Relative Efficiency Coursework

BNM862 Performance Analytics

Vijay Singh Bisht
220273880

Table of Contents

Introduction

The objective of this research is to create adequate suitable models for computing and analysing Premiership Footballers' performance relative efficiency as well as the nature of their returns to scale. There are two ways to analyse the relative efficiency: Parametric and Non Parametric. The parametric analysis is done using the econometrics software STATA17 and Non Parametric analysis to be done using PIM-DEA.

The data that we have provided with is for Footballers playing for different clubs in the Premier league for season 2021-2022. The data includes various information regarding players information and performance in the league.

Player Information:

1. Player – Name of the player
2. Nation – Country they belong to
3. Pos – Positions they play in
4. Squad – Name of the team they play for (CLUB)
5. Age and born – age and year born

Performance:

1. MP – Matches Played
2. Starts – Matches Started
3. Min – Minutes played total for all league matches
4. 90s – Matches played including fractions of a full 90 minutes
5. Gls – Goals Scored
6. Ast – Asists
7. npGls – Goals excluding penalty kicks
8. PK – Penalty kicks scored
9. PKatt – Penalty attempts
10. CrdY – Yellow card
11. CrdR – Red card
12. XG – Expected goals
13. Xnpxg – Expected non penalty goals
14. xAG – Expected asists
15. xNpGplusxA – Expected assists plus non penalty goals

Goal Scored (Gls) will be our output since we need to assess the player's efficiency and we will be looking at how the input factors influence the outcome (Goal scoring performance).

Descriptive Statistics

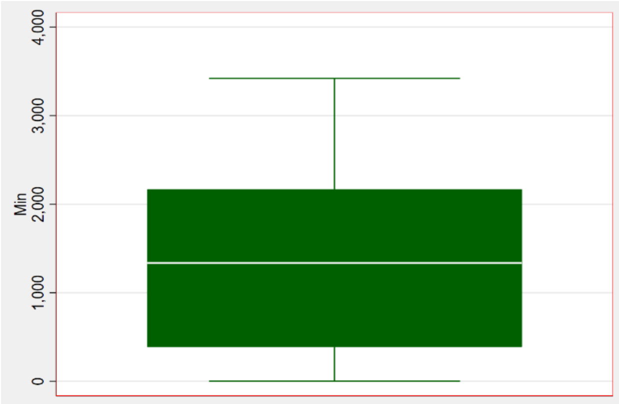
The Descriptive Statistics gives us the overall description of the data such as Number of observations, Mean, Std. dev, minimum and maximum values.

Index	Mean	Sum	Std. Dev.	Variance	Minimum	Maximum	Range
Age	25.36	12657	4.41	19.42	16	39	23
Born	1995.24	995623	4.41	19.42	1981	2004	23
MP	19.09	9526	11.75	137.98	1	38	37
Starts	15.23	7601	11.67	136.26	0	38	38
Min	1367.04	682155	1018.51	1037359.94	1	3420	3419
90s	15.19	7579.2	11.32	128.03	0	38	38
Gls	1.84	920	3.19	10.19	0	23	23
Ast	1.38	690	2.09	4.37	0	13	13
npGls	1.7	850	2.9	8.39	0	23	23
PK	0.14	70	0.64	0.41	0	6	6
PKatt	0.17	87	0.74	0.54	0	7	7
CrdY	2.46	1230	2.56	6.56	0	11	11
CrdR	0.08	40	0.29	0.08	0	2	2
xG	2.09	1041.6	3.18	10.09	0	23.7	23.7
xnpxg	1.95	973.2	2.86	8.21	0	18.8	18.8
xAG	1.51	755.3	1.93	3.74	0	12.9	12.9
xNpGplusxA	3.46	1728.1	4.48	20.07	0	28.4	28.4

Detailed Summary

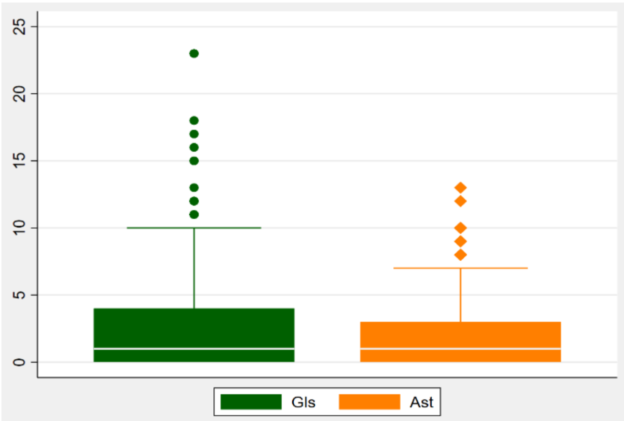

```
                            Gls
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs                 334
25%            0              0       Sum of wgt.         334

50%            1                      Mean           2.682635
                        Largest       Std. dev.      3.604451
75%            4             17
90%            7             18       Variance       12.99207
95%           10             23       Skewness       2.466014
99%           17             23       Kurtosis       10.82487
```

The following graph is a boxplot for Goals scored by the players. Most of the values lie from 0-5 and there are outliers in the upper part of the data.



Min				
	Percentiles	Smallest		
1%	308	186		
5%	572	223		
10%	790	301	Obs	334
25%	1077	308	Sum of wgt.	334
50%	1769		Mean	1778.775
		Largest	Std. dev.	788.8389
75%	2402	3253		
90%	2905	3321	Variance	622266.8
95%	3089	3361	Skewness	.1168415
99%	3253	3377	Kurtosis	2.003731

The above graph is a box plot for number of minutes played by the players. The box plot does not show any outliers in the data and the majority of data lies between 500Minutes to 2100Minutes.



