Additional Files: Mathematical Proofs of Optimization and Related Codes

1. Tables

Table 2: Revenue Simulation output when $\rho < 1$.In the table, all the units are in \$B. The optimal revenue across the years is obtained at ρ= 0.100 and s=1.11, **page 12**

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Max. Revenue
1996	62	5	1.11	0.1000	19511.9
1997	65	5	1.11	0.1000	20038.22
1998	62	10	1.11	0.1000	26579.25
1999	60	10	1.11	0.1000	26108.41
2000	65	10	1.11	0.1000	27274.01
2001	55	15	1.11	0.1000	30038.95
2002	45	15	1.11	0.1000	27008.5
2003	47	15	1.11	0.1000	27635.78
2004	50	20	1.11	0.1000	32723.23
2005	52	20	1.11	0.1000	33401.89
2006	55	20	1.11	0.1000	34399.27
2007	56	30	1.11	0.1000	42176.36
2008	57	30	1.11	0.1000	42563.18
2009	58	30	1.11	0.1000	42947.06
2010	59	40	1.11	0.1000	49839.75
2011	60	40	1.11	0.1000	50268.74
2012	60	40	1.11	0.1000	50268.74

Table 3: Revenue Simulation output when $\rho = 1$. 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 billion USD in 1996, while the difference rises to almost 6 billion USD in 2012, **page 13**

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Max. Revenue
1996	62	5	0	1	67
1997	65	5	0	1	70
1998	62	10	0	1	72
1999	60	10	0	1	70
2000	65	10	0	1	75
2001	55	15	0	1	70
2002	45	15	0	1	60
2003	47	15	0	1	62
2004	50	20	0	1	70
2005	52	20	0	1	72
2006	55	20	0	1	75
2007	56	30	0	1	86
2008	57	30	0	1	87
2009	58	30	0	1	88
2010	59	40	0	1	99
2011	60	40	0	1	100
2012	60	40	0	1	100

Table 4: Revenue Simulation output when $\rho > 1$. All the units are in \$B. The optimal revenue year-wise are obtained at $\rho = 1.1$ and s= -10, **page 13**

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Max. Revenue
1996	62	5	-10	1.1000	65.5239
1997	65	5	-10	1.1000	68.5078
1998	62	10	-10	1.1000	69.5306
1999	60	10	-10	1.1000	67.5538
2000	65	10	-10	1.1000	72.4972
2001	55	15	-10	1.1000	66.8546
2002	45	15	-10	1.1000	57.0674
2003	47	15	-10	1.1000	59.0214
2004	50	20	-10	1.1000	66.3455
2005	52	20	-10	1.1000	68.2907
2006	55	20	-10	1.1000	71.2122
2007	56	30	-10	1.1000	81.1220
2008	57	30	-10	1.1000	82.0859
2009	58	30	-10	1.1000	83.0503
2010	59	40	-10	1.1000	93.1262
2011	60	40	-10	1.1000	94.0819
2012	60	40	-10	1.1000	94.0819

Table 5: Revenue Simulation output for annual amortized dataset, when $\rho < 1$.Using fmincon function of matlab, the elasticity for maximum revenue is computed. As observed in the table , optimal revenues have been achieved at ρ=0.1 and s= 1.1. The details of the table is available on **page 12**.

Year	Server Cost	Energy cost	Infras-tructure	Annual I & E	Elasticity(s)	ρ	Max. Revenue
1992	1400	50	200	220	1.1	0.1	265159191.08
1995	1400	75	250	300	1.1	0.1	329860151.91
2000	1400	200	500	1200	1.1	0.1	691685894.7
2005	1400	1000	1000	2800	1.1	0.1	1644247792.93
2010	1400	1600	1500	3400	1.1	0.1	2029519563.1

Table 6: Revenue Simulation output when $\rho = 1$.Annual amortized dataset has been used for optimum revenue generation.The revenue generated is lower than ($\rho < 1$) case, but higher than the case, where $\rho > 1$. The revenue rises almost 4 times between 1992 and 2010, More details of the table can be found on **Page 13**.

