PUBLIC UTILITIES COMMISSION OF SRI LANKA

Cost (OPEX) benchmarking (Relative) of DLs

CP15/LIC/03

Licensing Division

31-Dec-15

This report provides an opportunity to identify relative performance of Distribution Licensees. Best and worst performers (Distribution Licenses) and overall cost (operational expenditure) improvement can be identified.

Table of Contents

Purpose of this Report

Objective of the proposed deliverable (CP15/LIC/03) is to carryout feasibility study on applying overall performance measuring methods as described in the “Report on electricity distribution Utilities Performance Indicators”, PUC/2013/158/EL/LI/05 *(PUCSL, 2013)*.

Aforementioned report has discussed about overall performance measuring methods and in this report these methods has been applied to calculate/estimate the overall performance relative to each DLs. Application of these benchmarking methods provide an opportunity to identify relative performance of DLs . Hence, Best and worst performers (DLs) and areas to cost (OPEX) improvement can be identified.

Measuring Overall OPEX Efficiency

Some of the commonly used electricity distribution performance indicators given below assess the performance with respect to some particular aspect of performance.

- Opex per Line Length
- Cost per Employee
- Consumer per Employee
- Sales per Employee
- Line length per Employee
- Opex per unit of sales

These indicators when taken in isolation do not provide an accurate picture of the overall efficiency or performance of the DL. These partial performance indicators can have following advantages and disadvantages *(PUCSL 2013).*

Advantages

- ✓ Easy to compute and understand
- ✓ Can be used to compare certain aspects of efficiency and productivity performance.
- ✓ Analysis can help identify trends, determine baselines and establish target performance.

Disadvantages

- × For many it cannot control for some differences in operating environment (eg: LECO vs CEB R1)
- × Can give misleading information regarding the overall economic performance of energy utilities producing multiple outputs and multiple inputs.
- × Cannot give an overall measure of potential for cost improvement.

Inputs taken to calculate these partial performance indicators or the calculated partial performance indicators itself can be used to measure overall performance of the DLs by using special methods that would produce single value giving overall efficiencies of the DLs. Following methods have been used by utility regulators worldwide *(PUCSL 2013)*.

Data Envelopment Analysis

Data Envelopment Analysis (DEA) uses linear programming to determine the efficiency frontier of the sample. The approach works by solving individual linear programming problems for each firm or observation, in which the firm's inputs and outputs are assigned a set of weights in order to maximize the ratio of weighted outputs to inputs (subject to the constraint that all efficiency scores are less than one). Under this approach, an efficient firm is one where no other firm– or linear combination of other firms - can produce more of all the outputs using less of any input. This means that the efficiency frontier is constructed from the 'envelope' of these linear combinations of input and output combinations.

A key step in DEA is the choice of appropriate input and output variables. The variables should, as far as possible, reflect the main aspects of resource-use in the activity concerned. DEA can also account for factors that are beyond the control of the firms and can affect their performance, e.g. customer density, authorized area of operation.

Note: Explanation on DEA is given in (PUCSL, 2013)

The DEA Model

The usual measure of efficiency,

$$\text{efficiency} = \frac{\text{output}}{\text{input}}$$

With multiple inputs and outputs, a common measure for efficiency is,

$$\text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}$$

Efficiency of the DL, **k**

$$\text{Efficiency of DL k} = \frac{\text{u1} \times \text{Y1} + \text{u2} \times \text{Y2} + \dots\dots\dots\dots}{\text{v1} \times \text{X1} + \text{v2} \times \text{X2} + \dots\dots\dots\dots}$$

Where,

u1- weight given to Output 1

v1 – weight given to Input 1

Y1 – Amount of Output 1 from DL k

X1 – Amount of Input 1 from DL k

$$Maximize\ Efficiency\ of\ DL\ k, \quad \frac{\sum u_i Y_{i_k}}{\sum v_i X_{i_k}},$$

$$Y_{i_k}, X_{i_k} - amount\ of\ output, input\ i\ from\ DL\ k$$

$$Subjected\ to, \quad \frac{\sum u_i Y_{i_j}}{\sum v_i X_{i_j}} \leq 1\ for\ all\ other\ DL\ j$$

$$u_i, v_i \geq 0$$

Above non linear model can be converted into a linear model. That is,

Maximize :

$$\sum u_i Y_{i_k}$$

Subjected to :

$$\sum v_i X_{i_k} = 1$$

$$\sum u_i Y_{i_j} - \sum v_i X_{i_j} \leq 0$$

$$u_i, v_i \geq 0$$

By solving above linear programming problem, the weights ui, vi can be obtained. Then the corresponding maximum efficiency of DLj with respect to other DLs can be calculated.

