indicates if managed support was adopted by customer i by time t. Customers who always used managed support have $ManagedSupport_{it} = 1$ in all periods. Thus, β identifies the effects on cloud demand of adopting and having access to managed support, and we expect it to be positive and significant per Hypothesis 1.

Similarly, $SwitchToBasic_{it}$ is a binary variable that is equal to one if the customer does not have access to managed support by the end of the focal period but was using managed support at the start of the focal period or in some prior period(s). If a customer never switches from managed to basic support, then $SwitchToBasic_{it} = 0$ for all periods. The γ coefficient identifies the durability of the effects of managed support. If customers retain the knowledge acquired during the time they used managed support then γ will be insignificant (suggesting behavior does not change) or negative and significant but with a very low value relative to β (suggesting the effects of managed support do not dissipate entirely). The sum $\beta + \gamma$ will describe the differences in behavior between basic support customers who accessed managed support in the past and those who exclusively accessed basic support. If Hypothesis 2 holds, and customers' prior access to managed support sets them apart from those who only used basic support, then $\beta + \gamma$ should be positive and significant.

Considering Experience

We model experience as the log of the number of days a customer has been a cloud customer by the time period t starts. Following standard practice, we add 1 to our regressor before computing its log since customers will have no prior experience in the period they adopt the cloud service. Experience is represented by the parameter $lnTenure_{it}$ $(= \ln(Tenure_{it} + 1))$ and we include it in our second model as follows:

$$\begin{aligned} y_{it} = \alpha &+ \beta_1\ ManagedSupport_{it} + \beta_2\ ManagedSupport_{it} \times lnTenure_{it} \\ &+ \gamma_1\ SwitchToBasic_{it} \quad + \gamma_2\ SwitchToBasic_{it} \times lnTenure_{it} \\ &+ \delta\ lnTenure_{it} + \mu_i + \tau_t + \varepsilon_{it} \end{aligned} \qquad (2)$$

Note that the new parameter $lnTenure_{it}$ captures the same effect as the lifetime period dummy l_{it} in model (1), only that the latter did not impose any functional form to the relationship. We have moved to a log-linear approach in this model for ease of interpretation on the interaction term and also have removed l_{it} from model (2) because l_{it} and $lnTenure_{it}$ capture similar variance in the data.

To test our third hypothesis, which argues that internal and external knowledge sources are substitutes, we interact the experience metric with the dummies for the adoption of and switch from managed support. If this hypothesis holds then the coefficient β_2 should be negative and significant.

Data

Sample

We have collected a unique data set on cloud infrastructure services consumption from a major public cloud provider. Our entire data set includes 54,706 customers that adopted the provider's services at some point between March 2009 and October 2011. The customers can freely choose if they rely only on the provider's basic support or if they pay additional fees to receive managed support. Customers can also switch from one type of support to another, and we observe the time in which such switching occurs.

From the observed customer base, for the purposes of our analysis, we exclude customers who consume very low levels of capacity or who do not perform changes to their architecture configuration. These are customers who are using the cloud as a low cost (fixed capacity) web hosting service (e.g., to host a small personal blog), and for whom the adoption of managed support would have no effect since they have no intentions of growing their IT capacity demand. Specifically, we exclude customers who (1) only accessed basic support and (2) averaged 512 MB RAM/hour or less during their first semester (first 6 months, excluding 1st month) or (3) had no scaling activity during their first semester (first 6 months, excluding 1st month). We do not consider their behavior during their 1st month in our threshold given that most customers are setting up their infrastructure during this time, and thus both IT capacity usage and scaling activity can be considered abnormal. This takes out 39,630 customers from the sample, though our results are robust to their inclusion (these results are available upon request).

Amongst the remaining 15,076 customers in our sample, 12,033 relied exclusively on basic support, 2,267 relied exclusively on managed support, and 776 accessed both types of support during their observed lifetimes. The data used in our analysis includes 194,488 customer-month observations. Table 1 provides descriptive statistics of our sample.

Table 1. Descriptive Statistics

Variable	Description	Observations	Mean	Std. Dev.	Min	Max
$Capacity_{it}$	Average GB RAM /hour	194,488	6.57	25.72	0	1,653.00
$lnCapacity_{it}$	$\ln(Capacity_{it} + 1)$	194,488	1.30	0.96	0.00	7.41
$Horizontal_{it}$	Indicator	194,488	0.22	0.41	0.00	1.00
$ManagedSupport_{it}$	Indicator	194,488	0.09	0.29	0.00	1.00
$SwitchToBasic_{it}$	Indicator	194,488	0.02	0.14	0.00	1.00
$Tenure_{it}$	Days as customer	194,488	237.59	202.90	0.00	923.00
$lnTenure_{it}$	$\ln(Tenure_{it} + 1)$	194,488	4.74	1.73	0.00	6.83

Dependent Variables: IT Capacity Usage and Architecture Complexity

We run all models with 2 different dependent variables (y_{it}): logged IT capacity demand ($lnCapacity_{it}$) and an indicator if horizontal scaling is being used or not ($Horizontal_{it}$).

