Characterizing the Usage Intensity of Public Cloud

Aadharsh Kannan[1], Jacob LaRiviere[1], and R. Preston McAfee[2]

[1] Microsoft AI & Research {akannan,jlariv}@microsoft.com
[2] preston@mcafee.cc

Abstract. This paper uses precise and novel data on country level Cloud IaaS and PaaS revenue to measure the intensive margin of technology diffusion across countries and within countries over time. We horse race diffusion models and find cloud diffusion exhibits both Log-Log and Logistic Growth Patterns. We use cross validation on nearly 100 features to determine what drives cross-country differences. We find that features impacting GDP,Internet Connectivity and Human Capital explain the intensity of cloud growth. We finally compare the relative impacts of these variables using a random coefficients model. While correlative, our algorithmic research design motivates data driven hypothesis generation and further causal work regarding how policymakers can encourage more cloud computing adoption and technology adoption more broadly.

1 Introduction

Cloud computing lowers the cost of new firms to enter markets. Instead of buying hardware, entrepreneurs can rent it from public cloud providers and further leverage platform as a service (PaaS) offerings to reduce spend to write and organize code. Because successful entrepreneurship creates value for new firms and its customers, understanding what can lead to increased entrepreneurship, such as cloud computing adoption, is important.

The rental market for cloud computing is a new technology less than ten years old. Technologies are often thought to diffuse throughout society in a "S-shape" wherein low levels of adoptions precede a sharp increase in adoption levels followed by a leveling off. The literature on technology diffusion provides a number of possible explanations for observed diffusion patterns including information asymmetry of adopters, firm or consumer heterogeneity in adoption costs or benefits and competitive forces which lead to supply side costs dynamics consistent with decreasing costs [8].

More recent work shows that at the country level there are two different technology adoption margins both of which vary predictably across nations and across technologies. The first is the extensive margin; some countries or regions adopt (a binary adopt or not adopt decision) technologies before others. The second is the intensive margin; a binary adopt or not adopt decision obscures *the level of intensity of adoption* at both a country and firm level [8]. Increasingly countries look more similar on the extensive margin but more varied on the intensive margin [8]. As a result, better data on intensive margin adoption decisions is needed to compare technological diffusion across countries. We provide that using a panel of Microsoft Azure (a major Cloud Provider) consumption data.

Cloud computing revenue is an ideal technology to study intensive margin diffusion of computation because cloud computing is a rental service. In cloud computing, customers are billed roughly per core hour (similar to wholesale expenditures on electricity). Other measures of computation penetration like personal computers per capita are more coarse because firms and consumers use computers with different intensities. For example, one may buy a personal computer for performing complex mathematical analysis (use it more intensively) or occasionally check mails (use it less intensively). Studying expenditure patterns of cloud computing at the individual country basis provides one of the highest possible signal to noise ratio of any computation metric available i.e. we have more nuanced data than other studies that permit us to drill deeper into the determinants of adoption and in particular identify how complementary goods play a role in adoption rates. Given that cloud computing is a rental service (metered usage), observed revenue stems from intensive usage.

In this paper we investigate expenditures on cloud computing of the two largest cloud providers in the world: Amazon Web Services (AWS) and Microsoft's Azure. In studying AWS we compare different diffusion models present in the literature using public revenue data from their SEC 10-Q filings. Using AWS revenue, we horse–race differing diffusion models selecting the one which fits the data best as our preferred model. We then estimate our preferred model on country level cloud expenditure data of Azure customers, obfuscating levels due to the proprietary data and focusing on across country differences. We use a wide variety of features to explain cross country differences in diffusion and leverage cross-validation to select which features best explain cross country differences, similar to [4]. Finally we include those features to in a random coefficients model to leverage both cross sectional and longitudinal variation in determining what explains the intensity of cloud adoption.

The main contribution of this paper is our use of cross-validation and our unique dataset to provide the first empirical evidence for what country level variables are associated with the intensity with which countries adopt it. We find that output levels (e.g., GDP, Local Supplier Quality), labor productivity (e.g., Availability of Engineers, Enrollment in Secondary Education) and different complementary technologies (e.g., last mile internet connectivity) all significantly correlate with cloud diffusion intensity. GDP, human capital and complementary technologies are all positively associated with growth (accounting for differences in adoption cycle), implying that subsidies or loans (a proxy for GDP) coupled with public good spending on complementary technology and increasing human capital are policies that could increase the rate and level of technology adoption. Because we only discuss correlations however, forming policy based upon these findings is not prudent until causal studies bear out these findings. As with any study leverage a non-causal research design, the correlates we find as statistically significant might not themselves be root causes. Similarly, because we only use data for a single public cloud provider additional evidence with additional data sources are required to determine the robustness of these empirical findings.

While our estimates are not causal, the novel data, model horse–racing, algorithmic feature selection and random coefficients model leveraging the panel structure of the data informs data–driven hypothesis generation for future theoretical and empirical studies to inform optimal policy. Further, we show that cross sectional country level variables like the ones used here combined with informed structural models can be used to predict technology diffusion while respecting historical "S-shaped" patterns. While the use case here is cloud the technique is broadly application for other technologies and therefore useful for important capital allocation functions like market sizing.

2 Aggregate Characteristics

We performed a statistical exercise to select the functional form of cloud revenue diffusion. The goal of this section is to determine which models are potentially inconsistent with aggregate data to avoid misspecification. We then take the most parsimonious model that remains to the country panel data of Azure Revenue. We include technical details in the Appendix and only discuss the high level approach here.

We fitted three diffusion models found in the literature: Log-Linear, Logistic Growth and Log-Log model ([7]). A distinguishing characteristic of these models is the differing implications for out–of–sample growth which we discuss below. However, because our sample period encompasses only the early stage of cloud adoption, we focus on in sample performance. We estimated the different models in the growth space (e.g., first–differenced versions of the aforementioned models).

To determine which model best describes the diffusion of cloud adoption we leverage publicly available quarterly AWS revenue data. We choose AWS revenue data because there is a longer time series as AWS was first to market. AWS and Azure both offer cloud-based IaaS but AWS started selling two years prior to Azure. As a result, AWS has more data and thus more signal for model selection. Previous literature shows that aggregate technology classes (e.g., cars, electricity, telephones, PCs, etc) can be summarized with a diffusion curve from the log family mitigating concerns that the different companies would be described by fundamentally different models ([7]). Further, according to a 2018 survey, the vast majority of large companies employing a multi-homing cloud strategy.See `https://www.virtustream.com/press-release/new-study-shifting-business-priorities-herald-the-era-of-multi-cloud`. Put another way, firms split their cloud workloads between providers like AWS and Azure. It seems likely that longer run trends for AWS will be similar to those of Azure.

We obtained AWS's quarterly revenue filings with the Securities and Exchange Commission[9], dating back to Q1 2014. AWS is a public cloud provider with significant market share [14] and a sufficiently long historical time series for choosing amongst different diffusion models. As a result, Figure 1 shows that all three models estimated from the revenue data accurately within–sample.

