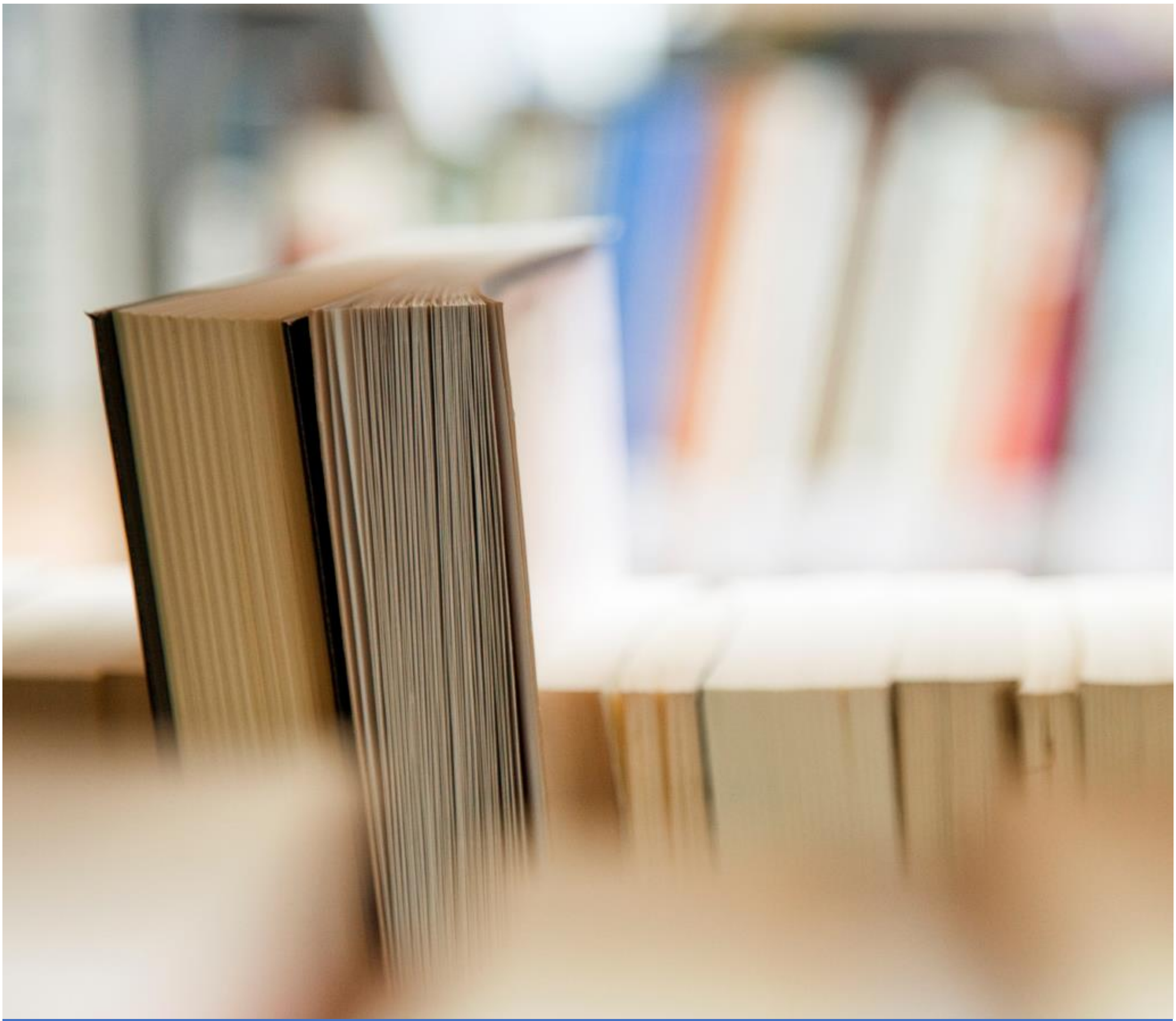


Performance analysis of electricity distribution companies

Parametric and non-parametric approach to understanding relative efficiency

Student number: 210197062

3/1/22

Performance Analytics

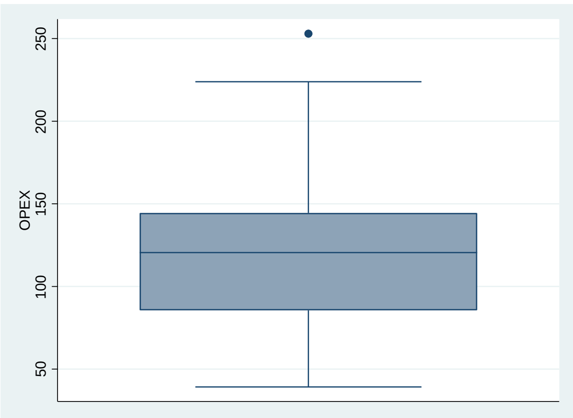
Table of Contents

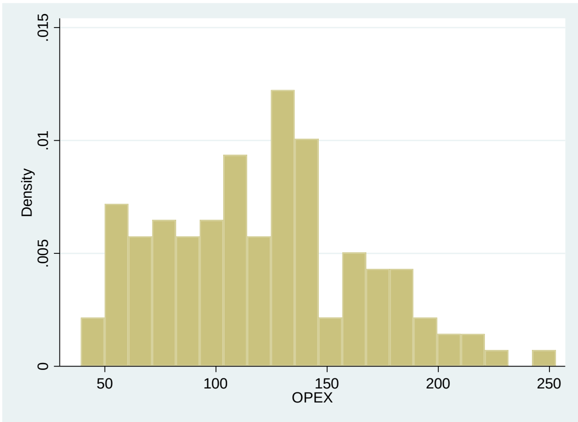
1. Understanding the variables

1.1. Operating expenditure (OPEX)

OPEX

	Percentiles	Smallest		
1%	41.62284	39.2133		
5%	55.2185	41.62284		
10%	60.14638	48.01018	Obs	130
25%	85.59116	51.01252	Sum of Wgt.	130
50%	120.4943		Mean	120.3071
		Largest	Std. Dev.	44.83498
75%	144.4473	212.4082		
90%	183.9182	215.7568	Variance	2010.176
95%	196.517	223.8748	Skewness	.3736408
99%	223.8748	252.9576	Kurtosis	2.639807