Gls				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	334
25%	0	0	Sum of wgt.	334
50%	1		Mean	2.682635
		Largest	Std. dev.	3.604451
75%	4	17		
90%	7	18	Variance	12.99207
95%	10	23	Skewness	2.466014
99%	17	23	Kurtosis	10.82487

Ast				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	334
25%	0	0	Sum of wgt.	334
50%	1		Mean	2.01497
		Largest	Std. dev.	2.29173
75%	3	10		
90%	5	10	Variance	5.252027
95%	7	12	Skewness	1.801686
99%	10	13	Kurtosis	6.792264

This is the box plot for goals scored and assists. There are outliers available in the data and we can observe that the data is right skewed.

Correlation

	Gls	Age	MP	Starts	Min	Ast	GPK	PK	PKatt	xG	xAG	npxGxAG
Gls	1.0000											
Age	0.0082	1.0000										
MP	0.4560	0.0240	1.0000									
Starts	0.3780	0.0763	0.8855	1.0000								
Min	0.3808	0.0729	0.9027	0.9938	1.0000							
Ast	0.5593	-0.0147	0.4035	0.3414	0.3381	1.0000						
GPK	0.9795	-0.0126	0.4565	0.3782	0.3787	0.5458	1.0000					
PK	0.5290	0.0913	0.1998	0.1664	0.1774	0.3048	0.3473	1.0000				
PKatt	0.5403	0.1084	0.1896	0.1577	0.1682	0.3140	0.3680	0.9654	1.0000			
xG	0.9228	0.0145	0.4871	0.3845	0.3906	0.5798	0.8835	0.5739	0.5987	1.0000		
xAG	0.6508	-0.0452	0.5331	0.4592	0.4593	0.8092	0.6412	0.3287	0.3456	0.6618	1.0000	
npxGxAG	0.8823	-0.0250	0.5610	0.4585	0.4613	0.7303	0.8747	0.4225	0.4429	0.9353	0.8685	1.0000

Cobb-Douglas Regression

Since the Cobb-Douglas approach includes linear parameters and because the variables are transformed into linear variables using log values, it is used in regression models.

we need to convert all input and output values into log values.

reg lnGls lnMP lnStarts lnMin lnPK lnPKatt lnxG

Source	SS	df	MS			
				Number of obs	=	33
				F(6, 26)	=	36.86
Model	19.1150355	6	3.18583926	Prob > F	=	0.0000
Residual	2.24733508	26	.086435965	R-squared	=	0.8948
				Adj R-squared	=	0.8705
Total	21.3623706	32	.667574082	Root MSE	=	.294

lnGls	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnMP	-1.004299	.6329679	-1.59	0.125	-2.305384	.2967849
lnStarts	-.2775832	1.191155	-0.23	0.818	-2.726037	2.170871
lnMin	1.088741	1.500804	0.73	0.475	-1.996206	4.173689
lnPK	.3425802	.3110956	1.10	0.281	-.296886	.9820463
lnPKatt	-.2274599	.2924167	-0.78	0.444	-.828531	.3736112
lnxG	.8524707	.1348211	6.32	0.000	.575342	1.129599
_cons	-3.925125	6.570349	-0.60	0.555	-17.43067	9.58042

In the first regression model, many values are insignificant to our analysis therefore we need to remove the most insignificant values one by one to achieve our best model where only significant values are available.

reg lnGls lnMin lnPK

Source	SS	df	MS			
				Number of obs	=	33
				F(2, 30)	=	38.46
Model	15.3689169	2	7.68445846	Prob > F	=	0.0000
Residual	5.9934537	30	.19978179	R-squared	=	0.7194
				Adj R-squared	=	0.7007
Total	21.3623706	32	.667574082	Root MSE	=	.44697

lnGls	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnMin	1.089467	.1651164	6.60	0.000	.7522539	1.426679
lnPK	.3340855	.1428314	2.34	0.026	.042385	.6257861
_cons	-6.604384	1.213903	-5.44	0.000	-9.083504	-4.125264

Running the regression again by removing the insignificant values and running the process again we get our most parsimonious model with 71% R^2 .

Cobb-Douglas Diagnostics

1. Heteroskedasticity test

```
Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of lnGls

H0: Constant variance

    chi2(1) =   0.32
Prob > chi2 = 0.5743
```

Here, the p values is 0.5743 which is >0.05. hence, we fail to reject the null-hypothesis. Also, variance is constant in the model.

2. White's test

```
White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

    chi2(5) =   6.32
Prob > chi2 = 0.2759
```

In the white's test, the p values is 0.2759, which is >0.05. So, we fail to reject the null hypothesis.

3. Ramsey RESET test

```
Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of lnGls

H0: Model has no omitted variables

F(3, 27) =   2.57
Prob > F = 0.0748
```

The p values for Ramsey test is 0.07, which is less but still >0.05. Hence, we fail to reject the null hypothesis. Also, we observe that there are no omitted variables.

vif

Variable	VIF	1/VIF
lnMin	1.23	0.811328
lnPK	1.23	0.811328
Mean VIF	1.23	

4. VIF

There is no multicollinearity in our model as the all values in VIF test are below 10.