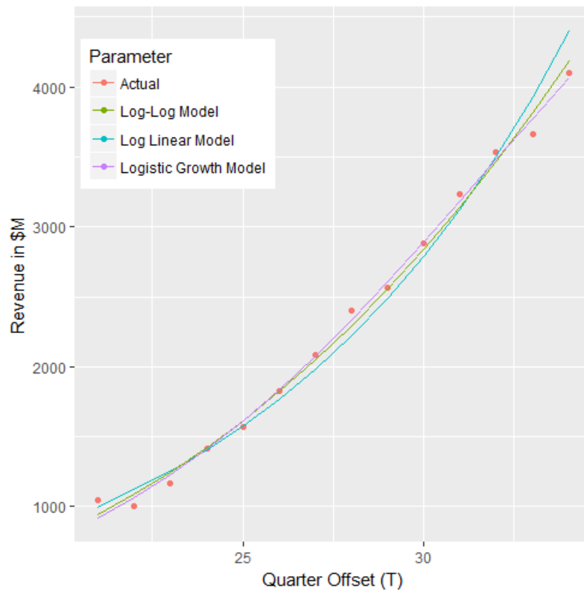


Fig. 1: Amazon Web Service - Revenue Data & Model Fit

The Appendix shows detailed model fitting we used to determine if any of the three empirical approaches fit the AWS data better than others. We leveraged the Kolmogorov Smirnov (KS) test to determine if the error structure implied by each model above can be rejected by the data. We reject the null hypothesis that the data in consistent with a Log Linear model. We fail to reject both the Logistic Growth model and the Log Log model. Given that the Log Log model is easier to work with and there is no compelling reason not to use it to compare in sample historical patterns, we use it for the remainder of the paper.

3 Country Level Characteristics

We estimate model of logged Azure revenue data on log time in the next section. We utilize the country level revenue data from Azure in order to understand the cloud usage across nations and characterize the systematic variation that we observe across nations as a function of country specific characteristics.[3] Specifically, we use LASSO to feature select what predicts intensity of cloud adoption. Finally we include those features in a random coefficients model to leverage both cross sectional and longitudinal variation in determining what explains the intensity of cloud adoption.

The external validity of insights from Azure data as being representative of cloud computing demand more broadly rely on Azure's scale, fierce competition for customers in the fast growing cloud industry and the nature of the cross-country analysis we perform in this study. It is likely that Azure buyers skew toward somewhat older, larger firms, due to those firms' established relationships with Microsoft. But because AWS

[3] Cloud resources are priced at a region level. For example, any customer can sign up for a particular VM family at the US West data center, at the Brazil data center or the India Central data center. The data we report in the paper is data for customers in a particular country. Since all customers have the same choice set, this is the appropriate unit of analysis for our exercise. Of course, a customer might have a preference and even be willing to pay a premium to put their workload in a nearby data center to their own physical location. This larger question about the preference of cloud customers for distance is its own research question that is beyond the scope of this paper. However, over our dataset's timeframe, most cloud providers had data centers which were roughly co-located (e.g., both AWS and Azure opened data centers in India, Brazil, etc).

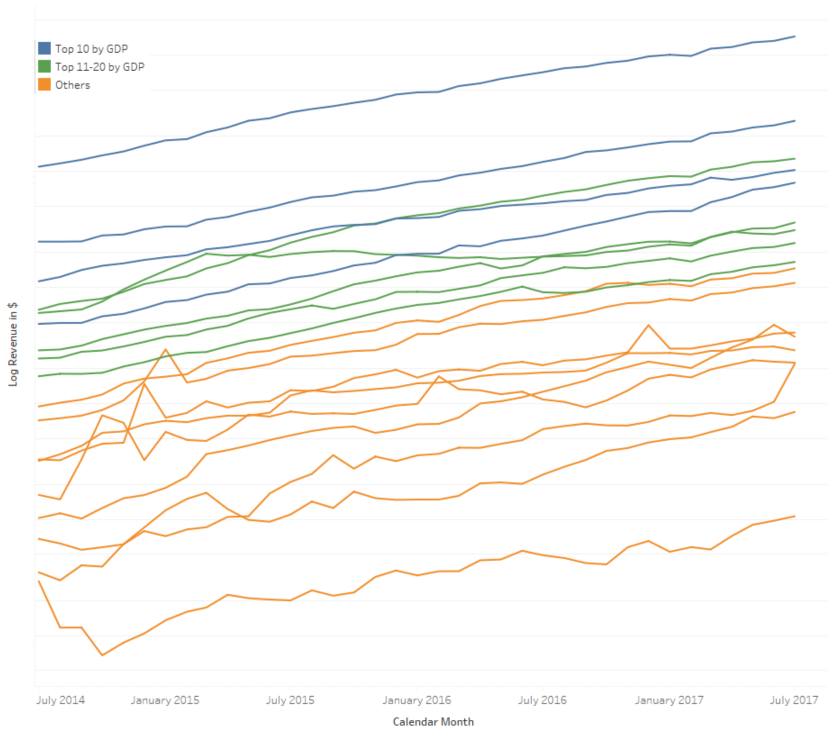


Fig. 2: Log Azure Revenue for specimen Countries

and Azure compete vigorously and successfully at the market level, as opposed to satisfying a niche buyer type, it is likely that the country-level determinants of Azure purchase reflect the cloud adoption market as a whole. Obviously it would be preferable to have all adoption data, or at least AWS as well as Azure, but Azure data is large enough to likely reflect the characteristics of technology adoption investigated by this study. That said, our paper speaks to correlates of Azure adoption.

We used monthly level data for 98 countries ranging from July 2014 to July 2017. A random subset of countries include Belgium, Czech Republic, Germany, Iceland, Israel, Latvia, Panama, Russia, Singapore, Slovakia an Trinidad & Tobago. In the LASSO specification below we use 111 unique right-hand side variables to explain country level growth rates between July 2015 to July 2017. Across all these countries the average growth during these two years was 66% (min 20% and max 100%). Of the 111 variables, 109 are taken from the World Economic Forum and 2 are from the Cisco connectivity index. Rather than describing all variables here, we describe the variables selected by LASSO below in detail.

Figure 2 shows the Azure Revenue in the Log scale over time. The y–axis labels have been intentionally obfuscated by not starting at zero and masking the labels throughout this section. We did this to obfuscate the actual revenue at the country level which is a proprietary information of Microsoft Corporation. Throughout this paper, care has been taken to prevent any inference of actual revenue numbers at the country level while still retaining the focus on the systematic differences that exists across countries. Visualized in the revenue scale, Figure 2, produces the familiar curvature that we witnessed in the aggregate AWS data as seen in Figure 1; naturally it follows a straight line in the Log Scale.

We analyze the variation using the Log-Log model presented in section A. We pick the Log-Log model because it has growth decay like Logistic Growth but is more straightforward to work with than Logistic Growth when comparing growth rates change as a function of country characteristics. Given that both the aforementioned models are consistent, the presented analysis can be repeated with any one of two. Put another way, as we compare diffusion across countries the analysis is consistent within any of the two models. We follow a data–driven approach to model selection that is described in this section.

