Revenue Forecasting in Technological Services: Evidence from Large Data Centers

Jyotirmoy Sarkar[a], Bidisha Goswami[b], Saibal Kar[c], Snehanshu Saha[b,1]

[a] *Wipro GE Healthcare,Bangalore, India*
[b] *Department of Computer Science and Engineering,PESIT-BSC, Bangalore, India*
[c] *Department of Economics, Calcutta University, India and IZA, Bonn*

Abstract

The global dependence on data centers has grown phenomenally in recent times. The demand for storage and maintenance of data is not limited to facilitators of information technology only, but has spread to all forms of businesses and service providers, private, public and individual. The economic scope and performance of the data centers, however, seem little discussed in the related literature. The rising cost of power supply, the crunch in storage space owing to high property prices, the difficulty of acquiring land for industrial use in various countries, etc., translate into important adjustment costs for data centers. Since the growth of business and competition leads to lower per unit prices, the rising costs offer considerable difficulty in arriving at the optimal revenue for large data centers. The results mainly show that for constant elasticity of scale production functions, the revenue and profit are maximized at low levels of elasticity. The firms can still cope with rising cost because the market for data center operations is fairly concentrated. We utilize the Constant Elasticity of Scale functions to derive conditions for cost minimization and revenue forecasting for large data centers. Further, this paper offers factor analysis in order to identify the precise contribution of each factor input in the overall cost function. The operational management in large data centers has important outcomes in view of considerable externality associated with it.

Keywords: Revenue management, Data centers, CES production

---

[1]Corresponding Author
Email Addresses:jyotirmoy208@gmail.com(Jyotirmoy Sarkar),
bidishagoswami@pes.edu(Bidisha Goswami),
snehanshusaha@pes.edu(Snehanshu Saha),
skar1801@yahoo.com(Saibal Kar)

functions, Concavity, Asia-Pacific.

---

1. Introduction

A data center is a facility, which houses a large number of computing equipments like the servers, the routers, the switches and the firewalls. Supporting components like the air conditioning, the backup equipments, and the fire suppression tools are also indispensable for activities in data centers. According to the broad classification, a data center can be 'complex' if it requires a dedicated building, or it can be 'simple' if it requires a smaller space, say a room, with fewer servers. Further, the facility may be shared by multiple organizations (shared data centers), or it may be owned by a single organization (private data center). It is well-known that following the emergence of cloud computing, data center services have become increasingly popular. The cloud computing service defined as a pool of computing resources cater computing functionality mainly as utility services. For example, companies like Microsoft, Google, Amazon, and IBM, that constitute the leading body of information technology (henceforth, IT) sector engaged with cloud computing, also invest significantly in data distribution and computational hosting services [2].

Indeed, in the recent years, a lot of investments have been made in data centers in order to support cloud computing by large organizations. However, casual empiricism suggests that this industry usually harbors a large number of firms and therefore deviations from least cost combinations of inputs owing to exogenous shocks could be potentially disastrous for many companies affecting the scale of operations. In many cases, cost reducing innovations and potential for flexibility are rather important, but these are often outcomes of sustained and costly research and development activities at the firm level. The larger firms are more likely to engage in such activities. The issue is particularly compelling since during the last decade the cost of servers, power and general maintenance of data centers rose drastically [20]. According to a survey by the Gartner Group, the energy consumptions accounts for up to 10% of a data center's operational expenses (henceforth, OPEX) and this may surge up to 50% in near future [3]. The power consumed by the computing system dissipates as heat and this is responsible for up to 70% of the total heat generated from the infrastructure of a typical data center [5]. Needless to mention, a powerful cooling system is therefore required for mitigating the amount of heat generated. The cost of the cooling system may range between \$2 to \$5 million per year for a conventional data center [4]. It is obvious that the failure to keep temperatures within technologically accepted limits may disrupt cloud services, which in turn may result in violation of Service Level Agreements(SLA).

Given this brief introduction, it may be useful to investigate the cost components and estimate the implications for operational reorganization within data centers in view of the sustenance of such facilities. This constitutes our main research question. Primarily, we use Hamilton to lend a general structure to the components of cost[1]. In this context, let us consider a data center which houses 50,000 servers and built with state-of-the-art techniques and equipments. Table 1 provides the major cost segments associated with such a data center. We assume that the costs are amortized, i.e., a one-time investment is allocated over a reasonable time-frame, with the opportunity cost of investment held at 5%. With this simple framework in

Table 1: Cost segmentation in Datacenter

Amortized Cost	Component	Sub-component
~45%	Servers	CPU, memory, storage systems
~25%	Infrastructures	Power distribution and cooling
15%	Power draw	Electrical utility cost
15 %	Network	Links, transit, equipment

mind, and given the possibility that firms could realign their energy-budget without sacrificing SLA we re-frame the purpose of this research as follows. *What is the optimal operational (price and size-wise) strategy for a firm engaged with large data centers and what is the most appropriate model of production, which would optimize revenue in the face of steadily escalating costs of equipments and maintenance?*

Notably, a large amount of data is generated every day due to business operations in the form of emails, messages, transactions, videos, etc. An organization needs to store the unstructured data somewhere in order to process it, such that valuable information can be extracted from this huge stock as and when necessary. Using this information, enterprises may be able to make important decisions regarding output, costs and generally profitability and subsequently forecast growth. It is easy to recognize that traditional infrastructure is not suitable for processing unstructured data. Hence scalable, cost effective cloud computing comes to the picture as a support for this data requirement. Estimates suggest that about 2.5 billion gigabytes of data is being created on a daily basis, which comprises of 200 million tweets and 30 billion pieces of contents shared on Facebook over a month alone, as part of all other activities. It has been projected that the amount of data will reach 43 trillion gigabytes, given that approximately 6 billion people shall possess cell phones in another 5 to 7 years[23]. Apparently, cloud computing and big data, both offer new paradigms and processes of the rapidly evolving technology. The speed with which companies adopt these is mainly driven by the fact that cloud computing in particular can be used as an utility service for big data analysis. The cost effectiveness of such actions needs a much better reconnaissance than is available in the literature. The present paper wishes to fill this gap.

Now, the reason behind choosing revenue optimization is straightforward. Every firm faces a set of fundamental questions related to the sales format, sales price, and sales volume of every product or service offered. In particular, for firms operating in competitive environment, the best answers to these questions lead to decisions that maximize revenue. The firms usually face a time constraint for addressing these questions, and that influences the optimal decision. The remainder of the paper is organized as follows. Section II discusses the related work from the available literature. In Section III, revenue optimization for large data centers is discussed. This section explains the mathematical foundation of the proposed model based on CES production function. Section IV discusses the results of experimentation with the CES model and highlights the possible implementation of the proposed model for different data sets. Section V provides detailed discussions about the concentration of firms in cloud computing services, thereby offering an insight into the competitiveness of this market. Section VI offers a factor analysis of the cost components and Section VII concludes.

2. RELATED WORK

The revenue forecasting models accommodate many allied considerations, including substantial uncertainty regarding the respective fundamentals at the firm level [12]. However, the importance of revenue forecasting for decision making within a given business is indisputable [13]. Such forecasting is associated with sales order recognition, which may be different from the operating revenue at a point in time. Indeed, the latter should be better treated as a predictor. Consequently, a number of approaches have developed, such that the possible deviations associated with the predictions and the observed values are minimized. To this end, the use of asymmetric loss function can isolate the forecast rationality and the costs associated with under-forecasting [14], [15]. With regard to data centers, James Hamilton [6] has earlier shown that power is not the largest component of cost.

This is true if the amortization cost of power, that of cooling infrastructure over 15 years and of new servers over 3 years are considered. This paper holds amortization at 5% per annum, and argues that the server hardware costs appear to be the largest. Nonetheless, it is quite possible that more efficient technology adopted in server and related equipments may help to lower the cost considerably as compared to power generation, which faces high demand from several other competing sectors of the economy and is exposed to various environmental constraints regarding sources and types.

Generally, a typical data center comprises of 100 fully loaded racks with the current generation 1U servers requiring $1.2 million for power and an additional $1.2 million for cooling infrastructure, per annum. Moreover, $1.8 million in cost is incurred for maintenance as well as amortization of power and cooling equipments [9]. Thus, power is the most significant variable cost of the data center while server hardware accounts for the biggest chunk of the total fixed cost. In a related context, Ghamkhari (2013) [21] investigates the trade-off between minimizing data centers energy expenditure and maximizing their revenue for various Internet and cloud computing services that they may offer. This paper proposes a novel profit maximization strategy for data centers for two different cases, with and without "behind-the-meter" renewable generators. In general, several issues in revenue management at the firm level has been discussed earlier (see, Chiang, Chen and Xu, 2007) [28]. In terms of theoretical advancement, Meissner and Strauss (2012) approximates the Markov value chain with a non-linear function and sets upper bounds on expected optimal revenue [29]. Further, Kemmer, Strauss and Winter (2012) break up the demand management problem into resource-level sub-problems and solve them simultaneously by generating dynamic marginal capacity value estimates [30]. Essentially, the management of large data centers and various important sub-components within it lends itself to similar operations management problem, which we investigate presently. We argue that the revenue management in large data centers display idiosyncratic features with veritable implications for economic activities in general. Several other methods and applications (viz. Weatherford and Kimes, 2003 discuss pick-up methods and regression reported lowest error in case of demand for hotels) [31] which have enriched the econometric and OR-based modeling of forecasts, but the outreach of the subject is certainly in need of more compelling estimates. Indeed, Fildes et al. (2007) clearly highlight the contributions of forecasting in OR, generally speaking, and discusses various applications that enrich the scope of empirical analysis [32]. Our paper is a contribution in this vein.

