APPLIED ECONOMETRICS

PROJECT 2

Spring, 2023
LESE207 Presentation

2021140655 Kim Sang Ock

Nerlove, **Returns to Scale in Electricity Supply**: Econometric Analysis

Table of Contents

Review: Nerlove, Returns to Scale in Electricity Supply

Summary of the derivation of the cost function for econometric analysis

- Attributes of Electricity Production: Absence of mass storage technology, Unionized industrial structure, Existence of regulatory agencies, Long-term contract-oriented, and Macro-in which personal utility does not have a significant impact.

- Modified Cobb-Douglas Production Function with dependent variable cost:

$$cost = k y^{\frac{1}{r}} P_l^{\frac{a_1}{r}} P_k^{\frac{a_2}{r}} P_f^{\frac{a_3}{r}} v$$

- whether $r = a_1 + a_2 + a_3$ is true determines the Return to Scale of Prices.

- Then we would estimate $\log(Cost) = \beta_1 + \beta_2 \log(Y) + \beta_3 \log(P_L) + \beta_4 \log(P_F) + \beta_5 \log(P_K) + u$ for test it and find some nonlinear relationship between log(cost) and log(y=output).

Econometric Analysis 1

Estimating Baseline Model

$$\log(Cost) = \beta_1 + \beta_2 \log(Y) + \beta_3 \log(P_L) + \beta_4 \log(P_F) + \beta_5 \log(P_K) + u$$

Dependent Variable: LOG(COSTS)
Method: Least Squares
Date: 05/17/23 Time: 17:00
Sample: 1 145
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.526503	1.774367	-1.987471	0.0488
LOG(KWH)	0.720394	0.017466	41.24448	0.0000
LOG(PL)	0.436341	0.291048	1.499209	0.1361
LOG(PF)	0.426517	0.100369	4.249483	0.0000
LOG(PK)	-0.219888	0.339429	-0.647819	0.5182

R-squared	0.925955	Mean dependent var	1.724663
Adjusted R-squared	0.923840	S.D. dependent var	1.421723
S.E. of regression	0.392356	Akaike info criterion	1.000578
Sum squared resid	21.55201	Schwarz criterion	1.103224
Log likelihood	-67.54189	Hannan-Quinn criter.	1.042286
F-statistic	437.6863	Durbin-Watson stat	1.013062
Prob(F-statistic)	0.000000		

Estimated Result:

$$\log(Cost) = \underset{(1.774367)}{-3.526503} + \underset{(0.017466)}{0.720394} \log(Y) + \underset{(0.291048)}{0.436341} \log(P_l)$$

$$+ \underset{(0.100369)}{0.426517} \log(P_f) - \underset{(0.339429)}{0.219888} \log(P_k) + u$$

$$R^2 = 0.925955$$

Econometric Analysis 1

Direct F−Test for $\boldsymbol{H_0: \beta_3 + \beta_4 + \beta_5 = 1 \; vs \; H_1: \beta_3 + \beta_4 + \beta_5 \neq 1}$

$$\log(Cost) = \beta_1 + \beta_2 \log(Y) + \beta_3 \log(P_L) + \beta_4 \log(P_F) + \beta_5 \log(P_K) + u$$

Wald Test:
Equation: EP2_MODEL1

Test Statistic	Value	df	Probability
t-statistic	-0.757404	140	0.4501
F-statistic	0.573660	(1, 140)	0.4501
Chi-square	0.573660	1	0.4488

Null Hypothesis: C(3) + C(4) + C(5) = 1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(3) + C(4) + C(5)	-0.357030	0.471387

Restrictions are linear in coefficients.

Result Interpretation:

- The F−statistic of this test is 0.573660 and the p−value is 0.4051 with following $F_{(1,140)}$.
- So, the null hypothesis is not rejected.
- We cannot rule out the possibility that the restriction $\beta_3 + \beta_4 + \beta_5 = 1$ of null hypothesis H_0 is true.