3.1 High Dimensional Sparse Model Selection

Many factors could predict a nation's intensity of cloud usage. However, many factors could be collinear. For example, Gross Domestic Product (GDP) is correlated with Enrollment in Secondary Education. In order to select a model from a large feature space that maximizes out of sample prediction but respects multicolinearity, we use LASSO to perform feature selection as in Belloni et al [4].

Table 1: High Dimensional Sparse Model Selection

Indicator	Estimate
log(GDP 2016 in Purchasing Power Parity Adj. US$)	0.917
Availability of scientists and engineers, 1-7 (best)	0.330
Strength of auditing and reporting standards,1-7 (best)	0.328
Willingness to delegate authority, 1-7 (best)	0.180
Local supplier quality, 1-7 (best)	0.159
Secondary education enrollment, gross %	0.027
Fixed telephone lines/100 pop.	0.014
Fixed broadband Internet subscriptions/100 pop.	0.002

The cost function being minimized is given by

$$\sum_{c\in C}\left(log(\Delta R_c) - \left(\alpha + \sum_{k\in F}\beta_k x_k\right)\right)^2 + \lambda\sum_{k\in F}|\beta_k|$$

ΔR_c – Azure revenue difference from 2015 to 2017 for Country c
C – Set of Countries i.e. {United States, United Kingdom, ... }.
F – Set of Country Economy Indicators

The goal of the exercise is to choose feature which have the most predictive power for cloud revenue over time. We utilized roughly 90 variables like $log(RealGDP_{2016}\$)$, Strength of auditing and reporting standards, Fixed telephone lines/100 population etc. about 100 countries given in the Global Competitive Index Report[12] for the set of candidate features to explain the intensity of cloud adoption. The GCIR contains a set of macroeconomic variables that capture various aspects of the economy for different nations ranking aggregate traits like corruption, transparency of government, etc.

The variables selected from the LASSO procedure are given in the Table 1. The coefficients reported in Table 1 are biased estimates of the relationship between these variables and outcomes since LASSO trades bias for variance reduction. We have intentionally obfuscated the revenue change $log(\Delta R_c)$ by not reporting the intercept (non-zero). We used k–fold cross validation (k=10) in order to pick the optimal λ by minimizing the out–of–sample prediction error (see figure in appendix).[4]

In the subsequent sub sections we give an economic interpretation to some of the indicators selected by the model under three major categories.

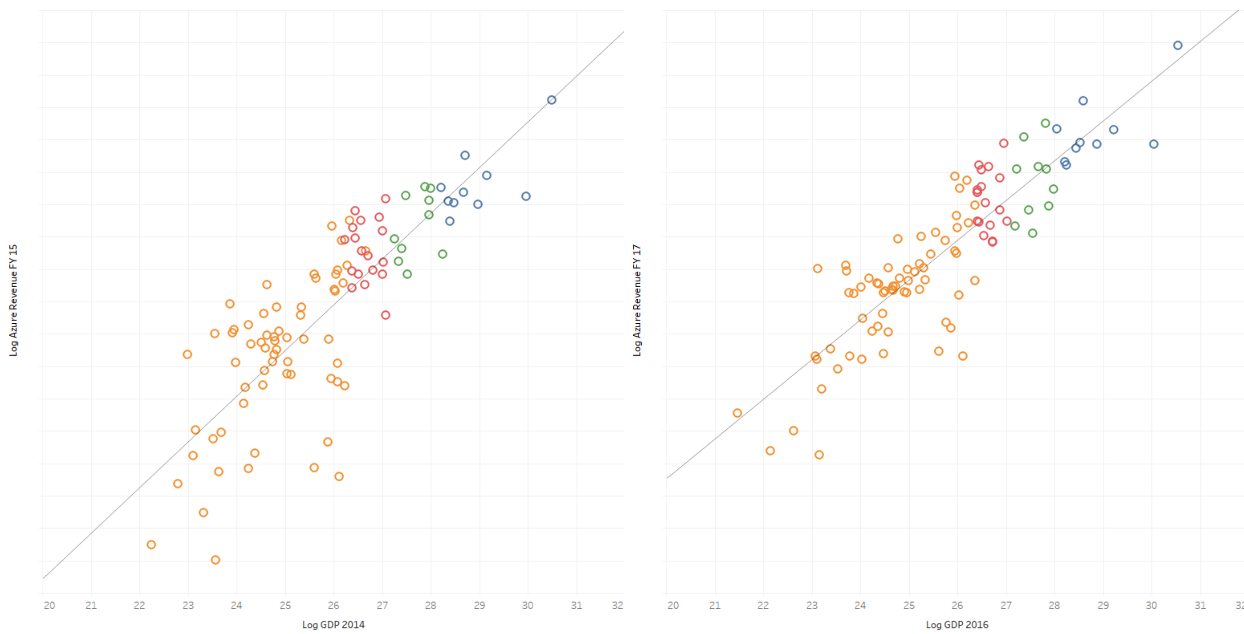
Gross Domestic Product (GDP) and Market Health GDP is a useful predictor of technology diffusion. Comin and Ferrer [8] show that wealth measures predict the long run adoption intensity of new technologies. We use the log of their product to examine the medium run trends in cloud uptake by using the cloud spend data from our provider at the country level.

Figure 3a and Figure 3b give us the relationship between Log Real GDP and Log Azure revenue for two different years.[5]. Looking within each panel, there is a strong linear correlation between the two variables

[4] Individuals using Internet % was dropped as it was not robust to stability checks (one standard error) around the minimum lambda. Further, the point estimate was less than .001.

[5] We create an explanatory variable x_{1c} where the subscript c stands for country as follows:

$$x_{1c} = log(\text{GDP 2016 in Purchasing Power Parity Adjusted US\$})$$



(a) 2014 Log GDP vs Log Cloud Usage FY-15

(b) 2016 Log GDP vs Log Cloud Usage FY-17

with some countries above versus below the average relationship. Looking across the two panels, the relative position of each country in Figure 3a to that of their corresponding positions in Figure 3b, countries move towards the right and upward. The movement towards the right is due to the fact that nations are getting richer (GDP has increased in 2016 since 2014). The movement upwards is a combined effect of Cloud adoption increasing and Azure market share changing in these individual nations. The Figures show that as GDP's rise, the relationship between cloud consumption and GDP gets tighter. Thus log GDP's predictive power has increased over time.

Other indicators of market health selected by the LASSO model (i.e. Strength of auditing and reporting standards, Willingness to delegate authority and Local supplier quality) address the extent to which the nation is fairly regulated and the strength of domestic market. Of course there is correlation between variables like local supplier quality and real GDP. In the next section we use cross sectional and longitudinal variation in cloud spend to identify the association of each on adoption.