Translog Regression

Translog is the enhanced Cobb-Douglas regression, therefore we use the second order and cross order variables in the regression needs to be converted and used.

```
reg lnGls lnMin lnPK Min PK minpk

      Source |       SS           df       MS      Number of obs   =        33
-------------+----------------------------------   F(5, 27)        =     13.92
       Model |  15.3929945         5  3.07859889   Prob > F        =    0.0000
    Residual |  5.96937616        27  .221088006   R-squared       =    0.7206
-------------+----------------------------------   Adj R-squared   =    0.6688
       Total |  21.3623706        32  .667574082   Root MSE        =     .4702

------------------------------------------------------------------------------
       lnGls | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       lnMin |   1.086331    .622315     1.75   0.092    -.1905538    2.363216
        lnPK |   1.518412   4.877095     0.31   0.758    -8.488561    11.52538
         Min |   .0000171   .0005175     0.03   0.974    -.0010448     .001079
          PK |  -.0158171   .2562839    -0.06   0.951    -.5416683     .510034
       minpk |  -.1480547   .6517202    -0.23   0.822    -1.485274    1.189165
       _cons |  -6.600752   3.686304    -1.79   0.085    -14.16442    .9629193
------------------------------------------------------------------------------
```

Here, the translog regression have many insignificant values, so after re running the regression model to remove insignificant values the R^2 value keeps on of decreasing.

PIM-DEA Analysis

PIM-DEA is Performance Improvement Management Software. PIM is a method for organising and disseminating product information throughout an organisation, whereas DEA is a technique for assessing the relative effectiveness of a group of decision-making units (DMUs).

Here, we will be taking Goals scored by the player (Gls) as our output in the software and the number of minutes played by the player as our input. We will be creating models of constant return to scale (CRS) and variable return to scale (VRS) for output-oriented models as the aim of our analysis is to increase the efficiency of the model which can be achieved by maximizing the goals scored by the player.

Output Oriented Model

In output-oriented model, the model assumes that the input values are fixed and focuses on increasing the output values. This indicates that the objective is to handle the greatest number of items or channels possible with the inputs available. The ratio of the actual outputs to the greatest outputs possible given the inputs is used to determine the efficiency score.

Selection for output oriented model:

The selection I have considered for my model is that the player should have played at least 600 minutes. This is to understand our analysis better.

CRS Output Oriented Model

In the following models, A set of decision-making units (DMUs) with the same scale of operations are compared for relative efficiency using the CRS (Constant Returns to Scale) model in Data Envelopment Analysis (DEA).

So, now will be creating CRS output-oriented model:

Input = Number of minutes played for all league matches (Min)

Output = Number of goals scored (Gls)

Efficiency for CRS model:

After creating an Output oriented CRS model, we examine the efficiency table output in order to identify the most and the least efficient DMUs from the model.

Here, we have the efficiency table for CRS model, we can observe that DMU 429 (Mohamed Salah) is the most efficient with 100% efficiency, and DMUs 215, 303, 418 are highly efficient with more than 85% efficiency.

Name	Efficiency	
P429	100	100%
P215	91.88	92%
P303	88.18	88%
P418	88.01	88%
P113	81.84	82%
P241	76.2	76%
P471	73.36	73%
P305	68.16	68%
P248	63.16	63%
P460	62.51	63%
P95	61.2	61%
P298	60.46	60%
P301	58.6	59%
P202	58.07	58%
P346	55.9	56%
P392	55.85	56%
P208	53.11	53%
P316	52.44	52%
P101	52.35	52%
P104	51.81	52%
P203	51.73	52%
P233	51.18	51%
P162	50.79	51%
P293	49.59	50%
P491	49.54	50%
P53	48.31	48%
P271	47.62	48%
P405	47.6	48%

Efficiency Table

Name	P429
(Frequencies)	202
P1	
P2	✓
P3	✓
P4	✓
P6	✓
P7	✓
P8	✓
P9	
P10	
P13	✓
P14	✓
P16	
P17	
P18	✓
P21	✓
P22	✓
P25	✓
P26	✓
P27	✓
P28	✓
P29	
P37	✓
P40	✓
P44	✓
P45	

Peer relationship table

According to the peer relationship table mentioned above for the most efficient DMU, the DMU 429 has an efficient relationship with 202 other DMUs. Peers are other DMUs that perform similar activities and operate under similar conditions.