Econometric Analysis 1

Indirect F-Test for $\boldsymbol{H_0: \beta_3 + \beta_4 + \beta_5 = 1 \ vs \ H_1: \beta_3 + \beta_4 + \beta_5 \neq 1}$

Derivation of Model A Under $\boldsymbol{H_0}$

$\log(Cost) = \beta_1 + \beta_2 \log(Y) + \beta_3 \log(P_L) + \beta_4 \log(P_F) + \boldsymbol{\beta_5} \log(P_K) + u$

$\Leftrightarrow \log(Cost) = \beta_1 + \beta_2 \log(Y) + \beta_3 \log(P_L) + \beta_4 \log(P_F) + (\boldsymbol{1 - \beta_3 - \beta_4}) \log(P_K) + u$

$\Leftrightarrow \log(Cost) - \boldsymbol{log(P_K)} = \beta_1 + \beta_2 \log(Y) + \beta_3 (\log(P_L) - \boldsymbol{log(P_K)}) + \beta_3 (\log(P_F) - \boldsymbol{log(P_K)}) + u$

= Model A

- Under H_0 , $r = a_1 + a_2 + a_3 \Leftrightarrow 1 = \frac{a_1}{r} + \frac{a_2}{r} + \frac{a_3}{r}$ holds for the theoretical model in the paper:

$$c = k y^{\frac{1}{r}} P_l^{\frac{a_1}{r}} P_k^{\frac{a_2}{r}} P_f^{\frac{a_3}{r}} v$$

- We can check whether the Constant Return to Scale of Prices holds by comparing Baseline Model and Model A.

Econometric Analysis 1

Estimating Model A & Getting F-stat

$$\log(Cost) - \log(P_K) = \beta_1 + \beta_2 \log(Y) + \beta_3(\log(P_L) - \log(P_K)) + \beta_3(\log(P_F) - \log(P_K)) + u$$

Dependent Variable: LOG(COSTS)-LOG(PK)
Method: Least Squares
Date: 05/17/23 Time: 11:52
Sample: 1 145
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.690789	0.884871	-5.301098	0.0000
LOG(KWH)	0.720688	0.017436	41.33398	0.0000
LOG(PL)-LOG(PK)	0.592910	0.204572	2.898291	0.0044
LOG(PF)-LOG(PK)	0.414471	0.098951	4.188644	0.0000

R-squared	0.927088	Mean dependent var	-3.432113
Adjusted R-squared	0.925537	S.D. dependent var	1.435663
S.E. of regression	0.391762	Akaike info criterion	0.990874
Sum squared resid	21.64032	Schwarz criterion	1.072991
Log likelihood	-67.83836	Hannan-Quinn criter.	1.024241
F-statistic	597.6166	Durbin-Watson stat	1.015369
Prob(F-statistic)	0.000000		

Result Interpretation:

- $F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-r)} = \frac{(21.64032-21.55201)/1}{21.55201/(145-5)} = 0.573654.$
- Its p-value is 0.450083 with following $F(1{,}140)$.
- we cannot reject H0 then there is no evidence that $\beta_3 + \beta_4 + \beta_5 = 1$ is not true.
- We can say that the evidence, which Constant Return to Prices ($r = a_1 + a_2 + a_3$) does not holds is not enough.

Econometric Analysis 2

Direct F–Test for $H_0: \beta_2 = 1 \ vs \ H_1: \beta_2 < 1$ (Model A)

$$\log(Cost) - \log(P_K) = \beta_1 + \beta_2 \log(Y) + \beta_3(\log(P_L) - \log(P_K)) + \beta_3(\log(P_F) - \log(P_K)) + u$$

Wald Test:
Equation: EQ2_MODEL2

Test Statistic	Value	df	Probability
t-statistic	-16.01956	141	0.0000
F-statistic	256.6262	(1, 141)	0.0000
Chi-square	256.6262	1	0.0000

Null Hypothesis: C(2)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(2)	-0.279312	0.017436

Restrictions are linear in coefficients.