We capture IT capacity demand, $Capacity_{it}$, as a customer i's average hourly consumption during period t measured in GB RAM/hour. For example, $Capacity_{i3} = 2.25$ means that customer i used, on average, a set of servers with a combined capacity of 2.25 GB RAM during period 3. This could be two 1GB RAM servers and a 256 MB RAM server, or 9 servers, each with 256 MB of RAM. Our measure reflects the way cloud infrastructure providers price their services: hourly rates that increase in servers' capacity. Given that the distribution of IT capacity demand is strongly skewed to the right, and that in some periods customers may not consume capacity, we use $lnCapacity_{it} = \ln(Capacity_{it} + 1)$ as our dependent variable.

Our analysis of the complexity of the customers' deployments is based on the automated analysis of the names given by them to the servers they were running at the end of every day within the period (month). Specifically, we run an algorithm that analyzes the names of the servers being run at midnight (at the provider's time zone) during every day in the customers' lifetimes – this has the consequence that our records will not reflect short experiments or applications that run for less than 24 hours and were not running at midnight, yet we know these events are extremely rare. Our assumption is that we can identify a horizontally scalable architecture (i.e., an architecture with high complexity) if we find two or more servers with very similar names. We assess the names' similarity by measuring the Levenshtein distance (the minimum number of character edits needed to transform one name into the other) between them. We assume that if the Levenshtein distance is less or equal to 2, then they are performing the same function in parallel. For example, servers with names `web.domain.com` and `sql.domain.com` (Levenshtein distance = 3) are assumed to be performing different functions, but `web1.domain.com` and `web2.domain.com` (Levenshtein distance = 1), are assumed to be working together in parallel. Similarly, we consider `worker`, `workerBB` and `workerCC` (Levenshtein distance = 2) to be 3 servers working together in parallel. In sum, if any two server names have a Levenshtein distance that is less or equal to 2 during period t, then $Horizontal_{it} = 1$, and is 0 otherwise.

Results

Effects of Adoption of Managed Support and Switching Back to Basic Support

We present our results for Model (1) in column (1) of Tables 2 and 3. Column (1) of Table 2 shows that customers who adopt and use managed support consume, on average, 110% (i.e., $e^{0.7404} - 1$) more IT capacity than customers who use basic support. Similarly, column (1) of Table 3 shows that these customers have a 15.50 percentage points greater likelihood of using a complex, horizontally scalable architecture. These results provide support for our first set of hypotheses (H1a and H1b) and suggest that the offering of managed support has a positive and measureable impact on both the IT capacity demand and the complexity of the infrastructure architecture employed.

On the other hand, column (1) of Table 2 shows that even after customers have switched from managed to basic support, we find that they continue consuming, on average, 90% (i.e., $e^{0.7404-0.1002} - 1$) more capacity than customers who never accessed managed support. Moreover, column (1) of Table 3 shows that customers are only 1.89 percentage points less likely to employ highly complex architectures after they switch. In general, basic support customers who accessed managed support in the past continue employing configurations with a greater complexity than customers who exclusively relied on basic support. Together these results provide support for our second set of hypotheses (H2a and H2b) which suggests that customers retain and leverage on the knowledge acquired during the time they accessed managed support, and the positive effects of technical support on IT capacity demand are durable.

Robustness Checks

Falsification Tests

As noted above, one concern is that omitted variable bias and reverse causality may influence our parameter estimates. To address these concerns, as a robustness check, we perform a falsification test and verify if there is any significant change in customer behavior in the periods immediately preceding the adoption of managed support. We examine whether customers' behavior before the adoption of managed support is similar among customers who will adopt managed support and those that will continue using basic support. For this, we add 2 variables to Model (1). Parameters $AdoptManagedIn2_{it}$ and $AdoptManagedIn4_{it}$ are dummy variables equal to 1 in the 2 and 4 periods (respectively) immediately before the adoption of managed support. Thus, for example, if a customer adopts managed support in $t = 10$, then $AdoptManagedIn4_{it} = 1$ for $t = 6, \dots, 9$, and is equal to 0 otherwise.

Table 2. Regression Results for Models (1), (3) and (4). $y_{it} = lnCapacity_{it}$.

Column	(1)	(2)	(3)	(4)	(5)
Model	Basic Model	Basic Model with Falsification Tests		Dynamic Panel Models	
				FE	GMM
$ManagedSupport_{it}$	0.7404*** (0.0368)	0.7880*** (0.0420)	0.8063*** (0.0466)	0.2888*** (0.0065)	0.0244** (0.0102)
$SwitchToBasic_{it}$	-0.1002*** (0.0153)	-0.1012*** (0.0153)	-0.1015*** (0.0153)	-0.0466*** (0.0064)	-0.0054 (0.0079)
$AdoptManagedIn2_{it}$		0.1525*** (0.0313)			
$AdoptManagedIn4_{it}$			0.1266*** (0.0350)		
$lnCapacity_{it-1}$				0.9153*** (0.0027)	1.0713*** (0.0393)
$lnCapacity_{it-2}$				-0.1808*** (0.0036)	-0.1251*** (0.0452)
$lnCapacity_{it-3}$				0.0381*** (0.0034)	0.0452 (0.0432)
$lnCapacity_{it-4}$				-0.0078*** (0.0023)	0.0245 (0.0322)
Constant	-2.4866*** (0.0704)	-2.4946*** (0.0704)	-2.4933*** (0.0704)	0.3499*** (0.0190)	-0.0567** (0.0272)
N	194,488	194,488	194,488	136,420	136,420
R^2	0.234	0.235	0.234	0.708	
Customers	15,076	15,076	15,076	13,009	13,009

