Data Envelopment Analysis and the implementation of their Applications in Python

R.Venugopal[1] and C.Veeramani[2]
{venusm80@gmail.com[1], veerasworld@yahoo.com[1]}

Research Scholar, Department of Applied Science (Mathematics),PSG College of Technology, Coimbatore, Tamil Nadu[1], Assistant Professor, Department of Applied Science (Mathematics), PSG College of Technology, Coimbatore - 641 004, Tamil Nadu[2]

Abstract. Data envelopment analysis (DEA) is highly used to rank the companies in the stock market. This study presents a DEA framework for ranking the companies in the stock market. In this regard, the data is collected from 24 companies listed in the NASDAQ stock market. Then the input and output measures are calculated from the collected data. Finally, the companies are ranked by using CCR, CCR-Aggressive, and CCR-Benvelovent DEA models based on input and output values.

Keywords: Data envelopment analysis, influence criteria, NASDAQ stock exchange.

1 Introduction

This The performance measurement methodology of DEA was proposed by Charnes et al. [2]. It is applied to assess the relative performance of DMUs when diverse inputs and outputs are present. DEA is a powerful decision-making tool that is applied in a variety of real-world problems ([3], [11], [7], [11]). Recently, Orkcu et al. [16] presented two-stage DEA manufacturing systems based on a neutral cross-efficiency evaluation. Oukil [14] built a comprehensive ranking algorithm based on DEA framework. The supper performance DEA model was used by Zamani et al. [18] to estimate the ideal stock in the Mumbai stock exchange. For the Latin American stock markets, Minutolo et al. [12] suggested a unique wavelet approach for portfolio selection. In a fuzzy context. Recently, many portfolio selection frameworks is developed based on the DEA approach ([4], [5], [9], [8], [1], [17]). To avoid the difficulties of choosing between the two alternative formulations, we suggest a DEA model for cross-efficiency evaluation in this study. The rest of the paper is organized into four sections explaining three DEA models with their model formulations. Implementation of all the three models in python and their output is displayed. Section 6 has a detailed case study of the stock exchange market dataset and its pseudo-code of CCR model and Cross-efficiency model. Final conclusion in the last section ranking or conclusion.

2. Proposed DEA framework

DEA employs a linear programming approach to tightly enclose observable input-output vectors. Without making any assumptions about data distribution, DEA permits numerous

ICCAP 2021, December 07-08, Chennai, India

DOI 10.4108/eai.7-12-2021.2314630

inputs–outputs to be considered simultaneously time. Efficiency is determined as a change of ratio in inputs and outputs in each scenario.

2.1 CCR Model

Consider DMUs have m inputs and s outputs. Let x_{ij}represents input, where $i = 1,2,\dots,m$ and y_{rj}denotes the output where $r = 1,2,\dots,s$ of DMU_j where $j = 1,\dots,n$. The efficiency values are determined as, θj $=\sum_{r=1}^{s} u_r y_{rj,} j = 1,\dots,n$

The CCR model stated as follows:

$$\text{Maximize } \theta_{kk} = \frac{\sum_{r=1}^{S} u_{rk} y_{r_k}}{\sum_{i=1}^{m} v_{ik} x_{i_k}}$$

$$\text{Subject to } \theta_{jk} = \frac{\sum_{r=1}^{S} u_{rk} y_{r_j}}{\sum_{i=1}^{m} v_{ik} x_{i_j}} \leq 1, j = 1, \dots, n$$

$u_{rk} \geq 0 , r = 1, \dots, s$

$v_{ik} \geq 0 , i = 1, \dots, m$

The above model can be turned into the following linear program:

$$\text{Maximize } \theta_{kk} = \sum_{r=1}^{S} u_{rk} y_{r_k}$$

$$\text{Subject to } \theta_{jk} = \sum_{i=1}^{m} v_{ik} x_{i_k} = 1$$

$$\sum_{r=1}^{s} u_{rk} y_{rj} - \sum_{r=1}^{s} v_{ik} x_{ij} \leq 0, \text{j} = 1, \dots, \text{n}$$