Result Interpretation:

- The F–statistic of this test is 256.6262 and the p–value is 0.0000··· with following $F_{(1,141)}$.
- So, the null hypothesis is rejected.
- We rule out the possibility that the restriction $\beta_2 = 1$ of null hypothesis H_0 is true.

Econometric Analysis 3

Estimating Model C

$$\log(Cost) - \log(P_K) = \beta_1 + \beta_2 \log(Y) + \beta_3 (\log(Y))^2 + \beta_4(\log(P_L) - \log(P_K)) + \beta_5(\log(P_F) - \log(P_K)) + u$$

Dependent Variable: LOG(COSTS)-LOG(PK)
Method: Least Squares
Date: 05/17/23 Time: 12:03
Sample: 1 145
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.764649	0.701727	-5.364837	0.0000
LOG(KWH)	0.152547	0.061860	2.465992	0.0149
LOG(KWH)^2	0.050514	0.005364	9.417823	0.0000
LOG(PL)-LOG(PK)	0.480586	0.161072	2.983663	0.0034
LOG(PF)-LOG(PK)	0.445248	0.077765	5.725535	0.0000

R-squared	0.955366	Mean dependent var	-3.432113
Adjusted R-squared	0.954091	S.D. dependent var	1.435663
S.E. of regression	0.307612	Akaike info criterion	0.513919
Sum squared resid	13.24751	Schwarz criterion	0.616565
Log likelihood	-32.25909	Hannan-Quinn criter.	0.555627
F-statistic	749.1534	Durbin-Watson stat	1.665259
Prob(F-statistic)	0.000000		

Estimated Result:

$$\log(Cost) - \log(P_k) = \underset{(0.701727)}{-3.764649} + \underset{(0.061860)}{0.1525478} \log(Y)$$

$$+ \underset{(0.005364)}{0.050514}(\log(Y))^2 + \underset{(0.161072)}{0.480586}(\log(P_l) - \log(P_k))$$

$$+ \underset{(0.077765)}{0.445248}\left(\log\left(P_f\right) - \log(P_k)\right) + u$$

$$R^2 = 0.955366$$

Econometric Analysis 3

Direct F−Test for $H_0: \beta_3 = 0 \ vs \ H_1: \beta_3 \neq 0$ (Model C)

$$\log(Cost) - \log(P_K) = \beta_1 + \beta_2 \log(Y) + \beta_3(\log(P_L) - \log(P_K)) + \beta_3(\log(P_F) - \log(P_K)) + u$$

Wald Test:
Equation: EQ2_MODEL3

Test Statistic	Value	df	Probability
t-statistic	9.417823	140	0.0000
F-statistic	88.69539	(1, 140)	0.0000
Chi-square	88.69539	1	0.0000

Null Hypothesis: C(3)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	0.050514	0.005364