All regressions include calendar (τ_t) and lifetime time dummies (l_{it}). Robust standard errors, clustered on customers, in parentheses. $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$. The GMM estimation in column (5) considers $ManagedSupport_{it}$ and $SwitchToBasic_{it}$ as endogenous. Given AR(2) in the errors, it uses the 3rd lag of $lnCapacity_{it}$ and the 3rd and 4th lags of $ManagedSupport_{it}$ and $SwitchToBasic_{it}$ as their instruments for the difference equation. It also uses the 2nd lag of the 3 variables' first difference as instruments for the levels equation. Total number of instruments is 186. Hansen (1982) specification test passed with $\chi^2(125) = 121.54,\ p = 0.571$. Robust standard errors use Windmeijer's (2005) finite sample correction.

We present our results with these new parameters in columns (2) and (3) of Tables 2 and 3. In terms of IT capacity demand, we find that customers tend to consume between 13% (i.e., $e^{0.1266} - 1$) and 16% (i.e., $e^{0.1525} - 1$) more capacity in the periods preceding the adoption of managed support. These coefficients are positive and significant. However, their magnitude is much lower compared to the magnitude of the coefficient for $ManagedSupport_{it}$, which indicates the change in behavior once managed support is adopted. A similar situation arises when considering the complexity of customers' architecture in the periods immediately preceding the adoption of managed support. Here we find that in these periods the likelihood of customers employing a horizontally scalable architecture increases between 3.57 and 3.59 percentage points. Nonetheless, these coefficients are again very low compared to the coefficient of $ManagedSupport_{it}$ which suggests that the adoption and continued access to managed support increases the likelihood of employing a horizontally scalable architecture by 16.62 to 17.37 percentage points.

Table 3. Regression Results for Models (1) and (3). $y_{it} = Horizontal_{it}$.

Column	(1)	(2)	(3)	(4)	(5)
Model	Basic Model	Basic Model with Falsification Tests		Dynamic Panel Models	
				FE	GMM
$ManagedSupport_{it}$	0.1550*** (0.0162)	0.1662*** (0.0184)	0.1737*** (0.0201)	0.0957*** (0.0054)	0.0055 (0.0051)
$SwitchToBasic_{it}$	-0.0189*** (0.0064)	-0.0191*** (0.0064)	-0.0193*** (0.0064)	-0.0051 (0.0054)	-0.0028 (0.0043)
$AdoptManagedIn2_{it}$		0.0357*** (0.0131)			
$AdoptManagedIn4_{it}$			0.0359*** (0.0138)		
$Horizontal_{it-1}$				0.5676*** (0.0028)	0.8253*** (0.0679)
$Horizontal_{it-2}$				0.0238*** (0.0031)	0.0609 (0.0569)
$Horizontal_{it-3}$				0.0054* (0.0030)	0.0814* (0.0448)
$Horizontal_{it-4}$				-0.0143*** (0.0025)	0.0258 (0.0417)
Constant	-0.5887*** (0.0424)	-0.5905*** (0.0424)	-0.5906*** (0.0424)	0.1122*** (0.0160)	-0.0445 (0.0292)
N	194,488	194,488	194,488	136,420	136,420
R^2	0.021	0.022	0.022	0.360	
Customers	15,076	15,076	15,076	13,009	13,009

All regressions include calendar (τ_t) and lifetime time dummies (l_{it}). Robust standard errors, clustered on customers, in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The GMM estimation in column (5) considers $ManagedSupport_{it}$ and $SwitchToBasic_{it}$ as endogenous. Given AR(2) in the errors, it uses the 3rd lag of $lnCapacity_{it}$ and the 3rd through 8th lags of $ManagedSupport_{it}$ and $SwitchToBasic_{it}$ as their instruments for the difference equation. It also uses the 2nd lag of the 3 variables' first difference as instruments for the levels equation. Total number of instruments is 260. Hansen (1982) specification test passed with $\chi^2(199) = 200.18,\ p = 0.463$. Robust standard errors use Windmeijer's (2005) finite sample correction.

In sum, these results indicate a significant increase in IT capacity demand and architecture complexity after the adoption of managed support that exceeds by far that in the periods preceding adoption. Thus, it is unlikely that our results solely reflect a demand for managed support due to increases in capacity in the period preceding adoption. However, two concerns remain: (1) that the current value of the dependent variable may depend in part on its past values and (2) that omitted variable bias may still be influencing our results. To address these concerns, we next implement a dynamic panel data model with instrumental variables.