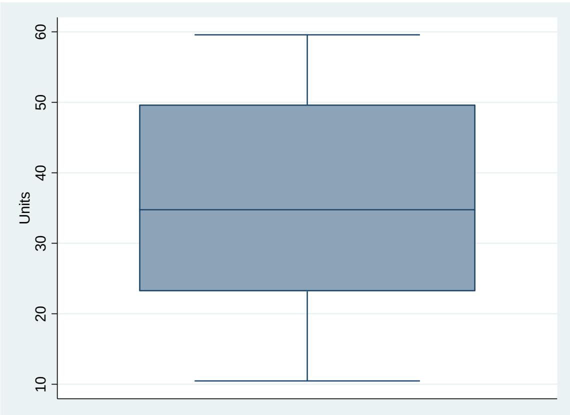


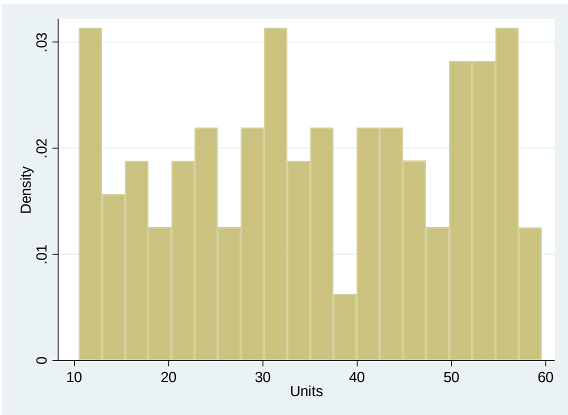
The initial observations based on the detailed descriptive statistics for the operating expenditure are as follows

1. The expenditure is spread to a large extend with values ranging from 50 to 250 showing the presence of electricity distribution companies ranging from relatively small to large players
2. Most of the companies have their expenditure in the range of 80 to 140 as seen in the boxplot
3. There is an outlier in the data which has very high operating expenditure as seen in the histogram this company seems to have 250 as its opex value which is far from the rest
4. The variation is relatively high due to extreme range of operating expenditure and moderately high coefficient of variation – 36.67%

1.2. Units

	Units			
	Percentiles	Smallest		
1%	10.87588	10.46846		
5%	12.26905	10.87588		
10%	13.97961	11.27567	Obs	130
25%	23.19315	11.4481	Sum of Wgt.	130
50%	34.75966		Mean	35.46294
		Largest	Std. Dev.	14.6778
75%	49.68169	57.7279		
90%	54.94995	57.8927	Variance	215.4377
95%	56.48274	58.6343	Skewness	-.0625031
99%	58.6343	59.57427	Kurtosis	1.755957



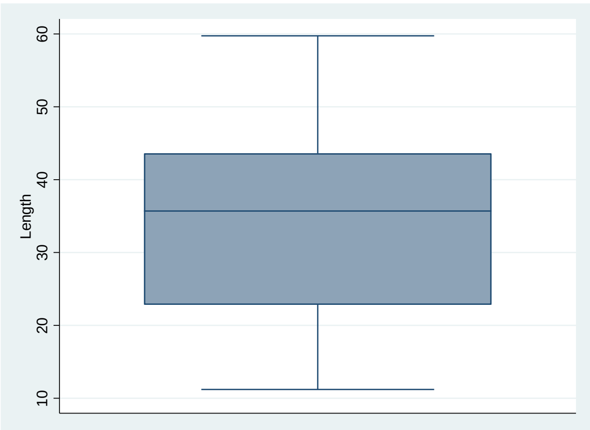


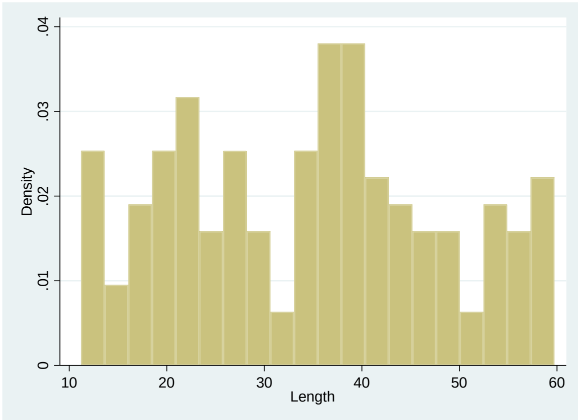
Observations

1. There is a heavy spread from the minimum to the maximum value of the units for the variable units, which shows the units handled by the companies are changing very dynamically
2. Even though with high range of spread, the concentration seems to be from around 22 units to 50 units
3. The mean value and the median value are almost overlapping

1.3. Length

```
. summ length, detail

                            Length
-------------------------------------------------------------
      Percentiles      Smallest
 1%       11.236        11.2022
 5%     13.00607         11.236
10%     16.40507       11.31535       Obs                 130
25%     22.82388       12.68258       Sum of Wgt.         130

50%      35.6943                      Mean            34.5977
                        Largest       Std. Dev.      13.41387
75%     43.62072         58.352
90%     54.53032       58.79757       Variance       179.9318
95%     57.55394       59.15311       Skewness       .0857663
99%     59.15311       59.74464       Kurtosis       2.026004
```





Observations

- All the major observations are consistent with the observation for units
- This pattern is consistent with electric companies' units delivered and length of the network being interdependent and dependent with the operating expenditure. We confirm this with correlation analysis to compare

1.4. Correlation analysis

	opex	units	length	ovshare	intens~y
opex	1.0000				
units	0.3954	1.0000			
length	0.7921	-0.0096	1.0000		
ovshare	-0.0542	-0.1557	0.0193	1.0000	
intensity	0.0062	-0.0560	0.0658	-0.1229	1.0000

Observations

- Length has the highest correlation out of all the variables against opex
- Units is moderately correlated with opex
- Ovshare and intensity has correlation coefficient value less than 0.1 which can be assumed to have relatively weak correlation with opex
- Ovshare and intensity can be considered as environmental conditions which marginally affect the opex variable

1.5. Summarizing descriptive statistics – Transformation process estimation

Based on the analysis conducted above we can say that the transformation process that can be assumed for this analysis of electricity distribution companies is Variable Returns to Scale. VRS model is preferred because of the variability of input variable is found to be high for all the independent variables and VRS also offers more flexibility and more comparable to the real-life scenario.