$u_{rk} \geq 0 , r = 1, \dots, s$

$v_{ik} \geq 0 , i = 1, \dots, m$

2.2 Cross-Efficiency Model

If the non-uniqueness of input and output values is not handled, the cross-efficiency evaluation will be harmed. To overcome this, Sexton et al. [15] has introduced a new method that secondary that optimizes the input and output values while keeping the CCR efficiency unchanged. There are two sorts of cross-efficiency models: aggressive and beneficent, which are discussed as follows:

Aggressive Model:

$$\text{Minimize} \sum_{r=1}^{S} u_{r_k} \left(\sum_{j=1 j \neq k}^{n} y_{rj} \right)$$

$$\text{Subject to, } \sum_{r=1}^{m} v_{i_k} \left(\sum_{j=1 j \neq k}^{n} x_{ij} \right) = 1$$

$\sum_{r=1}^{S} u_{rk} y_{r_k} - \theta_{kk}^{*} \sum_{i=1}^{s} v_{ik} x_{i_k} = 0$

$\sum_{r=1}^{s} u_{rk} y_{rj} - \sum_{r=1}^{s} v_{ik} x_{ij} \leq 0 , \text{j} = 1, \dots, \text{n} , j \neq k$

$u_{rk} \geq 0 , r = 1, \dots, s$

$v_{ik} \geq 0 , i = 1, \dots, m$

Benevolent Model:

Maximize $\sum_{r=1}^{S} u_{r_k} \left(\sum_{j=1 j\neq k}^{n} y_{rj} \right)$

subject to, $\sum_{r=1}^{m} v_{i_k} \left(\sum_{j=1 j\neq k}^{n} x_{ij} \right) = 1$

$\sum_{r=1}^{S} u_{rk} y_{r_k}$ - $\theta_{kk}^{*} \sum_{i=1}^{s} v_{ik} x_{i_k} = 0$

$\sum_{r=1}^{s} u_{rk} y_{rj} - \sum_{r=1}^{s} v_{ik} x_{ij} \leq 0$, j = 1, ..., n , $j \neq k$

$u_{rk} \geq 0 , r = 1, \dots , s$

$v_{ik} \geq 0 , i = 1, \dots , m$

where θ^{*}_{kk} is the efficiency of CCR model.

2.3 Python Programming

This section discusses the Python Algorithms for DEA. Here we will use python to call PULP which will then call solver to solve the linear programming inequalities. Pulp stands for Python Linear Programming. Pulp is a LP modeler and a free open source software. It is used to describe optimization problems as mathematical problems. Algorithm 1 explains about CCR code input will be the necessary excel or csv file using pandas library and we will be importing pulp library for solving LP problem. Then we will declare the variables and adding the constraints and objective function accordingly. Similarly Algorithm2 explains about the Cross efficiency model. Algorithms differ from the constraints and objective function. Locating each value from the csv or excel file is quite easy. Python algorithm for solving CCR and cross efficiency model is given algorithm 1 and algorithm 2 respectively.

Algorithm 1: CCR Code

```
input  : Csv /Excel file as df
output: Efficiency score
begin
    Create a empyt list for  i ∈ n do
        LpProblem(Maximize or Minimize)
        x ⟵ LpVariable('x')
        y ⟵ LpVariable('y')
        prob + Σx ⟵ df.loc[i,' ColumnName']
        prob + Σy ⟵ df.loc[i,' ColumnName'] == 1
        for j ∈ n do
            prob + Σx ⟵ df.loc[j,' ColumnName'] ≤ prob + Σy ⟵
            df.loc[j,' ColumnName']
            solve() append(objective)
        end
    end
    n determines the number of DMU
end
```