Dynamic Panel Estimation and Endogenous Adoption and Switching Decisions

Because of the nature of cloud infrastructure services, if customers do not change their behavior from one period to another, then the value of our dependent variables greatly depend on their values in prior periods. A first approach to account for this phenomenon is to include lagged values of the dependent variable as a regressor and estimate the model using standard fixed effects. The model has the following form:

$$y_{it} = \alpha + \sum_{n=1}^{p} \lambda_n y_{it-n} + \beta\ ManagedSupport_{it} + \gamma\ SwitchToBasic_{it} + \mu_i + \tau_t + l_{it} + \varepsilon_{it} \qquad (3)$$

We ran this fixed effects model using varying number of lags (i.e., $p = 1, \ldots, 5$), and attained qualitatively similar results. For reasons that will be explained below, we only present the results using 4 lags ($p = 4$) of $lnCapacity_{it}$ and $Horizontal_{it}$ in column (4) of Tables 2 and 3, respectively. With both dependent variables, $lnCapacity_{it}$ and $Horizontal_{it}$ we confirm our suspicion that the current value of the variable is strongly influenced by its past values; this is reflected in the large coefficients (close to 1 in the case of $lnCapacity_{it}$) for the lagged dependent variables. Nonetheless, even after controlling for this, column (4) of Tables 2 and 3 show that both IT capacity usage and the likelihood of employing a highly complex architecture still increase 33% (i.e., $e^{0.2888} - 1$) and 9.57 percentage points, respectively, with the adoption of and continued access to managed support ($ManagedSupport_{it}$). Furthermore, with respect to the effects of switching back from managed to basic support ($SwitchToBasic_{it}$), we also find that IT capacity usage only decreases marginally after customers switch from managed to basic support, and the likelihood of employing a horizontally scalable architecture does not have a statistically significant change after this event. Despite the magnitudes of the coefficients in column (4) are much lower than those in our baseline model in column (1), which was actually expected after including the lags of the dependent variables as regressors, the signs and statistical significance of the parameters of interest continue to hold.

The naïve usage of a fixed effect model to estimate model (3) opens the door to suffering from dynamic panel bias. Although this bias decreases in the number of periods (Nickell 1981), and we have a long panel with $T = 32$, the bias remains a concern. A solution to this issue involves using the System GMM and Difference GMM approaches that have evolved from the work of Anderson and Hsiao (1981), Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998), and that has been used recently in the IS literature (e.g., Archak et al. 2011; Ghose 2009). This approach has the added benefit that it allows us to consider $ManagedSupport_{it}$ and $SwitchToBasic_{it}$ as endogenous and use their lagged values and differences as instruments for them. We employ System GMM (Arellano and Bover 1995; Blundell and Bond 1998) while also applying the finite-sample correction proposed by Windmeijer (2005). We now describe separately the estimation procedure for each of the two dependent variables, starting with $lnCapacity_{it}$.

A first estimation procedure to make before running model (3) with System GMM is selecting the appropriate number of lags of the dependent variable to be included as regressor (p). Following a process similar to that executed by Chen et al. (2011), we started our lag number selection problem by choosing a number of lags that is consistent with our phenomena of interest and then test for serial correlation in the errors and the validity of the overidentifying restrictions. We chose to use 4 lags of $lnCapacity_{it}$ based on the provider's belief that it takes customers about 4 months to stabilize their behavior. In our first run, using all available instruments (from the 2nd lag to the end of the panel), the Arellano and Bond (1991) serial correlation test indicated that we do not only have the expected 1st order serial correlation but also have 2nd order serial correlation. As a result, we assume our errors follow a MA(1) process in which $\varepsilon_{it} = \eta_{it} + \rho\eta_{it-1}$, where $|\rho| < 1$ and $E[\eta_{it}] = 0$. Given this, we cannot use the 2nd lag of the variables as instruments for the difference equation (since $E[y_{it-2}\Delta\varepsilon_{it}] \neq 0$) nor the 1st lag of their first difference as instruments for the levels equation (since $E[\Delta y_{it-1}\varepsilon_{it}] \neq 0$). Nonetheless, we can still rely on the variables' 3rd (and posterior) lags and the 2nd lag of their first difference as instruments (Cameron and Trivedi 2010). That is, we assume $E[y_{it-3}\Delta\varepsilon_{it}] = E[\Delta y_{it-2}\varepsilon_{it}] = 0$. After this consideration, our model becomes

$$y_{it} = \alpha + \sum_{n=1}^{p} \lambda_n y_{it-n} + \beta\ ManagedSupport_{it} + \gamma\ SwitchToBasic_{it} + \mu_i + \tau_t + l_{it} + \eta_{it} + \rho\eta_{it-1}. \qquad (4)$$

We ran model (4) using all available instruments (from the 3rd lag to the end of the panel). We confirmed we still have 2nd order serial correlation and then used the Hansen (1982) J test to test the validity of our overidentifying restrictions. We passed the Hansen J test with $\chi^2(667) = 674.21, p = 0.415$. We then checked if we could use fewer lags of $lnCapacity_{it}$ as regressors, but found that we failed to pass the Hansen J test if we did so, and thus we kept the 4 lags.