1.6. Variable classification

Variables	Classification	Justification
Operating expenditure (OPEX)	Cost Dependent variable (Considered as Input for DEA process)	The is the cost expenditure to the electricity company, which is a form of cost
Units	Output	The opex is dependent on this variable as it is driven by this variable, from the perspective of opex optimization this is a direct output of any changes to opex
Length	Output	The opex is driven by this variable like units, from the perspective of opex optimization this is a direct output of any changes to opex
Ovshare	Output (Weak correlation with input, dropped from DEA analysis)	The opex is dependent on this relatively weakly, this can be considered as an indirect environmental variable which assists primary affecting variables like units and length
Intensity	Output (Weak correlation with input, dropped from DEA analysis)	The opex is dependent on this relatively weakly, this can be considered as an indirect environmental variable which assists primary affecting variables like units and length

2. Non – Parametric approach, Data Envelopment Analysis (DEA) for performance analysis

2.1. Model 1 - Base model creation and summary based on descriptive analysis

Model parameter	Selected parameter	Justification

Returns to scale	Variable Returns to Scale	Described in the descriptive analytics of variables in section 1
Inputs	Operating expenditure (opex)	Main consideration of the companies which needs to be optimized and on which efficiency is based on primarily
Outputs	Units, Length	Described in descriptive analytics, we don't consider intensity and overshare since they are very weakly correlated to input opex and might not contribute in our model in an efficient manner
Orientation	Input oriented	As the efficiency is primarily based on the operating expenditure, which is the input of this model, hence input orientation

2.2. Model 2 – Alternate assumption to base model

Returns to scale	Constant Returns to Scale
Inputs	Operating expenditure (opex)
Outputs	Units, Length
Orientation	Input oriented

2.3. Summary statistics comparison for the derived models

2.3.1. VRS model

Variable	Obs	Mean	Std. Dev.	Min	Max
vrslin2outio	130	$\boxed{78.00692}$	13.84569	47.63	100

The mean efficiency score for the VRS model seems to be moderately good with a value of 78%. As VRS includes more flexibility in the model this value is relatively less constrained compared to the CRS model

2.3.2. CRS model

Variable	Obs	Mean	Std. Dev.	Min	Max
crslin2outio	130	$\boxed{69.62185}$	11.49014	46.25	100

The mean efficiency score for the CRS model is relatively low compared to VRS, amounting up to 69.6 percentage. This is expected as CRS model is more constrained and stricter in calculating the efficiency scores.

2.4. Observation for VRS against CRS

There are some observations in the model which are worth noting for the VRS model

1. Efficient pairs

   There are around 4 – 100% DMU which has considered itself and less than 3 other DMUs to compare itself with, this is not a good indicator of the model as compared to CRS which has 2 – 100& efficient DMU which has strong number of comparators.

2. Weights

   VRS – There are around 14 DMU's which have 0 weights for units and there are around 3 DMU's which have 0 weights for length

   CRS – There are around 14 DMU's which have 0 weights for units and there are no DMU's which have 0 weights for length

3. Returns to Scale on the frontier

   VRS – The frontier includes all the RTS, i.e., increasing, decreasing and constant this is indicative of the fact that model is utilizing the flexible nature of VRS model

4. Fully efficient DMU's

   CRS – there are 2 DMU's, i.e., 1.54% of total DMU which are fully efficient

   VRS – there are 15 DMU's i.e., 11.54% of total DMU's which are deemed fully efficient

5. Correlation

   The correlation between CRS and VRS is found to be 0.79833 which is very good meaning both methods are producing similar result for efficiency scores

2.5. Selection of appropriate model under non-parametric approach. Data Envelopment Analysis

The model that we are going to be considering as the representative for the parametric approach is going to be **VRS**. The decision has been based on the fact that this model is shown to make good use of flexibility of VRS i.e., having fully formed frontier with all returns to scale. Good confidence can be assumed with VRS model as the correlation with CRS model is found to be around 80% which shows almost similar values of efficiency between each other. It is also worth acknowledging we don't have the perfect model formed with VRS as highlighted above in the comparison, however it is still worth considering as the correlation with CRS model is very close and there are good factors supporting the analysis.

2.6. Analysis for electricity company – 23 [Least efficient electric company – 47.63% efficient]

2.6.1. Expenditure target

Currently this company is handling 24.1377 units of GWhrs and a length of 35179.29 km with an expenditure of 160.90 million dollars. According to our model the best performing company which is – Company 12, can do the same job of handling the units and length with only the operating expenditure of 76.63 million dollars. Hence the recommendation for the least efficient company is 84.27 million dollars. This is a huge amount of expenditure to cut down.

2.6.2. Scale efficiency

The scale efficiency for this model is 97.09%, this number denotes how efficient the decision-making unit company 23 is against its own most optimal performance (CRS input efficiency / VRS input efficiency)

2.7. Analysis for electricity company – 112 [Semi efficient electric company – 72.68% efficient]

2.7.1. Expenditure target

Currently this company is handling 53.6369 units of GWhrs and a length of 34073 km with an expenditure of 139.0498 million dollars. According to our model the best performing company which is – Company 12 or 22, can do the same job of handling the units and length with only the operating expenditure of 101.0634 million dollars. Hence the recommendation for the least efficient company is 47.03 million dollars. This is a huge amount of expenditure to cut down.

2.7.2. Scale efficiency

The scale efficiency for this model is 88.29%, this number denotes how efficient the decision-making unit company 112 is against its own most optimal performance (CRS input efficiency / VRS input efficiency)

2.8. Benefits of efficiency scores and scale efficiency to understand a company performance

Efficiency scores and target – These scores are computed to provide a context among the peers in a multiple DMU scenario, so that practices from best performing peers can be understood and can be pushed to others as a benchmark to reach. The efficiency scores with the efficiency rating also provides the target that has to be reached which is calculated based on the peer performance. This is a very useful and important data to have

Scale efficiency – Is self-critical in nature where a DMU is judged based on its own most optimal performance. This shows the lack in its own performance against its own when returns are considered in a flexible manner.