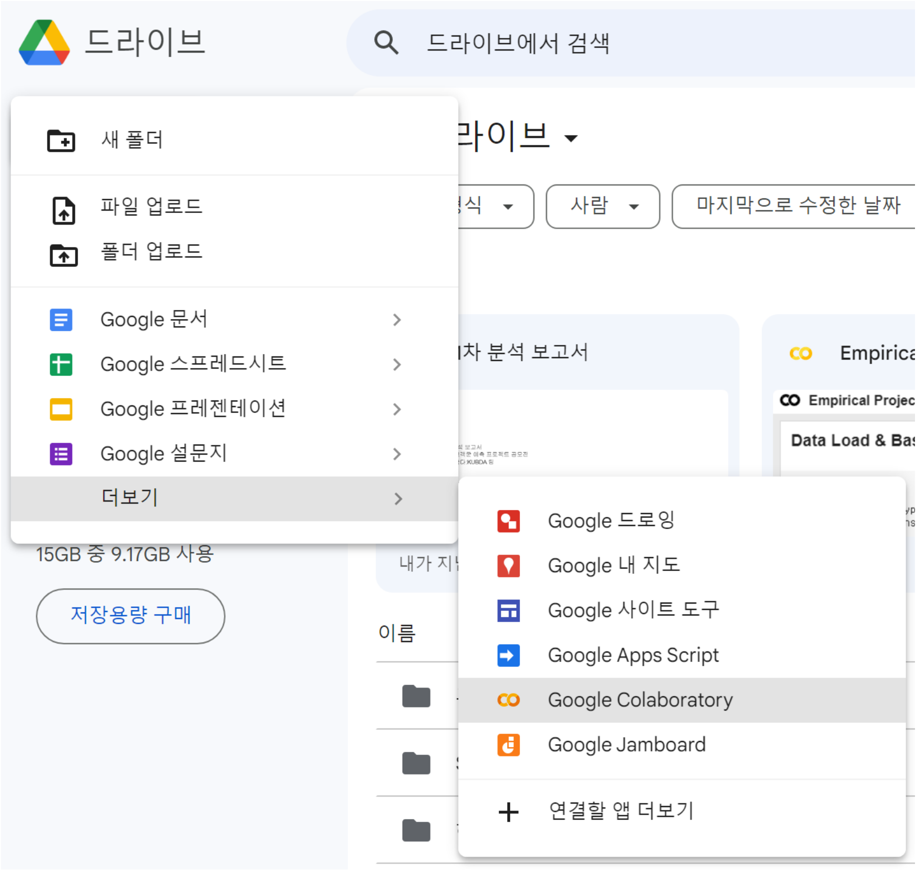
Restrictions are linear in coefficients.

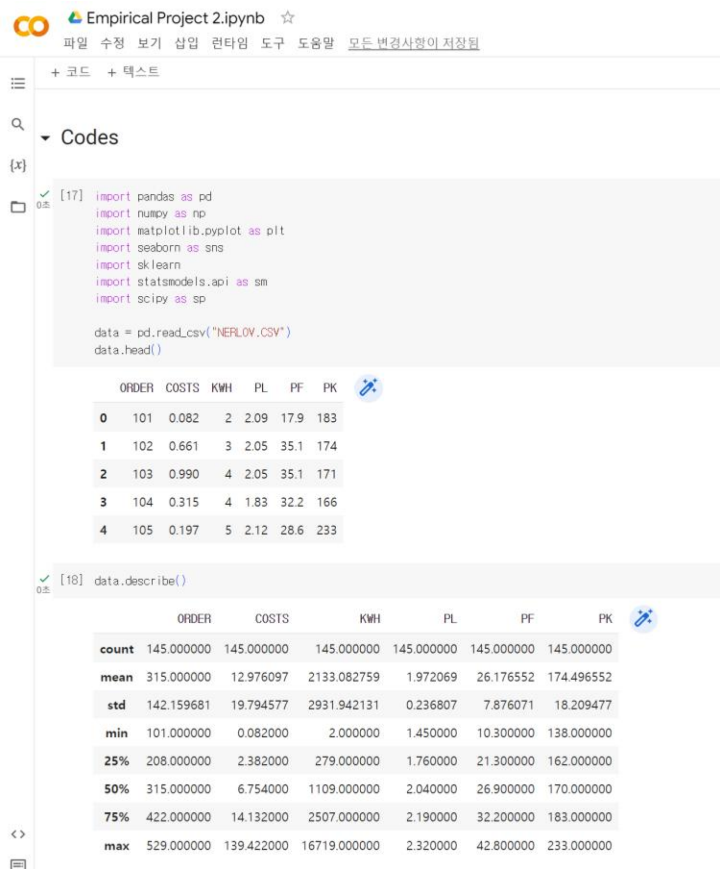
Result Interpretation:

- The F−statistic of this test is 88.69539 and the p−value is 0.0000⋯ with following $F_{(1,140)}$.
- So, the null hypothesis is rejected.
- We rule out the possibility that the restriction $\beta_3 = 0$ of null hypothesis H_0 is true.
- Then can estimate that there may exist a nonlinear relationship between the log(cost) and the log(output), which is convex because of (+) coef. of the 2nd term.