The second estimation decision to make is selecting an appropriate number of instruments to avoid the problem of over fitting the model with too many instruments (Roodman 2009). Again following Chen et al. (2011), we gradually reduced the number of lags used as instruments until we found the least number of instruments under which we still passed the instrument validity Hansen J test. We found that we can limit our model to the use of the 3rd lag of $lnCapacity_{it}$ and the 3rd and 4th lags of $ManagedSupport_{it}$ and $SwitchToBasic_{it}$ as instruments for the difference equation. We also use the 3rd lag of the 3 variables' first difference as instruments for the levels equation. When using such specification, we pass the Hansen J test with $\chi^2(125) = 121.54,\ p = 0.571$. Moreover, using only these instruments reduces the total number of instruments from 728 to 186.

The results for this System GMM specification are reported in column (5) of Table 2. As in column (4), we find that the values of $lnCapacity_{it}$ are strongly influenced by its prior values. In this model, while the magnitude of the coefficient from $ManagedSupport_{it}$ is reduced, it is still positive and significant with p-value = 0.017, which matches what was expected per H1a. The coefficient of $SwitchToBasic_{it}$ is no longer significant, which is different from what we had in our baseline model in column (1) and the fixed effects dynamic panel in column (4). Nonetheless, rather than contradicting H2a, this change provides better support for the hypothesis that basic support customers who accessed managed support in the past demand more IT capacity than those who never accessed managed support.

Turning to the model with $Horizontal_{it}$ as dependent variable, we also started the number of lags selection by using 4 lags of the dependent variable as regressors. Using all available instruments, the Arellano and Bond (1991) test also found 2nd order serial correlation, so we must again rely on the 3rd or later lags of the variables' values and the 3rd lag of their first difference as instruments. Similar to the case with $lnCapacity_{it}$, the lowest number of lags of $Horizontal_{it}$ we can include in our regression is 4. In terms of instruments, the specification with the smallest numbers of instruments employs the 3rd lag of $Horizontal_{it}$ and the 3rd through 8th lags of $ManagedSupport_{it}$ and $SwitchToBasic_{it}$ as instruments for the difference equation. We also use the 3rd lag of the 3 variables' first difference as instruments for the levels equation. With this specification we pass the Hansen (1982) instrument validity test with $\chi^2(199) = 200.18, p = 0.463$. Our total number of instruments is 260.

We report our results for this model in column (5) of Table 3. As expected, the coefficient of the first lag of the dependent variable in the GMM specification is larger than that in the FE specification ($0.8253 > 0.5676$). This strengthens the idea that a customer's use of a complex architecture in the focal period is strongly influenced by her use of such architecture in past periods. In part because of this, it is not a surprise that our coefficient describing the effects of employing managed support ($ManagedSupport_{it}$) is no longer statistically significant (p-value = 0.286). Despite this, since $ManagedSupport_{it}$ has a positive sign, we suggest that the insignificance of its coefficient in this model specification does not contradict our prior findings. The coefficient for $SwitchToBasic_{it}$ is not significant as in column (4), which again follows what is expected per H2b. In future work, we will investigate the robustness of our findings to the use of nonlinear dynamic panel data models that will better incorporate the variation in users' architecture complexity. For example, in addition to considering the usage of a horizontally scalable architecture we can also count the number of different sets of servers working in parallel.

How do Experience and Managed Support Interact to Shape Demand?

We now test our third set of hypotheses which suggest that there are substitution effects between customers' internal knowledge and external knowledge. The results of this test done with Model (2) are presented for both dependent variables in Table 4; we analyze the results with the models for $lnCapacity_{it}$ in columns (1) through (3) and $Horizontal_{it}$ in columns (4) through (6) respectively.

The coefficients on the interactions of $ManagedSupport_{it}$ and $lnTenure_{it}$ are not statistically significant in columns (2) and (3), so we fail to find support for Hypothesis 3a. Yet, the coefficients are negative and statistically significant in columns (5) and (6), which provides support for Hypothesis 3b. Together, we find weak support for the hypothesized substitution relationship between internal and external knowledge, whereby managed support has a stronger effect on younger customers. In order to better examine the implications of the interplay between the adoption of and continued access to managed support and customers' tenure, we evaluate the percentage changes on the likelihood of employing a horizontally scalable deployment due to turning the $ManagedSupport_{it}$ dummy on at different levels of tenure. Based

Table 4. Regression Results for Model (2)

Column	(1)	(2)	(3)	(4)	(5)	(6)
y_{it}	$lnCapacity_{it}$			$Horizontal_{it}$		
$ManagedSupport_{it}$	0.7407*** (0.0366)	0.7126*** (0.0430)	0.7044*** (0.0430)	0.1549*** (0.0162)	0.1801*** (0.0180)	0.1769*** (0.0180)
$SwitchToBasic_{it}$	-0.0949*** (0.0153)	-0.1023*** (0.0157)	-0.2244*** (0.0636)	-0.0178*** (0.0064)	-0.0111 (0.0068)	-0.0592** (0.0247)
$lnTenure_{it}$	0.1033*** (0.0020)	0.1028*** (0.0021)	0.1029*** (0.0021)	0.0166*** (0.0010)	0.0171*** (0.0010)	0.0172*** (0.0010)
$ManagedSupport_{it} \times lnTenure_{it}$		0.0052 (0.0054)	0.0058 (0.0055)		-0.0046** (0.0021)	-0.0044** (0.0021)
$SwitchToBasic_{it} \times lnTenure_{it}$			0.0240* (0.0131)			0.0095* (0.0053)
Constant	0.3840*** (0.0600)	0.3837*** (0.0600)	0.3850*** (0.0600)	0.0336 (0.0301)	0.0339 (0.0301)	0.0344 (0.0301)
N	194,488	194,488	194,488	194,488	194,488	194,488
R^2	0.231	0.231	0.231	0.021	0.021	0.021
Customers	15,076	15,076	15,076	15,076	15,076	15,076