Advantages of DEA

- Multi-dimensional method
- Inefficient firms are compared to actual firms (or linear combinations of these) rather than to some statistical measure
- Does not require the specification of a cost or production function.
- It does not require functional relationships between input and output factors
- DEA can be implemented on a small dataset (5 DLs in case of Sri Lanka), where regression analysis tends to require larger minimum sample size in order to stand up to statistical testing.

Disadvantages

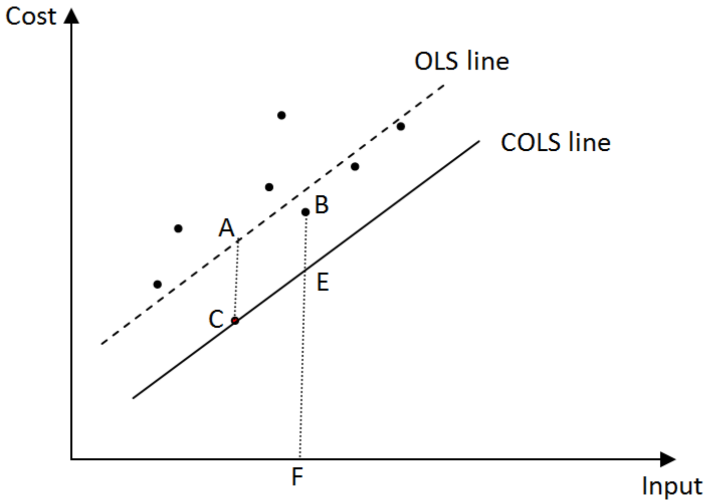
- The results could be influenced by random errors, measurement errors or extreme events
- In case of small samples and high number of input or/and output variables – danger of over-specification of model and “made-up” results for efficiency scores. As more variables are included in the model, the number of firms on the efficient frontier increases.

- The efficiency scores tend to be sensitive to the choice of input and output variables and, in some circumstances, inappropriate choices may lead to relatively inefficient firms defining the frontier.

Corrected Ordinary Least Squares Method

This method utilizes the standard regression technique, with the efficiency measures computed from the residuals. With this approach, the frontier is estimated (rather than calculated) using statistical techniques. A functional form for the production / cost function is specified, and this is estimated using ordinary least squares (OLS) techniques. The calculated line of best fit is then shifted to the efficient frontier by adding the absolute value of the largest negative estimated error to that of the other errors (for a cost function). This is therefore a 'corrected' form of OLS is used, COLS, rather than the standard form *(CEPA 2003).*

Following figure illustrates a COLS model with a single cost input C and one output Y. The efficient cost equation (COLS line) is estimated using Ordinary Least Squares (OLS) regression and then shifted by AC to on which the most efficient firm C lies. The efficiency score for an inefficient firm B is calculated as EF/BF.



Key Assumptions

- The COLS method requires specification of a cost or production function and therefore involves assumptions about technological properties of the firms' production process.
- It is assumed that all deviations from the frontier are due to inefficiency. There are therefore no measurement errors.

Advantages

- ✓ Easy to implement
- ✓ Allows statistical inference about which parameters to include in the frontier estimation.
- ✓ Requires no assumptions about the distribution of the inefficiency scores.

Disadvantages

- × The estimated parameters may not make engineering sense
- × The method makes no allowance for stochastic errors and relies heavily on the position of the single most efficient firm in the sample. (Stochastic conveys the idea of randomness. DLs internal environments can be affected by random events in the external environment)
- × Similar to DEA, COLS assumes that all deviations from the frontier are due to inefficiency.
- × It is not possible to identify firms to which inefficient firms are being compared in the same sense as DEA. All firms are being compared to a frontier defined by one frontier firm. However there may be no 'nearby' frontier firms.
- × Requires large data volume in order to create robust regression relationship
- × Sensitive to data quality (the company setting frontier could be an outlier)

Variables required to Assess OPEX Efficiency

The Opex must be efficiently utilized to provide the energy bought from Transmission Licensees to consumers. Therefore, care should be taken to select output variables that are Opex intensive or strong cost drivers. Relevant data should be accurate and importantly be practical to collect from the DLs timely. In regulators point of view, following factors were considered when selecting variables.

- Quality of the data
- Availability
- Ease of collection.
- Relevance to the business – i.e. electricity distribution business
- International Practices/ Reviews
- Reflecting the scale of operation.
- Cost drivers – variables having major influence on the cost of operation.