2.9. Recommendation to manager

- Identify opportunities for reducing expenditure like overhead costs
- See what is being done by the most efficient companies which are on the frontier line
- Reduce overhead network, since there is a high chance of failure of overhead networks this may be burdening company 23 by increasing expenditure. Hence recommendation is to reduce this so that company can be more efficient
- Manage high demand appropriately by distributing the load, as this can also be the reason for electric companies' failure, resulting in repair and reinstalment.

3. Econometric efficiency approximation

3.1. Model specification

3.1.1. Cobb Douglas model

Initial model is specified by considering the log transformation of all the input variables except the ovshare variable which is in proportions.

The resulting output is as below

Model type	Initial independent variables	Final independent variables	Degrees of freedom	Model performance – Adjusted R^2	Goodness of fit, f-statistics
Cobb Douglas ln(opex) – Dependent	ln(units), ln(length), ln(intensity), ovshare	ln(units), ln(length)	129	0.8729	0.0000
Interpretation		Prediction model based on 2 input variables	High amount of df equivalent to more accurate estimate	Model able to predict the dependent variable by a very good margin	Model adequately describes the data

3.1.2. Final model key parameters and statistics

```
. reg lnopex lnunits lnlength

      Source |       SS           df       MS      Number of obs   =       130
-------------+----------------------------------   F(2, 127)       =    444.08
       Model |  18.1355784         2  9.06778919   Prob > F        =    0.0000
    Residual |    2.593265       127  .020419409   R-squared       =    0.8749
-------------+----------------------------------   Adj R-squared   =    0.8729
       Total |  20.7288434       129  .160688708   Root MSE        =     .1429

------------------------------------------------------------------------------
      lnopex |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnunits |   .3277226   .0257845    12.71   0.000     .2766996    .3787455
    lnlength |   .7859432   .0286891    27.40   0.000     .7291727    .8427137
       _cons |   .8631644   .1364057     6.33   0.000     .5932422    1.133087
------------------------------------------------------------------------------
```

3.1.3. Diagnostics for Cobb-Douglas model

Test category	**Heteroskedasticity**		**Misspecification**		**Multicollinearity**
Test	B-P test	Whites test	Ramsey test	Ramsey test – RHS	Variance inflation factor
Null hypothesis	Constant variance	Homoskedasticity	Model has no omitted variable	Model has no omitted variable – RHS	-
Observation	p-value - 0.3375	p-value - 0.4483	p-value - 0.1972	p-value - 0.0916	1.0 for both input variables
Remarks	**Pass** – fail to reject null hypothesis	**Pass** – fail to reject null hypothesis	**Pass** – fail to reject null hypothesis	**Pass** – fail to reject null hypothesis	**Pass**, vif values less than 10

3.1.4. Returns to scale

Test is conducted to verify if the sum of the input variables is equal to 1 assuming the null hypothesis of Constant Returns to Scale

The result obtained is

```
( 1)   lnunits + lnlength = 1

       F(  1,    127) =     8.38
            Prob > F =     0.0045
```

Hence, we reject the null hypothesis of CRS to understand more if this is IRS or DRS under Variable Return to Scale, the sum of standardized Beta coefficient for the input variables is equal to

```
. display _b[lnunits] + _b[lnlength]
1.1136658
```

As we can see the sum is equal to approximately **1.1** hence concluding this is an **Increasing Returns to Scale model**

3.2. Translog model

The Translog model is constructed by considering all the second order term and the interaction term for the input variables. Carrying over from the Cobb Douglas model, we have 2 input variables – ln(units), ln(length) which is considered for creating the second order and interaction terms

Model type	**Initial independent variables**	**Final independent variables**	**Degrees of freedom**	**Model performance – Adjusted R^2**	**Goodness of fit, f-statistics**
Translog ln(opex) – Dependent	ln(units), ln(length), 0.5*ln(units)^2, 0.5*ln(length)^2, ln(units)*ln(length)	ln(units), ln(length), 0.5*ln(units)^2, 0.5*ln(length)^2	129	0.8801	0.0000
Interpretation		Prediction model based on 4 input variables	High amount of df equivalent to more accurate estimate	Model able to predict the dependent variable by a very good margin	Model adequately describes the data

3.2.1. Final model key parameters and statistics

```
. reg lnopex lnunits lnlength lnunit2 lnlength2

      Source |       SS           df       MS      Number of obs   =       130
-------------+----------------------------------   F(4, 125)       =    237.82
       Model |  18.321369         4  4.58034225   Prob > F        =    0.0000
    Residual |  2.40747436       125  .019259795   R-squared       =    0.8839
-------------+----------------------------------   Adj R-squared   =    0.8801
       Total |  20.7288434       129  .160688708   Root MSE        =    .13878

------------------------------------------------------------------------------
      lnopex |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnunits |   1.216971    .362998     3.35   0.001      .498553    1.935389
    lnlength |   1.670853   .4065831     4.11   0.000     .8661744    2.475531
     lnunit2 |  -.2700238   .1096642    -2.46   0.015     -.487063   -.0529847
   lnlength2 |  -.2622084   .1217197    -2.15   0.033    -.5031068     -.02131
       _cons |  -2.032492   .9476917    -2.14   0.034    -3.908091   -.1568925
------------------------------------------------------------------------------
```

3.2.2. Diagnostics for Translog model

Test category	Heteroskedasticity		Misspecification		Multicollinearity
Test	B-P test	Whites test	Ramsey test	Ramsey test – RHS	Variance inflation factor
Null hypothesis	Constant variance	Homoskedasticity	Model has no omitted variable	Model has no omitted variable – RHS	-
Observation	p-value - 0.2881	p-value - 0.6432	p-value - 0.5506	p-value - 0.0650	>200 for all variables
Remarks	**Pass** – fail to reject null hypothesis	**Pass** – fail to reject null hypothesis	**Pass** – fail to reject null hypothesis	**Pass** – fail to reject null hypothesis	**Fail**, vif values greater than 10, as this is expected since the input variables are second order term

3.3. Functional form selection through analysis

We select the Translog functional form due to reasons stated as below

- Two additional Translog variables compared to Cobb-Douglas model, which is supported by good number of observations

- Better adjusted R^2 for Translog function showing slightly better performance
- Translog in general being a more flexible functional form
- Diagnostics tests are passing in Translog function i.e., Heteroskedasticity & Misspecification. Multicollinearity is bound to fail as the input consists of interaction terms.