All regressions include calendar time dummies (no lifetime dummies). Robust standard errors, clustered on customers, in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 5. Changes in IT Capacity Demand and Likelihood of employing Horizontal Scaling from Adopting and Switching Back from Managed Support at Different Levels of Tenure

Column	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	$Capacity_{it}$ [a]			$Horizontal_{it}$		
Fixed level of $lnTenure_{it}$ [b]	$\bar{x}-\sigma^2$	$\bar{x}$	$\bar{x}+\sigma^2$	$\bar{x}-\sigma^2$	$\bar{x}$	$\bar{x}+\sigma^2$
$ManagedSupport_{it}=1$	105.80	107.86	109.94	16.36	15.60	14.84
$SwitchToBasic_{it}=1$	-14.11	-10.47	-6.68	-3.07	-1.44	0.19
$ManagedSupport_{it}=SwitchToBasic_{it}=1$	76.75	86.09	95.93	13.29	14.16	15.04

[a] Since $Capacity_{it} = e^{lnCapacity_{it}} - 1$, per model (2), the percentage change in $Capacity_{it}$ is given by $e^{\beta_1 ManagedSupport_{it} + \gamma_1 SwitchToBasic_{it} + lnTenure_{it}(\beta_2 ManagedSupport_{it} + \gamma_2 SwitchToBasic_{it})} - 1$. [b] $\bar{x}$ and σ^2 denote the sample mean and standard deviation of $lnTenure_{it}$, respectively. Results in columns (1) through (3) are based on column (3) of Table 4 and are expressed in percent change; results in columns (4) through (6) are based on column (6) of Table 4 and are expressed in changes in percentage points.

on the results in columns (3) and (6) of Table 4, Table 5 presents the percentage changes for both IT capacity ($Capaciy_{it}$) and architecture complexity ($Horizontal_{it}$). Although we include both dependent variables for completeness, we only analyze the results of the latter since the coefficients of the former are not statistically significant. The table shows that while managed support increases the likelihood of using horizontal scaling of a young customer (1 standard deviation below mean tenure) by 16.36 percentage points, it only increases it by 14.84 percentage points for an older customer (1 standard deviation above mean tenure). Table 5 also offers evidence of how switching from managed to basic support ($SwitchToBasic_{it}$) affects customers with different levels of experience. While young customers have, per column (4), a 3.07 percentage points decrease in their likelihood of employing a horizontally scalable deployment when they switch, older customers (column (6)) increase this likelihood by 0.19 percentage points. These results provide further evidence that customer's own experience ($lnTenure_{it}$) helps reduce their co-invention costs over time: The younger a customer is when she switches from managed to basic support, the greater the co-invention costs she must cope with on her own, and thus the greater the decrease in her ability to leverage on the cloud's scalability.

Conclusion

Cloud computing has been envisioned as a GPT that will serve as a catalyst for business innovation and an engine for economic growth. For example, cloud computing allows tens of thousands of small firms to have equal access to computing infrastructure as the large ones (Varian 2011). Understanding driving factors of cloud demand growth is of paramount importance towards generating insights on cloud computing economics.

Using a unique and rich data set on public cloud infrastructure services consumption by 15,076 firms over the period from March 2009 to October 2011, our study is the first to address the question of how a provider's technical support influences a client's cloud demand by directly observing and measuring how customers' access to support and their usage of IT infrastructure vary over time. We also explore the underlying mechanisms behind the positive relationship between technical support and IT capacity demand. Taking advantage of the near-commodity nature of the cloud, we provide evidence of factors that reduce the costs of adapting and customizing the GPT to the customers' business needs, thereby increasing the demand for the service.

Our estimates of the positive impact of offering technical support are economically significant. Customers who adopt and have accessed managed support, consume, on average, 110% more IT capacity than those who have only accessed basic support. The managed support customers are also 15.5 percentage points more likely to deploy more complex infrastructure architectures that make better use of the cloud's features (e.g., its scalability). We also find evidence that customers who switch back from managed to basic support continue consuming an average of 90% more IT capacity than customers who only had access to basic support throughout our entire sample period. Similarly, we find no evidence that customers who switch back significantly decrease the complexity of their architectures. These last two results indicate customers internalize and do not forget quickly what they have learned from the provider. Finally, we find some suggestive evidence that the adoption and continued access to managed support has a stronger positive influence on the behavior and the demand of newer customers, who presumably face stronger co-invention costs than customers with longer tenure.