Internet Connectivity Infrastructure Seldom do nations leapfrog in technology adoption[8]. Rather, adoption of newer technology is based on the presence of existing or base technology. A similar argument can be drawn for cloud adoption and underlying last mile Internet Connectivity Infrastructure. Fixed telephone lines/100 pop., Fixed Broadband Internet Subscriptions/100 pop. give us an indication of intensive adoption of complementary technologies for Cloud Computing. That LASSO selects these variables makes intuitive sense: better broadband coverage will impact latency and thus the quality of cloud.[6]

Human Capital Several studies have established a causal relationship between education and technology adoption ([21],[5],[16]). It is perhaps unsurprising that we find features indicative of better human capital (Availability of Scientists & Engineers and Secondary Education Enrollment) within a country to be predic-

[6] One might wonder if there is a mechanical correlation between broadband speeds and cloud revenue. For example, faster internet connection might allow tasks to be done more quickly and for a lower price. Put another way, cloud users must secure a VM and then upload a problem to it, so the upload time could in principle matter. However, most uses are powered down; the number of servers may vary but the uses don't look like "upload a problem, compute, download the answer" but rather "continuously operate on a database, serve customers, and scale up or down as needed." In the latter model, bandwidth speeds don't have a material impact on the time a VM is in use.

tors of intensive usage of cloud computing. Human capital is correlated with wealth and infrastructure as well motivating the random coefficients model below.

Supply Side Variables While the focus of this paper is to determine demand side covariates for cloud computing, it is interesting to consider possible supply side covariates. Any cloud customer can deploy their workload in any data center in the world. For example, in our sample only 10 unique countries had an Azure data center but we observed Azure revenue in almost every country. It is possible, though, that even accounting for country demand shifters there would still be signal in supply side variables like proximity to data centers. The simplest supply side variable, then, is the count of data centers in a country. Of course this variable is far from exogenous as Azure didn't randomly build data centers but it is instructive: in this early stage of cloud computing, does proximity matter even controlling for aggregate demand shifters.

To investigate the importance of different supply side variables, we extended the LASSO to includ the following: the count of data centers in a country, the interaction of a DC indicator (e.g., $1DCCount > 0$) and price of the most popular SKU over the sample, the interaction of a DC indicator and the age of the first DC in the country, and an indicator variable for AWS having a DC in the country. The 1 standard error LASSO model did not select any of these as features. While the out of sample MSE minimizing lambda did select AWS compete and Azure price, neither was close to statistical significant in the post LASSO OLS regression of 2015 to 2017 growth rates out they weren't close to being statistically significant (p-values of .29 and .87 respectively) when we included them out of curiousity. While these results are clearly not causal, this is evidence that demand drivers have primary explanatory power for endogenously determined cloud revenue over our sample. We don't include data center counts in the random coefficient model because it seems only marginally predictive of revenue growth over our sample.

There are two possible reasons why supply side covariates would not be statistically significant correlates of cloud spend at the country level. First latency rather than national borders matter. For example, a customer in Tijuana might prefer to deploy a workload in a California data center over one near Mexico City. Second, presence of a domestic DC could matter but not for the type of variation most prevalent in our data. For example, comparing African countries to US and EU countries implies that socio-demographics matter relatively more. More generally, our results don't necessarily suggest that having a domestic DC doesn't matter. Rather it suggests it isn't a primary driver of cross-country differences in cloud adoption over our sample.

3.2 Random Coefficients Model

In order to understand the relationship between these aforementioned factors and cloud usage intensity, we must decouple the within-nation variation (e.g., systematic growth in revenue within a country over time) and cross-sectional variation (e.g., systematic across country variation). To do so we use a Random Coefficients model.

There are two effects at play in characterizing the cloud usage of individual nations 1. Inherent trend of a country characterized by idiosyncratic features 2. Common trend of a country characterized by factors like GDP, Connectivity, and Human Capital. The random coefficients model (RCM) specification [10] helps us determine the relationship between factors like GDP, human capital and last mile connectivity with the Cloud usage while controlling for the inherent trends that are idiosyncratic. We carry forward notations introduced earlier (section A).

Given that $\tilde{F_c}$ are features identified in section 3.1, the RCM specification[7] is given below:

$$log(R_{cT}) = (\bar{\alpha} + u_{0c}) + \left(\bar{\beta_c} + u_{1c}\right) log(T) + \epsilon_{cT} \tag{1}$$

$$\epsilon_{cT} \sim N(0, \sigma_{cT})$$

[7] Model shown here can be decomposed into a mixed model which contains both country level fixed effects and random effects for α and β. A detailed discussion motivating the choice of model is presented in the Appendix D.

$$\bar{\beta_c} = \gamma_0 + \sum_{k \in \tilde{F_c}} \gamma_k x_{ck} + \epsilon_\beta \tag{2}$$

$\tilde{F_c}$ – non-zero features from Model Selection in country c
$\epsilon_\beta \sim N(0, \sigma_\beta)$

There is a subscript c to $\bar{\beta_c}$ and not to $\bar{\alpha}$. The term $\bar{\beta_c}$ (common trend) is defined by systematic relationships to GDP, Human Capital and Network Infrastructure in the equation (2) with an error of ϵ_β. The expression $\beta = \bar{\beta_c} + u_{1c}$ is what we call the "intrinsic" growth rate of a country since the actual growth rate varies with time as per the Log–Log model. The term u_{1c} is the country specific component of the "intrinsic" growth (inherent trend). Similarly, for intercept, $\alpha = \bar{\alpha} + u_{0c}$ is assumed to have a common mean $\bar{\alpha}$ but country specific component u_{0c}.

The reason for letting α and β vary by country is that countries start adopting cloud in different time periods but we estimate the model in common time periods;[8] Put another way, a nation that adopts in January 2011 is 18 months later in their cloud adoption curve than on who adopts in July 2012. This specification accounts for that. Nations with higher amount of cloud spend during the first period of observation will have lower growth in percentage terms. u_{0c} is thus required to handle the structure of our data. As a result, we expect a negative correlation between u_0 and u_1. The adjustments, u_0 and u_1 are assumed to be normally distributed with a mean zero and variances σ^2_{u0} and σ^2_{u1} respectively. We have also allowed for correlation between u_0 and u_1 denoted by ρ_u.

The specification of u_0 and u_1 are given below:

$$\Sigma_u = \begin{pmatrix} \sigma^2_{u0} & \rho_u \sigma_{u0} \sigma_{u1} \\ \rho_u \sigma_{u0} \sigma_{u1} & \sigma^2_{u1} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u \right)$$

3.3 Estimation and Results

A model specification given above can be estimated using Monte Carlo Markov Chain [18] or any other simulated maximum likelihood mechanism [15]. We used the Monte Carlo Markov Chain procedure to estimate our model. Given that we had to respect the data confidentiality of the revenue of individual countries, we have reported only those parameters that are relevant to characterizing the systematic variation.