4. Deterministic frontier

4.1. COLS frontier

The corrected OLS frontier or COLS is a method which approximates the inefficiency by which the frontier envelopes all the observations above it. This means that COLS assumes all the residuals is due to inefficiency only without consideration of stochastic element.

The summary of the efficiency scores estimated using COLS frontier is as follows:

```
. sum CE_cols, detail

                           CE_cols
-------------------------------------------------------------
      Percentiles      Smallest
 1%      .574647       .5560023
 5%     .5989652        .574647
10%     .6148771       .5787959       Obs                 130
25%     .6856813       .5832774       Sum of Wgt.         130

50%     .7858989                      Mean           .7728605
                        Largest       Std. Dev.      .1023966
75%     .8404677       .9576212
90%     .8982156       .9617656       Variance       .0104851
95%      .923686       .9648663       Skewness       -.215245
99%     .9648663              1       Kurtosis       2.278367
```

As we can see the mean efficiency score using deterministic frontier COLS is 0.7728 which is considerably low, this is mainly because this is a very constricted approach and very strictly estimates the efficiency which can be related to the assumptions such as enveloping of all the observations under the frontier and all residuals due to inefficiency.

4.1.1. Ranking electricity distribution units under COLS

Top 5 most efficient   Bottom 5 least efficient

unit	CE_cols	r_cols
68	1	1
22	.9648663	2
46	.9617656	3
49	.9576212	4
87	.9286239	5

unit	CE_cols	r_cols
23	.5560023	130
69	.574647	129
29	.5787959	128
55	.5832774	127
51	.5838338	126

4.2. MOLS frontier

In MOLS the frontier estimation is not as extreme as COLS, where we shift the regression line to form the frontier based on some of least squared. MOLS is done based on two distributions. MOLS does not assume all the deviations in residuals are due to inefficiencies, not as critical as COLS

4.2.1. Half normal distribution of inefficiency

Efficiency score estimation is carried out based on half normally distributed inefficiency. Initially the inefficiency scores calculated based on rmse times numerical factor (less than 1) and residuals and then this is converted to efficiency scores. The summary is as below

```
. sum CE_mols_hn, detail

                         CE_mols_hn
-------------------------------------------------------------
      Percentiles      Smallest
 1%      .671653       .6498609
 5%     .7000763        .671653
10%     .7186744       .6765022       Obs                 130
25%     .8014309       .6817403       Sum of Wgt.         130

50%     .9185663                      Mean           .9033269
                        Largest       Std. Dev.      .1196822
75%     .9823468       1.119277
90%     1.049843       1.124121       Variance       .0143238
95%     1.079613       1.127745       Skewness       -.215245
99%     1.127745        1.16881       Kurtosis       2.278367
```

As we can see the mean value is found to be 0.9033 which is substantially higher than the value obtained from COLS approach. This is expected as MOLS is not a severe approach like COLS, where MOLS does not estimate all inefficiency is due to residual deviations

4.2.1.1. Ranking electricity distribution units under MOLS – half normal

<u>Top 5 most efficient</u>

<u>Bottom 5 least efficient</u>

unit	CE_mol~n	r_mols~n
68	1.16881	1
22	1.127745	2
46	1.124121	3
49	1.119277	4
87	1.085385	5

unit	CE_mol~n	r_mols~n
23	.6498609	130
69	.671653	129
29	.6765022	128
55	.6817403	127
51	.6823907	126

4.2.2. Exponential distribution of inefficiency

Efficiency score estimation is carried out based on Exponentially distributed inefficiency. Initially the inefficiency scores calculated based on RMSE and residuals and then this is converted to efficiency scores. The summary is as below

```
. sum CE_mols_exp, detail

                         CE_mols_exp
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .6530752       .6318859
 5%     .6807123       .6530752
10%      .698796       .6577903       Obs                 130
25%     .7792636       .6628835       Sum of Wgt.         130

50%      .893159                      Mean           .8783411
                        Largest       Std. Dev.      .1163718
75%     .9551753       1.088318
90%     1.020805       1.093028       Variance       .0135424
95%     1.049751       1.096552       Skewness      -.2152451
99%     1.096552       1.136481       Kurtosis       2.278367
```

As explained above due to MOLS assumptions we observe relatively higher value of mean, however we see this value lesser than MOLS for half normal since we consider a higher value of estimated u which is equal to RMSE value

4.2.2.1. Ranking electricity distribution units under MOLS - exponential

Top 5 most efficient

unit	CE_mol~p	r_mols~p
68	1.136481	1
22	1.096552	2
46	1.093028	3
49	1.088318	4
87	1.055363	5

Bottom 5 least efficient

unit	CE_mol~p	r_mols~p
23	.6318859	130
69	.6530752	129
29	.6577903	128
55	.6628835	127
51	.6635159	126

5. Stochastic Frontier Analysis

This is a non-deterministic approach for calculating the efficiency scores for the DMU, here the advantage is there is consideration of noise factor or the random element such as luck involved.

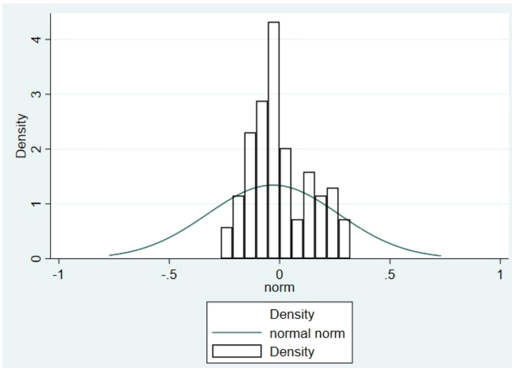
5.1. Skewness and noise level

The summarization of residuals shows a positive skew which is amounting to 0.457 shows the presence of inefficiency and the statistical test is run to confirm zero value of skew result is as below, whose p- value rejects null hypothesis and confirms a zero value for skewness.