Future Research and Limitations

Although, to the best of our knowledge, this is the first study that empirically examines drivers of demand for cloud infrastructure services, doing so does not come without some inherent limitations. These limitations, nonetheless, may be overcome by future research through additional data collection.

First, even though we have visibility regarding customers' IT capacity demand, and we have made an effort to infer the complexity of their deployments by analyzing the names they have given to their servers, we do not observe with complete accuracy how customers are using their servers. In particular, we do not observe what applications they are running on the servers or the specific technicalities of their configurations. This, in turn, limits our ability to assess how well or how efficiently customers are using cloud infrastructure services. Further research may employ surveys and interviews with cloud customers to overcome this limitation.

An additional limitation is that our approach does not at present control for time-varying factors that might influence both the decision to contract managed support and IT capacity demand. Future work in this area should include the use of additional firm-level controls or employ instrumental variables analysis to further explore how such omitted variables might influence results.

Third, one significant limitation of this study is our inability to directly observe the value that customers derive from their utilization of cloud infrastructure services. For example, since we do not observe customer firms' financial and operational performance, we cannot follow prior literature on IT value (e.g., Aral et al. 2006; Brynjolfsson and Hitt 1995; Brynjolfsson and Hitt 1996) in capturing the impact of the adoption and usage of cloud services on firm performance. Similarly, while one of the most commonly mentioned benefits of cloud computing is its ability to reduce idle IT capacity waste (Armbrust et al. 2009; Harms and Yamartino 2010), we cannot capture these IT capacity savings since we do not observe customers' IT investments outside of the provider's cloud. In our research, we observe customers' revealed preferences to employ cloud infrastructure services over some other IT infrastructure alternative, such as a self-run data center. Under the assumption that customers are economically rational entities, we correspondingly also assume that customers' demand for IT capacity in the cloud is a proxy for the value they derive from it. This presents an exciting future research opportunity that combines cloud usage data with metrics on firm performance and in-house IT infrastructure.

References

Anderson, T.W., and Hsiao, C. 1981. "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association* (76:375), pp 598-606.

Aral, S., Brynjolfsson, E., and Wu, D.J. 2006. "Which Came First, It or Productivity? Virtuous Cycle of Investment and Use in Enterprise Systems," *27th International Conference on Information Systems (ICIS)*, Milwaukee, Wisconsin, pp. 1819-1839.

Archak, N., Ghose, A., and Ipeirotis, P.G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (57:8), August 1, 2011, pp 1485-1509.

Arellano, M., and Bond, S. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies* (58:2), pp 277-297.

Arellano, M., and Bover, O. 1995. "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics* (68:1), pp 29-51.

Argote, L., and Miron-Spektor, E. 2011. "Organizational Learning: From Experience to Knowledge," *Organization Science* (22:5), September/October 2011, pp 1123-1137.

Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. 2009. "Above the Clouds: A Berkeley View of Cloud Computing," Technical Report No. UCB/EECS-2009-28, UC Berkeley Reliable Adaptive Distributed Systems Laboratory.

Arora, A., and Gambardella, A. 1994. "Evaluating Technological Information and Utilizing It: Scientific Knowledge, Technological Capability, and External Linkages in Biotechnology," *Journal of Economic Behavior and Organization* (24:1), pp 91-114.

Attewell, P. 1992. "Technology Diffusion and Organizational Learning: The Case of Business Computing," *Organization Science* (3:1), pp 1-19.

Blundell, R., and Bond, S. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics* (87:1), pp 115-143.

Boone, T., Ganeshan, R., and Hicks, R.L. 2008. "Learning and Knowledge Depreciation in Professional Services," *Management Science* (54:7), July 1, 2008, pp 1231-1236.

Bresnahan, T., and Trajtenberg, M. 1995. "General Purpose Technologies: 'Engines of Growth?'," *Journal of Econometrics* (65:83-108).

Bresnahan, T.F., and Greenstein, S. 1997. "Technical Progress and Co-Invention in Computing and in the Uses of Computers," *Brookings Papers on Economic Activity: Microeconomics*), 1996 Special Issue Microeconomics, pp 1-78.

Bresnahan, T.F., and Greenstein, S. 2001. "The Economic Contribution of Information Technology: Towards Comparative and User Studies," *Journal of Evolutionary Economics* (11:1), pp 95-118.

Brynjolfsson, E., and Hitt, L.M. 1995. "Information Technology as a Factor of Production: The Role of Differences among Firms," *Economics of Innovation and New Technology* (3:4), pp 183-200.

Brynjolfsson, E., and Hitt, L.M. 1996. "Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending," *Management Science* (42:4), pp 541-558.

Brynjolfsson, E., Hofmann, P., and Jordan, J. 2010. "Cloud Computing and Electricity: Beyond the Utility Model," *Communications of the ACM* (53:5), pp 32-34.

Cameron, A.C., and Trivedi, P.K. 2010. *Microeconometrics Using Stata, Revised Edition*. Texas: Stata Press.

Casalicchio, E., and Colajanni, M. 2000. "Scalable Web Clusters with Static and Dynamic Contents," *IEEE International Conference on Cluster Computing, 2000.*, pp. 170-177.