Following variables are identified as the important.

- Energy Purchased and Sold (GWh)
- Total number of consumers
- No. of employees
- Distribution lines length (km) – This includes MV and LV network length
- No. of substations
- Authorized operation area (km^2) –This is a constant for each licensee.
- Operational Expenditure (LKR Million) – with breakdown

Note that, in international practices, the use of supply/service quality as a variable is rare. Most of the countries reviewed run separately a quality-of-service reward/penalty regime *(USAID, 2004)*. In Sri Lanka, the supply/service quality is to be determined according to the drafted Electricity Distribution performance regulations, where penalties have been introduced for underperformance.

Data Sources

The lifeblood of this exercise is the data. In order to obtain accurate data following sources were considered to be the most effective as these sources are publicly available or submitted to PUCSL by DLs as a routine submission.

- Annual Reports published by CEB
- Annual Reports published by LECO
- Annual Statistical Digests published by CEB
- Annual Statistical Digests published by LECO
- Biennial MV Distribution Development Plans submitted by DLs (CEB/LECO) as a requirement of Distribution Code.
- Financial Accounts submitted by DLs (CEB/LECO) as required by Distribution License.

Further, since these sources are compiled by Licensees they have no reason to argue on the accuracy of the data which may affect the final efficiency results. In addition PUCSL require minimal effort to extract these data, prohibiting the requirement for exclusive inquiry from DLs. Mainly, information on following input/output variables which are useful for determining relative efficiencies (OPEX efficiency) could be extracted from aforementioned data sources.

Information	Unit
Units Sold	GWh
total number of consumers	Nos
No. of employees	Nos
Total MV line length	km
Total LV line length	km
LV distribution substations	Nos
OPEX	LKR(Mil)

Following table depicts the data set that was used for this analysis.

DL-Year	Opex-Adjusted to year 2014 (LKR Mil)	Units Sold (GWh)	Number of consumers	No. of employees	Network Length (km)
DL1-2014	5,877.61	3,047	1,504,453	2,888	50,622
DL1-2013	Not Available	2,928	1,432,024	Not Available	47,819
DL1-2012	5,762.83	2,892	1,351,767	Not Available	42,998
DL1-2011	5,412.37	2,797	1,265,463	Not Available	41,169
DL2-2014	7,777.49	3,377	1,872,836	3,876	45,259
DL2-2013	Not Available	3,127	1,807,970	3,914	43,555
DL2-2012	6,539.07	3,003	1,563,349	Not Available	37,940
DL2-2011	Not Available	2,844	1,488,745	Not Available	36,639
DL3-2014	4,307.56	1,828	1,106,161	2,379	34,988
DL3-2013	Not Available	1,891	1,062,848	2,274	33,564
DL3-2012	3,474.97	1,938	1,188,156	2,522	34,134
DL3-2011	4,416.18	1,846	1,116,039	Not Available	32,113
DL4-2014	3,691.91	1,458	934,080	1,977	28,069
DL4-2013	Not Available	1,367	907,917	1,975	27,536
DL4-2012	3,474.97	1,338	876,588	2,015	26,260
DL4-2011	3,467.61	1,269	847,199	2,031	26,137
LECO-2014	2,181.00	1,272	526,990	1,474	4,337
LECO-2013	1,778.80	1,221	520,676	1,462	4,296
LECO-2012	1,801.31	1,216	500,783	1,463	4,247
LECO-2011	1,818.57	1,184	491,042	1,451	4,340
LECO-2010	1,567.53	1,123	473,079	1,338	4,151

