# A Review of Methods of Comparing Programmatic Efficiency Between two or more Groups of DMUs in Data Envelopment Analysis.

Dr Gary Simpson

Aston Business School
Aston University
Birmingham B4 7ET
Tel +44 (0) 121 204 3241
Email: g.p.m.simpson@aston.ac.uk

## *Abstract*

In some applications of Data Envelopment Analysis (DEA) there may be doubt as to whether all the DMUs form a single group with a common efficiency distribution. The Mann-Whitney rank statistic has been used to evaluate if two groups of DMUs come from a common efficiency distribution under the assumption of them sharing a common frontier and to test if the two groups have a common frontier. These procedures have subsequently been extended using the Kruskal-Wallis rank statistic to consider more than two groups. This paper identifies problems with the second of these applications of both the Mann-Whtney and Kruskal-Wallis rank statistics. It also considers possible alternative methods of testing if groups have a common frontier, and the difficulties of disaggregating managerial and programmatic efficiency within a non-parametric framework.

*Keywords*: Data Envelopment Analysis (DEA); Statistics, Programmatic Efficiency.

In one of the first empirical applications of the DEA methodology Charnes, Cooper and Rhodes[1] introduce a distinction between two types of efficiency which are of interest to policymakers. These two types of efficiency are managerial efficiency and programmatic efficiency.  The managerial efficiency is the classic DEA efficiency and measures the performance of an individual decision making unit (DMU) in comparison with the observed production possibility frontier. The programmatic efficiency recognises that different groups of the DMUs may not have the same production possibility frontier because of programmatic differences and seeks to reveal potential efficiency differences between the productive programmes and to test the relative efficiency of each programme irrespective of potentially different distributions of managerial efficiency between them.

Brockett and Golany (1996) re-analysed the data of Charnes, Cooper and Rhodes (1981) and proposed a process using the Mann-Whitney rank statistic to test if a group of DMUs representing on program is more efficient than another by the nature of the program as opposed to the efficiencies of the individual DMUs within it. These ideas have been subsequently extended to more than two groups by using a more general rank sum test, the Kruskal-Wallis rank test (Sueyoshi T, Aoki S (2001)).

The following steps are proposed by Brockett and Golany (1996) to estimate the programmatic efficiency (as opposed to managerial efficiency).

1. Split the DMUs into the two groups according to their respective programmes and run DEA separately for each the two groups;
2. In each of the two groups project the inefficient DMUs to the efficient frontier of this group, thus attempting to eliminate the effects of managerial inefficiencies within a programme.
3. Run DEA on the combined set of the projected DMUs from both groups; which includes all DMUs.
4. Apply a statistical test to test if the two groups have the same distribution of efficiency values

First we will describe the difficulties in separating programmatic and managerial inefficiency within a DEA framework, and then we will demonstrate that the process used by Brockett and Golany (1996) is inappropriate and will invalidate the results of the statistical test. This is because of the process at step 2 which will depend on how the efficient DMUs are distributed between the programmes. It is also demonstrated that it produces biased results, particularly favouring the larger of two unequally sized programmes. We then show that these problems will persist when more than two groups are considered. Finally we make some suggestions as to how these problems maybe addressed.

## *1. Managerial Efficiency vs Programmatic Efficiency*

If different programmes transform the same inputs into the same outputs there is said to be no programmatic inefficiency if they share a common production possibility frontier. That is the outputs that can be obtained for a given set of inputs are the same for an efficient DMU each of the programmes. Managerial inefficiency refers to any short fall outputs or over consumption of inputs of a particular DMU relative to the production possibility frontier.

So if we consider figure 1, then if the DMU A shown is a member of Programme A its managerial efficiency is given by OA/OA' and the programmatic efficiency of Programme A for DMUs with this output mix would be given by OA'/OA''. Whereas, if DMU A shown is a member of Programme B its managerial efficiency is given by OA/OA'' and the programmatic efficiency of Programme A for DMUs with this output mix would be given by OA''/OA''=1. Hence, this DMU provides no evidence of programmatic inefficiency for programme B but does provide evidence of programmatic inefficiency for programme A.