In view of the problem at hand, Fan *et al.* (2007) [8] considers the characteristics of aggregate power usage of large collections of servers (up to 15 thousand) for different classes of applications over a period of approximately six months. The study concludes that the opportunities for power and energy savings are significant, but the benefits are greater at the cluster-level (thousands of servers) than at the rack-level (tens). Puschel *et al.* (2016) [24] have adopted a new technique, namely, the policy-based service admission control model, to optimize revenue of cloud providers while taking informational uncertainty regarding resource requirements into account. Notwithstanding, Bodenstein *et al.* have proposed an energy efficient and intelligent system allocation mechanism to reduce power cost by 40%[27]. Arguably, there are many more complications associated with such practices. Thus, Gurumurthi *et al.*[7] discuss an optimal trade-off between energy efficiency and service performance over a set of distributed Internet data centers with dynamic demand. The review suggests that reducing variable costs is particularly difficult for computing services as discussed in some of the previous studies. Therefore, like most production plans, other strategic adjustments need to be considered in order to maximize revenue.

3. REVENUE OPTIMIZATION IN DATA CENTERS

In the extant literature, convex optimization principle has been used in the recent past to solve fundamental problems in science [25], [26]. Notwithstanding, optimization associated with techniques related to performance in data centers and the problem of cost minimization remains an important issue. We are aware of the cross-effects of reducing the use of one factor vis-a-vis another and apply the CES production function to estimate the impact of reducing the cost of the contributing factors on revenue maximization at the data centers. It is well-known that CES belongs to the family of neoclassical production functions (see [25], [26]). The CES production function for two inputs can be represented in the form below:

$$Q(L, K) = (\alpha L^{\rho} + (1-\alpha)K^{\rho})^{1/\rho} \tag{1}$$

where
Q= Quantity of output
L, K =Labor and capital, respectively
$\rho = \frac{s-1}{s}$
$s = \frac{1}{1-\rho}$, Elasticity of substitution
and α = Share parameter

[2]

3.1. *The Analytical Structure*

Consider an enterprise that has to choose its input bundle (S, I, P, N) where S, I, P and N are the number of servers, the investment in infrastructure, the cost of power and the networking cost of a cloud data center, respectively. We determine the (global) cost minimizing and profit maximizing input bundles for such a production outlay. The enterprise wants to maximize its production, subject to the cost constraint.

The CES function is written as:

$$f(S, I, P, N) = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} \tag{2}$$

Let m be the upper bound of cost of the inputs.

$$w_1 S + w_2 I + w_3 P + w_4 N = m \tag{3}$$

w_1: Unit cost of servers
w_2: Unit cost of infrastructure
w_3: Unit cost of power
w_4: Unit cost of network

The Optimization problem for production maximization is peceived as:
max f(S, I, P, N) subject to m
The following values of S, I, P and N thus obtained are the values for which the data center has maximum production of satisfying the given constraints

---

[2]Further, the elasticity of substitution is constant for the CES production function. More specifically, the elasticity of substitution measures the percentage change in the factor ratio divided by the percentage change in the technical rate of substitution, while holding output fixed.

Case I: Linear production function:
When we set $\rho = 1$, the production function becomes: $y = K + L$, where the two inputs, capital and labor are perfect substitutes.

Case 2: Cobb Douglas production function:
When ρ tends to 0, i.e. $\lim_{\rho \to 0} y$, the isoquants of the CES production function look very much like those of the CobbDouglas production function. This can be shown in a variety of different ways, but the easiest is to compute the technical rate of substitution. As such, the two inputs in this case are imperfect substitutes, leading to standard isoquants.

Case 3: Leontieff Production function:
When ρ tends to $-\infty$, i.e. $\lim_{\rho \to -\infty} y$, the isoquants are L shaped, which we associate with the perfect complements case for inputs.

on the total investment.

$$S^* = \frac{m w_1{}^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{4}$$

$$I^* = \frac{m w_2{}^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{5}$$

$$N^* = \frac{m w_3{}^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{6}$$

$$P^* = \frac{m w_4{}^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{7}$$

These results are proved in Appendix A.

3.2. Cost Minimization

Consider an enterprise that sets a target level of output by investing a minimum amount. The CES function is of the form:

$$y_{tar} = f(S, I, P, N) = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} \tag{8}$$

y_{tar} is the target output of the firm that needs to be achieved and w_1, w_2, w_3 and w_4 are unit prices of servers, infrastructure, power and network respectively. The cost minimization problem may be written as:

$$\min_{S,I,P,N} w_1 S + w_2 I + w_3 P + w_4 N \textit{ subject to } y_{tar} \tag{9}$$

The cost for producing y_{tar} units in cheapest way is c, where

$$c = w_1 S + w_2 I + w_3 P + w_4 N \tag{10}$$

where c can be written as:

$$c = \left(\frac{y_{tar}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}}\right)^{\frac{1}{\rho}-1} \tag{11}$$

The results, (10)-(11) are proved in Appendix B.

3.2.1. Global Minima for Cost minimization: A heuristic approach

We use the Gradient Descent method to retrieve the values of elasticity ensuring cost minimization. For simplification of equations, let us consider two cost segments X and Y. w_1 and w_2 are unit prices of X and Y. Rewriting the cost function using the newly selected variables, we obtain

$$c = w_1 X + w_2 Y \tag{12}$$

The CES function thus formed is,

$$\begin{aligned} y_{tar} &= (X^\rho + Y^\rho)^{\frac{1}{\rho}} \\ y_{tar}^\rho &= X^\rho + Y^\rho \\ X^\rho &= y_{tar}^\rho - Y^\rho \\ X &= (y_{tar}^\rho - Y^\rho)^{\frac{1}{\rho}} \end{aligned}$$

where y_{tar} is the optimal output level.Substituting the value of X in the cost function equation i.e.(12), we obtain

$$c = w_1 \left(y_{tar}^\rho - Y^\rho\right)^{\frac{1}{\rho}} + w_2 Y$$

Therefore, differentiating with respect to the elasticity of substitution, we get

$$\frac{\partial c}{\partial \rho} = \frac{-w_1 \left(y_{tar}^\rho - Y^\rho\right)^{\frac{1}{\rho}}}{\rho^2} ln \left(y_{tar}^\rho - Y^\rho\right) \\ \left(y_{tar}^\rho ln y_{tar} - Y^\rho ln Y\right)$$

The partial derivative is used in gradient descent method for cost minimization.

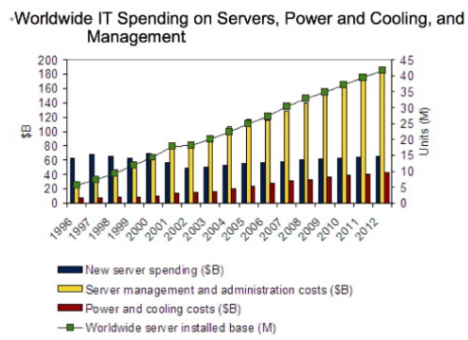
Gradient Descent Algorithm:

1. **procedure** GRADIENTDESCENT()
2. $\frac{\partial c}{\partial \rho} = \frac{-w_1 \left(y_{tar}^\rho - Y^\rho\right)^{\frac{1}{\rho}}}{\rho^2} ln \left(y_{tar}^\rho - Y^\rho\right)$ $\left(y_{tar}^\rho ln y_{tar} - Y^\rho ln Y\right)$
3. **repeat**
4. $\rho_{n+1} \leftarrow \rho_n - \delta \frac{\partial c}{\partial \rho}$
5. $\rho_n \leftarrow \rho_{n+1}$
6. **until** $(\rho_{n+1} > 0)$
7. **end procedure**

Using the above algorithm, the optimal values of α and cost have been computed(cf. Section 4). The dual is of course, profit maximization as we show below.

4. Computation of Revenue and Profit from data

As mentioned earlier, the server and power/cooling costs form the biggest chunk of the total cost. These two inputs are considered for computing the values of the elasticity using 3D plots. Revenue maximization is demonstrated graphically. All simulation results have been generated by Matlab.



The above figure captures two types of costs, namely, administrative cost and input (power/cooling) cost. The optimal elasticity of each input and maximum revenue for each year using Matlab code are obtained.

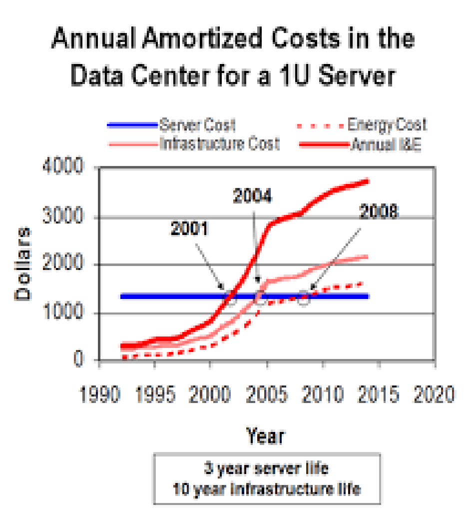


Figure 1: Graph representing the annual amortized costs for a 1U size server

Fig. 1 is the graphical representation of Annual Amortization Costs in data centers for 1U server [16]. All units are in $. The data is fairly accurate and

represented in a tabular format on a yearly basis. Maximum revenue and optimal elasticity constants are demonstrated in the same table, available in Additional file [22].
The experiment has been conducted for three scenarios for Revenue.
1. $\rho < 1$
2. $\rho = 1$
3. $\rho > 1$

4.1. Case 1: $\rho < 1$

Applying the constraints $\rho < 1$ and $\rho > 0$ to the function, $f = (x^{\rho} + y^{\rho})^{\frac{1}{\rho}}$ and using fmincon function of matlab(elaborated in footnote), the values of elasticity for which revenue is maximized for each year are obtained. [3] In table 2, all units are in \$B. The optimal revenue across the years is obtained at ρ= 0.100 and s=1.11. In Fig. 2a, the X axis represents ρ and the Y axis represents Revenue. It is quite clear from the graph that the