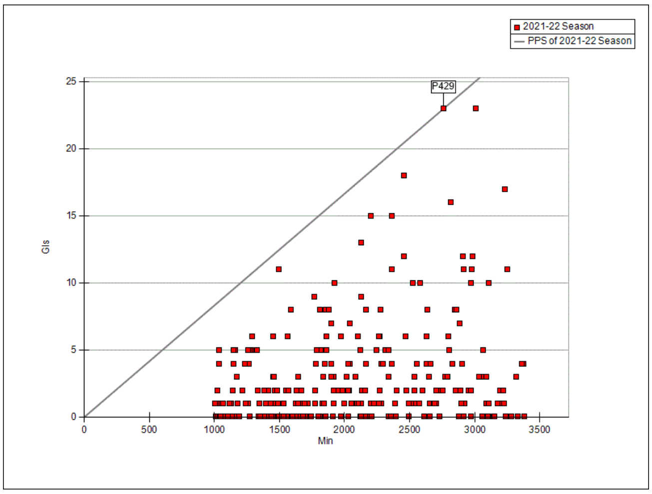
VRS Output Oriented Model

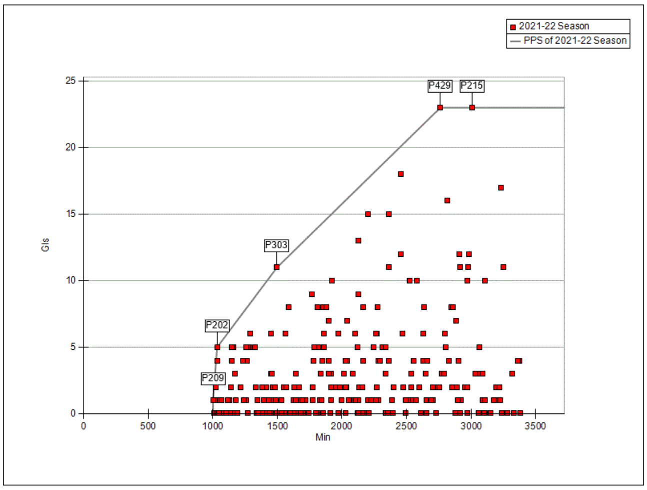
Name	Efficiency	
P215	100	100%
P372	100	100%
P429	100	100%
P340	88.15	88%
P418	87.07	87%
P303	82.19	82%
P113	80.08	80%
P241	75.1	75%
P248	73.91	74%
P471	71.52	72%
P305	69.57	70%
P212	61.97	62%
P460	60.23	60%
P361	59.95	60%
P41	59.46	59%
P95	58.35	58%
P301	57.99	58%
P298	56.83	57%
P346	55.1	55%
P159	52.52	53%
P53	52.17	52%
P491	52.17	52%
P392	50.84	51%

Here, we have the efficiency table for VRS model, we can see that there are lot of efficient DMUs in this VRS input model which is very promising.

The DMU 202, DMU 209, DMU 303, and DMU 429 are fully efficient DMUs with 100% efficiency and DMU 25, DMU 73, and DMU 86 are highly efficient with 99% efficiency, while DMU 256, DMU 61, DMU 120, DMU 63, DMU 335, and DMU 371 also close with 96-98% efficiency.

PPS Chart





seeing the Output-oriented model's PPS graph. In both the CRS output orientation model and the VRS output orientation model as DMU 429 is efficient is on the efficient frontier, indicating that it is an MPSS.

Multi-variate Analysis

Now we will be considering multi-variate output oriented model with one input and two output to obtain more efficient DMUs from our dataset. We will be making CRS and VRS for the same.

CRS Multi-variate (1 Input 2 Output)

Input = Minutes played

Output = Goals, Assist

Name	Efficiency	
P271	100	100%
P388	100	100%
P429	100	100%
P215	91.88	92%
P303	88.18	88%
P418	88.01	88%
P113	81.84	82%
P346	76.34	76%
P241	76.2	76%
P229	75.6	76%
P37	74.84	75%
P233	74.84	75%
P471	73.36	73%
P305	68.16	68%
P367	67.59	68%
P223	65.67	66%
P301	64.92	65%
P8	63.4	63%
P248	63.16	63%
P460	62.51	63%
P53	62.08	62%
P272	61.6	62%

Here, For the multivariate, the efficiency table for CRS model with 1 input and 2 output we can observe that we get more efficient DMUs for this CRS model compared the previous CRS model. We have DMUs

271, 388 and 429 with 100% efficiency.

Name	P271	P388	P429
(Frequencies)	99	83	167
P1		✓	
P2			✓
P3	✓	✓	
P4	✓	✓	
P6	✓	✓	
P7	✓	✓	
P8	✓	✓	
P9		✓	
P10		✓	
P13			✓
P14	✓		✓

From peer relationship table, we can observe that DMU 429 has an efficient relationship with 167 DMUs and DMU 388 and DMU 271 has efficient relationship with 83 and 99 DMUs respectively.

VRS Multi (1 Input 2 Output)

Input = Minutes played

Output = Goals, Assist

Name	Efficiency	
P120	100	100%
P202	100	100%
P209	100	100%
P215	100	100%
P271	100	100%
P303	100	100%
P367	100	100%
P388	100	100%
P429	100	100%
P8	92.31	92%
P37	89.86	90%
P418	89.57	90%
P86	87.5	88%
P104	86.21	86%
P229	86.06	86%
P25	85.18	85%
P113	84.87	85%
P346	84.24	84%
P101	81	81%
P410	80.89	81%
P241	78.04	78%
P316	77.7	78%
P233	77.38	77%
P53	76.92	77%

For VRS model with 1 input and 2 output we got more efficient DMUs for the VRS model as DMU 120, DMU 202, DMU 209, DMU 215, DMU 271, DMU 303, DMU 367, DMU 388, DMU 429. We got 9 DMUs with 100% efficiency in CRS compared to 3 in CRS for multi-variate model.

Name	P120	P202	P209	P215	P271	P303	P367	P388	P429
(Frequencies)	10	28	5	0	44	91	8	93	206
P1									✓
P2						✓			✓
P3								✓	✓
P4								✓	✓
P6	✓	✓			✓				
P7								✓	✓
P8									✓
P9									✓
P10								✓	✓
P13						✓			✓
P14								✓	✓

Also, In the peer relationship table, we got peer relation for our most efficient DMUs. DMU 429 has most peer relations with 206 other DMUs and DMU 215 has least with zero DMUs.

DMU summary:

A DMU summary in PIM-DEA includes various performance measurements and information regarding the efficiency ratings of each decision-making unit.