We have also reported the Median Absolute Percentage Contribution (referred to as $MdAPC_\beta$ and $MdAPC_\alpha$) of $\bar{\beta_c}$ and $\bar{\alpha}$ as a measure of the explanatory power of $\bar{\beta_c}$ and $\bar{\alpha}$. Such measures are robust to outliers [19]. The values of the $MdAPC_\beta$ and $MdAPC_\alpha$ give the degree to which the growth and the initial level (as observed in the first period of the data) can be explained using the macro–economic indicators discussed previously. The term $MdAPC_\beta$ is defined to be the median, across all nations of the quotient $\frac{|\gamma_0 + \sum_{k \in \tilde{F_c}} \gamma_k x_{ck}|}{|\gamma_0 + \sum_{k \in \tilde{F_c}} \gamma_k x_{ck} + u_{1c}|}$. Put another way, it is the median absolute contribution to growth rate of $x_k \forall k \in \tilde{F}$ across all nations. Similarly we can define $MdAPC_\alpha$ as a median of $\frac{|\bar{\alpha}|}{|\bar{\alpha} + u0c|}$.

Table 2 shows estimated values from the random coefficients model. σ_{cT} is the unexplained variation in predicting cloud usage at the country level. We find it is <5% relative to the parameters for the median country. Put another way the coefficient of variation of σ_{cT} relative to $\bar{\alpha} + u_{0c}$ and $\bar{\beta_c} + u_{1c}$ is less than 5%. We thus conclude that a very large share of cloud revenue changes are systematic rather than unpredictable.

[8] An alternative way to handle this problem would be to rescale the time dimension for each country so that the first adopting period of that country takes the value "one", the second "two" and so one. We leave the data in its most simple form choosing instead to relax our statistical model.

Table 2: Random Coefficients Model Estimates

Indicator	Coefficient	Estimate (Std.Err)
Std. Deviation of Revenue Equation	σ_{cT}	0.817 (0.001)
Correlation between u_{0c} & u_{1c}	ρ_u	-0.987 (1e-10)
log(GDP 2016 in Purchasing Power Parity Adj. US$)	γ_1	0.218 (0.001)
Availability of scientists and engineers, 1-7 (best)	γ_2	0.03 (0.002)
Strength of auditing and reporting standards,1-7 (best)	γ_3	0.11 (0.004)
Willingness to delegate authority, 1-7 (best)	γ_4	0.024 (0.003)
Local supplier quality, 1-7 (best)	γ_5	0.031 (0.004)
Secondary education enrollment, gross %	γ_6	0.003 (0.0005)
Fixed telephone lines/100 pop.	γ_7	0.003 (0.0007)
Fixed broadband Internet subscriptions/100 pop.	γ_8	0.005 (0.001)
	$MdAPE_\beta$	**50.45%**
	$MdAPE_\alpha$	**97.11%**

ρ_u is the correlation between growth and the initial revenue level (recall that we use only a left censored subset of the universe of all revenue data in this study). Its value (-0.987) shows a strong negative correlation between initial value and growth. The implication is that later adopting countries will have a lower growth rate later in the life–cycle as their adoption is in a high initial growth phase (u_{1c} is high) and their initial revenue level is low (u_{0c} is low). This is consistent with standard technology diffusion processes.[9]

The coefficients on GDP, Connectivity, Human Capital indicators are all positive and statistically significant. The positive coefficients mean that revenue growth is higher in countries with higher GDP, higher human capital index and better connectivity. The coefficients on availability of engineers, auditing, willingness to delegate and local supplier quality all have the same support so are directly comparable. Availability of skilled technical labor and local supplier quality have the largest coefficients at a roughly 30% higher than delegation and a triple the coefficient of auditing and reporting. This is a reasonable result: ultimately both are indicative of quality labor supply. This is consistent with a theory that employees drive adoption of new technologies which help them to be more productive *controlling for income and other factors*. Similarly delegation is an outcome of norms in a labor market. We were somewhat surprised at the import of labor supply features on cloud adoption. Government policy insofar as it impacts education attainment and telecommunications infrastructure is also very important.

While correlative, these results imply that improved access to income (perhaps through loans), increased availability of skilled human capital and investment in complementary public goods like fiber infrastructure (improvement to last mile connectivity) may increase technology adoption for cloud. Of course, this is not a causal result since there is no exogenous variation in our data. Thus we view this as motivation for future causal empirical studies and theoretical models to explain these findings.

The importance of the selected macroeconomic indicators can be seen in the $MdAPC_\beta$ measure. We find that for the median country in our data these factors explain roughly 50% of the variation in growth rates. This controls for differential timing of adoption, censorship in our observation and other random monthly variations. Coefficient of determination ($R^2_{\beta_c}$) on equation 2 explaining $\bar{\beta_c}$, is 98.77% which means that less than 1.2% of the variation (unexplained by the macroeconomic indicators) in $\bar{\beta_c}$ is random. In other words, the estimates of γ_0 through γ_8 are accurate.

We think non-experimental nature of our work limits our ability to provide deeper insights for marketing or even economic modeling. Of course, correlations provide restrictions on models, but because there is substantial correlation among the explanatory variables, we can't identify what variables are driving the outcomes. Instead, we produce insight into the kind of countries that adopt cloud. This is valuable not so much for marketing purposes but for understanding the determinants of future growth, which are tightly bound to the adoption of improved technology. Moreover, our findings may suggest economic policies to promote growth, e.g. increasing the number of engineers, though we emphasize that is correlation, not

[9] γ_0 (not reported) has a negative value and is uninformative of within and across country comparisons' evaluation.

causation, at this point. Finally, cloud adoption is likely to correlate with adoption of other technologies, such as IoT and 5G.

Insofar as Azure revenue patterns reflect broader cloud patterns this analysis is broadly applicable to cloud computing generally. It is credible that Azure revenue patterns reflect broader cloud consumption patterns. Between 2015 and 2019, Azure has been the second largest cloud provider in the world behind AWS but ahead of Google Cloud Platform, IBM, Alibaba and others.[10] Further, as mentioned above, almost 90% of large companies multi-home their cloud workloads.[11] Because firms split their cloud workloads between providers like AWS and Azure, it seems likely that longer run trends for Azure represent broader industry trends. Put another way, correlates of Azure market share are not likely to be correlated with country level features because its reasonable that fluctuations in Azure market share are likely second order to fluctuations in total cloud usage across countries.

While we leverage some within country across time data, the majority of the variation which identifies the relationships between Azure revenue and country features is cross sectional. If there is systematic variation in Azure market share relative to other cloud providers and those features, our research design will have biased coefficient estimates. For example, if countries with a large number of engineers have higher market share of Azure, then those features will receive relatively higher weight than if the dependent variable were the sum of cloud revenue across all providers. We don't view this as a large problem for two reasons. First, it is unlikely the sign of covariate coefficients will flip; number of engineers likely isn't good for Azure and bad for AWS. Second, as mentioned previously there are several reasons why we don't expect there to be large market share differences at the country level since most firms multi-home and Azure is the second largest cloud provider. Future work along the lines of [20] which tries to interpolate market shares leveraging the rollout of new data centers could mitigate this concern insofar as it is an issue. In lieu of that, a globally representative stratified survey would be required to determine total cloud spend by country.