In Data Envelopment Analysis the observed best practice frontier is used rather than the true but unknown production possibility frontier. Because, the observed frontiers of two randomly selected groups of DMUs having a common production possibility frontier would be expected to differ, there is a desire to ascertain if the observed differences provide sufficient evidence at a given level of significance to reject the null hypothesis at the production possibility frontiers are the same.

In some circumstances there is an expectation that the managerial efficiencies of the programmes will differ, for example if one is a new initiative which was thought to have attracted the better managers, and so it is desirable to allow for different distributions of managerial efficiency.  It is such a process that Brockett and Golany proposed; correcting for managerial inefficiency by projecting to the within programme frontier. Unfortunately because the this projection step is in general not equally effective for each of the programmes at removing the managerial inefficiency their test will generally be biased against programmes represented by fewer DMUs.

We can at this point also note that there is an intrinsic difficulty in disaggregating managerial and programmatic efficiency. If we are consider the DEA frontier to be a statistical estimator of the true production possibility frontier on the basis of its asymptotic properties under the assumption of the distribution of managerial inefficiency being sufficiently dense that there is a reasonable expectation that DMUs will appear on the frontier, then unless we make further assumptions about how the managers are distributed between the programmes there is no guarantee that the distributions of managerial efficiency within programmes will both have this property.  For example consider the case where one programme has selected its managers from the top 20% of managers overall before the introduction of the programme, while the new programme will have the desired property, the old programme will have lost many of its best managers and so this expectation may no longer hold.

## 2.Problems with Brockett & Golany's programme evaluation procedure

We will demonstrate that this procedure results in seriously erroneous results by considering a simple case. If we have 100 DMUs which all share a common production possibility frontier and are drawn from a common managerial efficiency distribution. We divide the set of DMUs into two groups at random. Hence we know there is no difference in programmatic efficiency between the two groups. We then apply the procedure proposed is as follows:

I. Split the group of all DMUs into two groups (A and B) containing $n_a$ and $n_b$ DMUs respectively. Run DEA separately for each of the two groups.

II. In each of the two groups separately, adjust each inefficient DMU to its "level if efficient" value by projecting it onto the efficiency frontier for that group.

III. Run a pooled DEA with all the projected DMUs

IV. Apply Mann-Whitney Rank Test to the pooled DEA results. Computing

$$Z = \frac{\frac{n_a}{2}\left(n_a + n_b + 1\right) - \sum_{GroupA} Ranks}{\sqrt{\frac{n_a \cdot n_b \left(n_a + n_b + 1\right)}{12}}}$$

Considering initially the two randomly selected groups to be Group A consisting of 24 DMUs and Group B consisting of the remaining 76 DMUs and that 10 of the DMUs define the efficient frontier of the whole set of 100 DMUs.

The probability that all 10 of these globally efficient DMUs happen to have been put in Group B by chance is given by

Probability = $^{76}C_{10} / {}^{100}C_{10} = 0.0551$

Now, if we project DMUs in each of the Groups A and B to their own respective frontiers and then recalculate the efficiencies of the pooled DMUs.

All 76 DMUs in Group B will appear efficient as the ten globally efficient DMUs all appear in this group and so also define its frontier. Hence all 66 of the inefficient DMUs will be projected to this globally efficient frontier.

However none of the DMUs in Group A will appear on the joint efficient frontier. So none of the projected DMUs in this group will be projected to the joint frontier.

Hence, the DMUs in Group B will occupy ranks 1 to 76, and have a Rank Sum = 2926.

Whereas the DMUs in Group will A occupy ranks 77 to 100 and have a Rank Sum = 2124

So calculating the Mann-Whitney Rank Statistic gives Z=-7.36 hence suggests P=0.0000.

That is the test statistic suggests that this distribution is highly significant indicating that there is practically no chance of the groups sharing a common frontier but in fact we know it will occur greater that 5% of the time when there is no difference between the two groups either in terms of the production frontier or managerial efficiencies!

