Paper 1872-2014

Efficiency Estimation using a Hybrid of Data Envelopment Analysis and Linear Regression

John Dilip Raj, GE Capital Retail Finance

ABSTRACT

Literature suggests two main approaches, parametric and non-parametric, for constructing efficiency frontiers using which efficiency scores of other units can be based. Parametric functions can be either deterministic or stochastic in nature. However, when multiple inputs and outputs are encountered, Data Envelopment Analysis (DEA), a non-parametric approach, is a powerful tool used for decades in measurement of productivity/efficiency with wide range of applications. Both approaches have its advantages and limitations. This paper attempts to further explore and validate a hybrid approach, taking the best of both DEA and parametric approach, in order to estimate efficiency of Decision Making Units (DMUs) in an even better way.

INTRODUCTION

Measuring how productive/efficient a person or organization is happens almost daily, consciously or unconsciously. It can be said that every individual, branches within an organization or organizations themselves are measured in terms of their productivity/output. Scientists have developed many ways to measure productivity. DEA is one very widely used technique to measure the same. Farrell (1957) is widely credited for providing a good measure of productive efficiency, a concept furthered with DEA, which became very popular with the work of Charnes, Cooper and Rhodes (1978). It is not just nonparametric approaches that are used for the estimation of production frontiers, but also parametric approaches. Lovell and Schmidt (1998) provide a comparison of these two approaches.

DEA

A review of the evolution and growth of DEA from the publication of Charnes, Cooper and Rhodes (1978) till about 17 year years, with extensive bibliography is provided by Seiford (1996). DEA is a nonparametric approach and measures productivity of homogenous DMUs with the help of an efficiency frontier. An advantage with this technique is that each relatively inefficient (less than 100% efficiency) is not just compared with one ideal DMU but is benchmarked only with units can be said to be similar to it and yet efficient. Additionally, a path is also given by which the relatively inefficient units can become efficient.

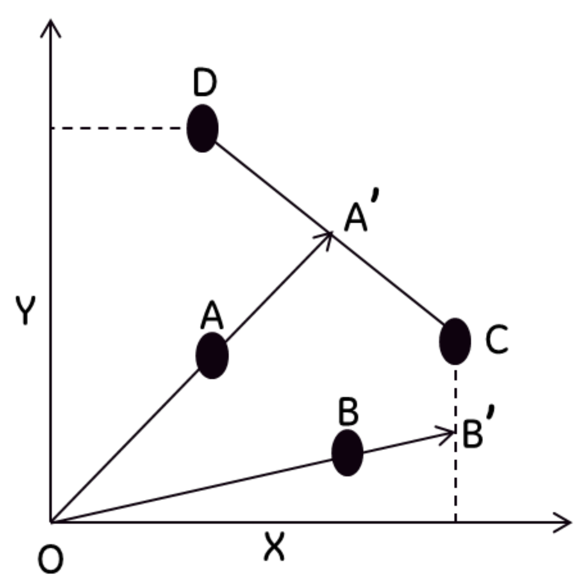
Sadiq (2011) contributed immensely to the SAS® community by providing codes using the OPTMODEL procedure to calculate efficiency of decision making units.

SOME LIMITATIONS IN THE EFFICIENCY FRONTIER IN DEA

When relative efficiency is computed, it is important that the right benchmark units against which the others are evaluated are derived. In DEA, there is an efficiency frontier on which the DMUs that are said to be 100% efficient fall. But are these really the benchmark units by which the relatively inefficient DMUs should be measured? Khezrimotlagh, Salleh and Mohsenpour (2012) note that the technical efficient DMUs that fall on frontier need not be either efficient or more efficient than the ones that do not fall there. They also proposed a method to overcome the super-efficiency method of ranking of the DMUs.

Tofallis (2001) cites another reason, the problem of slack that occurs in DEA. Consider Figure 1, where ‘C’ and ‘D’ are the DMUs that fall on the efficiency frontier. DMUs ‘A’ and ‘B’ are the relatively inefficient units. The line connecting DMUs ‘C’ and ‘D’ is the real efficiency frontier. However, the efficiency frontier is closed by drawing lines from C and D to the respective axes. This, represented by the dotted line in the figure, is not the real efficiency frontier but only one forced in order to form the envelope. DMU ‘A’ is correctly benchmarked, in that the line passing from the origin meets at the frontier connecting the efficient (100%) units of' ‘C’ and ‘D’. However, the path for DMU ‘B’ does not meet at the real efficiency frontier, but rather the dotted line connecting the last point of the frontier to the axis. There is therefore a possibility that DMU ‘B’ may not be as efficient or inefficient as it is shown to be due to the frontier being forcefully closed.