---

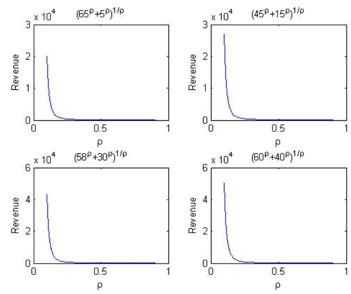
[3] ***Matlab Code for Fmincon***

The matlab fmincon code when $\rho < 1$:

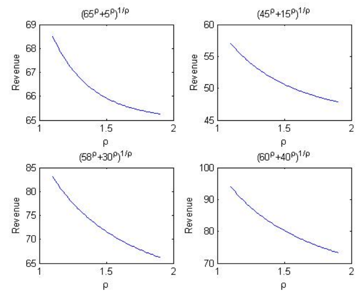
$$A = [1; -1];$$
$$b = [0.9; -0.1];$$
$$x0 = [0.4];$$
$$[x, fval] = fmincon(@myCES, x0, A, b)$$
$$function f = myCES(x)$$
$$pow = 1/x(1);$$
$$f = -(62^x(1) + 5^x(1)).^pow;$$
$$end$$

The matlab fmincon code when $\rho > 1$:

$$A = [1; -1];$$
$$b = [1.9; -1.1];$$
$$x0 = [0.4];$$
$$[x, fval] = fmincon(@myCES, x0, A, b)$$
$$function f = myCES(x)$$
$$pow = 1/x(1);$$
$$f = -(62^x(1) + 5^x(1)).^pow;$$
$$end$$

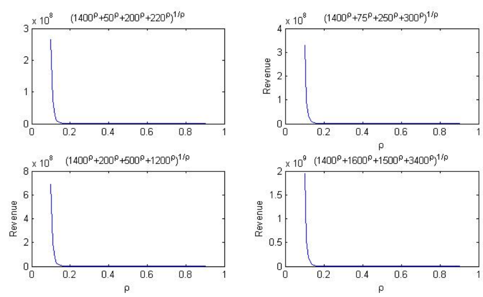


(a) Revenue over the years when $\rho < 1$

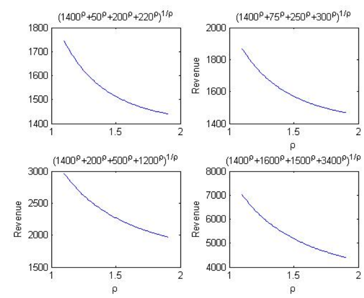


(b) Revenue over the years, when $\rho > 1$

Figure 2: Revenue against years



(a) Revenue against years when $\rho < 1$



(b) Revenue against years, when $\rho > 1$

Figure 3: Revenue against years

maximum revenue is achieved when the value of ρ is 0.100. The graphs display the same optimal values as represented in table 2. All the graphs depict a common pattern, where the revenue falls sharply beyond $\rho > 0.100$. Another pattern, which is observed throughout the 4 graphs is that there is no major fluctuation in revenue between $0.100 < \rho < 0.9$.

Applying the same constraints on the dataset of Fig 1 and using fmincon function of matlab [discussed in footnote3], optimal elasticity for maximum revenue is computed. As observed in table 5 of the additional file[22], optimal revenues have been achieved at ρ=0.1 and s= 1.1.

Cost data for the years 1992, 1995, 2000 and 2005 have been used for generating the graphs Fig. 3a. The graphs do not display any features associated with concavity or convexity. The common pattern, that the revenue falls sharply after ρ, has already been observed for the previous data set. The maximum revenue is attained in the region which is closer to ρ= 0.1.

4.2. Case 2: $\rho = 1$

This sub-section deliberates on the effect of CES function on revenue generation in data centers when $\rho = 1$. As there is no variation in ρ, contrary to other cases, the application of 'fmincon' function is redundant. We will explore revenue optimization and the pattern of revenue generation when $\rho = 1$. The scenario exhibits a linear production function, where the response to revenue is the sum of the factor inputs. We apply this on the dataset (table 3, ADDITIONAL FILE [22]) that captures IT spending globally. We observe slightly higher revenue, after comparing with the case of $\rho > 1$. The difference in revenue between these two cases stood at 1.5 B USD in 1996, while the difference rises to almost 6 B USD in 2012. However, in the case of the Annual Amortization Dataset, the CES function behaves linearly such that the revenue is the sum of all factor inputs. Subsequently, a singular ρ value is applied to the data set and revenues have been calculated. The revenue generated is lower than the previous case, but higher than the case, where $\rho > 1$. The revenue rises almost 4 times between 1992 and 2010 (Table 6 of the additional file [22]).

4.3. Case 3: $\rho > 1$

Applying the constraints $\rho < 2$ and $\rho > 1$ to the function $f = (x^\rho + y^\rho)^{\frac{1}{\rho}}$ and using fmincon function of Matlab [Appendix H], the values of elasticity for maximum year-wise revenue are obtained (Table 4 of the additional file [22], all units are in \$B). The optimal revenue year-wise are obtained at ρ = 1.1 and s= -10. Fig.3 represents the discussion visually.

In Fig. 2b, X axis represents ρ and the Y axis represents revenue. Maximum revenue lies in the neighborhood of ρ = 1.1. Apart from the first figure, there is no sharp fall in revenue beyond ρ =1.1. The graphs are predominantly concave down. Identical constraints are applied on the annual amortization dataset and fmincon function of Matlab [Appendix H] has been used to obtain optimal elasticity.

It is observed that the revenue rises 4 folds between 1992 and 2010 (table 7 of additional file [22]). The maximum revenue is attained at ρ=1.1. Using the above data set, 2D simulation graphs are produced. Fig. 3b represents the graphs produced using the cost data of 1992, 1995, 2000 and 2005. We observe two distinctly visible trends in the above graphs, across the data sets whenever $\rho > 1$.

- the concave decreasing trend
- maximum revenue is attained at $\rho = 1.1$

Gradient descent method has been applied to worldwide IT spending data set to compute optimal elasticity for cost minimization. The initial value of ρ has been assumed to be 1.2 whereas step size for each iteration has been set to 0.001 (Table 8 of additional file [22]). We assumed target revenue as \$240B and unit cost of new server installation as 0.6B for the ascent algorithm. Subsequently, we explore the behavior of profits using the proposed model. Two possibilities for better comprehension of the behavior of the profit function are considered. The worldwide IT spending data set is considered for profit analysis. The data set consists of two cost components, namely, New Server installation cost and Power & Cooling cost. The weights of these cost segments are incorporated as 0.3 and 0.4 respectively.

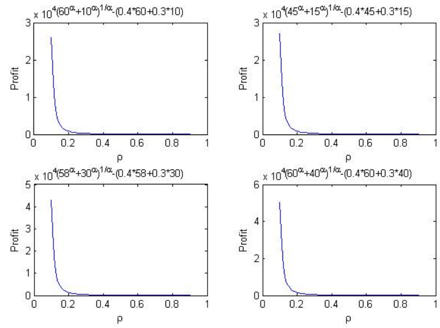
4.4. Case 1: $\rho < 1$

The constraints $\rho < 1$ & $\rho > 0$ are applied to the function $f = (x^\rho + y^\rho)^{\frac{1}{\rho}}$. The Fmincon function of Matlab is used to converge to the optimal ρ value for the maximum profit to be obtained. New server cost and Power & Cooling cost derived from world wide IT spending data along with optimal profit have been displayed in table 9 of additional file [22]. The table shows that maximum profit has been obtained at $\rho = 0.1$, similar to the scenario for maximum revenue. Cost data for the years 1999, 2002, 2009 and 2012 have been used to generate the graph (Fig. 4a). X and Y axes represent Profit and ρ, respectively. The graphs do not exhibit any concavity or convexity. The common pattern of revenue and profit falling sharply after $\rho = 0.1$ is observed for revenue graphs. Maximum profit is attained in the neighborhood of $\rho = 0.1$. The graphs of Profit vs. years are demonstrated along with 'Profit Vs ρ' graphs. This is done to faciliate deeper insight into the results. The profits from year 1999 to 2012 have been depicted in Fig. 4b. Profits rise between 1999 and 2001, while falling for the year 2001, captured by a sudden dip in the graph. Profits rise sharply, post 2001 till 2012 and the biggest jump in profit is observed between 2006 and 2008.

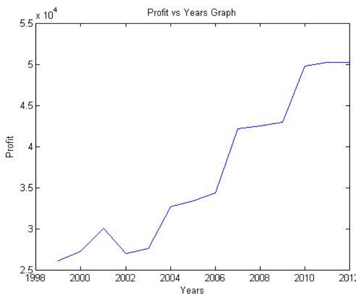
4.5. Case 2: $\rho > 1$

Next, we apply the constraints $\rho > 1$ & $\rho < 2$ to the function $f = (x^\rho + y^\rho)^{\frac{1}{\rho}}$.

Similar to the earlier case, Fmincon function is utilized to determine optimal ρ. Maximum profit, optimal ρ and the two cost segments are shown in Table 10 of additional file [22]. The graphs (Fig. 5a) made use of cost data for the years 1999, 2002, 2009, and 2012. These graphs show common trend of being concave down. Maximum revenue is attained at $\rho = 1.1$.