Year	Server Cost	Energy cost	Infras-tructure	Annual I & E	Elasticity(s)	ρ	Max. Revenue
1992	1400	50	200	220	0	1	1870
1995	1400	75	250	300	0	1	2025
2000	1400	200	500	1200	0	1	3300
2005	1400	1000	1000	2800	0	1	6200
2010	1400	1600	1500	3400	0	1	7900

Table 7: Revenue Simulation output for annual amortized dataset, when $\rho > 1$,It is observed that the revenue rises 4 folds between 1992 and 2010. The maximum revenue is attained at ρ=1.1, **page 13**

Year	Server Cost	Energy cost	Infras-truc-ture	Annual I & E	Elasticity(s)	ρ	Max. Revenue
1992	1400	50	200	220	-10	1.1	1619.02
1995	1400	75	250	300	-10	1.1	1733.20
2000	1400	200	500	1200	-10	1.1	2738.19
2005	1400	1000	1000	2800	-10	1.1	5066.52
2010	1400	1600	1500	3400	-10	1.1	6424.43

Table 8: Gradient Descent output for cost minimization. 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, **page 14**

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Min. Cost
1996	62	5	1.007	0.0078	49.6
1997	65	5	1.21	0.1764	52.32
1998	62	10	1.007	0.0078	49.6
1999	60	10	1.03	0.0305	48
2000	65	10	1.21	0.1764	52.32
2001	55	15	1.09	0.0835	44
2002	45	15	.99	0.0076	36
2003	47	15	1.19	0.1600	37.61
2004	50	20	1.15	0.1323	40
2005	52	20	1.12	0.1132	41.6
2006	55	20	1.09	0.0835	44
2007	56	30	1.079	0.0733	44.8
2008	57	30	1.06	0.0629	45.6
2009	58	30	1.055	0.0523	46.4
2010	59	40	1.04	0.0415	47.2
2011	60	40	1.03	0.0305	48
2012	60	40	1.03	0.0305	48

Table 9: Simulation output for profit maximization when $\rho < 1$. New server cost and Power & Cooling cost derived from world wide IT spending data along with optimal profit have been displayed in the table. The table shows that maximum profit has been obtained at $\rho = 0.1$,The reference of the table can be found on **page 14**.

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Max. Profit
1996	62	5	1.1	0.1	19486
1997	65	5	1.1	0.1	20011
1998	62	10	1.1	0.1	26551
1999	60	10	1.1	0.1	26081
2000	65	10	1.1	0.1	27245
2001	55	15	1.1	0.1	30012
2002	45	15	1.1	0.1	26986
2003	47	15	1.1	0.1	27612
2004	50	20	1.1	0.1	32697
2005	52	20	1.1	0.1	33375
2006	55	20	1.1	0.1	34371
2007	56	30	1.1	0.1	42145
2008	57	30	1.1	0.1	42531
2009	58	30	1.1	0.1	42915
2010	59	40	1.1	0.1	49804
2011	60	40	1.1	0.1	50233
2012	60	40	1.1	0.1	50233

Table 10: Simulation output for profit maximization when $\rho > 1$.Fmincon function is utilized to determine optimal ρ. Maximum profit, optimal ρ and the two cost segments are shown in the Table.The reference of the table is available on **page 15**.