Figure 1: The Problem of Slack in DEA



Literature also shows such issues arising when weight restrictions are not there. Charnes, Cooper and Huang (1990) found that some inefficient banks are tagged as efficient and therefore DEA weight restrictions had to be imposed to get more realistic efficiencies.

A smaller, yet potential issue witnessed while using the DEA macros in SAS is shown in Table 1. The last three columns show efficiency scores for 8 DMUs based only the first output, the second output and a combination of the two outputs respectively. All the three efficiencies appear to be calculated well, in that the DMUs using fewer inputs to produce the outputs are ranked higher. However, the efficiency scores obtained when both the outputs are considered together takes values on the higher side uniformly. This becomes a little difficult to explain to mangers and people on the ground as to how a DMU with an efficiency of 16% on one metric and 32% on another has an overall efficiency of 32%. This is not so much as a technical problem but one in being able to effectively sell the technique easily. Even so, it will be good to address this in future research.

Table 1: Comparison of Efficiencies using Single and Multiple Outputs using DEA

DMU	Input1	Input2	Output1	Output2	Efficiency based on Output1 alone	Efficiency based on Output2 alone	Efficiency based on Outputs 1 & 2
1	48	3551	537	95444	16%	32%	32%
2	43	3973	518	101798	17%	37%	37%
3	47	4114	563	135337	17%	45%	45%
4	41	4121	517	121034	18%	45%	45%
5	56	4921	378	140768	9%	39%	39%
6	58	4321	376	116887	9%	32%	32%
7	46	4154	537	71756	16%	24%	27%
8	43	3772	1098	54362	36%	20%	38%

In this paper, we restrict the scope to presenting one possible solution for the issue described using Figure 1.

COMBINING DEA WITH PARAMETRIC APPROACHES

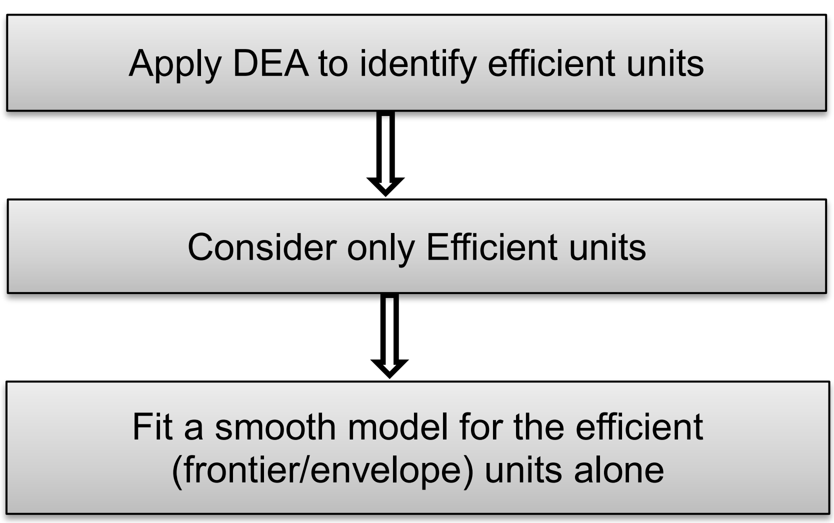
Given that the estimation of frontier is of paramount importance, various approaches have been suggested. The core frontier approaches though are either parametric or non-parametric. Murillo-Zamorano (2004) provides a detailed, critical review of both these two frontier approaches and also touches upon more recent advances in frontier efficiency measurement techniques like Bayesian, bootstrapping, duality theory. In addition to parametric and non-parametric approaches, semi-parametric models too have been attempted, with Burt (1993) outlining the benefits and limitations of these. Similarly, parametric and nonparametric approaches have their own benefits and limitations. This is true for DEA as well, which uses a non-parametric approach.

While frontiers have been developed within parametric and non-parametric approaches, would a combination of these approaches help in better estimation of efficiencies? Tofallis (2001) outlines one such approach, where the limitations of both DEA and parametric approaches are highlighted and a hybrid approach presented, combining the best of both.

STEPS INVOLVED IN THE COMBINED APPROACH

One method of combining DEA and parametric approach is as shown in Figure 2. Here, the efficiency frontier is first calculated using DEA. The DMUs that fall on the envelope (relatively efficient units) are alone taken. A model is then fitted using these efficient units and efficiency scores for all DMUs are then calculated.