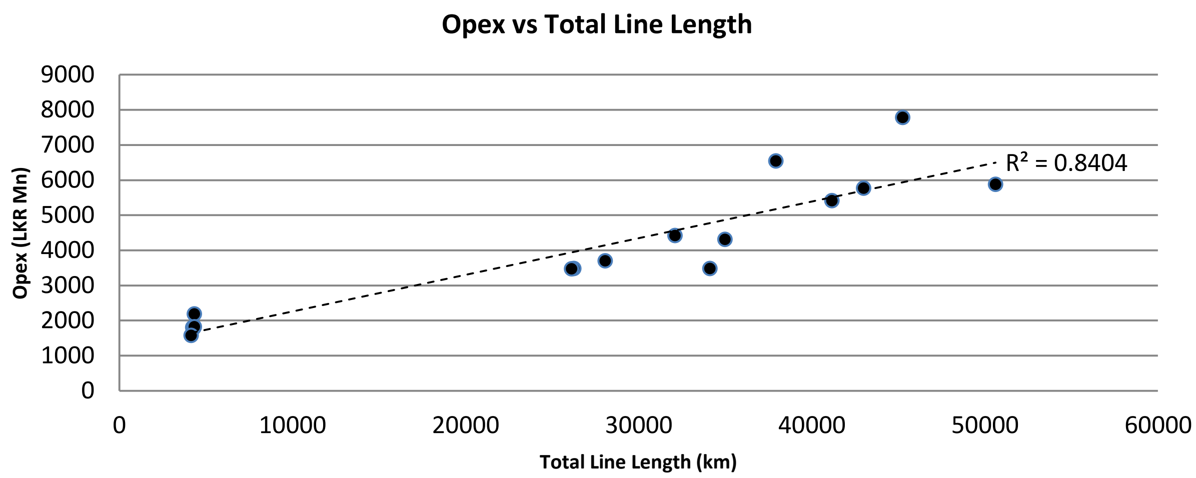
Opex values are adjusted to present values using CPPI. Relevant CCPI values are given in following table.

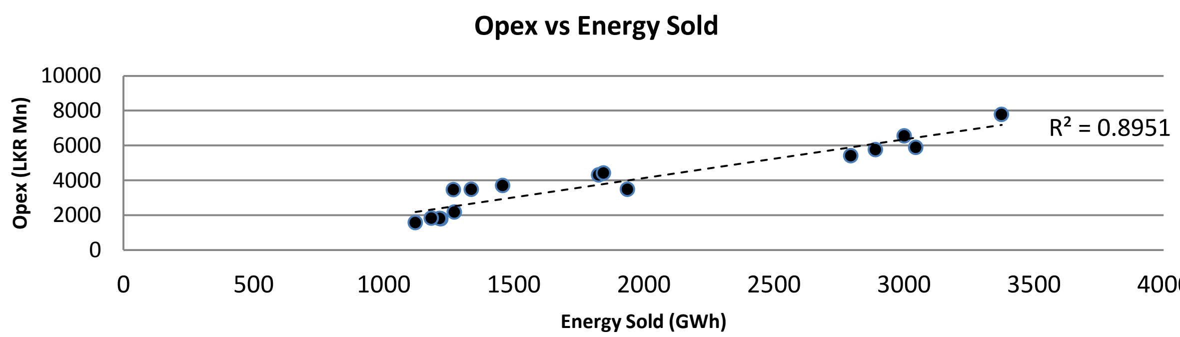
Year	Avg. CCPI
2014	179.8583
2013	174.15
2012	162.8917
2011	151.4667
2010	141.9333

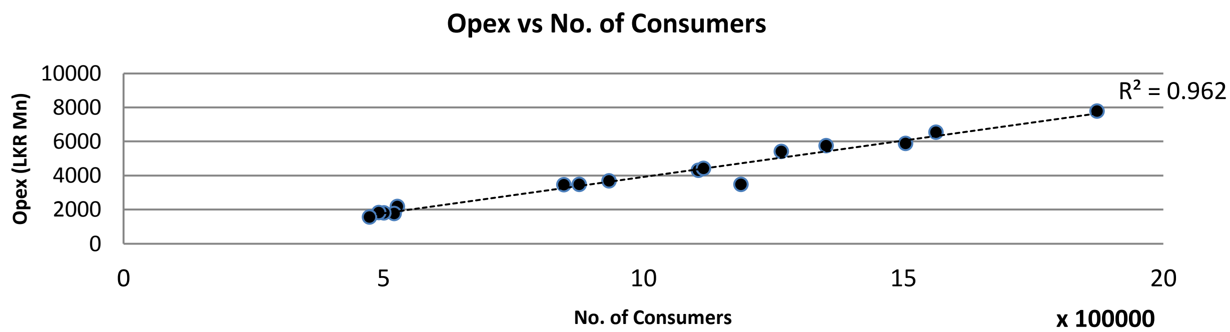
(CCPI, 2006)