This problem will persist when the groups are of more similar sizes and when not all of the efficient DMUs fall in one of the groups, as we will illustrate in the following example.

Again considering 100 DMUs but now divided into two randomly selected groups; Group A consisting of 37 DMUs and Group B consisting of the remaining 63 DMUs and again 10 of the DMUs define the observed efficient frontier of the whole set of 100 DMUs.

The probability that all 10 of these globally efficient DMUs happen to have been put in Group B by chance is now given by

Probability = ${}^{63}C_{10}$ / ${}^{100}C_{10}$ = 0.0074

Which is quite small, but still considerably larger that the p-value that would be obtained by applying the procedure of Brocket and Golany , but let us now consider the probability that 9 of these globally efficient DMUs happen to have been put in Group B and 1 is placed in group A.

Probability = ${}^{63}C_9\ {}^{37}C_1 / {}^{100}C_{10} = 0.0506$

We now consider what happens when we project the DMUs to their respective frontiers. For simplicities sake we will consider a simple one input two output constant returns to scale DEA model as shown in figure 2. If we define $\eta_i$ as the number of inefficient DMUs in group B that project onto the $i^{th}$ segment of frontier (numbering the segments in order of increasing ratio of output2 to output1). Then, for segments 2 to 10 (that is all the segments on the non-dominated part of the frontier) the inefficient DMUs in Group B will project to the joint efficient frontier if and only if both of the DMUs defining the segment are in Group B.

The probability of this occurring is 8/10, but for segments 1 and 11 only a single DMU defines the segment so the probability becomes 9/10

Hence we expect $9+\frac{8}{10}\sum_{i=2}^{10}\eta_i+\frac{9}{10}(\eta_1+\eta_{11})\geq 9+\frac{8}{10}\sum_{i=1}^{11}\eta_i=9+\frac{8}{10}(63-9)=52.2$ of the DMUs in Group B to appear on to the joint efficient frontier.

Whereas inefficient DMUs from Group A will only project to the joint frontier if they are un-enveloped and the efficient DMU in Group A is the closest one as shown in figure 2.

So we expect $1+\frac{0}{10}\sum_{i=2}^{10}\eta_i+\frac{1}{10}(\eta_1+\eta_{11})\leq 1+\frac{1}{10}\sum_{i=1}^{11}\eta_i=1+\frac{1}{10}(37-1)=4.6$ DMUs from Group A to appear on the joint efficient frontier.

That is a total of about 57 DMUs are expected to appear on the joint efficient frontier and to tie for the top rank and will each be given a rank number of 29. The remaining 43 DMUs will share the remaining ranks (58 to 100) with an average rank number of 79.

So the expected rank sum for group A is 4.6x29+32.4x79 =2693

This would give Z =-5.89 and again P=0.0000.

So once again the test has suggested a highly significant difference when the result is to be expected in more than 5% of the cases.

## *3 More than two groups: the Kruskal and Wallis Rank Test*

As already stated Sueyoshi & Aoki extend on the ideas of Brockett & Golany in line with the suggested extensions in their original paper to consider more than two groups of DMUs and use it to consider how a frontier may shift over time.

Their procedure works in the same way projecting DMUs to the within group frontier then pooling the projected DMUs and ranking them using DEA. But now the sums of the ranks are compared using the Kruskal Wallis Rank test.

If there are K groups and the $j^{th}$ group has $n_j$ DMUs and Rank Sums $R_j$ and $N=\Sigma n_j$

Then $H = \left(\frac{12}{N(N+1)}\right)\sum_{j=1}^{K}\frac{R_j^2}{n_j} - 3(N+1)$ Should follow a $\chi^2$ distribution with *K-1* degrees of freedom.

We will again demonstrate that this procedure will be problematic by way of an illustrative example. If we have 60 DMUs which all share a common production possibility frontier and are drawn from a common managerial efficiency distribution. We divide the set of DMUs into three groups at random. Hence we know there is no difference in programmatic efficiency between the three groups.