Figure 2: Steps Combining DEA & Parametric Approach to Calculate Efficiencies



We present a simple approach of calculating efficiency when only one output is present. The approach adopted is that of finding efficient units using DEA and then using linear regression to obtain efficiency scores. The data presented is on a modified dataset for confidentiality reasons but comparable to the actual results obtained. Table 2 shows results of just two of the many DMUs, the second (DMU 30) of which uses far more inputs that the first (DMU 27) and produces more outputs, though not proportionally more. Using DEA alone, there was difference in the efficiency score (90% for DMU 27 and 81% for DMU 28) but not much in terms of ranking. The DEA plus linear regression model approach though further separated the efficiency scores and the ranks. In the overall data, it was also witnessed that the spread of efficiency scores was more through a parametric approach than DEA.

Table 2: Comparison of Efficiencies based on DEA and Linear Regression

DMU	Input1	Input2	Input3	Input4	Input5	Output	Efficiency based on DEA	DEA Rank	DEA + Linear Regression Efficiency	DEA + Linear Regression Rank
27	56	1699	10	141	22,269	1,600,619	90%	27	83%	18
30	86	2116	11	306	35,908	1,900,843	81%	28	62%	56

The approach detailed above can be done when there are significant numbers of DMUs that are tagged as efficient. If this is not the case, the possibility of using DMUs with greater than 90 or 95% efficiency can be explored.

When efficiency of DMUs for more than one output needs to be calculated too, the approach suggested by Tofallis (2001) can be employed.

CONCLUSION

DEA has proved to be a very powerful tool in benchmarking DMUs. Among different standalone techniques in calculating efficiencies of DMUs, DEA is quite superior to most, if not all. However, there may be situations, as ones detailed in this paper, where it is better to use DEA with other techniques. This paper used one such possible approach on the lines of Tofallis (2001) by combining DEA with a parametric technique to obtain better and more accurate results.

REFERENCES

Bult, J. R. (1993), 'Semiparametric versus Parametric Classification Models: An Application to Direct Marketing', *Journal of Marketing Research*, 30, 380–390.

Charnes, A., Cooper, W. W. & Rhodes, E. (1978), 'Measuring the Inefficiency of Decision Making Units', *European Journal of Operational Research*, 2(6), 429-444.

Charnes, A., Cooper, W. W., Huang, Z. M. & Sun, D. B. (1990), 'Polyhedral Cone-Ratio DEA with an Illustrative Application to Large Commercial Banks, *Journal of Econometrics*, 46(1-2), 73-91.

Darehmiraki, M. & Behdani, Z. (2013), 'A New DEA Approach to Rank Alternatives in MCDA', *Journal of Data Envelopment Analysis and Decision Science*, 2013, 1-7.

Farrell, M. J. (1957), 'The Measurement of Productive Efficiency', *Journal of the Royal Statistical Society*, 120(3), 253-290.

Kao, C. (2010), 'Weight Determination for Consistently Ranking Alternatives in Multiple Criteria Decision Analysis', *Applied Mathematics Modeling*, 34, 1779-1787.

Khezrimotlagh, D., Salleh, S. & Mohsenpour, Z. (2012), 'A New Method in Data Envelopment Analysis to Find Efficient Decision Making Units and Rank both Technical Efficient and Inefficient DMUs together', *Applied Mathematical Sciences*, 6(93), 4609 – 4615.

Lovell, C. A. L., & Schmidt, P. (1988), 'A Comparison of Alternative Approaches to the Measurement of Productive Efficiency', in Dogramaci, A., & R. Färe (eds.) Applications of Modern Production Theory: Efficiency and Productivity, Kluwer: Boston.

Murillo-Zamorano, L. R. (2004), 'Economic Efficiency and Frontier Techniques', *Journal of Economic Surveys*, 18(1), 33-77.

Sadiq, S. (2011), 'The Final Frontier: A SAS® Approach to Data Envelopment Analysis', *SAS Global Forum*, Paper 198-2011.

Seiford, L. M. (1996), 'Data Envelopment Analysis: The Evolution of the State of the Art (1978–1995)', *Journal of Productivity Analysis*, 7(2-3), 99-137.

Tofallis, C. (2001), 'Combining Two Approaches to Efficiency Assessment', *Journal of the Operational Research Society*, Vol. 52(11), 1225-1231.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

John Dilip Raj
GE
iLabs Building, Block 1A, Hitech City,
Madhapur, Hyderabad – 500081
India
Email: dilip.john@ge.com

DISCLAIMER: "No copies or reprints are permitted without the **author's** express written permission, and this notice shall also be included in reproductions of the Work."