Cassiman, B., and Veugelers, R. 2006. "In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition," *Management Science* (52:1), January 1, 2006, pp 68-82.

Ceccagnoli, M., Higgins, M.J., and Palermo, V. 2011. "Internal Versus External R&D: Complements or Substitutes?," *SSRN eLibrary*).

Chen, H., De, P., and Hu, J. 2011. "IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists' Activities and Music Sales," *Workshop on Information Systems and Economics*, Shanghai, China.

Cherkasova, L. 2000. "Flex: Load Balancing and Management Strategy for Scalable Web Hosting Service," *Fifth IEEE Symposium on Computers and Communications (ISCC 2000)*, Antibes, France, pp. 8-8.

Cohen, W.M., and Levinthal, D.A. 1989. "Innovation and Learning: The Two Faces of R & D," *The Economic Journal* (99:397), pp 569-596.

Cohen, W.M., and Levinthal, D.A. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly* (35:1), pp 128-152.

Darr, E.D., and Kurtzberg, T.R. 2000. "An Investigation of Partner Similarity Dimensions on Knowledge Transfer," *Organizational Behavior and Human Decision Processes* (82:1), pp 28-44.

Dutton, J.M., and Thomas, A. 1984. "Treating Progress Functions as a Managerial Opportunity," *Academy of Management Review* (9:2), pp 235-247.

Fichman, R.G., and Kemerer, C.F. 1999. "The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps," *Information Systems Research* (10:3), pp 255-275.

Forman, C., Goldfarb, A., and Greenstein, S. 2008. "Understanding the Inputs into Innovation: Do Cities Substitute for Internal Firm Resources?," *Journal of Economics & Management Strategy* (17:2), pp 295-316.

Garcia, D.F., Rodrigo, G., Entrialgo, J., Garcia, J., and Garcia, M. 2008. "Experimental Evaluation of Horizontal and Vertical Scalability of Cluster-Based Application Servers for Transactional Workloads," in: *8th International Conference on Applied Informatics and Communications (AIC'08)*. Rhodes, Greece: World Scientific and Engineering Academy and Society (WSEAS), pp. 29-34.

Ghose, A. 2009. "Internet Exchanges for Used Goods: An Empirical Analysis of Trade Patterns and Adverse Selection," *MIS Quarterly* (33:2), pp 263-291.

Hansen, L.P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* (50:4), pp 1029-1054.

Harms, R., and Yamartino, M. 2010. "The Economics of the Cloud," Microsoft.

Hitt, L.M., Wu, D.J., and Zhou, X. 2002. "Investment in Enterprise Resource Planning: Business Impact and Productivity Measures," *Journal of Management Information Systems* (19:1), pp 71-98.

Ko, D.-G., Kirsch, L.J., and King, W.R. 2005. "Antecedents of Knowledge Transfer from Consultants to Clients in Enterprise System Implementations," *MIS Quarterly* (29:1), pp 59-85.

Mell, P., and Grance, T. 2011. "The Nist Definition of Cloud Computing," National Institute of Standards and Technology Information Technology Laboratory (ed.). Gaithersburg, MD.

Menon, T., and Pfeffer, J. 2003. "Valuing Internal Vs. External Knowledge: Explaining the Preference for Outsiders," *Management Science* (49:4), April 1, 2003, pp 497-513.

Michael, M., Moreira, J.E., Shiloach, D., and Wisniewski, R.W. 2007. "Scale-up X Scale-Out: A Case Study Using Nutch/Lucene," *Parallel and Distributed Processing Symposium (IPDPS 2007). IEEE International*, pp. 1-8.

Microsoft, and Edge Strategies. 2011. "SMB Cloud Adoption Study Dec 2010 - Global Report," http://www.edgestrategies.com.

Nickell, S. 1981. "Biases in Dynamic Models with Fixed Effects," *Econometrica* (49:6), pp 1417-1426.

Reese, G. 2009. *Cloud Application Architectures: Building Applications and Infrastructure in the Cloud*. O'Reilly Media.

Roodman, D. 2009. "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics & Statistics* (71:1), pp 135-158.

SearchDataCenter.com. 2011. "Data Center Decisions 2011 Survey Special Report," TechTarget.

Symantec. 2011. "State of the Cloud Survey," Symantec.

Varian, H.R. 2010. "Computer Mediated Transactions," *American Economic Review* (100:2), pp 1-10.

Varian, H.R. 2011. "Micromultinationals Will Run the World," in: *Foreign Policy*. http://www.foreignpolicy.com/articles/2011/08/15/micromultinationals_will_run_the_world.

Vega-Jurado, J., Gutiérrez-Gracia, A., and Fernández-de-Lucio, I. 2009. "Does External Knowledge Sourcing Matter for Innovation? Evidence from the Spanish Manufacturing Industry," *Industrial and Corporate Change* (18:4), August 1, 2009, pp 637-670.

Windmeijer, F. 2005. "A Finite Sample Correction for the Variance of Linear Efficient Two-Step Gmm Estimators," *Journal of Econometrics* (126:1), pp 25-51.