CRS for Multi-variate (1 Input 2 Output):

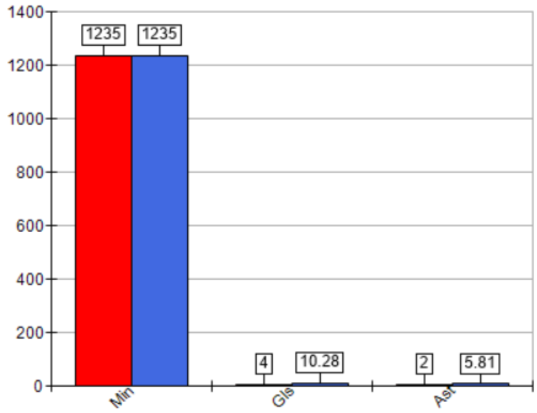
Here, from our DMU summary for CRS model we can examine the efficiency for inefficient players in the league.

We will be observing the effciency of DMU 398 and DMU 25.

DMU 398 – Marcus Rashford

DMU P398
Efficiency 38.89%
Omega 0
Peers P429 (0.45)

Name	Min	Gls	Ast
Type	Input	Output	Output
Actual	1235	4	2
Target	1235	10.28	5.81
Slack	0	0	0.67
Weight	0	0.25	0

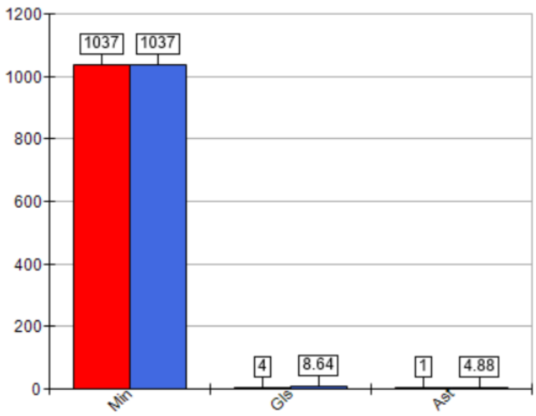


Looking at the DMU summary for DMU 398, We see that DMU 398 is Marcus Rashford, a stiker from Manchester United has an efficiency score of 38.89%. He has played 1235 Minutes and has 4 Goals and 2 Assists. Now, for him to be fully effcient (100%) Marcus Rashford has to score about 10 Goals (10.28) and 6 Assists (5.81) in the same playing time (1235 mins).

DMU 25 – Pierre-Emerick Aubameyang

DMU P25
Efficiency 46.32%
Omega 0
Peers P429 (0.38)

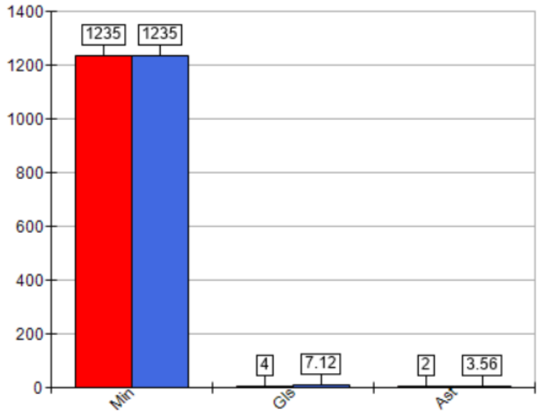
Name	Min	Gls	Ast
Type	Input	Output	Output
Actual	1037	4	1
Target	1037	8.64	4.88
Slack	0	0	2.72
Weight	0	0.25	0



Also, for DMU 25, the player is Pierre-Emerick Aubameyang, striker from Arsenal has currently player for 1037 Minutes, has scored 4 Goals and 1 Assist and have the efficiency of 46.32%. For Aubameyang to be fully efficient (100%) he has to score about 9 Goals(8.64) and 5 Assists (4.88) in the same playing time (1037 mins).

DMU P398
Efficiency 56.19%
Omega 1.41
Peers P202 (0.48), P271 (0.16), P303 (0.35)

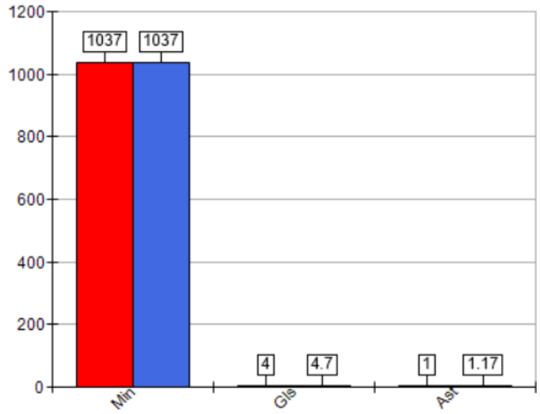
Name	Min	Gls	Ast
Type	Input	Output	Output
Actual	1235	4	2
Target	1235	7.12	3.56
Slack	0	0	0
Weight	0	0.19	0.11



VRS for Multi-variate (1 Input 2 Output):

DMU P25
Efficiency 85.18%
Omega 12.15
Peers P120 (0.06), P202 (0.92), P271 (0.02)

Name	Min	Gls	Ast
Type	Input	Output	Output
Actual	1037	4	1
Target	1037	4.7	1.17
Slack	0	0	0
Weight	0.01	0.13	0.48



Similarly, For VRS multi-variate output oriented model we get the following effciency, actual value and target value for the DMU summary are:

VRS efficicny for DMU 398 is 56.19% and for DMU 25 is 85.18%. We can see that there have been increase in players efficiecny when we move from CRS to VRS in multi variate output oriented model.