. sktest res

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2) (joint)	Prob>chi2 (joint)
res	130	0.0317	0.0595	7.48	0.0238



For noise level analysis we consider the standard deviation values in stochastic analysis

Stoc. frontier normal/exponential model — Number of obs = 130; Wald chi2(4) = 1338.97; Log likelihood = 78.156617; Prob > chi2 = 0.0000

lnopex	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnunits	1.152615	.302799	3.81	0.000	.5591396	1.74609
lnlength	1.364382	.3520737	3.88	0.000	.6743298	2.054433
lnunit2	-.2575677	.0916573	-2.81	0.005	-.4372126	-.0779227
lnlength2	-.1728751	.1055175	-1.64	0.101	-.3796856	.0339355
_cons	-1.499622	.818947	-1.83	0.067	-3.104729	.1054843
/lnsig2v	-5.305675	.3700965	-14.34	0.000	-6.031051	-4.5803
/lnsig2u	-4.065402	.2986104	-13.61	0.000	-4.650668	-3.480136
sigma_v	.070451	.0130368			.0490201	.1012513
sigma_u	.1309813	.0195562			.0977508	.1755084
sigma2	.0221194	.0043027			.0136862	.0305527
lambda	1.859182	.029211			1.80193	1.916435

LR test of sigma_u=0: chibar2(01) = 6.67 Prob >= chibar2 = 0.005

As we can see the model incorporates some noise element even though it is a bit low, it is slightly significant and hence this model is a good model. The likelihood ratio seems to be passing as it is equal 0. The lambda value is 1.85 showing the ratio of inefficiency to noise to be slightly high, however worth consideration.

Efficiency is estimated using JMLS and BC calculates efficiencies based on half normally distributed efficiencies and exponentially distributed efficiencies

5.2. Ranking electricity distribution units under SFA – BC – exponential

Top 5 most efficient

unit	CE_sfa..	r_~C_exp
68	.9756435	1
22	.9754132	2
46	.971827	3
49	.9717987	4
20	.9690906	5

Bottom 5 least efficient

unit	CE_sfa..	r_~C_exp
23	.6579043	130
69	.66991	129
29	.6811193	128
51	.683438	127
55	.6887261	126

5.3. Ranking electricity distribution units under SFA – JMLS – Exponential

Top 5 most efficient

unit	CE~S_exp	r_~S_exp
68	.97589	1
22	.9756637	2
46	.9721422	3
49	.9721143	4
20	.9694573	5

Bottom 5 least efficient

unit	CE~S_exp	r_~S_exp
23	.659539	130
69	.6715745	129
29	.6828117	128
51	.6851361	127
55	.6904374	126

5.4. Ranking electricity distribution units under SFA – BC – Half normal

Top 5 most efficient

unit	CE_sfa..	r_s~C_hn
68	.9765612	1
22	.9744003	2
49	.9712986	3
46	.9710667	4
20	.9676871	5

Bottom 5 least efficient

unit	CE_sfa..	r_s~C_hn
23	.6315734	130
69	.6402825	129
29	.65306	128
51	.6544893	127
55	.6587658	126

5.5. Ranking electricity distribution units under SFA – JMLS – Half normal

Top 5 most efficient

unit	CE_~S_hn	r_s~S_hn
68	.9767718	1
22	.9746426	2
49	.9715879	3
46	.9713596	4
20	.9680321	5

Bottom 5 least efficient

unit	CE_~S_hn	r_s~S_hn
23	.6324555	130
69	.6411766	129
29	.653972	128
51	.6554033	127
55	.6596858	126

5.6. Summary of Efficiency score for SFA analysis

Summary statistics for all the various efficiency calculated using JMLS and BC estimates show a very similar trend as we can see almost all the statistics namely – Mean, std deviation, min and max values are very near to each other, we create correlation table to further confirm the closeness of the efficiency scores

. sum CE_sfa_JMLS_exp CE_sfa_BC_exp CE_sfa_JMLS_hn CE_sfa_BC_hn

Variable	Obs	Mean	Std. Dev.	Min	Max
CE_sfa_JML~p	130	.8817618	.0863421	.659539	.97589
CE_sfa_BC_~p	130	.8804583	.0867579	.6579043	.9756435
CE_sfa_JML~n	130	.8504919	.094304	.6324555	.9767718
CE_sfa_BC_hn	130	.8495394	.0944103	.6315734	.9765612

As suspected correlation scores also show very strong correlation between the estimators from both the approaches, as per the general practice we consider **BC estimate using exponential distribution**, since half normal distribution must be taken into consideration with suspicion.

	CE~S_exp	CE_sfa..	CE_~S_hn	CE_sfa..
CE_sfa_JML~p	1.0000			
CE_sfa_BC_~p	1.0000	1.0000		
CE_sfa_JML~n	0.9896	0.9902	1.0000	
CE_sfa_BC_hn	0.9893	0.9899	1.0000	1.0000

5.7. Comparison between deterministic and stochastic frontier efficiency estimates

. sum CE_cols CE_mols_exp CE_mols_hn CE_sfa_BC_exp

Variable	Obs	Mean	Std. Dev.	Min	Max
CE_cols	130	.7728605	.1023966	.5560023	1
CE_mols_exp	130	.8783411	.1163718	.6318859	1.136481
CE_mols_hn	130	.9033269	.1196822	.6498609	1.16881
CE_sfa_BC_~p	130	.8804583	.0867579	.6579043	.9756435

The key observations that can be understood from the summary table is as below

- COLS frontier is the most critical approach with mean efficiency lowest as expected since COLS considers strong assumption of all residual deviations due to inefficiency
- MOLS – half normal is the most laid-back approach for efficiency estimation with mean efficiency lowest as it is less restrictive in its assumption of inefficiency.
- Stochastic frontier lies midway in between as it considers the inefficiency and noise or luck element, this may be an ideal efficiency estimate for consideration