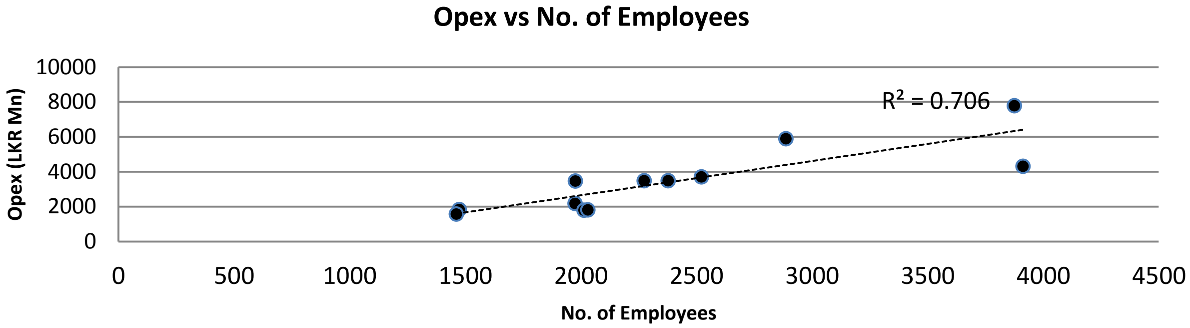
Correlation

The correlation of each variable that may contribute to Opex was checked by calculating the coefficient of determination (R^2). Here we have used scatter plots of each variable against Opex (Adjusted to present value). The R^2 is interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variable.







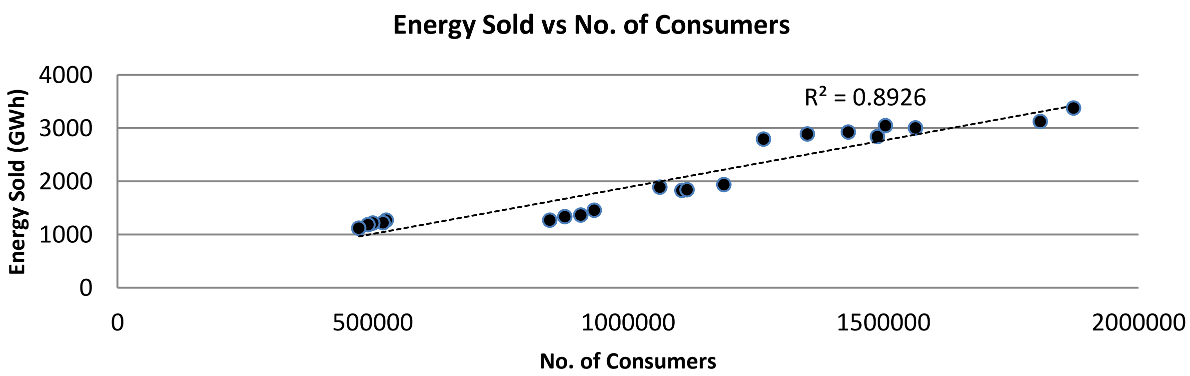


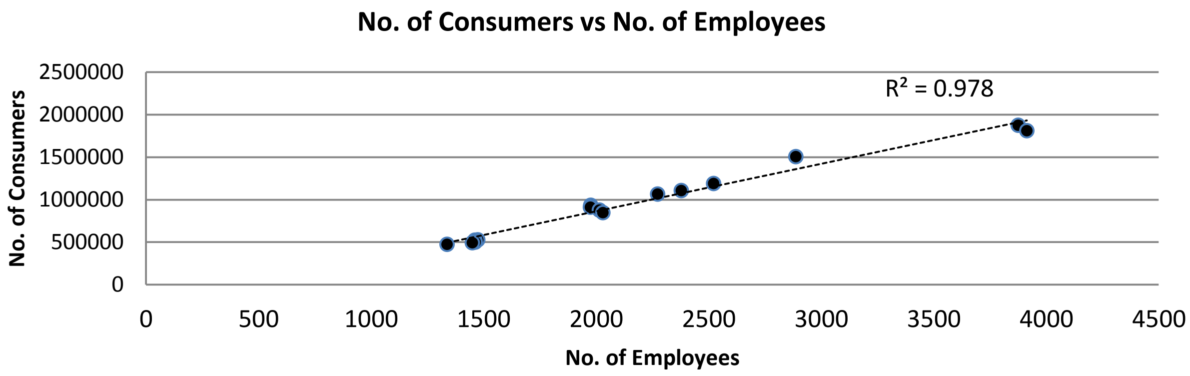
Variable	Adjusted-R^2 (Association with Opex)
Energy Sold	0.895
No of Consumers	0.962
No. of Employees	0.760
Total Network Route Length	0.840