Scale Efficiency

Scale Efficiency = CRS efficiency / VRS efficiency

Therefor, Scale Efficiency for both the players is

1. DMU 25
   Scale efficiecncy = 46.32 / 85.18 = 0.5437
   The scale efficiency for Pierre-Emerick Aubameyang is 54%

2. DMU 398
   Scale efficiecncy = 38.89 / 56.19 = 0.6921
   The scale efficiency for Marcus Rashford is 69%

STATA

The efficiency of a system may be determined using deterministic frontiers, such as the COLS (Corrected OLS) and MOLS (Modified OLS) functions, which are simple to estimate and simple to understand.

The regression line in COLS is to be moved down, and depending on the kind of function, the observation will either appear above or below the line. Any values other than those on the frontier appear above the line for the production function.

Our model is Production function, so running the COLS and MOLS analysis based on the the Production fuction.

sum maxres

Variable	Obs	Mean	Std. dev.	Min	Max
maxres	334	.6442196	0	.6442196	.6442196

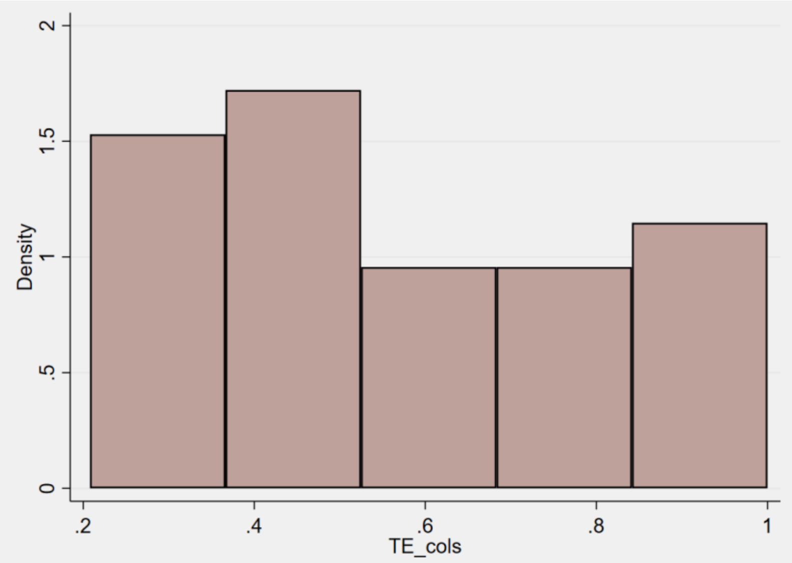
Here, standard deviation is zero and as maximum residual is a constant, mean, min and max values are all same.

COLS (Corrected OLS)

lnE_cols

	Percentiles	Smallest		
1%	0	0		
5%	.0342922	.0342922		
10%	.0773767	.0534768	Obs	33
25%	.3210355	.0773767	Sum of wgt.	33
50%	.649097		Mean	.6442196
		Largest	Std. dev.	.4323188
75%	.9870285	1.244984		
90%	1.244984	1.311518	Variance	.1868996
95%	1.338784	1.338784	Skewness	.2216274
99%	1.570431	1.570431	Kurtosis	2.09219

From the CRS model, we see that the technical efficiency is with mean 64% and std. dev is 43%. We will also be examining our model visualization to get more information.



From the Histogram here, we can see that majority data lies between 0.2-0.4.

MOLS (Modified OLS)

We will be calculating the estimate of E(u) assuming exponential distributed efficiency and half-normally distributed efficiency and after that we will calculate technical efficiency.

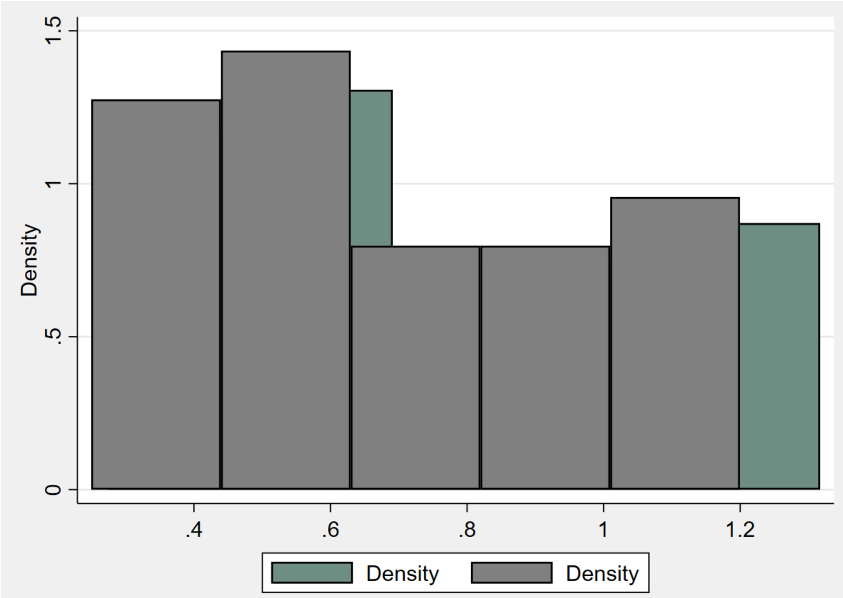
TE_mols_exp

	Percentiles	Smallest		
1%	.2494793	.2494793		
5%	.3145123	.3145123		
10%	.3454413	.3232056	Obs	33
25%	.4470982	.3454413	Sum of wgt.	33
50%	.6268514		Mean	.6869312
		Largest	Std. dev.	.2815965
75%	.8702418	1.110349		
90%	1.110349	1.137206	Variance	.0792966
95%	1.159234	1.159234	Skewness	.3583444
99%	1.199676	1.199676	Kurtosis	1.935255

TE_mols_hn

	Percentiles	Smallest		
1%	.2739066	.2739066		
5%	.3453072	.3453072		
10%	.3792646	.3548517	Obs	33
25%	.490875	.3792646	Sum of wgt.	33
50%	.6882285		Mean	.7541909
		Largest	Std. dev.	.3091685
75%	.9554501	1.219067		
90%	1.219067	1.248554	Variance	.0955852
95%	1.272738	1.272738	Skewness	.3583444
99%	1.31714	1.31714	Kurtosis	1.935255

From the detailed summary of MOLS half normal and Exponential, we get various stastics to summarize our findings. We can also see that values is above 100%.