Algorithm 2: Cross-Efficiency Code

```
input  : Csv /Excel file as df
output: Efficiency scores and Rank
begin
    Create a empyt list for  i ∈ n do
        LpProblem(Maximize or Minimize)
        x ⟵ LpVariable('x')
        y ⟵ LpVariable('y')
        prob + Σx ⟵ df.loc[i,' ColumnName']
        prob + Σy ⟵ df.loc[i,' ColumnName'] == 1
        for j ∈ n,j≠i do
            prob + Σx ⟵ df.loc[j,' ColumnName'] ≤ CCR − Score * (prob + Σy ⟵
            df.loc[j,' ColumnName'])
            solve() append(objective)
        end
    end
    n determines the number of DMU
end
```

3. Case Study

In this section we rank the companies in the NASDAQ stock exchange based on their performance, which is listed in Table 1. In this regard, the data collected form 24 firms listed in the NASDAQ stocks for the financial year (period 1 June 2020 to 31 December 2020). The 24 firms listed in the NASDAQ stocks are shows in Table 3. The 24 firms financial ratio consider as criteria on inputs and outputs. Parameters followed by Lim et al, [10] represent a range of performances perspectives of a firm: Profitability, assetutilization, liquidity, leverage and growth. Here we consider

Inputs variables [10] : Receivable turnover(C1), Inventory turnover(C2), Asset turnover(C3), Current ratio(C4), Quick ratio(C5), Debt to equity ratio(C6), Leverage ratio(C7), Solvency ratio-I(C8), Solvency ratio-II(C9).

Outputs variables [10]: Return on equity(A1), Return on assets(A2), Net profit margin(A3), Earnings per share(A4).

The input parameter values and output parameter values are calculated by using the collected data, which is given in Table 2. Table 2 considered as a decision matrix. For every firm CCR efficiency and Cross-efficiency in both Aggressive and Benevolent way ranking is shown in Table 3.

Table 1: List of Firm Details

S.NO	Name
1	MSFT Microsoft Corporation

2	AAPL Apple Inc.
3	Amazon
4	GOOGL COMPUTER SOFTWARE
5	Adobe Inc.
6	FIRST SOLAR
7	PEPSICO
8	SANOFI
9	NVIDIA
10	Mondelez
11	Qualcomm
12	Celgene
13	T-Mobile
14	CSX
15	Walgreens
16	Vodafone group
17	Pharmaceuticals
18	Micro technology
19	Amazon.com, Inc.
20	TexasInstruments Incorporated
21	Biogen Inc
22	Illumina, Inc
23	Maruti Suzuki India Ltd
24	Advanced Micro Devices, Inc.

Table 2: Input and Output values (Decision Matrix)

S.NO	C1	C2	C3	C4	C5	C6	C7	C8	C9	A1	A2	A3	A4
1	46.94	137.24	5.60	25.02	103.90	4.33	22.14	5.31	12.14	288.72	138.13	241.56	23.04
2	136.06	799.41	12.35	25.69	62.40	3.60	27.63	6.42	13.63	451.17	232.44	272.49	66.58
3	277.30	117.80	26.27	16.94	33.38	14.33	73.11	10.67	0.00	336.35	56.00	29.30	42.26
4	60.87	713.94	5.31	54.20	7.16	0.46	12.48	1.97	2.48	150.75	121.28	226.50	206.15
5	60.89	42.53	5.32	46.24	174.35	0.00	11.58	1.36	1.58	159.75	138.06	260.34	50.55
6	87.76	0.00	6.93	51.65	146.79	0.00	11.37	1.16	1.37	232.30	204.17	288.85	41.53
7	0.00	2.34	0.00	0.00	98.54	59.75	8.05	49.75	0.00	16.89	69.89	38.74	0.00
8	70.18	67.37	4.93	35.63	129.33	0.82	14.09	2.83	4.09	57.56	44.69	75.78	22.76
9	33.84	5.51	0.41	33.72	172.43	0.32	19.50	2.99	9.11	-1940.32	-291.03	-57.87	-15.19
10	132.67	43.66	6.05	22.48	73.55	2.49	15.85	3.61	5.85	197.23	125.37	205.53	22.42
11	84.25	91.75	8.91	12.05	40.03	18.88	44.41	7.45	34.41	418.00	98.46	109.38	44.41
12	68.60	0.00	4.41	24.35	109.25	2.62	16.36	3.84	6.36	130.42	77.69	170.78	18.74
13	48.23	16.93	3.52	36.15	19.99	2.67	17.36	4.23	7.36	103.10	60.13	152.37	24.50
14	94.10	49.10	7.45	51.32	876.86	1.96	14.97	3.26	4.97	182.99	121.32	159.99	19.11
15	68.51	57.83	4.47	8.13	163.59	6.64	24.36	5.88	14.36	94.28	39.64	111.80	21.22