4 Conclusions

We use very precise data (Cloud IaaS and PaaS revenue) to measure the intensive margin of technology diffusion over time and across countries. Consistent with [7], we find evidence that intensive margin log diffusion consistent with AWS revenue data are also consistent with country level Azure revenue data. GDP, last mile connectivity and human capital explain roughly 50% of the "intrinsic" growth of cloud. Labor supply and human capital variables were particularly important even conditioning on wealth levels and infrastructure.

There are two larger implications of this work. First, we show that cross sectional country level variables like the ones used here combined with informed structural models can be used to predict technology diffusion. While the use case here is cloud, the technique is broadly application for other technologies and therefore useful for important capital allocation functions like market sizing.

Second, if these results are confirmed with subsequent causal estimates, the policy implication is that governments may be able to increase the intensive margin of technology adoption by mimicking increases in GDP (e.g., IT loan programs). Supporting STEM education and fluid labor market (e.g., delegation of tasks) and clear enforceable reporting policies and improving infrastructure could lead to more rapid adoption of new technologies. Thus we view this as motivation for subsequent causal empirical work and theoretical modeling which can explain these systematic findings.

Acknowledgement. The authors would like to thank Microsoft Corporation for providing their country level cloud consumption data. The authors would like to thank members of the Office of the Chief Economist at Microsoft for their invaluable guidance. The authors would also like to thank Mark Russinovich (CTO of Azure) and Marcus Fontoura (Distinguished Engineer, Azure) for reviewing our work and providing insightful comments and feedback. Positions presented in this paper are strictly that of the authors and does not represent the official or unofficial position of Microsoft Corporation.

[10] Quarterly aggregate revenue data from large cloud providers is publicly available in 10-K and 10-Q filings of US listed companies.

[11] See `https://www.virtustream.com/press-release/new-study-shifting-business-priorities-herald-the-era-of-multi-clo`

References

1. Abril, P.S., Plant, R.: Mathematical researches into the law of population growth increase. Nouveaux Memoires de l'Academie Royale des Sciences et Belles-Lettres de Bruxelles. 18, 1–42 (1845)
2. Ayres, R.: Technological Transformations and Long Waves. Ph.D. thesis, International Institute for Applied Systems Analysis, Laxenburg, Austria (Feb 1989), http://webarchive.iiasa.ac.at/Admin/PUB/Documents/RR-89-001.pdf
3. Bauke, H.: Parameter estimation for power-law distributions by maximum likelihood methods. The European Physical Journal B 58(2) (Aug 2007), https://doi.org/10.1140/epjb/e2007-00219-y
4. Belloni, A., Chernozhukov, V., et al.: Least squares after model selection in high-dimensional sparse models. Bernoulli 19(2), 521–547 (2013)
5. Benhabib, J., Spiegel, M.: Human capital and technology diffusion. In: Aghion, P., Durlauf, S. (eds.) Handbook of Economic Growth, vol. 1, Part A, chap. 13, pp. 935–966. Elsevier, 1 edn. (2005), https://EconPapers.repec.org/RePEc:eee:grochp:1-13
6. Clauset, A., Rohilla, C., Newman, M.: Power law distribution in empirical data. Society for Industrial and Applied Math 51(4) (Nov 2009), https://doi.org/10.1137/070710111
7. Comin, D., Hobijn, B.: An exploration of technology diffusion. American Economic Review 100(5), 2031–59 (2010)
8. Comin, D.A., Ferrer, M.M.: If technology has arrived everywhere, why has income diverged? Working Paper 19010, National Bureau of Economic Research (May 2013), http://www.nber.org/papers/w19010
9. Filing, A.Q.R.: Securities and exchange commission filings. Tech. rep. (2017), http://phx.corporate-ir.net/phoenix.zhtml?c=97664&p=irol-sec&control_selectgroup=Annual%20Filings
10. Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. (eds.): Longitudinal Data Analysis, chap. 9, pp. 309–362. Chapman and Hall CRC, Boca Ranton, FL, 2nd edn. (2009), www.stat.ncsu.edu/people/davidian/courses/st732/notes/chap9.pdf
11. Forecast, G.: Gartner says ww public cloud services market to grow 18 percent in 2017. Tech. rep. (2017), http://www.gartner.com/newsroom/id/3616417
12. Forum, W.E.: The global competitiveness report 2017–2018. Tech. rep. (2017-2018), https://www.weforum.org/reports/the-global-competitiveness-report-2017-2018
13. Fox, J., Weisberg, S.: An R Companion to Applied Regression, chap. Appendix Material, p. Appendix. SAGE, Washington DC, 2nd edn. (2010), http://socserv.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Nonlinear-Regression.pdf
14. Novet, J.: Amazon's cloud growth might be slowing, but it's still way ahead of everybody else. Tech. rep. (2017), https://www.cnbc.com/2017/07/28/aws-still-rules-the-cloud-despite-growth-slump.html
15. Park, S., Gupta, S.: Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. Journal of Marketing Research 46(4), 531–542 (2009), http://www.jstor.org/stable/20618914
16. Riddell, W.C., Song, X.: The role of education in technology use and adoption: Evidence from the canadian workplace and employee survey. ILR Review 70(5), 1219–1253 (2017), https://doi.org/10.1177/0019793916687719
17. Skinner, J., Staiger, D.: Technology adoption from hybrid corn to beta blockers. National Bureau of Economic Research No. w11251 (2005)
18. Sorensen, T., Vashisht, S.: Bayesian linear mixed models using stan. Quantitative Methods for Psychology (Jun 2016), https://arxiv.org/abs/1506.06201
19. Tayman, J., Swanson, D.A.: On the validity of mape as a measure of population forecast accuracy. Population Research and Policy Review 18(4), 299–322 (Aug 1999), https://doi.org/10.1023/A:1006166418051
20. Wang, S., LaRiviere, J., Kannan, A.: Spatial competition and missing data: an application to cloud computing. NBER Technology Workshop, Stanford University (2019)
21. Welch, F.: Education in production. Journal of Political Economy 78(1), 35–59 (1970), https://EconPapers.repec.org/RePEc:ucp:jpolec:v:78:y:1970:i:1:p:35-59

Apendix

A Model Fitting

There are a few common variables and notation used in this section. Following [8] we define a quarter offset T the number of quarters since AWS was available as a public cloud offering: $T = t - t_0$ where t denotes the current quarter and t_0 denotes the first quarter AWS was sold.[12] Given we observe the data on a quarterly basis, T is an integer number. R_T denotes the revenue in millions of dollars (denoted as \$M) for a given quarter offset T. Also unless specifically noted, the term ϵ is an error term with a mean zero normal distribution.

Table 3: Parameter estimates and consistency checks

Model	Estimates (Std.Err)**	R^2
Log Linear		0.9846
	α_1 4.513 (0.110)	
	β_1 0.114 (0.004)	
Logistic		-
	θ_1 6854.53 (811.11)	
	θ_2 -5.490 (0.191)	
	θ_3 0.173 (0.013)	
Log Log		0.9894
	α_2 -2.573 (0.294)	
	β_2 3.094 (0.089)	

**significant at 0.001

Log Linear Model The Log Linear model assumes a linear relationship between $log(R_T)$ and Quarter offset T. The relationship and the growth implications are given below:

$$log(R_T) = \alpha_1 + \beta_1 T + \epsilon \tag{3}$$

raising both sides to the e power we get

$$R_T = e^{\alpha_1} e^{\beta_1 T} e^{\epsilon}$$

$$\Rightarrow \frac{\partial R_T}{R_T} = \beta_1 \partial T \tag{4}$$

We estimated the parameters through regression equation (3). The resulting coefficient and the corresponding R^2 have been shown in Table 3; both parameter estimates are highly significant and the R^2 is 0.985.