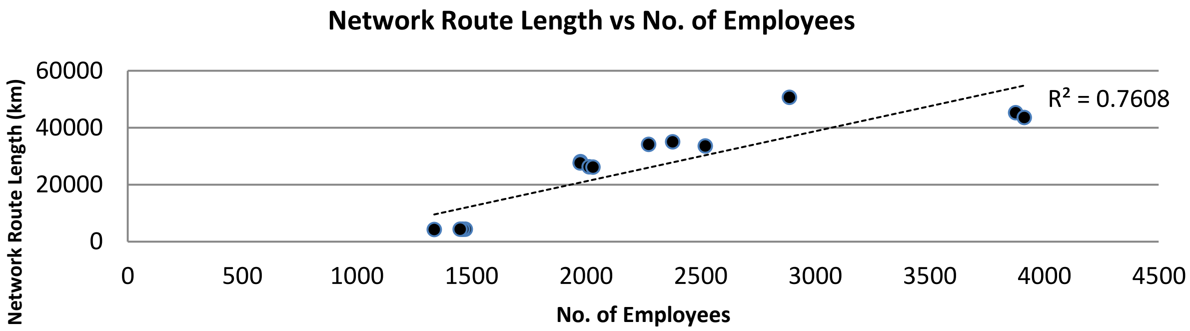
Scatter plots and respective R^2 values indicate that Energy sold, No. of consumers, number of employees and network line length are having a relatively strong linear association with operational expenditure. Hence these variables can be treated as **cost** (Opex) **drivers**. Note that Number of Distribution substation could not be taken for analysis since unavailability of data relevant to CEB DLs for years 2011, 2012 and 2013.

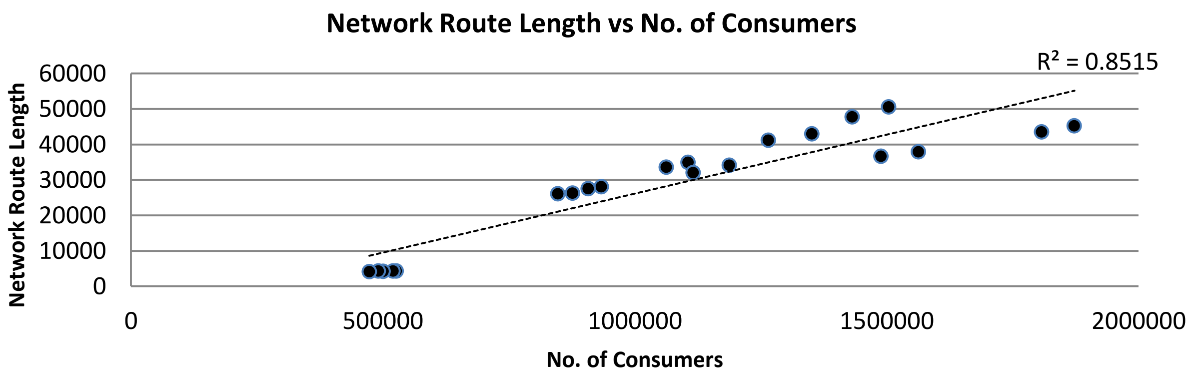
Independent Predictors

It is important to check the correlation between predictor variables as well. They should be independent of each other. If mutually dependent predictors are included, then the model can become complicated. Predictor variables (for regression models) are Energy sold, number of consumers, number of employees and network route length. This checking was done using following scatter plots.









From above scatter plots, following relations could be observed.

Predictor variables	Coefficient of Determination (R^2)
Energy Sold and No of Consumers	0.8926
No. of Consumers and No. of Employees	0.978
No. of Employees and Network Rout Length	0.7608
Network Route Length and No of Consumers	0.8515

According to above table it can be concluded that ”number of consumers” and “Network Route Length” are the best predictor variables (relatively weak correlation) that could be used in a regression model for estimating the Opex.

COLS Model Selection

Starting with the full model, (i.e. the model containing all the variables) the backward elimination process was carried out until the model with highest Adjusted R^2 is reached (eliminating one variable at a time). The adjusted-R^2 values for each linear model are given below. It depicts that the highest Adjusted-R^2 is when the number of consumers are taken as predictor variable.

Model No.	Units Sold (GWh)	Number of consumers	No. of employees	Network Length (km)	Adjusted R-Squared
1	X	X	X	X	0.9414
2	X	X	X		0.9469
3	X	X		X	0.9442
4	X		X	X	0.9485
5		X	X	X	0.9462
6	X	X			0.9497
7	X			X	0.9142
8	X		X		0.9477
9		X	X		0.9521
10		X		X	0.9502
11			X	X	0.9511
12	X				0.8624
13		X			**0.9539**
14			X		0.9506
15				X	0.8146

Even though the parsimonious model (highlighted in above table) indicated that taking the “No of Consumers” along is sufficient for determining Opex, it is paramount to incorporate other cost drivers into the model. This has been the practice in international benchmarking studies.