16	118.09	65.15	4.67	32.13	876.54	19.16	48.57	3.73	38.57	-384.86	83.53	197.70	22.86
17	71.64	0.00	3.21	10.88	273.68	1.89	67.12	8.36	57.12	279.28	41.78	136.45	30.21
18	62.42	12.67	4.34	42.08	1251.3	13.11	30.78	5.91	20.78	303.92	105.23	239.42	22.50
19	457.39	171.34	7.02	16.62	509.03	16.49	35.65	7.11	25.65	-219.74	-52.43	-25.73	-9.03
20	112.52	140.58	3.57	12.04	254.94	9.70	30.47	6.69	20.47	203.10	68.65	196.72	23.19
21	238.00	82.64	20.78	13.45	158.24	3.35	20.37	5.00	10.37	142.23	70.40	33.96	30.82
22	98.76	0.00	0.62	10.61	287.21	1.20	26.37	5.90	16.37	70.99	27.14	455.93	45.61
23	46.29	618.52	3.09	7.67	171.27	5.24	18.72	4.63	8.72	63.34	36.10	210.36	50.09
24	92.95	31.73	4.29	33.88	955.83	2.98	21.32	5.05	11.32	-314.50	-117.47	-1306.58	-6.60

Table 3: Efficiency Score and rank of the DMUs

DMU	CCR	Aggressive	Rank	Benevolent	Rank
1	1.0000	0.3750	16	0.6801	20
2	1.0000	0.2803	12	0.7609	22
3	1.0000	0.0369	5	0.3270	7
4	1.0000	0.3580	15	0.7128	21
5	1.0000	0.5993	21	0.8009	23
6	0.4639	-0.0246	4	0.1736	6
7	1.0000	0.2729	11	0.6451	17
8	1.0000	0.4227	17	0.4628	13
9	0.7609	0.2602	9	0.3869	10
10	1.0000	0.2728	10	0.3588	9
11	0.7622	0.0845	7	-0.3371	2
12	1.0000	0.5392	20	0.4498	12
13	0.0209	-0.1351	3	-0.2720	3
14	1.0000	0.2860	14	0.4681	14
15	0.8641	0.1645	8	0.3301	8
16	1.0000	0.7956	22	0.6430	16
17	0.3320	-1.2404	1	-1.0298	1
18	1.0000	-0.3938	2	0.6656	19
19	1.0000	0.8163	23	0.5473	15
20	1.0000	0.4599	18	0.6510	18
21	1.0000	0.8435	24	0.8445	24
22	0.8899	0.4911	19	0.4210	11
23	1.0000	0.0626	6	0.0682	5
24	1.0000	0.2859	13	-0.0372	4

4. Conclusion and future work

For comparing and assessing DMUs, the DEA approach is a useful tool. Thus CCR and Cross-efficiency models are implemented successfully in python and each model gives us efficiencies of each DMU. One model is more efficient than the other in comparison with another one. The stock exchange market dataset is taken under consideration for the application of these models and found the best companies. The experimental of 24 firm data outcome of the results implemented these models and order the firms. This paper proposed a portfolio selection strategy that can use to pick stocks in major stock markets. The CCR models can optimize the input weights and output weights but there is no guarantee that these two models will lead to the same efficiency ranking for all the DMUs. Each model has its own disadvantage where each model overcomes the cons of another model successfully provides us in finding the most efficient Decision-Making Unit (DMU). The future work that may be an extension of this study will be the study of the neutral DEA method. As from the study Aggressive and Benevolent may not provide us the best efficient DMU whereas its disadvantages are resolved in Neutral DEA. Further, the scoring value of the DEA model is considered as the weights of the criteria. Based on this weight, one can rank the companies by using different MCDMs and competitive studies can carry out.