We will now consider each of the three randomly selected groups to consist of 20 DMUs, as this should minimise the problems with the procedure, and that 5 of the DMUs define the observed efficient frontier of the whole set of 60 DMUs.

Firstly we note that the probability that all 5 globally efficient DMUs are in one of the groups is given by : Probability $=3\ {}^{20}C_5 / {}^{60}C_5 = 0.00852$

The group which possessed all the globally efficient DMUs would take up ranks 1 to 19 and the remaining ranks would be shared by the other two groups. Hence giving a minimum value of H when the remaining ranks are equally shared between these two groups of:

$$H = \left(\frac{12}{60(60+1)}\right)\left(\frac{210^2 + 810^2 + 810^2}{20}\right) - 3(60+1) = 39.34$$

which would give a p-value of 0.00000. So the probability of rejecting a true null hypothesis is again considerably larger than the test suggests.

We also note that the if only 4 of the globally efficient DMUs are in one of the groups, 1 of them in another and none in the final group is given by as we would expect in more than 10% of the cases if the groups where randomly selected ( Probability = 6 ${}^{20}C_4\, {}^{20}C_1 / {}^{60}C_5$ = 0.10645) the problems will persist. This is illustrated in figure 4 and using the same logic as we did for the two group case.

So for Group A (the group with none of the efficient units) clearly none of its inefficient DMUs will project to the joint frontier.

For Group B (the group with only one of the efficient units) the expected number of its inefficient DMUs that will project to the joint frontier is given by

$$\frac{0}{5}\sum_{i=2}^{5}\eta_i + \frac{1}{5}(\eta_1 + \eta_6) \le \frac{1}{5}\sum_{i=1}^{6}\eta_i = \frac{1}{5}(20-1) = 3.8$$

So in total we expect less than 4.8 DMUs from this group on the joint frontier.

But for Group C the expected number of its inefficient DMUs that will project to the joint frontier is given by

$$\frac{3}{5}\sum_{i=2}^{5}\eta_i + \frac{4}{5}(\eta_1 + \eta_6) \ge \frac{3}{5}\sum_{i=1}^{6}\eta_i = \frac{3}{5}(20-4) = 9.6$$

So in total we expect at least 4+9.6=13.6 of the DMUs from this group on the joint frontier

In total this gives an expected 18.4 DMUs on the joint frontier (hence an average rank of 9.7) with the remaining 41.6 DMUs sharing the remaining ranks (with an average rank of 39.7). This gives the following expected rank sums, under the conservative assumption that the remaining ranks are shared equally between the programmes.

Group A Rank Sum = 20x39.7 =794

Group B Rank Sum = 4.8x9.7 +15.2x39.7 = 650

Group C Rank Sum = 13.6x9.7+6.4x39.7 = 386

Then the expected value of H is greater than

$$H = \left(\frac{12}{60(60+1)}\right)\left(\frac{386^2 + 650^2 + 794^2}{20}\right) - 3(60+1) = 14.0327$$

which would give a p-value of less than 0.000895.

So again the procedure has produced a highly significant result, for an occurrence which has a probability of greater than 10% of occurring when the null hypothesis is true! In this case, there is not bias against any particular programme, as each, being the same size are equally likely to be under represented in the number of DMUs on the joint frontier prior to projection, but there is a bias against the programme(s) which are under represented in this way because of the over sensitivity of the test.

## 4 Conclusions & Way Ahead

It is clear that the current tests do not properly separate managerial and programmatic efficiency.

The root cause of the problems is in the projection of units to the within programme frontier, which does not treat the programmes equivalently, particularly when the number of DMUs in the programs are unequal.