Our study on selecting the best model also indicated that the Adjusted-R2 for each model considered above, does not vary significantly (varying around 0.94 ~ 0.95 in most cases). Therefore it is decided to consider maximum possible the cost driving predictor variables.

In the regression output of the model 1(see above table), the coefficient of the predictor variable "No of consumers" indicates a negative value. This is not practical, as the variable "no of consumers" is a cost driver. Hence Model 1 is rejected.

	Coefficients
Intercept	-1283.86609
Units Sold (GWh)	0.464119124
Number of consumers	-0.001052876
No. of employees	2.080015729
Network Length (km)	0.025821865

The Model 4 is the next best model we could pick, having higher adjusted-R^2 and most of the cost drivers as predictor variables (i.e. Units sold, No. of employees and Network length). Coefficients are as follows.

	Coefficients
Intercept	-1120.160126
Units Sold (GWh)	0.375428594
No. of employees	1.693658905
Network Length (km)	0.018435715

RESIDUAL OUTPUT

Observation (DL-Year)	*Predicted Opex-Adjusted (LKR Mil)*	*Residuals*
DL1-2014	5,848.310	29.299511
DL2-2014	7,546.666	230.819332
DL3-2014	4,240.367	67.188508
DL3-2012	4,508.113	-1033.146216
DL4-2014	3,293.051	398.857183
DL4-2012	3,279.008	195.958827
DL4-2011	3,277.934	189.674695
LECO-2014	1,933.951	247.049324
LECO-2013	1,893.559	-114.759978
LECO-2012	1,892.481	-91.172842
LECO-2011	1,861.777	-43.204596
LECO-2010	1,644.094	-76.563747

For observations in the year 2014, the corresponding OLS line is represented by,

$$OPEX\ in\ LKR\ Mn = -1120.16 + 0.3754 * (Units\ Sold\ in\ GWh) + 1.6937 * (No.of\ Employees) + 0.01844 * (Network\ Length\ in\ km)$$

By shifting the OLS line to cross the most Opex efficient point (DL1, for 2014 values) the Corrected OLS line can be formed. It is given by the following line.

$$OPEX\ in\ LKR\ Mn = (-1120.16 + \mathbf{29.2995}) + \ 0.3754 * (Units\ Sold\ in\ GWh) + 1.6937 * (No.of\ Employees) + 0.01844 * (Network\ Length\ in\ km)$$

Now the efficient Opex and relative efficiencies are given as follows (according to the new COLS line). Efficiency is given by,

$$Relative\ Efficiency = \frac{Predicted\ Efficient\ Opex\ by\ COLS}{Actual\ Opex} * 100$$

Observation (2014)	**Actual Opex (LKR Mil)**	***Eficient Opex (LKR Mil)***	***Relative Efficiency (relative to most efficient DL , i.e. DL1)***
DL1	5,877.61	5,877.61	100.0
DL2	7,777.49	7,575.97	97.4
DL3	4,307.56	4,269.67	99.1
DL4	3,691.91	3,322.35	90.0
DL5	2,181.00	1,963.25	90.0

Implementation of Data Envelopment Analysis.

For the DEA Model , Input and Output variables were taken in following manner.

Input Variable	Reasons
Opex	Using Opex as input the DL carry out its business of distribution of electricity. No. of employees was not considered as an input since Opex itself carries more than 50 % as employee cost.

Output Variables	Reason to Consider
Energy Sold	Main product of the Distribution business. A cost driver. Failure to maintain the system properly causes reduction in reliability of supply and consequently reducing the output (Energy sold).
No of Consumers	A cost driver. Portion of Opex used for managing the consumer base.
Total Line Length	A cost driver. Portion of Opex is used for maintain the infrastructure.

When implementing DEA care has to be taken to verify the results with other methods. A rough rule of thumb which can provide guidance is to choose a value of n that satisfies

$n \geq \max\{m \times s, 3(m + s)\}$ *(Cooper, 2001)*

where n = number of DLs, m = number of inputs and s = number of outputs. Otherwise all the DLs would get closer to 100% efficiency and discrimination could be difficult. With small sample and high number of input / output variables there is a danger of receiving made-up results for efficiency scores (*ERRA 2002*).