ACKNOLEDGEMENT

The author thanks the PSG College of Technology management and principal, for generous support to carry out the research work.

References

[1] Amin, G.R. and Hajjami, M., 2016. Application of optimistic and pessimistic OWA and DEA methods in stock selection. International Journal of Intelligent Systems, 31(12), pp.1220-1233.
[2] Charnes, A., Cooper, W.W. and Rhodes, E., 1978. Measuring the efficiency of decision making units. European journal of operational research, 2(6), pp.429-444.
[3] Cook, W.D. and Seiford, L.M., 2009. Data envelopment analysis (DEA)Thirty years on. European journal of operational research, 192(1), pp.1-17
[4] Chen, H.H., 2008. Stock selection using data envelopment analysis. Industrial Management & Data Systems
[5] Dia, M., 2009. A portfolio selection methodology based on data envelopment analysis. INFOR: Information Systems and Operational Research, 47(1), pp.71-79.
[6] Emrouznejad, A., 2014. Advances in data envelopment analysis. Annals of operations research, 214(1), pp.1-4.
[7] Emrouznejad, A. and Yang, G.L., 2018. A survey and analysis of the first 40 years of scholarly literature in DEA: 19782016. Socio-economic planning sciences, 61, pp.4-8.
[8] Huang, C.Y., Chiou, C.C., Wu, T.H. and Yang, S.C., 2015. An integrated DEA-MODM methodology for portfolio optimization. Operational Research, 15(1), pp.115-136.
[9] Ho, C.T.B. and Oh, K.B., 2010. Selecting internet company stocks using a combined DEA and AHP approach. International Journal of Systems Science, 41(3), pp.325-336.
[10] Lim, S., Oh, K.W. and Zhu, J., 2014. Use of DEA cross-efficiency evaluation in portfolio selection: An application to Korean stock market. European Journal of Operational Research, 236(1), pp.361-368.
[11] Lu, W.M. and Lo, S.F., 2007. A closer look at the economic-environmental disparities for regional development in China. European Journal of Operational Research, 183(2), pp.882-894.

[12] Minutolo, M.C., Sierra, A. and Kristjanpoller, W.D., 2018. Short and Long-Term Dynamic Dependencies of Main Latin American Stock Indexes and Commodity Prices: A Wavelet Approach. Journal of Accounting Finance (2158-3625), 18(6).

[13] Oral, M., Kettani, O. and Lang, P., 1991. A methodology for collective evaluation and selection of industrial R&D projects. Management science, 37(7), pp.871-885.

[14] Oukil, A., 2020. Exploiting value system multiplicity and preference voting for robust ranking. Omega, 94, p.102048.

[15] Sexton, T.R., Silkman, R.H. and Hogan, A.J., 1986. Data envelopment analysis: Critique and extensions. New Directions for Program Evaluation, 1986(32), pp.73-105.

[16] Tavana, M., Keramatpour, M., Santos-Arteaga, F.J. and Ghorbaniane, E., 2015. A fuzzy hybrid project portfolio selection method using data envelopment analysis, TOPSIS and integer programming. Expert Systems with Applications, 42(22), pp.8432-8444.

[17] Tarnaud, A.C. and Leleu, H., 2018. Portfolio analysis with DEA: Prior to choosing a model. Omega, 75, pp.57-76.

[18] Zamani, L., Beegam, R. and Borzoian, S., 2014. Portfolio selection using Data Envelopment Analysis (DEA): A case of select Indian investment companies. International Journal of Current Research and Academic Review, 2(4), pp.50-55.