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Max. Profit
1996	62	5	-10	1.1	39.2239
1997	65	5	-10	1.1	41.0078
1998	62	10	-10	1.1	41.7306
1999	60	10	-10	1.1	40.5538
2000	65	10	-10	1.1	43.4972
2001	55	15	-10	1.1	40.3546
2002	45	15	-10	1.1	34.5674
2003	47	15	-10	1.1	35.7214
2004	50	20	-10	1.1	40.3455
2005	52	20	-10	1.1	41.4907
2006	55	20	-10	1.1	43.2122
2007	56	30	-10	1.1	49.7220
2008	57	30	-10	1.1	50.2859
2009	58	30	-10	1.1	50.8503
2010	59	40	-10	1.1	57.5262
2011	60	40	-10	1.1	58.0819
2012	60	40	-10	1.1	58.0819

2. Replication:

A replication experiment is performed to estimate the imprecision or random error of the analytical method. Methods of measurements are almost always subject to some random variation. Repeat measurements will usually reveal slightly different results, sometimes a little higher, sometimes a little lower. Determining the amount of random error is usually one of the first steps in a method validation study[1]. To measure the variability associated with an experiment, repetitions are performed in various fields such as engineering, science, and statistics. As we are going to elaborate, at least two replications need to be considered in 3^2 factorial design.**The 3^2 factorial analysis methodology has been discussed under section 6 on pages 21-24.** The result of the replication experiment has been displayed in Table 19. The difference with the previous 3^2 experimental design in terms of variations is visible. The interaction factor, which was the second most significant thing in the case of $\rho < 1$, has reduced significantly and goes down to 2.25%. Still,

the new server is the most major contributor and power and cooling contributing 9.27 % can be ignored. In the case of $\rho > 1$, the interaction factor still is the least significant factor. It is observed that the experiment error is quite high.
In statistics, we perform a variety of interval studies to understand the behavior of the produced result. Out of these, the confidence interval is widely used. A confidence interval defines a range of values, which has been derived from a sample data set. The range may contain the value of the unknown parameter. Due to the random characteristic, it is unlikely that two samples collected from a population yield an identical confidence interval. But by repeating a sample many times, we may find a confidence interval, which holds the unknown parameter. We apply the same concept to each of the coefficients (q_A, q_B, q_AB). Each coefficient value falls in the range with 90 percent confidence as shown in table 19.

Table 19: Percentage variations

	$\rho < 1$	$\rho > 1$
New Server	63.1	47.74
Power and Cooling	9.27	27.417
Interaction	2.25	4.10
Error	25.2	20.73

Table 20: Percentage variations

	$\rho < 1$ CI	$\rho > 1$ CI
q_0	33553.785-37008.065	68.204-74.516
q_A	4200.75-7649.03	8.1265-14.4385
q_B	2760.63-6214.91	1.171-7.4835
q_AB	9.395-9.395	-0.8985- 5.4135

3. Quadratic Least Square fit of the CES model

After the linear regression fitting(section 5 of main file, pages 18-20), it is required to investigate the quadratic least square fitting of the CES production function. Let us consider the CES production function again.

$$y = (A^\rho + B^\rho)^{\frac{1}{\rho}}$$

If we have N number of data points, the equation can be rewritten as below.

$$y_1 = (A_1^\rho + B_1^\rho)^{\frac{1}{\rho}}$$

$$y_N = (A_N^\rho + B_N^\rho)^{\frac{1}{\rho}}$$

As the number of the data points is greater than the number of unknowns, it is an over-determined system. We will discuss two instances of least square approaches.

No Constraints:

As the CES function is non-linear, we will explore non-linear least square method. Our objective is to fit a set of observation with our proposed model, which is non-linear. We assume that there is no restriction or constraints on the parameters. The basis of the method is to approximate the model by a linear one and to refine the parameters by successive iterations. We consider m data points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$, and the model function as $y = f(x, \beta)$. The variable x is dependent on n parameters, $(\beta_1, \beta_2, \ldots \beta_n)$. So the expectation is to find the vector ofβ such that the curve is a best fit for the observed data set, means the sum of squares

$$S = \sum_{i=0}^{n} r_i^2$$

is minimized, where the residuals (errors) r_i are given by

$$r_i = y_i - f(x_i, \beta)$$
$$i = 1, 2 \ldots m$$

The minimum value will be attained where gradient is 0. There will be n gradient equations

$$\frac{\partial S}{\partial \beta_j} = 2 \sum_i r_i \frac{\partial r}{\partial \beta_j} (j = 1 \ldots n)$$