The implication of equation (4) is that there is a constant Quarter–on–Quarter growth rate of β_1. Although the estimated model is consistent with the data (high $R^2 = 0.9846$), it is unrealistic (the growth rate cannot be constant in the long run). Such tight fit arises from two reasons 1. log transformation of R_T 2. serial correlation present in R_T. Even large differences in levels are small differences after log transformations are made and revenue is serially correlated across time. Indeed the revenue growth rate has been steadily decreasing for the past several years. For this reason we estimate this, and all other diffusion models, in growth space below after introducing them in more intuitive level space.

[12] Robustness checks show the models presented here are not very sensitive to the starting point t_0). This is expected to give low initial quarter revenue of AWS sales relative to later sale, despite the log transform.

Logistic Growth Model The Logistic Growth model has historically been used to model complex growth process like population growth [1] but also technology diffusion [2], [17]. The functional form of the model is:

$$R_T = \frac{\theta_1}{1 + e^{-\theta_2 - \theta_3 T}} + \epsilon$$

$$\Rightarrow \frac{\partial R_T}{R_T} = \theta_3 (1 - \frac{R_T}{\theta_1}) \partial T \tag{5}$$

Equation (5) indicates the growth rate reaches 0 as time goes to ∞ and $R_T \to \theta_1$. The parametric assumptions impose the "S–shaped" diffusion associated with technological diffusion. The KS test and the R^2 both indicate good within sample fit.

The estimated accuracy of θ_1 is hence crucial for the model's long run performance. We estimated the model using the non linear regression estimation strategy prescribed by Fox & Weisberg[13]. The parameter θ_1 estimates as shown in Table 3 was 6854.53. Accounting for 95% confidence intervals, the model predicts AWS revenue reaching steady state at about \$24 – \$31 Billion/Yr. Independent Public Cloud market forecasts place the overall market at about \$86 Billion/Yr by 2020 for the infrastructure and platform cloud services[11]. Given AWS's current market share, that translates to \$30 Billion/Yr in revenue[14]. Thus the logistic model gives more conservative out of sample projections due to its parametric assumptions.

Log Log Model The Log-Log model forms a linear relationship as shown in equation (6) between $log(R_T)$ and $log(T)$. The model and its growth implications are given below:

$$log(R_T) = \alpha_2 + \beta_2 log(T) + \epsilon \tag{6}$$

raising both sides to the e power we get

$$R_T = e^{\alpha_2} T^{\beta_2} e^{\epsilon}$$

$$\Rightarrow \frac{\partial R_T}{R_T} = \frac{\beta_2}{T} \partial T \tag{7}$$

Therefore for a unit change in time ∂T, we find that the growth rate decays at the rate of β_2 i.e. $\frac{\beta_2}{T}$ is inversely proportional to time. We estimate[13] this model by running a regression as described in equation (6). The decay rate of the growth,as shown in equation(7), is such that $\lim_{T \to \infty} \frac{\beta_2}{T} \to 0$ and offers a much slower growth decay than the Logistic Growth process described in equation (5). Like the other two models, R^2 (0.9894) of this model indicate good within sample properties.

Growth Models As noted above Log transformation of revenue leads to tight fits in the levels space, we hence estimate the aforementioned models in growth space. One advantage of using growth space models is that growth is more noisy than level changes, which provides value from a model selection standpoint. If the actual revenue generating process is consistent with the model in the level space, we should observe consistency in the (noisier) growth space as well.

Given the lack of data, rather than performing out–of–sample MSE to compare models we use the Kolmogorov Smirnov statistical test (KS going forward). The KS is a non-parametric test that takes as the null hypothesis that a dataset is distributed subject to a distribution specified by the statistician. A

Table 4: Growth estimates and consistency checks

Model	Estimates (Std.Err)*	KS D	p-value
Log Linear $\frac{\partial R_T}{R_T} = \zeta$	$\zeta = 0.125$ (0.0138)	0.583	0.0337
Logistic Growth $\frac{\partial R_T}{R_T} = \xi + \lambda R_T$	$\xi = 0.207$ (0.00293) $\lambda =$ -3.593e-05 (1.199e-05)	0.167	**0.998**
Log Log $\frac{\partial R_T}{R_T} = \frac{\phi}{T}$	$\phi = 3.339$ (0.308)	0.3333	**0.536**

*all estimates are significant at 0.05

high p-value thus means we fail to reject the null hypothesis. All models are parametric, therefore imposing structure with growth implications on the data, which we discussed above.

Table 4 shows the the KS test (high p-value) for the equivalent growth space models. In equivalent cases of the Logistic Growth and Log Log Model, we fail to reject the null hypothesis whereas we reject the null hypothesis in case of the Log Linear Model.

Ambiguity in growth characteristic The Log–Log and Logistic Growth models are hence shown to be consistent with the data. We can rule out the Log Linear model using growth space (Table 4). Economic intuition supports this case as the Log Linear model is unlikely to occur given it requires a sustained growth well into the future.

As seen in Figure 4, two projections of the revenue process from Log–Log and Logistic Growth model show very different future growth paths. Given that we are still in the initial phase of the cloud adoption and that we have not observed an inflection point yet, it is hard to draw a conclusion with high confidence on which one of the two projected states is more likely to occur. Given that the Log Log model is much easier to work with and the primary goal of this analysis is what leads to cloud adoption industry's early stage we opt for it for the remainder of the paper.

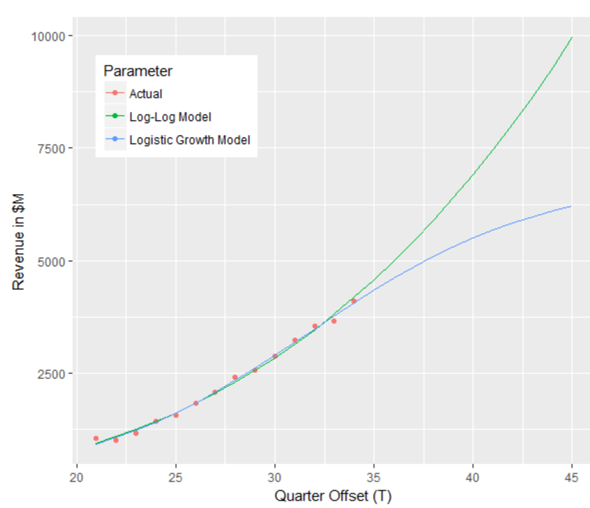


Fig. 4: Future projection of AWS Quarterly Revenue using the Log Log and Logistic Growth Process

[13] Estimate is unbiased (for this case) in spite of Log Transformation of independent variable. See appendix A for more details.