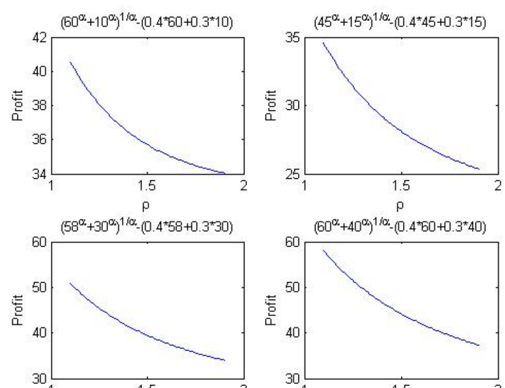


(a) Profit of four different year against ρ, when $\rho < 1$

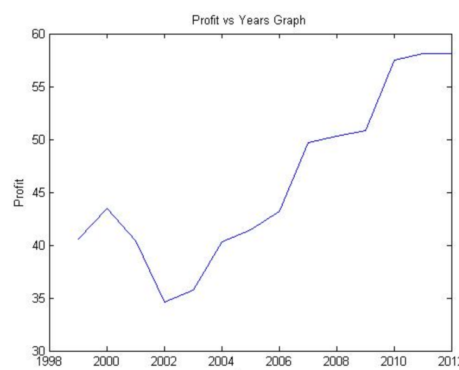


(b) Profit against years when $\rho < 1$

Figure 4: Profit against years



(a) profit of four different year against ρ, when $\rho > 1$



(b) Profit against years, when $\rho > 1$

Figure 5: Profit against years

The levels of profit between 1999 and 2012 have been depicted in Fig. 5b. In fact, the level of profit since 2000 declines significantly to reach the lowest in 2002. It rises after 2002 secularly, and reaches the highest level in 2012. The 2X2 design space is defined over years and level of profit, respectively.

5. Concentration of firms in Data Center Industry

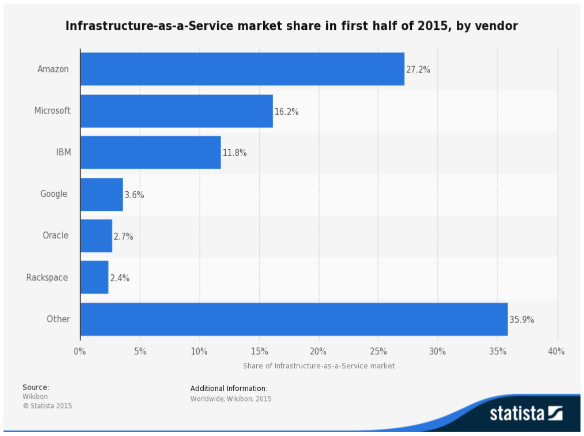
As seen in the previous section, the level of profit rises in all the cases despite increase in cost. This is generally perplexing, unless some changes in the market share is responsible for such outcomes. Therefore, it is natural to investigate the level of concentration of firms that belong to the data center facilities. It is well known that for setting up a data center, a firm needs huge amount of initial investments, which favors relatively large organizations to maintain their own data centers. For example, bigger firms such as Amazon, Google, Microsoft, etc., have constructed massive computing infrastructure to support their websites and related business services. These organizations have also started to rent out their infrastructure to developers and small firms, who do not have their own data centers. But, does that necessarily imply that firms in this sector are getting more and more concentrated and therefore account for monopoly profit? We employ Hirschman-Herfindahl Index (HHI) to measure the level of competition or concentration of such firms.

The Hirschman-Herfindahl Index (HHI) is a widely used technique for measuring the degree of market concentration. It is calculated by summing the squared market share of each firm competing in a given market. The value of HHI can vary between approximately zero to 10,000. The HHI is expressed as:

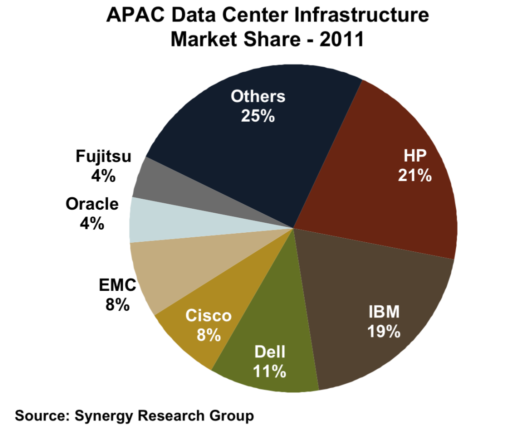
$$HHI = s_1^2 + s_2^2 + s_3^2 + ... + s_n^2 \tag{13}$$

Here s_n is the market share of the ith firm.

High HHI implies a few firms control the business. Thus, new cost outlay compensates the firms for such investments more than proportionately. This raises revenue and profit. Conversely, if the market is shared by a large number of firms, the HHI value shall be low and additional cost outlay could be disastrous. Interestingly, for a global distribution of technology firms, the spatial variations in concentration and profitability could serve wider interest. To begin with, let us discuss the degree of market concentration



(a) Infrastructure-as-a-Service Market share in 2015 first half



(b) Asia pacific region Data Center Market share in 2011

Figure 6: Data Center Market Share

for firms present in the Asia-Pacific region. The region has generated just over USD 20 billion in data center infrastructure revenues for the worlds leading technology vendors and the market and has grown by 23% from the previous year, according to data from Synergy Research Group [18].

$$HHI = 21^2 + 19^2 + 11^2 + 8^2 + 8^2 + 4^2 + 4^2 + 25^2 = 1708$$

Since, this also captures a legal dimension, the U.S. Department of Justice considers a market to be competitive if it shows a HHI score less than 1000. Conversely, if the score is between 1000 and 1800, the market is deemed as moderately concentrated, while a score above 1800 suggests a highly concentrated marketplace [17]. In the Asian-Pacific Economic Cooperation (APEC) zone, the concentration is moderate and tending towards a highly concentrated marketplace. Next, we try to find the level of firm concentration in Infrastructure as a Service (IaaS). The IaaS market share data is collected from Business Insider (2016) [19]. The HHI for IaaS market share is given below

$$\begin{aligned} HHI &= 27.2^2 + 16.6^2 + 11.8^2 + 3.6^2 + 2.7^2 + 2.4^2 + 35.9^2 \\ &= 2456.34 \end{aligned}$$

If we exclude 'others' (rest of the firms apart from Amazon, IBM, Oracle, Google and RackSpace highlighted in fig 6a) from HHI calculation, it becomes 1167.53. And yet, it cannot be considered as a competitive market. Overall, only a few firms seem to control the major share of the market for infrastructure as a service.

6. Factor Analysis for Data Centers using CES

The CES production function can accommodate any number of factors as it can be expanded to 'n' number of input variables. In our discussion, we have thus far considered two factor inputs and extended up to four inputs generating the optimal revenue. In this section, we shall discuss how these factors and their interactions contribute towards revenue optimization. Indeed, we hypothesize that the results of the factor analysis shall suggest that some factors have low output elasticity while others are more productive. For ease of calculation, we have considered two cost segments, namely, the New Server Cost, and Power & Cooling Cost, as factor inputs. We need to identify the high and low points of each factor. It may be useful to point out, subsequently we considered two data sets with 2 and 4 factors but here the worldwide IT spending data set is used for obtaining a set of estimates.

6.1. Factor Analysis for Proposed Model

The objective of the factor analysis is to identify the significant factors, in particular the various costs associated with data centers in order to attain maximum revenue. The contribution of each factor helps to determine how factors impact the optimization problem. It identifies the contribution (in percentage terms) of each factor, and points out the scope of intervention.

6.1.1. The 3^2 Design

The factors, which have been considered for 3 level factorial design are listed below:

- New Server installation cost
- Power and Cooling

Instead of pointing out a single value, we have considered a range for high, medium and low levels for each factor.

In both the tables, all units are in$ billion. Let us define two variables x_A and x_B as representative of the New Server cost and Power & Cooling cost. The mapping of high, medium and low levels of each factors are

Table 11: Factor 1 level

Level	Range
Low	5-15
Medium	16-29
High	30-40

Table 12: Factor 3 Level

Level	Range
Low	45-52
Medium	53-59
High	60-65

demonstrated in table 13.

Table 13: Factor Initialization

Factor	High	Medium	Low
x_A	2	1	0
x_B	2	1	0

The Revenue y can now be formulated on x_A and x_B using a nonlinear regression model of the form:

$$y = q_0 + q_1 x_A + q_2 x_B + q_{12} x_{AB} + q_{11} x_A^2 + q_{22} x_B^2 \tag{14}$$

The effect of each factor is calculated as the proportion of the total variation in the output, as contributed by the factor. The sum of contribution (SC) can be represented by equation (20):

$$SC = q_1 + q_2 + q_{11} + q_{12} + q_{22} \tag{15}$$

Where: $q_1 + q_{11}$ is the portion of SC that is contributed by New Server cost. $q_2 + q_{22}$ is the portion of SC that is contributed by Power & Cooling cost. q_{12} is the portion of SC that is contributed by the interactions of New Server cost and Power & Cooling cost.

$$SC = SCA + SCB + SCB \tag{16}$$

Fraction of variation contributed by $A = SCA/SC$
Fraction of variation contributed by $B = SCB/SC$

Fraction of variation contributed by $AB = SCAB/SC$
The objective behind applying this methodology is to identify the major cost factors to be accommodated in the proposed CES model.

6.2. Experimental Observations

6.2.1. Experiment 1 : With two factors $\rho < 1$

Table 14: Experiment 1: With two factors $\rho < 1$

	Power and Cooling		
Server	Low	medium	High
Low	287322.14	33062.56	
Medium	30038.95	34399.27	44381.5875
High	23902.358		50268.74

After solving all the equations, the regression relation stands as:

$$\begin{aligned} y = 27322.14 + 7143.511x_A - 4462.551x_B \\ -1380.1x_Ax_B - 4426.701x_A^2 + 10202.971x_B^2 \end{aligned} \tag{17}$$

Substituting the parameter in equation (14), we obtain
$SC = SCA + SCB + SCB$
=7143.51 + 4462.551 + 1380.1 + 4426.701 + 10202.971
=27615.834
The above analysis suggests the following. The contribution of New Server cost to revenue is 41.89%. The contribution of Power & Cooling cost to revenue is 53.10% and the contribution by the interaction of New Server cost and Power & Cooling is 4.99%.