The gradient equations are the combinations of the independent variables and parameters in the non-linear system. These equations don't have any closed form solution. Hence, the values are obtained by successive approximation during each iteration.

$$\beta_j \approx \beta_j^{k+1} = \beta_j^k + \Delta\beta_j$$

K is the number of iterations and $\Delta\beta$ is the increment. Talyor series has been used to linearize the model at each iteration. We have used lsqnonlin function of Matlab, which deals with non-linear least square approach. The details of this approach are available in Appendix F. The results after applying n-linear least square has been shown in the following table 17.

Table 17: Least Square Result without constraints

	$\rho < 1$	$\rho > 1$
ρ	0.100	1.100

We have utilized the same Matlab lsqnonlin function for the case with constraints. The details have been elaborated in Appendix F. Table 18 is obtained after applying least square method with constraints.

Table 18: Least Square Result with constraints

	$\rho < 1$	$\rho > 1$
ρ	0.100	1.100

4. Experiments

The experiments are the continuation of the subsection 6.2 from the main file. Two different scenarios, which are not discussed in details, are elaborated in this section.

4.0.1. With two factors $\rho > 1$

Replacing the observations in the model, we get following equations:

$58.0444 = q_0$
$67.3181 = q_0 + q_2 + q_{22}$
$66.8546 = q_0 + q_1 + q_{11}$
$71.2122 = q_0 + q_1 + q_2 + q_{11} + q_{12} + q_{22}$

Table 14: 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

$84.8461 = q_0 + q_1 + 2q_2 + q_{11} + 2q_{12} + 4q_{22}$
$68.72266 = q_0 + 2q_1 + 4q_{11}$
$94.0819 = q_0 + 2q_1 + 2q_2 + 4q_{11} + 4q_{12} + 4q_{22}$
After solving all the equations, the regression equation is:

$$y = 58.0444 + 12.2814x_A + 19.17105x_B - 4.9161x_Ax_B \\ -3.4712x_A^2 - 9.89735x_B^2 \tag{1}$$

4.0.2. With two factors $\rho = 1$

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

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

Replacing the observations in the model, we get following equations:
$61 = q_0$
$71 = q_0 + q_2 + q_{22}$
$70 = q_0 + q_1 + q_{11}$
$75 = q_0 + q_1 + q_2 + q_{11} + q_{12} + q_{22}$
$90 = q_0 + q_1 + 2q_2 + q_{11} + 2q_{12} + 4q_{22}$
$70.8 = q_0 + 2q_1 + 4q_{11}$
$100 = q_0 + 2q_1 + 2q_2 + 4q_{11} + 4q_{12} + 4q_{22}$

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

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

5. References

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

Appendix A. Proof of Revenue Maximization

The Lagrangian function for optimization problem is:

$$\mathcal{L} = y - \lambda(w_1S + w_2I + w_3P + w_4N - m)$$

$$\mathcal{L} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} - \lambda(w_1S + w_2I + w_3P + w_4N - m)$$

The first order conditions are:

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

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

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

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

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(w_1S + w_2I + w_3P + w_4N - m) = 0 \tag{A.5}$$

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 :

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

Appendix C. Proof of Profit Maximization

The profit function is written below
Profit=$(S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} - (w_1S + w_2I + w_3P + w_4N)$
We want to maximize the profit subject to $w_1S + w_2I + w_3P + w_4N = c_{thresh}$
Using Lagrange multiplier,

$$\begin{aligned}\mathcal{L} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}} - (w_1S + w_2I + w_3P + w_4N) \\ +\lambda(w_1S + w_2I + w_3P + w_4N - c_{thresh})\end{aligned} \tag{C.1}$$