	TE_cols	TE_mol~p	TE_mol~n
TE_cols	1.0000		
TE_mols_exp	1.0000	1.0000	
TE_mols_hn	1.0000	1.0000	1.0000

Variable	Obs	Mean	Std. dev.	Min	Max
TE_mols_exp	334	.9667006	.1260321	.2494793	1
TE_mols_hn	334	.971048	.1149538	.2739066	1
TE_cols	33	.5725974	.2347271	.2079556	1

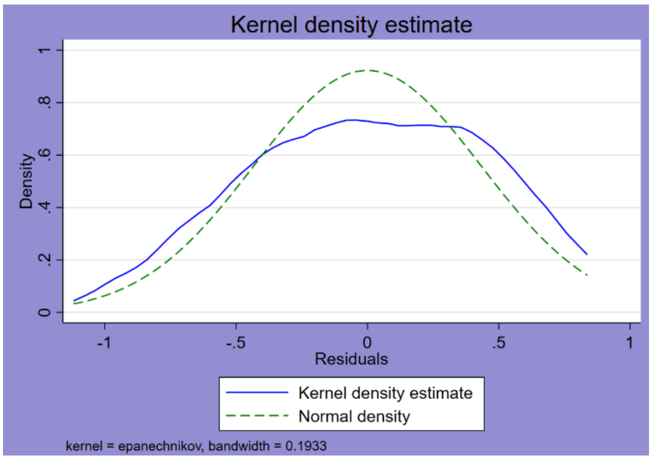
Looking at all three distributions, the mean is highest for MOLS half-normal. Therefore, most lenient model is MOLS half normal.

Stochastic Frontier Analysis (STATA)

method used to estimate the technical efficiency of a production process is known as Stochastic Frontier Analysis (SFA).

```
. sum res, detail

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -.9262109      -.9262109
 5%    -.6945638      -.6945638
10%    -.6007643      -.6672986       Obs                  33
25%    -.3428088      -.6007643       Sum of wgt.          33

50%    -.0048774                      Mean           1.14e-09
                        Largest       Std. dev.      .4323188
75%     .3231841       .5668429
90%     .5668429       .5907428       Variance       .1868996
95%     .6099275       .6099275       Skewness      -.2216275
99%     .6442196       .6442196       Kurtosis        2.09219
```

```
. sktest res

Skewness and kurtosis tests for normality
                                                       ----- Joint test -----
    Variable |        Obs   Pr(skewness)   Pr(kurtosis)   Adj chi2(2)  Prob>chi2
-------------+-----------------------------------------------------------------
         res |         33      0.5496         0.1667          2.46      0.2926
```



Here, the kernel density estimate is following the normal curve very closely. Therefore, our data estimation was correct.

Stochastic Frontier Analysis Half-Normal Model

```
Iteration 26:  log likelihood = -15.102936
Iteration 27:  log likelihood = -15.102932
Iteration 28:  log likelihood = -15.102928
Iteration 29:  log likelihood = -15.102925
Iteration 30:  log likelihood = -15.102923
Iteration 31:  log likelihood = -15.102922

Stoc. frontier normal/half-normal model              Number of obs =        33
                                                     Wald chi2(2)  =  1.39e+10
Log likelihood = -15.102922                          Prob > chi2   =    0.0000

------------------------------------------------------------------------------
       lnGls | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       lnMin |   1.158446   .0000184  6.3e+04   0.000      1.15841    1.158482
        lnPK |   .2789968   6.89e-06  4.0e+04   0.000     .2789833    .2790103
       _cons |  -6.459299   .0001379 -4.7e+04   0.000    -6.459569   -6.459028
-------------+----------------------------------------------------------------
    /lnsig2v |  -36.52059   557.4108    -0.07   0.948    -1129.026    1055.985
    /lnsig2u |  -.5362541    .246183    -2.18   0.029    -1.018764   -.0537444
-------------+----------------------------------------------------------------
     sigma_v |   1.17e-08   3.27e-06                      6.8e-246    2.0e+229
     sigma_u |   .7648106   .0941417                      .6008668    .9734857
      sigma2 |   .5849352   .1440011                      .3026983    .8671722
      lambda |   6.51e+07   .0941417                      6.51e+07    6.51e+07
------------------------------------------------------------------------------
LR test of sigma_u=0: chibar2(01) = 7.15                Prob >= chibar2 = 0.004
```

From the above analysis, we observe that sigma_v is very low and sigma_u values is 0.76481 therefore as lamda is the ratio of sigma_v and sigma_u so the value is very low. Also the values of mean, std dev, min and max are very comparable.