6.2.2. Experiment 2 : With two factors $\rho = 1$

Table 15: Experiment 2: With two factors $\rho = 1$

	Power and Cooling		
Server	Low	Medium	High
Low	61	71	
Medium	70	75	90
High	70.8		100

After solving all the equations, the regression relation appears as:

$$y = 61 + 13.1x_A - 4.6x_B - 5x_Ax_B - 4.6x_A^2 + 14.6x_B^2 \tag{18}$$

Substituting the parameter in equation (14), we obtain
$SC = q_1 + q_2 + q_{11} + q_{12} + q_{22}$
$=4.1 + 13.1 + 5 + 14.6 + 4.6$
$=41.4$

Following this, we can interpret the effect of each factor as follows: The contribution of New Server cost to revenue is 41.54%. The contribution of Power & Cooling cost to revenue is 46.37%. Finally, the contribution by the interaction between New Server cost and Power & Cooling cost is 12.07%.

6.2.3. Experiment 3 : With two factors $\rho > 1$

Table 16: Experiment 3: With two factors$\rho > 1$

	Power and Cooling		
Server	Low	Medium	High
Low	58.0444	67.3181	
Medium	66.8546	71.2122	84.8461
High	68.72266		94.0819

After solving all the equations, the regression equation is:

$$\begin{aligned} y = 58.0444 + 12.2814x_A + 19.17105x_B - 4.9161x_Ax_B \\ -3.4712x_A^2 - 9.89735x_B^2 \end{aligned} \tag{19}$$

Substituting the parameter in equation (14), we obtain
$SC = q_1 + q_2 + q_{11} + q_{12} + q_{22}$
$=12.2814 + 19.17105 + 4.9161 + 3.4712 + 9.89735$
$=49.7371$
We may interpret the effect of each factor as follows: The contribution of New Server cost to revenue is 31.67%. The contribution of Power & Cooling cost to revenue is 58.44% and the contribution by the interaction between New Server cost and Power & Cooling cost is 9.8841%. The interaction term is too small to be neglected.

7. Discussion and Conclusion

Revenue optimization in data centers around the world is not a well researched topic, despite the fact that the outreach of information technology into daily activities has been enormous. The inability of large and small service providers to sustain under escalating costs of equipments, power usage, etc should not only imply potential disruption for the firms alone, but shall evidently jeopardize almost all other activities globally. A detailed analysis of the cost components and their role in revenue optimization for technology service-providing and server maintaining firms is therefore imperative and timely. This paper identified the major cost components influencing optimal revenue generation. Assuming CES production function for such firms, it was established that firms register maximum profit closer to the point where the elasticity of substitution is low, and close to 0.1. In this regard, we considered the server cost, power& cooling, and infrastructure cost as inputs to the cost function - notably the physical space required for each of these are crucial considerations, which most discussion related to service sector seem to undermine. We used the model designed to minimize cost by maintaining a target revenue. As observed in the result and discussion section, the proposed model has been applied on a data set collected from various sources displaying optimal elasticity value, ρ that is, for maximum revenue obtained. Subsequently, the flexibility of CES functions allowed us to incorporate additional input variables. Indeed, the model was applied to real-time data set on server cost and various components of it. We showed that the optimization estimates are quite consistent with what is experienced on the cost structure in several such firms across important regions hosting the service sector facilities.

Further, we have demonstrated the range of elasticity for which the optimization remains valid. Graphs based on the real-time data set indicate this succinctly. We used factor analysis briefly to accommodate any other factor which could technically show potential for profit maximization in technology service firms. And, finally, in order to explain the tentative rise in costs and profits concomitantly in these firms, we employed well known HHI to measure the degree of market concentration. It seems that the market is high to moderately concentrated and could absorb rising costs much more effectively, eventually translating that into rising prices.

It should be natural to ponder over the point that the prices have also been falling globally, for all such services. One of the major explanations to this end would possibly be the technological innovations that have been lowering costs significantly by raising the subscriber base phenomenally. Therefore, experiments are designed to identify major factors contributing significantly towards potential response variables. In the first case, we considered two levels in cost and in the other case assumed three levels, exploring three scenarios in each case. In the first case, the server installation cost turns out to be the most significant factor, whereas in the second case, Annual I&E and Infrastructure cost was the most significant factor. Interaction factors were found to be not significant at 5% level and have been ignored. Non-parametric[Appendix J] estimation has been performed over original and replicated data sets to reaffirm our conclusion regarding interaction among factors. In case of, $\rho < 1$ only, there is some evidence of interaction. The experiments[Supplementary file] were performed on random data. The confidence interval has been computed for each of the effects for the replication experiment. None of the confidence intervals includes zero. This implies that all the effects are significantly different from zero at a 90% confidence level. Multiple linear regression[Appendix K] technique has also been applied on the worldwide IT spending data to facilitate the prediction of the output response, revenue in our case, if *a priori* cost (predictor variables) are known. The optimal ρ value obtained by the least square approach endorses the findings. The results obtained from the least square test (with k =1 or otherwise) and that from the multiple linear regression, matches closely with the assumptions made (A detailed least square formulation for consistent and over-determined systems may be found in the supplementary file B). Subsequent calculations were carried out using active set solver to estimate the elasticity values which support the initial assumptions. Note that, we obtained the optimal revenue in all cases. It was concluded that by controlling the elasticity values, the expectations about the target revenue may in fact be attained. The scale factor and the precise technological interventions that create this cost balancing effect shall contribute to our future research interest.

8. References

[1] Greenberg, A., Hamilton, J., Maltz, A. D., Patel, P., 2009, The Cost of a Cloud: Research Problems in Data Center Networks, ACM SIGCOMM Computer Communication Review, 39(1), 68-73.

[2] B. P. Rimal, Eunmi Choi, I. Lumb, 2009, A Taxonomy and Survey of Cloud Computing Systems, The Fifth International Joint Conference on INC, IMS and IDC, pp. 44-51.

[3] Gartner Group, available at: http://www.gartner.com/; accessed on 22/1/2016

[4] J. Moore, J. Chase, P. Ranganathan and R. Sharma, 2005, Making Scheduling Cool: Temperature-Aware Workload Placement in Data

Centers, Proceedings of the annual conference on USENIX Annual Technical Conference, CA, USA.

[5] N. Rasmussen, Calculating Total Cooling Requirements for Data Centers, white paper, APC Legendary Reliability, available at: http://www.ptsdcs.com/whitepapers/23.pdf; accessed on 20/6/2016.

[6] Hamilton, J., 2008. Cost of Power in Large-Scale Data Centers, available at: http://perspectives.mvdirona.com/2008/11/cost-of-power-in-large-scale-data-centers/, accessed on 17/8/2016.

[7] Gurumurthi, M. K. S., Sivasubramaniam A., Franke, H., 2003, Reducing disk power consumption in servers with drpm, Computer, 36(12), 59-66.

[8] Fan, X., Weber, W.D., Barroso, L.A., 2007, Power provisioning for a warehouse-sized computer. Proceedings of the 34th annual international symposium on Computer architecture, ACM, New York, 13-23.

[9] Patel, C., Shah, A., 2005. Cost Model for Planning, Development and Operation of a Data Center, Internet Systems and Storage Laboratory, HP Laboratories Technical Report, HPL-2005-107R1, Palo Alto, CA

[10] https://www.westgard.com/lesson22.htm, accessed on 2/2/2016

[11] Jain, R., The Art of Computer Systems Performance Analysis, 1991, Wiley Professional Computing, ISBN: 978-0-471-50336-1.

[12] Buettner, T, Kauder, B, 2010, Revenue Forecasting Practices: Differences across Countries and Consequences for Forecasting Performance, Fiscal Studies, 31(3), 313-340.

[13] Whitfield, R. I., Duffy, A. H. B., 2013, Extended Revenue Forecasting within a Service Industry, International Journal of Production Economics, 141(2), 505-518.

[14] Saha, S., Sarkar, J., Dwivedi A., Dwivedi, N., Narasimhamurthy A. M., Roy, R., Rao, S., 2016, A novel revenue optimization model to address the operation and maintenance cost of a data center, Journal of Cloud Computing: Advances, Systems and Applications, 5(1), DOI: 10.1186/s13677-015-0050-8, 1-23.

[15] Ray, B.K., 1993, Long-Range Forecasting of IBM Product Revenues Using a Seasonal Fractionally Differenced ARMA Model, International Journal of Forecasting, 9(2), 255-269.

[16] Christian L., Belady, P.E., 2007, In the data center, power and cooling costs more than the it equipment it supports. Available at: https://www.electronics-cooling.com/2007/02/in-the-data-center-power-and-cooling-costs-more-than-the-it-equipment-it-supports, accessed on 10/09/2016.

[17] http://www.investopedia.com/terms/h/hhi.asp, accessed on 22/1/2016

[18] https://www.srgresearch.com/articles/hp-ibm-and-dell-lead-burgeoning-apac-data-center-infrastructure-market, accessed on 20/10/2016.

[19] http://www.businessinsider.in/The-cloud-wars-explained-Why-nobody-can-catch-up-with-Amazon/articleshow/49706488.cms, accessed on 10/1/2016.

[20] Arrow K. J., Chenery, H. B., Minhas, B. S., Solo,. R. M., 1961, Capital-Labor Substitution and Economic Efficiency, The Review of Economics and Statistics, 43(3), 225-250.

[21] Ghamkhari, M., Mohsenian-Rad, H., 2013, Energy and Performance Management of Green Data Centers: A Profit Maximization Approach, IEEE Trans. Smart Grid, 4(2), 1017-1025.

[22] Additional file, https://github.com/jyotirmoy208/CES, accessed on 16/2/2017

[23] https://www.ibm.com/blogs/cloud-computing/2014/02/cloud-computing-and-big-data-an-ideal-combination/, accessed on 10/1/2017

[24] Puschel, T., Schryen, G., Hristova, D., Neumann D., 2015, Revenue Management for Cloud Computing Providers: Decision Models for Service Admission Control under Non-probabilistic Uncertainty. European Journal of Operational Research, 244(2), 637-647.

[25] Bora, K., Saha, S., Agrawal, S., Safonova, M., Routh, S., Narasimhamurthy, A., 2016, CD-HPF: New Habitability Score Via Data Analytic Modeling. Astronomy and Computing, 17, 129-143, Elsevier.