One can however note that if instead of projecting the units before applying the non-parametric tests we simply applied them to the un-projected DMUs we have a test that will fairly detect shift in the programme frontier if the distribution of managerial efficiency is the same in the two groups. Unfortunately, it is this assumption of a common distribution of managerial efficiency between the programs or groups that we wish to avoid, but if we have data for several time periods this approach may prove fruitful. For example, if we have a single programme into which a new initiative is introduced into a subset of the DMUs , we could assume the overall distribution of managerial efficiency is the same before and after the introduction. Applying the Mann-Whitney test to the whole data sets before and after would then test for a shift in the frontier caused by the introduction of the new initiative. A similar test of the subset that introduced the initiative before and after would give the combined effect of the shift in frontier and any advantage in selecting the better managers. Alternatively, if we consider that if there is no programmatic inefficiency we do not expect there to be an association between which facets of the frontier are defined by DMUs in a particular programme and the facets that the inefficient units in that program will project to. So if we expect the input/output mixes to have the same distribution for the two programmes we could proceed with a test on these lines. Simpson (2004) has suggested such a test for the two programme case based on this approach and further work may allow a generalisation to the k programme case.

References

Charnes A, Cooper WW, Rhodes E, (1981), "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through" Management Science, 27, 6 668-697.

Brockett PL, Golany B, (1996), "Using Rank Statistics for Determining Programmatic Efficiency Differences in Data Envelopment Analysis" Management Science., 42, 3 466-472.

Sueyoshi T, Aoki S, (2001), "A use of a nonparametric statistic for DEA frontier shift: the Kruskal and Wallis rank test" Omega Int J Manage Sci, 29 1-18.

Simpson G, (2004), Programmatic Efficiency Comparisons Between Unequally Sized Groups of DMUs in DEA, *Aston Business School Working Paper*