6. Efficiency comparisons

6.1. Parametric efficiency scores vs nonparametric efficiency scores

Considered efficiency scores

Non-Parametric: VRS input oriented model

Parametric: COLS frontier, MOLS frontier(exponential), SFA frontier (BC-estimation-exp)

	vrslin~o	CE_cols	CE_mol~p	CE_sfa..
vrslin2outio	1.0000			
CE_cols	0.6721	1.0000		
CE_mols_exp	0.6721	1.0000	1.0000	
CE_sfa_BC_~p	0.7084	0.9528	0.9528	1.0000

Variable	Obs	Mean	Std. Dev.	Min	Max
vrslin2outio	130	78.00692	13.84569	47.63	100
CE_cols	130	.7728605	.1023966	.5560023	1
CE_mols_exp	130	.8783411	.1163718	.6318859	1.136481
CE_sfa_BC_~p	130	.8804583	.0867579	.6579043	.9756435

Interpretation

- It can be said that all the efficiency scores that can be considered to be significantly correlated
- VRS from non-parametric and SFA frontier are 2 best estimates from each category, this claim is made on the basis that
- They are more representative of the real-life situation in both estimates, where VRS involves flexibility which is reflected by the Translog model considered for analysis and SFA analysis which considers the luck, environmental or noise element into our analysis
- The efficiency scores from both of these methods are significantly correlated up to around 71%, which is a good sign saying their estimates are almost similar
- It is also worth noting the estimates from the parametric approach are very closely related to each other, with over 90% correlation, this is a good sign of the strength of the model with closely related values
- The mean efficiency value from COLS is found to be the most critical evaluation as expected since COLS considered most strict assumption out of all which is that of all residual deviation being due to inefficiency values, while the most relaxed assumption is due to SFA analysis, due to its flexibility in the consideration of inefficiency with noise elements.

6.2. Deterministic frontiers vs Stochastic frontiers

Considered efficiency scores

Deterministic frontiers: COLS frontier, MOLS frontier (exponential)

Stochastic frontiers: SFA frontier (BC estimation – exponential)

	CE_cols	CE_mol~p	CE_sfa..
CE_cols	1.0000		
CE_mols_exp	1.0000	1.0000	
CE_sfa_BC_~p	0.9528	0.9528	1.0000

Variable	Obs	Mean	Std. Dev.	Min	Max
CE_cols	130	.7728605	.1023966	.5560023	1
CE_mols_exp	130	.8783411	.1163718	.6318859	1.136481
CE_sfa_BC_~p	130	.8804583	.0867579	.6579043	.9756435

Interpretation

- The efficiency values from the COLS frontier are correlated 100% with MOLS exponential which shows the closeness of values for both of the deterministic approach
- The correlation of the SFA is also found to be very close to the deterministic approaches which is also a good sign of the model quality
- Out of the considered efficiency scores the least mean is from COLS frontier and most is from SFA frontier as explained above
- MOLS is also a laid-back approach for efficiency estimation with mean efficiency very high as it is less restrictive in its assumption of inefficiency.
- Stochastic frontier lies midway in between as it considers the inefficiency and noise or luck element, this is an ideal efficiency estimate for consideration

6.3. Disagreements at company level

Non-realistic target estimates

As we saw with non-parametric estimates, some DMU's have non realistic targets of over 50% reduction of operating expenditure, this may not be feasible as such a sudden reduction can have serious ramifications to the company.

All inputs not considered for analysis

The locally affecting inputs for each of the DMU, such as location which can heavily affect the operating expenditure of the electric company for example the overhead costs such as rent and resource cost maybe much varied from one location to another or buying some resource in bulk at a large distribution company maybe costing cheaper compared to smaller company which may render improvement of efficiency not possible.

Non-discretionary factors

There maybe non-discretionary input factors which may have been missed in the analysis leading to biased efficiency scores. In our analysis no environmental variables are considered, this can lead to wrong values which maybe disagreed by company level where the model shows one efficiency and the reality maybe different

6.4. Reliability of results

6.4.1. Non – Parametric approach

- The model that we are going to be considering as the representative for the parametric approach is going to be **VRS**.
- The decision has been based on the fact that this model is shown to make good use of flexibility of VRS i.e., having fully formed frontier with all returns to scale.

- Good confidence can be assumed with VRS model as the correlation with CRS model is found to be around 80% which shows almost similar values of efficiency between each other.
- It is also worth acknowledging we don’t have the perfect model formed with VRS as highlighted above in the comparison, however it is still worth considering as the correlation with CRS model is very close and there are good factors supporting the analysis.

6.4.2. Parametric approach

- The COLS efficiency is less reliable because it considers purely inefficiency without consideration for any environmental variables
- MOLS efficiency is relatively better since the model assumptions are not as restrictive as COLS, however this is not able to justify for the noise or the environment variable
- SFA is the most realistic approach since it is more comparable to the real life scenario, this is the most reliable score as it represents reality while also being closely correlated to the COLS and MOLS efficiency score.

7. Appendix

7.1. Initial model – Cobb Douglas

```
. reg lnopex lnunits lnlength lnintensity ovshare
```

Source	SS	df	MS
Model	18.1407797	4	4.53519493
Residual	2.58806365	125	.020704509
Total	20.7288434	129	.160688708

Number of obs	=	130
F(4, 125)	=	219.04
Prob > F	=	0.0000
R-squared	=	0.8751
Adj R-squared	=	0.8712
Root MSE	=	.14389

lnopex	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnunits	.3260337	.0263705	12.36	0.000	.2738432	.3782243
lnlength	.7858752	.0290433	27.06	0.000	.728395	.8433555
lnintensity	.0040882	.0251496	0.16	0.871	-.0456859	.0538624
ovshare	-.0234593	.0534719	-0.44	0.662	-.1292868	.0823681
_cons	.8811341	.1472531	5.98	0.000	.589702	1.172566

7.2. Initial model - Translog

```
. reg lnopex lnunits lnlength lnunit2 lnlength2 lnunitslength
```

Source	SS	df	MS
Model	18.3250001	5	3.66500002
Residual	2.4038433	124	.019385833
Total	20.7288434	129	.160688708