[26] Ginde, G., Saha,S., Mathur, A., Venkatagiri, S., Vadakkepat, S., Narasimhamurthy A., Daya Sagar, B. S., 2016, ScientoBASE: A Framework and Model for Computing Scholastic Indicators of non-local influ-

ence of Journals via Native Data Acquisition algorithms. ScientoMetrics(Springer), 108(3), 1479-1529.

[27] Bodensteina, C., Schryenb, G., Neumanna D., 2012, Energy-aware workload management models for operation cost reduction in data centers. European Journal of Operational Research, 222(1), 157-167.

[28] Chiang, W. C., Chen, C.H.J.,Xu, X., 2007, An overview of research on revenue management: current issues and future research. International Journal of Revenue Management,1(1), DOI: 10.1504/IJRM.2007.011196,97-128.

[29] Meissner, J.,Strauss, A., 2012, Network revenue management with inventory-sensitive bid prices and customer choice. European Journal of Operational Research, 216 (2), 459-468.

[30] Kemmer,P.,Strauss, A.,Winter, T., 2012, Dynamic Simultaneous Fare Proration for Large-Scale Network Revenue Management. Journal of The Operational Research Society, 63(10), 1336-1350.

[31] Weatherford, L. R., & Kimes, S. E. 2003, A comparison of forecasting methods for hotel revenue management. International Journal of Forecasting, 19(3), 401-415.

[32] Fildes, R., Nikolopoulos, K.,Crone, S. F., Syntetos, A. A., 2008, Forecasting and operational research: a review, Journal of the Operational Research Society, 59, 1150-1172.

Appendix A. Proof of Revenue Maximization

The Lagrangian function for optimization problem is:

$$\mathcal{L} = y - \lambda(w_1 S + w_2 I + w_3 P + w_4 N - m)$$
$$\mathcal{L} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} - \lambda(w_1 S + w_2 I + w_3 P + w_4 N - m)$$

The first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial S} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} S^{\rho-1} - \lambda w_1 = 0 \tag{A.1}$$

$$\frac{\partial \mathcal{L}}{\partial I} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} I^{\rho-1} - \lambda w_2 = 0 \tag{A.2}$$

$$\frac{\partial \mathcal{L}}{\partial N} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} N^{\rho-1} - \lambda w_3 = 0 \tag{A.3}$$

$$\frac{\partial \mathcal{L}}{\partial I} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} P^{\rho-1} - \lambda w_3 = 0 \tag{A.4}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(w_1 S + w_2 I + w_3 P + w_4 N - m) = 0 \tag{A.5}$$

Dividing (A.2),(A.3),(A.4) by (A.1)

$$\frac{w_2}{w_1} = (\frac{I}{S})^{\rho-1}$$

Similarly,

$$I = \sqrt[\rho-1]{\frac{w_2}{w_1} S} \tag{A.6}$$

$$P = \sqrt[\rho-1]{\frac{w_3}{w_1} S} \tag{A.7}$$

$$N = \sqrt[\rho-1]{\frac{w_4}{w_1} S} \tag{A.8}$$

Substituting these values in equation ((A.5)), we obtain

$$w_1 S + W_2 \sqrt[\rho-1]{\frac{w_2}{w_1} S} + W_3 \sqrt[\rho-1]{\frac{w_3}{w_1} S} + w_4 \sqrt[\rho-1]{\frac{w_4}{w_1} S} - m = 0$$

$$S = \frac{m w_1 \frac{1}{\rho-1}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{A.9}$$

Similarly

$$I = \frac{m w_2 \frac{1}{\rho-1}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{A.10}$$

$$N = \frac{m w_3 \frac{1}{\rho-1}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{A.11}$$

$$P = \frac{m w_4 \frac{1}{\rho-1}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{A.12}$$

Appendix B. Proof of Cost Optimization

$$\mathcal{L} = w_1 S + w_2 I + w_3 P + w_4 N - \lambda((S^\rho + I^\rho + P^\rho + N^\rho) - y_{tar})$$

The first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial S} = w_1 - \lambda S^{\rho-1}(S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} = 0 \tag{B.1}$$

$$\frac{\partial \mathcal{L}}{\partial I} = w_2 - \lambda I^{\rho-1}(S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} = 0 \tag{B.2}$$

$$\frac{\partial \mathcal{L}}{\partial P} = w_3 - \lambda P^{\rho-1}(S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} = 0 \tag{B.3}$$

$$\frac{\partial \mathcal{L}}{\partial N} = w_4 - \lambda N^{\rho-1}(S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} = 0 \tag{B.4}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} - y_{tar} = 0 \tag{B.5}$$

Dividing (B.2), (B.3), (B.4) by (B.1)

$$I = \sqrt[\rho-1]{\frac{w_2}{w_1}} S$$

$$P = \sqrt[\rho-1]{\frac{w_3}{w_1}} S$$

$$N = \sqrt[\rho-1]{\frac{w_4}{w_1}} S$$

Substituting the values in CES function

$$y_{tar} = (S^\rho + (\frac{w_2}{w_1}^{\frac{\rho}{\rho-1}} S^\rho) + (\frac{w_3}{w_1}^{\frac{\rho}{\rho-1}} S^\rho) + (\frac{w_4}{w_1}^{\frac{\rho}{\rho-1}} S^\rho))$$

$$S = \frac{y_{tar} w_1^{\frac{1}{\rho-1}}}{(w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}})^{\frac{1}{\rho}}} \tag{B.6}$$

$$I = \frac{y_{tar} w_2^{\frac{1}{\rho-1}}}{(w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}})^{\frac{1}{\rho}}} \tag{B.7}$$

$$P = \frac{y_{tar} w_3^{\frac{1}{\rho-1}}}{(w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}})^{\frac{1}{\rho}}} \tag{B.8}$$

The cost function can be rewritten as :

$$c = \left(\frac{y_{tar}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}}\right)^{\frac{1}{\rho}-1}$$

Appendix C. Proof of Profit Maximization

The profit function is written below
Profit=$(S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} - (w_1 S + w_2 I + w_3 P + w_4 N)$
We want to maximize the profit subject to $w_1 S + w_2 I + w_3 P + w_4 N = c_{thresh}$
Using Lagrange multiplier,

$$\begin{aligned} \mathcal{L} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} - (w_1 S + w_2 I + w_3 P + w_4 N) \\ +\lambda(w_1 S + w_2 I + w_3 P + w_4 N - c_{thresh}) \end{aligned} \tag{C.1}$$

$$\frac{\partial \mathcal{L}}{\partial S} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} S^{\rho-1} - w_1 + \lambda w_1 = 0 \tag{C.2}$$

$$\frac{\partial \mathcal{L}}{\partial I} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} I^{\rho-1} - w_2 + \lambda w_2 = 0 \tag{C.3}$$

$$\frac{\partial \mathcal{L}}{\partial P} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} P^{\rho-1} - w_3 + \lambda w_3 = 0 \tag{C.4}$$

$$\frac{\partial \mathcal{L}}{\partial N} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} N^{\rho-1} - w_4 + \lambda w_4 = 0 \tag{C.5}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = w_1 S + w_2 I + w_3 P + w_4 N - c_{thresh} = 0 \tag{C.6}$$

Comparing the λ value from equation (C.2) an (C.3), We obtain

$$I = \frac{w_2}{w_1}^{\frac{1}{\rho-1}} S$$

Similarly,

$$P = \frac{w_3}{w_1}^{\frac{1}{\rho-1}} S$$

$$N = \frac{w_4}{w_1}^{\frac{1}{\rho-1}} S$$

Putting the values of I, P, N in equation (C.6)

$$w_1 + w_2 \frac{w_2}{w_1}^{\frac{1}{\rho-1}} S + w_3 \frac{w_3}{w_1}^{\frac{1}{\rho-1}} S + w_4 \frac{w_4}{w_1}^{\frac{1}{\rho-1}} S - c_{thresh} = 0$$

$$S = \frac{c_{thresh} w_1^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}}$$

Similarly,

$$I = \frac{c_{thresh} w_2^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}}$$

$$P = \frac{c_{thresh} w_3^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}}$$

$$N = \frac{c_{thresh} w_4^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}}$$

Appendix D. Curvature Characteristic of CES

CES function:

$$y = (K^\rho + L^\rho)^{\frac{1}{\rho}}$$

$$\frac{\partial y}{\partial K} = \frac{1}{\rho}(K^\rho + L^\rho)^{\frac{1}{\rho}-1} \rho K^{\rho-1}$$

$$\frac{\partial y}{\partial L} = \frac{1}{\rho}(K^\rho + L^\rho)^{\frac{1}{\rho}-1} \rho L^{\rho-1}$$

$$\frac{\partial y}{\partial K \partial L} = \rho K^{\rho-1} L^{\rho-1} (K^\rho + L^\rho)^{\frac{1}{\rho}-2}$$

$$\frac{\partial y}{\partial L \partial K} = \rho K^{\rho-1} L^{\rho-1} (K^\rho + L^\rho)^{\frac{1}{\rho}-2}$$

$$\frac{\partial^2 y}{\partial^2 K} = \rho(K^\rho - 1)^2 (K^\rho + L^\rho)^{\frac{1}{\rho}-2} + (\rho-1) K^{\rho-2} (K^\rho + L^\rho)^{\frac{1}{\rho}-1}$$

$$\frac{\partial^2 y}{\partial^2 L} = \rho(L^{\rho-1})^2 (K^\rho + L^\rho)^{\frac{1}{\rho}-2} + (\rho-1) L^{\rho-2} (K^\rho + L^\rho)^{\frac{1}{\rho}-1}$$