Figure 1: Managerial Vs Programmatic Efficiency

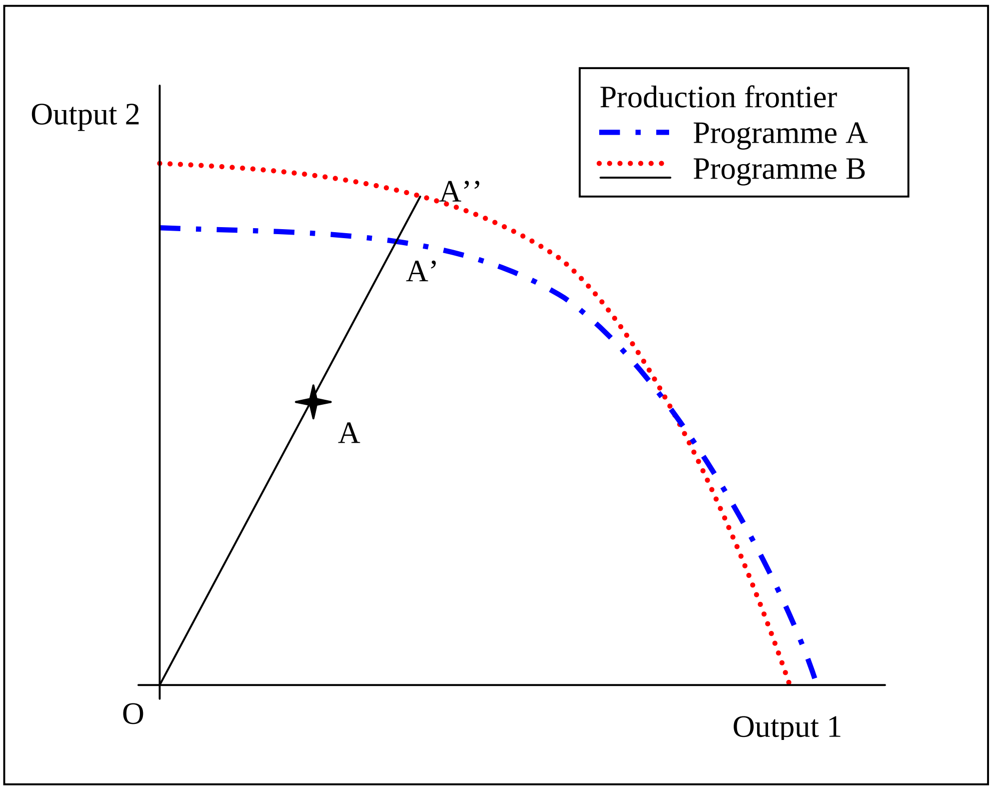



Figure 2: Segments of the DEA frontier

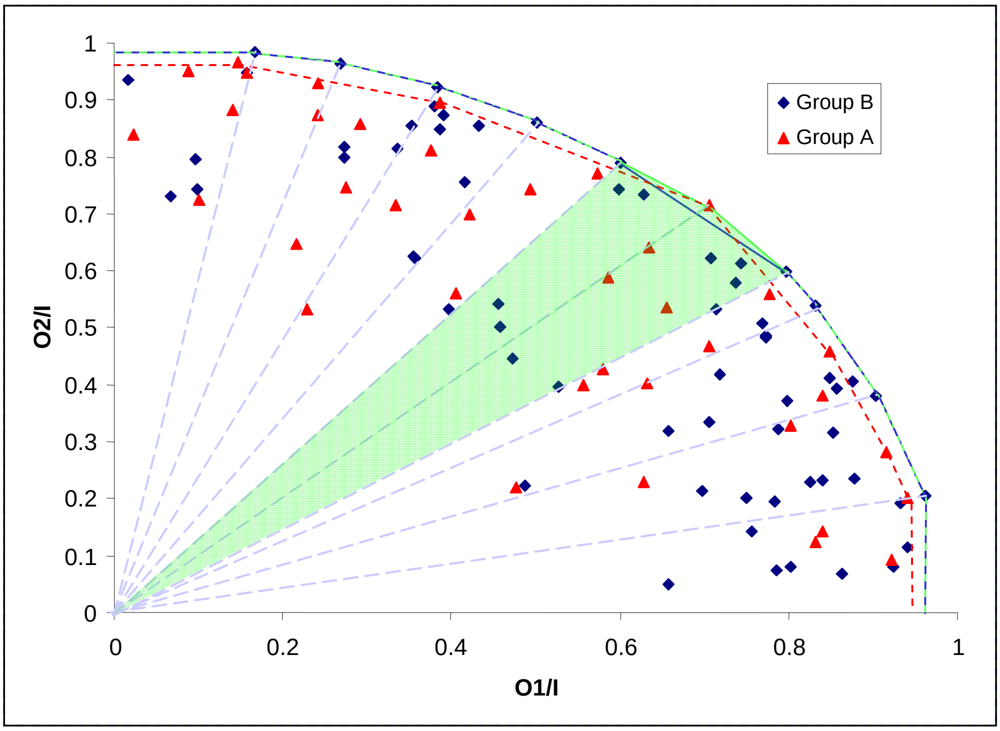


Figure 3: A Segment of the DEA frontier where units from Group A project to the joint frontier