$$\frac{\partial \mathcal{L}}{\partial S} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} S^{\rho-1} - w_1 + \lambda w_1 = 0 \tag{C.2}$$
$$\frac{\partial \mathcal{L}}{\partial I} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} I^{\rho-1} - w_2 + \lambda w_2 = 0 \tag{C.3}$$
$$\frac{\partial \mathcal{L}}{\partial P} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} P^{\rho-1} - w_3 + \lambda w_3 = 0 \tag{C.4}$$
$$\frac{\partial \mathcal{L}}{\partial N} = (S^\rho + I^\rho + P^\rho + N^\rho)^{\frac{1}{\rho}-1} N^{\rho-1} - w_4 + \lambda w_4 = 0 \tag{C.5}$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = w_1 S + w_2 I + w_3 P + w_4 N - c_{thresh} = 0 \tag{C.6}$$

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,

$$
\begin{aligned}
I &= \frac{c_{thresh} w_2^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \\
P &= \frac{c_{thresh} w_3^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}} \\
N &= \frac{c_{thresh} w_4^{\frac{1}{\rho-1}}}{w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} + w_3^{\frac{\rho}{\rho-1}} + w_4^{\frac{\rho}{\rho-1}}}
\end{aligned}
$$

Appendix D. Curvature Characteristic of CES

CES function:

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

Hessian Matrix

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

$\Delta_1 = (K^\rho + L^\rho)^{\frac{1}{\rho}-1} K^{\rho-1}(\frac{\rho K^{\rho-1}}{K^\rho+L^\rho} + \frac{\rho-1}{K})$
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)$\frac{\rho-1}{K} \geq \frac{\rho K^{\rho-1}}{K^\rho+L^\rho}$

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

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

$(\rho - 1) > 0$
$=> \rho > 1$

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

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

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.1 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.

Table 21: Simulation output for Random data, when $\rho < 1$

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Max. Revenue
1996	78	15	1.11	0.1000	36235.29
1997	60	35	1.11	0.1000	47096.27
1998	55	18	1.11	0.1000	32725.54
1999	44	20	1.11	0.1000	30613.65
2000	75	30	1.11	0.1000	49084.85
2001	62	15	1.11	0.1000	32023.18
2002	55	25	1.11	0.1000	38267.07
2003	49	18	1.11	0.1000	30794.72
2004	55	25	1.11	0.1000	38267.07
2005	57	24	1.11	0.1000	38229.9
2006	65	35	1.11	0.1000	49076.16
2007	46	38	1.11	0.1000	42832.04
2008	77	27	1.11	0.1000	47335.4
2009	68	32	1.11	0.1000	48107.5
2010	59	16	1.11	0.1000	32138.37
2011	48	10	1.11	0.1000	23134.73
2012	60	30	1.11	0.1000	43706.2

Table 22: Simulation output for Random data, when $\rho > 1$

Year	New Server	Power and Cooling	Elasticity(s)	ρ	Max. Revenue
1996	78	15	-10	1.1000	85.55
1997	60	35	-10	1.1000	85.57
1998	55	18	-10	1.1000	66.57
1999	44	20	-10	1.1000	58.08
2000	75	30	-10	1.1000	95.04
2001	62	15	-10	1.1000	70.62
2002	55	25	-10	1.1000	72.4
2003	49	18	-10	1.1000	61.01
2004	55	25	-10	1.1000	72.4
2005	57	24	-10	1.1000	73.4
2006	65	35	-10	1.1000	90.12
2007	46	38	-10	1.1000	75.5
2008	77	27	-10	1.1000	94.38
2009	68	32	-10	1.1000	90.3
2010	59	16	-10	1.1000	68.6
2011	48	10	-10	1.1000	53.5
2012	60	30	-10	1.1000	81.2

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.