Hessian Matrix

$$\begin{bmatrix} \begin{matrix} \rho(K^{\rho}-1)^2(K^{\rho}+L^{\rho})^{\frac{1}{\rho}-2} \\ +(\rho-1)K^{\rho-2}(K^{\rho}+L^{\rho})^{\frac{1}{\rho}-1} \end{matrix} & \rho K^{\rho-1}L^{\rho-1}(K^{\rho}+L^{\rho})^{\frac{1}{\rho}-2} \\ \\ \rho K^{\rho-1}L^{\rho-1}(K^{\rho}+L^{\rho})^{\frac{1}{\rho}-2} & \begin{matrix} \rho(L^{\rho-1})^2(K^{\rho}+L^{\rho})^{\frac{1}{\rho}-2} \\ +(\rho-1)L^{\rho-2}(K^{\rho}+L^{\rho})^{\frac{1}{\rho}-1} \end{matrix} \end{bmatrix}$$

$\Delta_1 = (K^{\rho}+L^{\rho})^{\frac{1}{\rho}-1}K^{\rho-1}(\frac{\rho K^{\rho-1}}{K^{\rho}+L^{\rho}} + \frac{\rho-1}{K})$
As K, L, $\rho > 0$ $\Delta_1 > 0$;
$\Delta_2 = \rho(\rho-1)(K^{\rho-1})^2(L^{\rho-2})(K^{\rho}+L^{\rho})^{\frac{2}{\rho}-3} + \rho(\rho-1)(L^{\rho-1})^2(K^{\rho-2})(K^{\rho}+L^{\rho})^{\frac{2}{\rho}-3} + (\rho-1)^2(K^{\rho-2})(L^{\rho-2})(K^{\rho}+L^{\rho})^{\frac{2}{\rho}-2}$
$\Delta_2 \geq 0$ in case $\rho \geq 1$
As $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$ in case $\rho \geq 1$. It will produce concave graph.
When $\rho < 1$, $\Delta_1 \geq 0$ and $\Delta_2 \leq 0$.
It is neither concave or convex.

Appendix E. Positivity of Δ_1 of CES Hessian Matrix

Let us now explain the reason why Δ_1 from previous appendix is always positive.
Considering Δ_1 value again:
$\Delta_1 = (K^{\rho}+L^{\rho})^{\frac{1}{\rho}-1}K^{\rho-1}(\frac{\rho K^{\rho-1}}{K^{\rho}+L^{\rho}} + \frac{\rho-1}{K})$
Δ_1 will be negative if below two conditions are satisfied.
1)$\rho < 1$
2)$\frac{\rho-1}{K} \geq \frac{\rho K^{\rho-1}}{K^{\rho}+L^{\rho}}$

$$
\begin{aligned}
&\frac{\rho-1}{K} \geq \frac{\rho K^{\rho-1}}{K^\rho + L^\rho} \\
&=> (\rho-1)(K^\rho + L^\rho) \geq \rho K^\rho \\
&=> \frac{\rho-1}{\rho} \geq \frac{K^\rho}{K^\rho + L^\rho} \\
&=> 1 - \frac{1}{\rho} \geq \frac{K^\rho}{K^\rho + L^\rho} \\
&=> 1 - \frac{K^\rho}{K^\rho + L^\rho} \geq \frac{1}{\rho} \\
&=> \frac{L^\rho}{K^\rho + L^\rho} \geq \frac{1}{\rho} \\
&=> \rho L^\rho \geq K^\rho + L^\rho \\
&=> (\rho-1) \geq \frac{K^\rho}{L^\rho} \\
&=> \rho - 1 \geq (\frac{K}{L})^\rho
\end{aligned}
$$

As K, $L > 0$, hence $(\frac{K}{L})^\rho$ will be always positive.

$$
\begin{aligned}
&(\rho - 1) > 0 \\
&=> \rho > 1
\end{aligned}
$$

Which is contradicting the first condition $\rho < 1$.
Hence Δ_1 will be always positive.

Appendix F. Least Square Approach

The proposed model is fitted using least square approach.

$$(k_1^r + k_2^r)^{\frac{1}{r}}$$

Matlab code for $r < 1$: The datsets k_1, k_2 and observed data ydata are given as

$$
\begin{aligned}
&k_2 = [5, 5, 10, 10, 10, 15, 15, 15, 20, 20, 20, 30, 30, 30, 40, 40, 40] \\
&k_1 = [62, 65, 62, 60, 65, 55, 45, 47, 50, 52, 55, 56, 57, 58, 59, 60, 60] \\
&ydata = [19511.9, 20038.22, 26579.25, 26108.41, 27274.01, \\
&30038.95, 27008.5, 27635.78, 32723.23, 33401.89, 34399.27, \\
&42176.36, 42563.18, 42947.06, 49839.75, 50268.74, 50268.74]
\end{aligned}
$$

The function which is needed to pass in the matlab lsqnonlin function is $fun = @(r)(k_1^r + k_2^r)^{\frac{1}{r}} - ydata$ In the case of without constraints, no upper and lower bound of r need to pass in the function. We will assume a initial value for r as 0.4. Next calling the lsqnonlin function
x = lsqnonlin(fun,x0)
x =0.1000; So we obtained the least square r value as 0.1000. Say the upper bound for r is 0.9 and lower bound as 0.1. We will pass the same information in lsqnonlin function.

$$x = lsqnonlin(fun, x_0, 0, 0.9);$$
$$x = 0.1000;$$

We get the same output as we have obtained without constraints.
Matlab code for $r > 1$:

$$k_2 = [5, 5, 10, 10, 10, 15, 15, 15, 20, 20, 20, 30, 30, 30, 40, 40, 40]$$
$$k_1 = [62, 65, 62, 60, 65, 55, 45, 47, 50, 52, 55, 56, 57, 58, 59, 60, 60]$$
$$ydata = [65.5239, 68.5078, 69.5306, 67.5538, 72.4972,$$
$$66.8546, 57.0674, 59.0214, 66.3455, 68.2907, 71.2122, 81.1220,$$
$$82.0859, 83.0503, 93.1262, 94.0819, 94.0819]$$
$$fun = @(r)(k_1^r + k_2^r)^{\frac{1}{r}} - ydata;$$
$$x_0 = 1.4;$$
$$x = lsqnonlin(fun, x_0)$$
$$x = 1.1000;$$

Say the upper bound for r is 1.9 and lower bound as 1.

$$x = lsqnonlin(fun, x_0, 1, 1.9);$$
$$x = 1.1000;$$

Appendix G. Goodness of Fit Test

Shapiro-Wilk test is test of normality, which is frequently used in statistics. It utilizes the null hypothesis principle to test a sample dataset belongs to normally distributed population.The null-hypothesis of this test is that the population is normally distributed. If the p-value is less than the threshold alpha level, then the null hypothesis is rejected and the tested data are not from normally distributed population. In other words, the data are not normal. On the contrary, if the p-value is greater than the chosen alpha level, then the null hypothesis that the data came from a normally distributed population cannot be rejected.

Code for Shapiro-Wilk goodness-of-fit test

Mupad command has been used to generate the code.

$data := [62, 65, 62, 60, 65, 55, 45, 47, 50, 52, 55, 56, 57, 58, 59, 60, 60] :$

$stats :: swGOFT(data)$

$[PValue = 0.4011881607, StatValue = 0.9463348025]$

$data := [78, 60, 55, 44, 75, 62, 55, 49, 55, 57, 65, 46, 77, 68, 59, 48, 60] :$

$stats :: swGOFT(data)$

$[PValue = 0.3841895654, StatValue = 0.9451314104]$

$data := [5, 5, 10, 10, 10, 15, 15, 15, 20, 20, 20, 30, 30, 30, 40, 40, 40] :$

$stats :: swGOFT(data)$

$[PValue = 0.07626140514, StatValue = 0.9030010795]$

$data := [15, 35, 18, 20, 30, 15, 25, 18, 25, 24, 35, 38, 27, 32, 16, 10, 30] :$

$stats :: swGOFT(data)$

$[PValue = 0.6499020509, StatValue = 0.9609763672]$

0.05 is the threshold of pvalue. As all the observed Pvalue are greater than the threshold. The null hypothesis is accepted and datasets are belong to normal distribution.

A chi-square test is applied on a sample data from a population to test whether the data is consistent with a hypothesized distribution.

Code for chi-square:

data := [62,65,62,60,65,55,45,47,50,52,55,56,57,58,59,60,60];

h = chi2gof(data);

In case of normal distribution h=0 otherwise h=1.

Appendix H. Matlab Code for Fmincon

The matlab fmincon code when $\rho < 1$:

$$A = [1; -1];$$
$$b = [0.9; -0.1];$$
$$x0 = [0.4];$$
$$[x, fval] = fmincon(@myCES, x0, A, b)$$
$$function f = myCES(x)$$
$$pow = 1/x(1);$$
$$f = -(62^x(1) + 5^x(1)).^pow;$$
$$end$$

The matlab fmincon code when $\rho > 1$:

$$A = [1; -1];$$
$$b = [1.9; -1.1];$$
$$x0 = [0.4];$$
$$[x, fval] = fmincon(@myCES, x0, A, b)$$
$$function f = myCES(x)$$
$$pow = 1/x(1);$$
$$f = -(62^x(1) + 5^x(1)).^pow;$$
$$end$$

Appendix I. Randomization of Data

The data collected from various sources are not sufficient enough to identify the effect of factors on the revenue. So we have to find the probability distribution of the original data set and random data set need to generate which will follow the same distribution of real data set. We have found through experiment that original data set follows normal distribution. Fig.I.7 depicts the normal distribution of the server cost of original and random data and the same displays the normal distribution of power & cooling cost. The maximum revenue for random data has been calculated and presented in table 21 and 22 of additional file [22]. Shapiro-Wilk Original Test and Chi Square- Goodness have been conducted on the original and actual data set to identify the normal distribution behavior. The Null Hypotheses

h_0: After adding noise to the original data set, the data follows normal Distribution. If $h_0 = 1$, the null hypothesis is rejected at 5% significance level. if h0=0, the null hypothesis is accepted at 5% significance level. After the experiment, we found that data set follows a normal distribution with 95% confidence level i.e. $h_0 = 0$. The details of the Shapiro-Wilk Original Test and the Chi Square- Goodness have been elaborated in Appendix G.