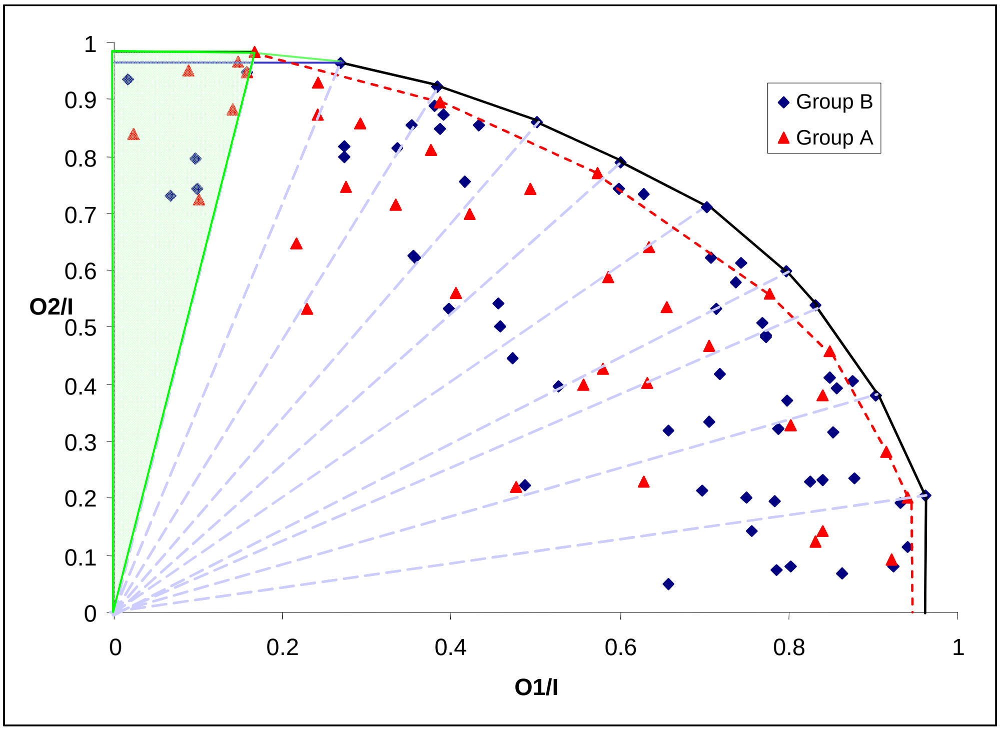
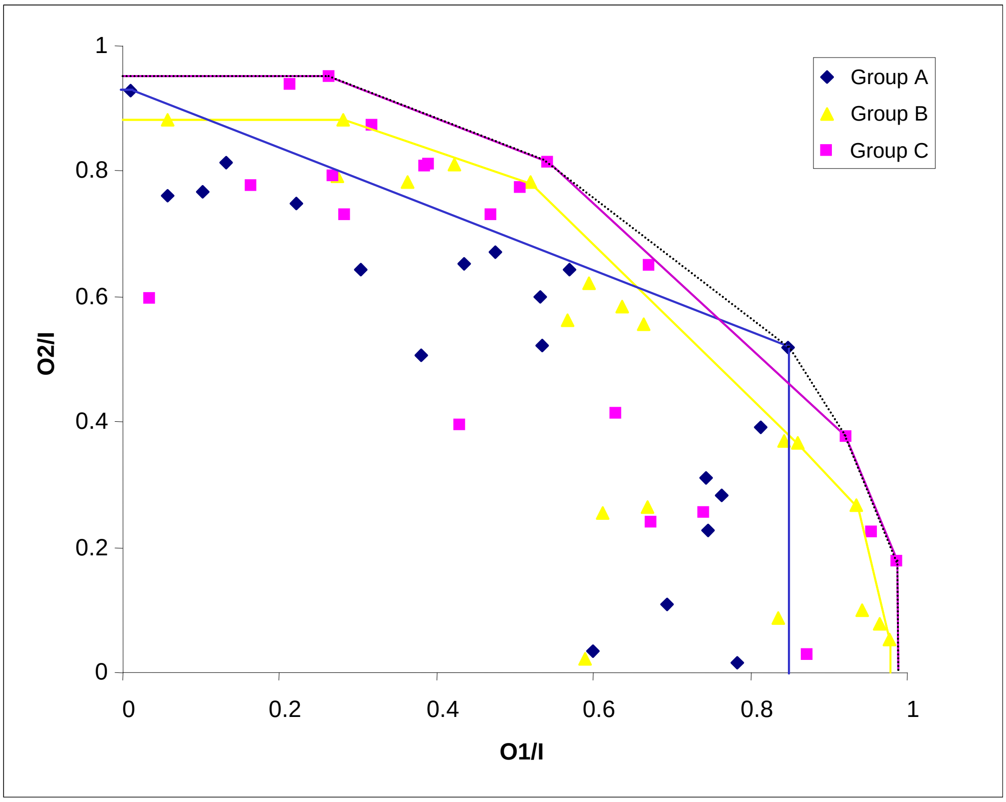



Figure 4:


Group A
Group B
Group C
O2/I
O1/I
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1