Number of obs	=	130
F(5, 124)	=	189.06
Prob > F	=	0.0000
R-squared	=	0.8840
Adj R-squared	=	0.8794
Root MSE	=	.13923

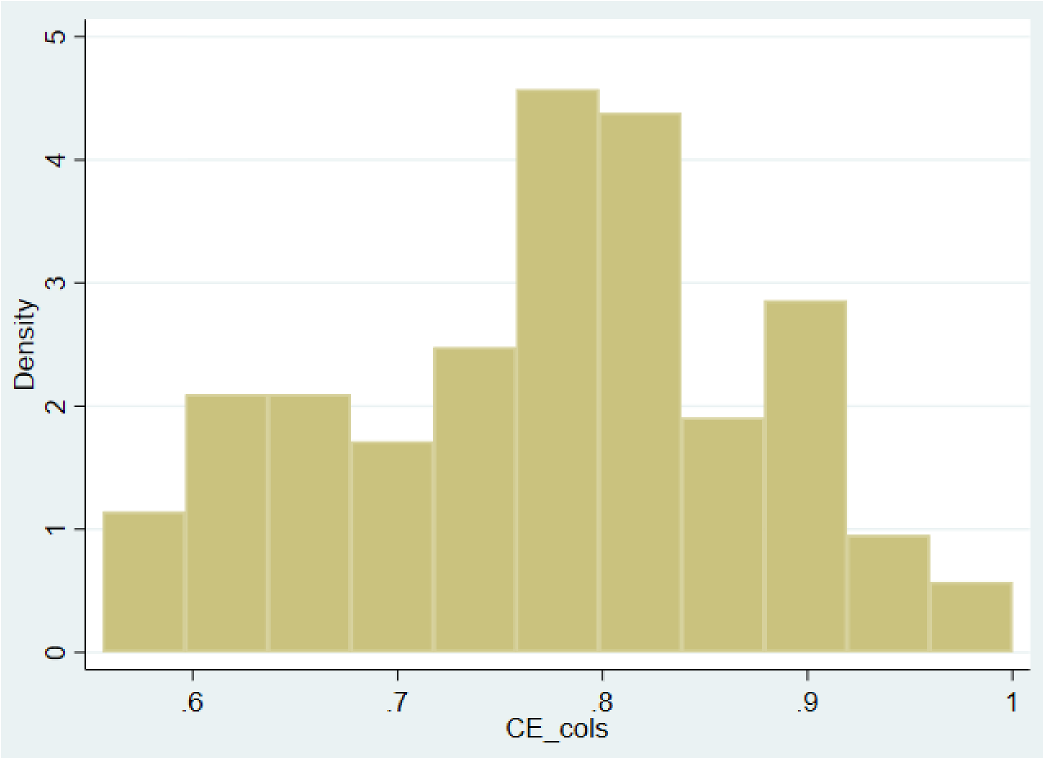
lnopex	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnunits	1.091808	.465046	2.35	0.020	.1713517	2.012264
lnlength	1.588627	.4499873	3.53	0.001	.6979755	2.479278
lnunit2	-.2630029	.111212	-2.36	0.020	-.4831227	-.0428831
lnlength2	-.2674465	.1227156	-2.18	0.031	-.5103351	-.0245578
lnunitslength	.0287818	.0665032	0.43	0.666	-.1028468	.1604103
_cons	-1.670268	1.266686	-1.32	0.190	-4.177394	.8368582

7.3. Test for elasticity

```
. test lnunits + lnlength + lnunit2 + lnlength2 = 1

 ( 1)  lnunits + lnlength + lnunit2 + lnlength2 = 1

       F(  1,   125) =   11.24
            Prob > F =    0.0011

.
end of do-file

. do "C:\Users\210197~1\AppData\Local\Temp\STD2300_000000.tmp"

. test lnunits = lnlength

 ( 1)  lnunits - lnlength = 0

       F(  1,   125) =    0.79
            Prob > F =    0.3769

.
end of do-file

. do "C:\Users\210197~1\AppData\Local\Temp\STD2300_000000.tmp"

. test lnunits = lnunit2

 ( 1)  lnunits - lnunit2 = 0

       F(  1,   125) =    9.91
            Prob > F =    0.0021
```

7.4. COLS detailed summary

```
. sum CE_cols, detail

                           CE_cols
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .574647       .5560023
 5%    .5989652        .574647
10%    .6148771       .5787959       Obs                 130
25%    .6856813       .5832774       Sum of Wgt.         130

50%    .7858989                      Mean           .7728605
                        Largest       Std. Dev.      .1023966
75%    .8404677       .9576212
90%    .8982156       .9617656       Variance       .0104851
95%     .923686       .9648663       Skewness       -.215245
99%    .9648663              1       Kurtosis       2.278367
```



7.5. Detailed summary of MOLS – EXP & HN efficiency

```
. sum CE_mols_hn, detail

                          CE_mols_hn
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .671653       .6498609
 5%    .7000763        .671653
10%    .7186744       .6765022       Obs                 130
25%    .8014309       .6817403       Sum of Wgt.         130

50%    .9185663                      Mean           .9033269
                        Largest      Std. Dev.      .1196822
75%    .9823468       1.119277
90%    1.049843       1.124121       Variance       .0143238
95%    1.079613       1.127745       Skewness      -.215245
99%    1.127745        1.16881       Kurtosis       2.278367
```

```
. sum CE_mols_exp, detail

                          CE_mols_exp
-------------------------------------------------------------
      Percentiles      Smallest
 1%    .6530752       .6318859
 5%    .6807123       .6530752
10%     .698796       .6577903       Obs                 130
25%    .7792636       .6628835       Sum of Wgt.         130

50%     .893159                      Mean           .8783411
                        Largest      Std. Dev.      .1163718
75%    .9551753       1.088318
90%    1.020805       1.093028       Variance       .0135424
95%    1.049751       1.096552       Skewness      -.2152451
99%    1.096552       1.136481       Kurtosis       2.278367
```

Aston University
Birmingham