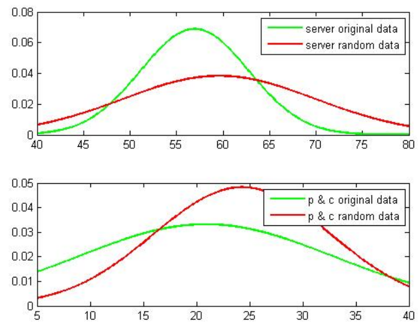
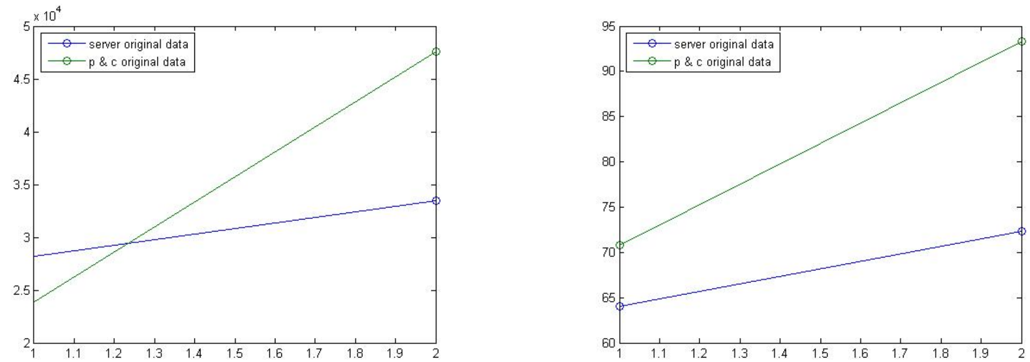


Figure I.7: The Original and Generated Server,Power & Cooling Data that follows Normal Distribution

Appendix J. Non-parametric Estimation

The Non-parametric statistic does not belong to the family of probability distributions. It can be both descriptive and statistical. No assumption about the probability distribution of the sample data has been made in non-parametric estimation. The typical parameters mean, variance are accessed in the estimation. The typical parameters are the mean, variance, etc. Unlike parametric statistics, non-parametric statistics make no assumptions about the probability distributions of the variables being assessed. In parametric, there is no specific distinction between the true models and fitted models. In contrast, non-parametric methods are able to distinguish between the true and fitted models. The drawbacks of non-parametric tests in comparison to parametric tests are that these are less powerful . We have performed non-parametric estimation on the data set ignoring whether the data set is part of a probability distribution. Fig. J.8a and J.8b showcase the result of non-parametric estimation on the original data set. Both the cost segments, Server and power & cooling have been displayed in the figure. Fig.J.8a suggests an interaction between the two cost components. Fig.J.8b reveals no interaction between the factors. The non-parametric estimation



(a) Non parametric estimation on orignial data, when $\rho < 1$

(b) Non parametric estimation on orignial data when $\rho = 1$

Figure J.8: Non parametric estimation on orignial data

on generated data has been shown in Fig.J.10a and J.10b. None of the figures demonstrate any interaction between factors.

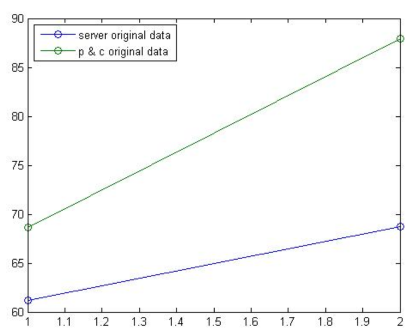
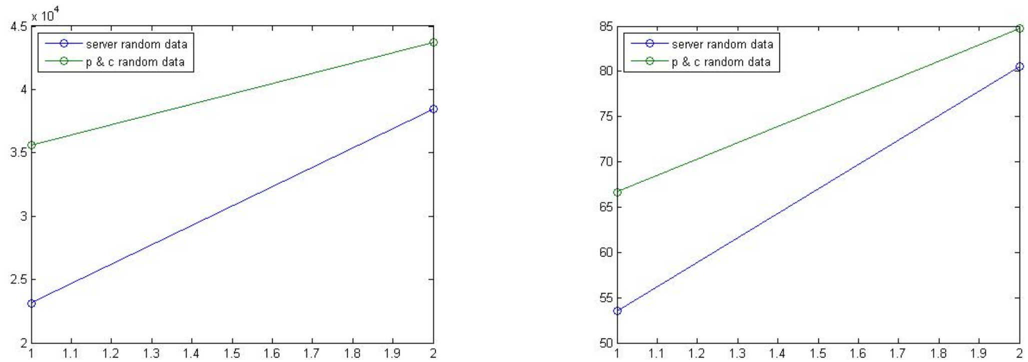


Figure J.9: Non parametric estimation on orignial data when $\rho > 1$

Appendix K. PREDICTION AND FORECASTING

The linear regression models are restricted in three ways. First, only one predictor variable needs to be considered. Second, the predictor variable should be quantitative and third response must be a linear function of the predictor. Multiple linear regression is a technique, where more than one predictor variables may be considered [11]. Multiple linear regression helps



(a) Non parametric estimation on random data when $\rho < 1$

(b) Non parametric estimation on random data when $\rho > 1$

Figure J.10: Non parametric estimation on random data

one to predict a response variable y as a function of k predictor variables $x_1, x_2, x_3 \cdots x_k$ using a linear model of the following form.

$$y = b_0 + b_1x_1 + b_2x_2 + \cdot + b_kx_k + e$$

Here, $b_0, b_1, \cdot, b_k$ are the $k+1$ fixed parameters and e is the error term. Given a sample data set $(x_11, x_21, \cdot, x_k1, y_1), \cdot, (x_1n, x_2n, \cdot, x_kn, y_n)$ of n observations, the model consists of the following n equations:

$$\begin{aligned} y_1 &= b_0 + b_1x_{11} + b_2x_{21} + \cdot + b_kx_{k1} + e_1 \\ y_2 &= b_0 + b_1x_{12} + b_2x_{22} + \cdot + b_kx_{k2} + e_2 \\ &\vdots \\ y_n &= b_0 + b_1x_{1n} + b_2x_{2n} + \cdot + b_kx_{kn} + e_n \end{aligned}$$

The equations above can be rewritten in vector notation as

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} b_0 \\ \vdots \\ b_k \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

The above reads $y = Xb + e$ in simplified notation.

- b = A column vector with k+1 elements are $b_0, b_1, ..., b_k$

- y is a column vector of n observed values of $y = y_1, ..., y_n$
- X= An n by K+1 matrix whose $(i, j+1)^{th}$ element $X_{i,j+1}$ is 1 if j is 0; else x_{ij}
- e is a column vector of n error terms $e_1, e_2 \cdot, e_n$

Parameter estimation:

$$b = (X^T X)^{-1}(X^T y) \tag{K.1}$$

Allocation of variation:

$$SSY = \sum_{i=1}^{n} y_i^2 \tag{K.2}$$

$$SS0 = n\overline{y}^2 \tag{K.3}$$

$$SST = SSY - SS0 \tag{K.4}$$

$$SSE = y^T y - b^T X^T y \tag{K.5}$$

$$SSR = SST - SSE \tag{K.6}$$

where, SSY = sum of squares of Y, SST = total sum of squares, $SS0$ = sum of squares of y, SSE = sum of squared errrors and SSR = sum of squares given by regression. Coefficient of determination is calculated as

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} \tag{K.7}$$

Coefficient of multiple correlation is defined as

$$R = \sqrt{\frac{SSR}{SST}} \tag{K.8}$$

The multiple regression has been applied on the data set, for $\rho < 1$ and $\rho > 1$ respectively. The findings of different parameters and coefficient of determination have been demonstrated in table 23. The R squared test implies how well the linear regression model can explain the data set. It is also known as the coefficient of determination and the coefficient of multiple determination for multiple regression. The linear expressions produced by the data sets fitted well as the R squared are 99.08 % and 99.99 %.

$$y = 10073 + 113.85x_1 + 848.85x_2 \tag{K.9}$$

$$y = -1.5782 + 1.0072x_1 + 0.8782x_2 \tag{K.10}$$

Table 23: Multiple Linear Regression Results

	$\rho < 1$	$\rho > 1$
SSY	$2.1652 * 10^{10}$	$9.5578 * 10^{4}$
SSO	$1.9979 * 10^{10}$	$9.3381 * 10^{4}$
SST	$1.6738 * 10^{9}$	$2.1970 * 10^{3}$
SSR	$1.6584 * 10^{9}$	$2.1967 * 10^{3}$
SSE	$1.5397 * 10^{7}$	0.3271
R squared	0.9908	.9999

Appendix L. Profit Maximization

Following above description, the profit function is written as,

$$Profit = (S^{\rho} + I^{\rho} + P^{\rho} + N^{\rho})^{\frac{1}{\rho}} - (w_1 S + w_2 I + w_3 P + w_4 N) \tag{L.1}$$

Say, the cost should not surpass the threshold c_{thresh}. This requires us to maximize $(S^{\rho} + I^{\rho} + P^{\rho} + N^{\rho}) - (w_1 S + w_2 I + w_3 P + w_4 N)$ subject to $w_1 S + w_2 I + w_3 P + w_4 N = c_{thresh}$.
The following values of S, I, P and N thus obtained are the values for which the data center can attain maximum profit by not violating the constraints.

$$S = \frac{c_{thresh} w_1^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{L.2}$$

$$I = \frac{c_{thresh} w_2^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{L.3}$$

$$P = \frac{c_{thresh} w_3^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{L.4}$$

$$N = \frac{c_{thresh} w_4^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \tag{L.5}$$

The results, (16)-(19) are analytically verified in Appendix C.