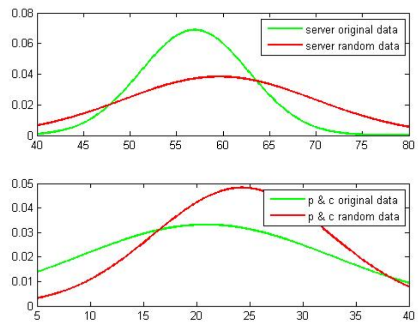


Figure I.1: The Original and Generated Server,Power & Cooling Data that follows Normal Distribution

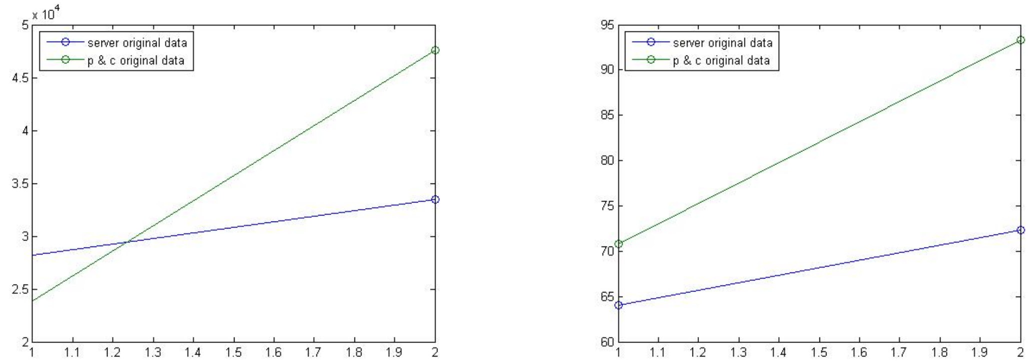
Appendix J. Non-parametric Estimation

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 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.2a and J.2b 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.2a suggests an interaction between the two cost components. Fig.J.2b reveals no interaction between the factors. The non-parametric estimation on generated data has been shown in Fig.J.4a and J.4b. None of the figures demonstrate any interaction between factors.

Appendix K. Profit Maximization

Following above description, the profit function is written as,

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



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

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

Figure J.2: Non parametric estimation on orignial data

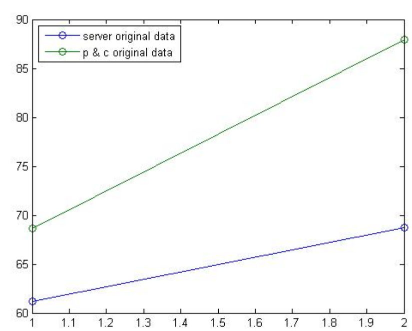
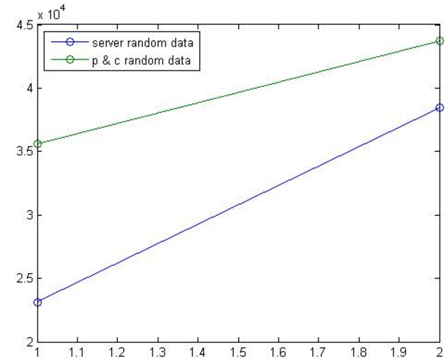


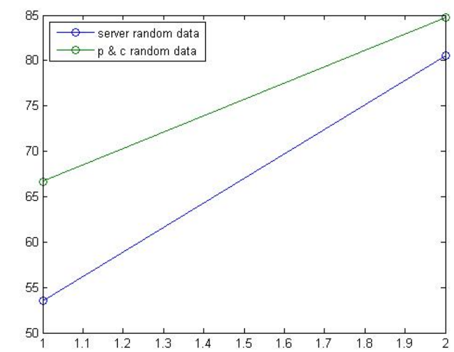
Figure J.3: Non parametric estimation on orignial data when $\rho > 1$

Say, the cost should not surpass the threshold c_{thresh}. This requires us to maximize $(S^\rho + I^\rho + P^\rho + N^\rho) - (w_1S + w_2I + w_3P + w_4N)$ subject to $w_1S + w_2I + w_3P + w_4N = c_{thresh}$.
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.

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



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



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

Figure J.4: Non parametric estimation on random data

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

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

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

The results, K.2-K.5 are analytically verified in Appendix C.