Stochastic Frontier Analysis Exponential Model

```
Iteration 2:   log likelihood = -18.673746
Iteration 3:   log likelihood = -18.671862
Iteration 4:   log likelihood = -18.671495
Iteration 5:   log likelihood = -18.671482
Iteration 6:   log likelihood = -18.671482

Stoc. frontier normal/exponential model             Number of obs =     33
                                                    Wald chi2(2)  =  80.59
Log likelihood = -18.671482                         Prob > chi2   = 0.0000
```

lnGls	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
lnMin	1.087236	.1591588	6.83	0.000	.7752901	1.399181
lnPK	.3334082	.1363595	2.45	0.014	.0661484	.6006679
_cons	-6.480853	1.29972	-4.99	0.000	-9.028256	-3.933449
/lnsig2v	-1.769797	.5439109	-3.25	0.001	-2.835842	-.703751
/lnsig2u	-4.482114	7.560835	-0.59	0.553	-19.30108	10.33685
sigma_v	.4127561	.1122513			.242217	.7033677
sigma_u	.106346	.4020325			.0000644	175.638
sigma2	.1816771	.0450102			.0934587	.2698956
lambda	.2576486	.5032305			-.728665	1.243962

```
LR test of sigma_u=0: chibar2(01) = 0.01              Prob >= chibar2 = 0.452
```

Variable	Obs	Mean	Std. dev.	Min	Max
TE_sfa_JML~e	33	.899443	.0244984	.8291125	.9283898
TE_sfa_BC_e	33	.9038589	.0225609	.8395385	.9307088

From the above analysis, we observe that sigma_u is 0.106346 and sigma_v values is 0.4127561 therefore as lamda is the ratio of sigma_v and sigma_u which is 0.2576486 here.

Here, the value of lamda is much higher in Exponential SFA than in Half-normal SFA.

Therefore, our exponential SFA model is better because of better lamda value.

Variable	Obs	Mean	Std. dev.	Min	Max
TE_sfa_JML~p	33	.9930929	.0001972	.9927084	.9936123
TE_sfa_BC_~p	33	.993069	.0001986	.9926817	.9935918
TE_sfa_JML~n	33	.5792741	.2408948	.2149923	.9999999
TE_sfa_BC_hn	33	.5792741	.2408948	.2149923	.9999999

Here, We see that all the SFA results for half-normal and exponential, JMLS exponential model has the heighest mean value and statdard deviation value for BC and JML half normal is same.

	TE~S_exp	TE_sfa..	TE_~S_hn	TE_sfa..
TE_sfa_JML~p	1.0000			
TE_sfa_BC_~p	1.0000	1.0000		
TE_sfa_JML~n	-0.4971	-0.4972	1.0000	
TE_sfa_BC_hn	-0.4971	-0.4972	1.0000	1.0000

SFA Model Comaprison with the deterministic models

Variable	Obs	Mean	Std. dev.	Min	Max
TE_sfa_BC_~p	33	.993069	.0001986	.9926817	.9935918
TE_cols	33	.5725974	.2347271	.2079556	1
TE_mols_exp	334	.9667006	.1260321	.2494793	1
TE_mols_hn	334	.971048	.1149538	.2739066	1

The mean value of COLS model is less compared to the sfa_BC when we look at the disctriptive statistics of the models.

. corr TE_mols_exp TE_sfa_BC_exp TE_cols TE_mols_hn
(obs=33)

	TE_mol~p	TE_sfa..	TE_cols	TE_mol~n
TE_mols_exp	1.0000			
TE_sfa_BC_~p	-0.5279	1.0000		
TE_cols	0.9889	-0.4871	1.0000	
TE_mols_hn	0.9933	-0.5474	0.9691	1.0000

Now, from the correalation we can see that MOLS exponential is highly similar and therefore, It is wise to conclude that we go with MOLS_EXP model for our analysis. Also, MOLS exponential has high mean of 96%.

Comparisons of DEA and Econometrics Model

After doing our analysis in PIM DEA with CRS and VRS and in STATA with COLS, MOLS and SFA models we obtained different results to compare and conclude our findings and finalize our best practice.

DMU	Inefficient Players	Econometric - STATA					PIM-DEA	
		lnPE_mols_exp	lnPE_COLS	lnPE_mols_hn	TE_sfa_JMLS_hn	PE_sfa_JMLS_exp	CRS(%)	VRS(%)
DMU398	Marcus Rashford	0.75846	-0.14412	0.34758	0.61245	0.60312	38.89%	56.19%
DMU25	Pierre-Emerick Aubameyang	0.85124	-0.91245	0.69854	0.71038	0.74815	46.32%	85.18%

The table above shows the comparison between the two of our inefficient players in Econamatric – STATA and PIM-DEA. As the values of model is good in Mols Exponential with 0.75846 and 0.85124. Also, the mean is better for the model as well as the standard deviation values are good as well.Therefore, This is the best model from Econometrics analysis.

The multivariate VRS output oriented model with values for the two inefficienct players is 56% and 85%, which is far better than CRS values. Hence, the variable return to scale is the best model from PIM-DEA analysis.

The variables in the SFA model underwent the Translog and Cobb-Douglas regressions, where the values were transformed to log for linear results and a functional structure with attributes of global returns to scale. The DEA VRS model is usually a flexible model because it enables local returns to scale.

The DEA VRS output orientation model appears to be the most fair estimator of them all when taking into account the constraints present in the modelling of both the econometric and DEA models. This is because it is flexible in computing the analysis without the log values, accepts flexibility in the changing parameters, and takes discrimination into account.

In conclusion, both the analysis are best although DEA analysis provides useful information for improvement and a better realiability for analysis.