Using the Frontier Analyst software application by Banxia Software the DEA analysis was carries out and the efficiency scores are given in following table. In following table the right most column indicates the relative efficiencies of DLs in year 2014 where efficiency of DL1 was taken as 100% since it is the DL having highest efficiency (91.5%) among all DLs for 2014.

	Input	Output				
DL-Year	Opex-Adjusted to year 2014 (LKR Mil)	Units Sold (GWh)	Number of consumers	Network Length (km)	Relative Efficiency Score	Relative Efficiencies for 2014
DL1-2014	5,877.61	3,047	1,504,453	50,622	91.5	100.0
DL2-2014	7,777.49	3,377	1,872,836	45,259	72.6	79.3
DL3-2014	4,307.56	1,828	1,106,161	34,988	82.7	90.4
DL3-2012	3,474.97	1,938	1,188,156	34,134	100.0	N/A
DL4-2014	3,691.91	1,458	934,080	28,069	77.4	84.6
DL4-2012	3,474.97	1,338	876,588	26,260	76.9	N/A
DL4-2011	3,467.61	1,269	847,199	26,137	76.7	N/A
LECO-2014	2,181.00	1,272	526,990	4,337	81.4	89.0
LECO-2013	1,778.80	1,221	520,676	4,296	96.5	N/A
LECO-2012	1,801.31	1,216	500,783	4,247	94.2	N/A
LECO-2011	1,818.57	1,184	491,042	4,340	90.9	N/A
LECO-2010	1,567.53	1,123	473,079	4,151	100.0	N/A

Conclusion and Recommendations.

According to both DEA and COLS methods of evaluation, DL1 is the one having highest efficiency. Relative to DL1 (CEB Region 1) LECO is approximately 90% efficient. In both methods CEB Region1, CEB Region 3, CEB Region 4 have got their ranking as 1,2 and 4 respectively. Different ranks have observed in CEB Region 2 and LECO when evaluated using these two different methods. Following two tables depicts the results obtained through DEA and COLS methods.

DL	Efficiency Score	
	DEA	COLS
CEB Reg 1	100.0	100.0
CEB Reg 2	79.3	97.4
CEB Reg 3	90.4	99.1
CEB Reg 4	84.6	90.0
LECO	89.0	90.0

DL	Rank	
	DEA	COLS
CEB Reg 1	1	1
CEB Reg 2	5	3
CEB Reg 3	2	2
CEB Reg 4	4	4
LECO	3	4

Differences in efficiency score may have been caused by lack of data such as number of employees and Opex of several years. By getting cleaned data more robust scores would have been obtained. Further COLS method ideally requires large data volume in order to create robust regression relationship. For small sample sizes (5 DLs) DEA is more appealing.

As the regulator these methods/results can be taken as inputs to future price reviews. Above results offer conclusive evidence that DL1 i.e. CEB Region 1 can be taken as the efficient frontier. Cost (Opex) evaluation of other DLs can be done using DL1 as an overall benchmark.

It is important to note that development of functional forms for Opex is difficult since it has to derive from relatively small sample size. Therefore DEA method would be more appealing for future studies. Checking for robustness could be done using the COLS method.

Reference

PUCSL 2013, Report on electricity distribution Utilities Performance Indicators”, PUC/2013/158/EL/LI/05

USAID 2004, Concept Paper; Performance benchmarks for electricity distribution companies in South Asia, Nov 2004; USAID SARI/Energy program, www.sari-energy.org (South Asia Regional Initiative for Energy)

ACCC/AER 2012, Benchmarking OPEX and CAPEX in Energy Networks, ACCC/AER Working paper series, working paper No.06, May 2012.

CEPA 2003, Final report titled “Background to work on assessing efficiency for the 2005 price control review” prepared for Ofgem by Cambridge Economic Policy Associates (CEPA), 2003.

Talluri 2000, Data envelopment analysis : Models and extensions, Srinivas Talluri; Decision Line, May 2000.

Cooper 2001, Sensitivity and Stability Analysis in DEA: Some Recent Developments,W. W. Cooper, Shanling Li, L. M. Seiford, Ph.D., kaoru tone, R. M. Thrall, J. Zhu, Journal of Productivity Analysis, 15, 217–246, 2001

ERRA 2002, Benchmarking in ERRA (Energy Regulators Regional Association) Countries, Issue Paper 2002, ERRA Licensing and Competition committee.

CCPI 2006,
http://www.statistics.gov.lk/price/ccpi%28new%29/Movements%20of%20CCPI%28200607%29.pdf