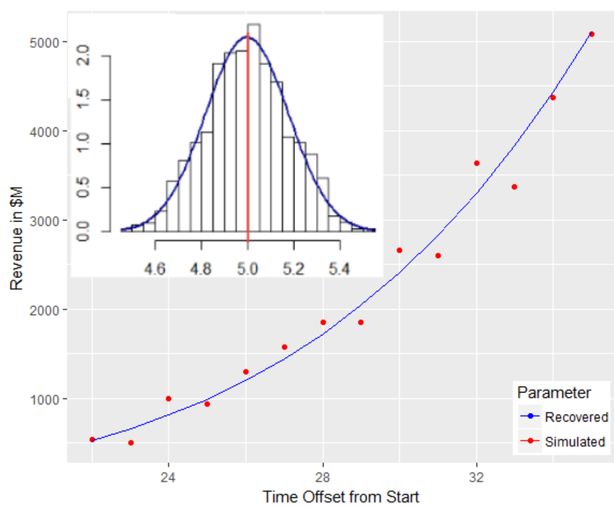
B Simulation to show unbiased estimates of Log Transformation

Under general circumstances one has to be cautious in estimating the growth parameter β_2 using a log transformed regression, as in equation (6), due to possibility of an inherent bias being introduced in the parameter estimates [6]. The estimation strategy suggested by Aaron et. all [6] cannot be applied at $\alpha < 1$ i.e in our case $\beta_2 > 1$. However as noted by Bauke [3] given that there is no uncertainty in the independent variable, as T is in our case, we can use the binning method (in our estimate $b_{i+1} = b_i + 1$) which gives an unbiased estimate. We have performed a simulated parameter retrieval to show that the aforementioned constraints are true in our setting.

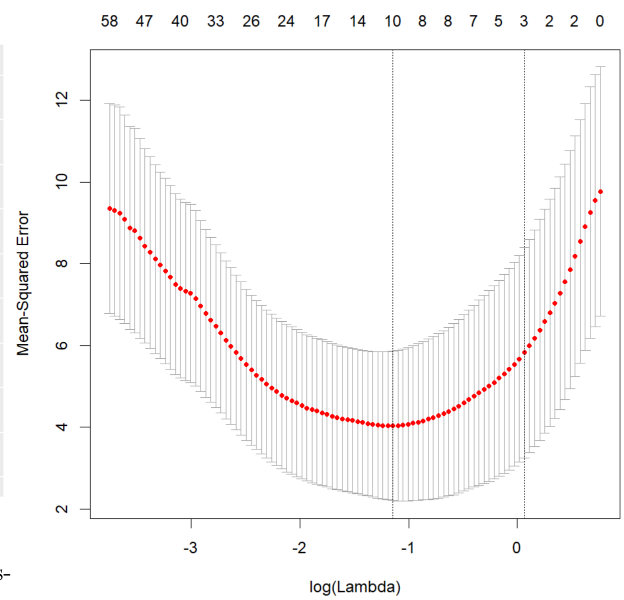
Given a true parameter of $\alpha_r = 100$ and $\beta_r = 5$, and a simulated revenue S_τ (same number of points as observed in the original model) and a discrete time offset τ with the relationship between the two described as follows:

$$S_\tau = e^{\alpha_r} \tau^{\beta_r} e^{\epsilon} \tag{8}$$

Here ϵ is a $N(0, \sigma)$ and we selected $\sigma = 0.1$ as it results in a $R^2 \in (0.95, 0.98)$ similar to original regression results as shown in Table 3. We generated $N = 1000$ processes using equation 8. We then recovered the parameter estimate $\hat{\alpha_r^i}$ and $\hat{\beta_r^i}$ where i is the simulation subscript using the Log Log regression $log(S_\tau) = \alpha_r + \beta_r log(\tau) + \epsilon$. We found that $\sum_{i=1}^{N} \frac{\hat{\beta_r^i}}{N} = \beta_r$ (true for α_r^i). This confirms that the estimation technique is unbiased. Figure 5a shows a specimen fit and the distribution of the $\hat{\beta_r^i}$



(a) Specimen Simulated Draw and Fit - Inset has the distribution of $\hat{\beta_r^i}$ with mean at β_r



(b) Cross Validation Results for High Dimensional Sparse Model Selection

C Combinations of Models

As noted in the section A, there is ambiguity around the choice of a consistent model in projecting the expected revenue out into the future. One solution in such an environment would be to learn a model that trades–off between a set of consistent models. An illustrative model would be as follows:

$$R_T = \lambda e^{\alpha_2} T^{\beta_2} + (1-\lambda)\frac{\theta_1}{1+e^{-\theta_2-\theta_3 T}} + \epsilon$$

$$\text{where } 0 \leq \lambda \leq 1$$

Given that we only posses few (14) observations of revenue, we need to resort to simulated maximum likelihood estimation strategies. Also, while devising such a model with relatively many (6) parameters, we run the risk of over-fitting the data.

D Random Coefficients Model Motivation

Given that we make Normality assumptions about the error term ϵ_{cT} in the Random Coefficients Model, one could conceive a simple Ordinary Least Squares regression that includes the macroeconomic indicators to achieve the same unbiased estimates. In this section, we explain why that notion is insufficient in accurately estimating the coefficients. Consider the aforementioned model that we proposed.

$$log(R_{cT}) = (\bar{\alpha} + u_{0c}) + \left(\bar{\beta}_c + u_{1c}\right) log(T) + \epsilon_{cT} \tag{9}$$

$$\bar{\beta}_c = \gamma_0 + \sum_{k \in \tilde{F}_c} \gamma_k x_{ck} + \epsilon_\beta \tag{10}$$

$\tilde{F}_c$ – non-zero features from Model Selection in country c
Substituting equation 10 in equation 9 yields

$$log(R_{cT}) = (\bar{\alpha} + u_{0c}) + \left(\gamma_0 + \sum_{k \in \tilde{F}_c} \gamma_k x_{ck} + u_{1c}\right) log(T) + \xi_{cT} \tag{11}$$

Naturally $\xi_{cT} = \epsilon_{cT} + \epsilon_\beta log(T)$

There are a few issues with the approach of trying to estimate this combined model. We can only estimate u_{0c} with a country fixed effect. This means we will obtain a model with combined intercept of $\bar{\alpha}$ and u_{0c} with one nation used as a reference nation (full collinearity); the same problem persist with γ_0 and u_{1c}. Such coefficients don't have an independent interpretation as their estimates depend on the reference nation. We cannot estimate ρ_u without making further error assumptions about $\hat{\alpha}$ and $\hat{\beta}$. Additionally, we cannot estimate equation 11 using Ordinary Least Squares, because ξcT is not independent of log(T), this means we need to make additional assumptions about the distribution of error's interaction with $log(T)$. The proposed model is a solution to these very issues and produces an accurate estimate ($R^2_{\beta_c}$) for $\bar{\beta}_c$.