Conducting Econometric Analysis 1 in Python

Introduction: Using Google Colab





Conducting Econometric Analysis 1 in Python

Estimating Baseline Model

Dependent Variable: LOG(COSTS)
Method: Least Squares
Date: 05/17/23 Time: 17:00
Sample: 1 145
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.526503	1.774367	-1.987471	0.0488
LOG(KWH)	0.720394	0.017466	41.24448	0.0000
LOG(PL)	0.436341	0.291048	1.499209	0.1361
LOG(PF)	0.426517	0.100369	4.249483	0.0000
LOG(PK)	-0.219888	0.339429	-0.647819	0.5182

R-squared	0.925955	Mean dependent var	1.724663
Adjusted R-squared	0.923840	S.D. dependent var	1.421723
S.E. of regression	0.392356	Akaike info criterion	1.000578
Sum squared resid	21.55201	Schwarz criterion	1.103224
Log likelihood	-67.54189	Hannan-Quinn criter.	1.042286
F-statistic	437.6863	Durbin-Watson stat	1.013062
Prob(F-statistic)	0.000000		

0초 [19]

```
Baseline_data = np.log(data[['COSTS','KWH','PL','PF','PK']])
Baseline_data.columns = ['Log(Cost)','Log(Y(kWh))','Log(Pl)','Log(Pf)','Log(Pk)']
Baseline_data.head()
```

	Log(Cost)	Log(Y(kWh))	Log(Pl)	Log(Pf)	Log(Pk)
0	-2.501036	0.693147	0.737164	2.884801	5.209486
1	-0.414001	1.098612	0.717840	3.558201	5.159055
2	-0.010050	1.386294	0.717840	3.558201	5.141664
3	-1.155183	1.386294	0.604316	3.471966	5.111988
4	-1.624552	1.609438	0.751416	3.353407	5.451038

0초 [20]

```
Baseline = sm.OLS(Baseline_data['Log(Cost)'], sm.add_constant(Baseline_data[['Log(Y(kWh))','Log(Pl)','Log(Pf)','Log(Pk)']])).fit()
print(Baseline.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              Log(Cost)   R-squared:                       0.926
Model:                            OLS   Adj. R-squared:                  0.924
Method:                 Least Squares   F-statistic:                     437.7
Date:                Wed, 24 May 2023   Prob (F-statistic):           4.82e-78
Time:                        03:00:51   Log-Likelihood:                -67.542
No. Observations:                 145   AIC:                             145.1
Df Residuals:                     140   BIC:                             160.0
Df Model:                           4
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const          -3.5265      1.774     -1.987      0.049      -7.035      -0.018
Log(Y(kWh))     0.7204      0.017     41.244      0.000       0.686       0.755
Log(Pl)         0.4363      0.291      1.499      0.136      -0.139       1.012
Log(Pf)         0.4265      0.100      4.249      0.000       0.228       0.625
Log(Pk)        -0.2199      0.339     -0.648      0.518      -0.891       0.451
==============================================================================
Omnibus:                       51.403   Durbin-Watson:                   1.013
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              175.700
Skew:                           1.303   Prob(JB):                     7.03e-39
Kurtosis:                       7.721   Cond. No.                         506.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Conducting Econometric Analysis 1 in Python

Estimating Model A

Dependent Variable: LOG(COSTS)-LOG(PK)
Method: Least Squares
Date: 05/17/23 Time: 11:52
Sample: 1 145
Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.690789	0.884871	-5.301098	0.0000
LOG(KWH)	0.720688	0.017436	41.33398	0.0000
LOG(PL)-LOG(PK)	0.592910	0.204572	2.898291	0.0044
LOG(PF)-LOG(PK)	0.414471	0.098951	4.188644	0.0000

R-squared	0.927088	Mean dependent var	-3.432113
Adjusted R-squared	0.925537	S.D. dependent var	1.435663
S.E. of regression	0.391762	Akaike info criterion	0.990874
Sum squared resid	21.64032	Schwarz criterion	1.072991
Log likelihood	-67.83836	Hannan-Quinn criter.	1.024241
F-statistic	597.6166	Durbin-Watson stat	1.015369
Prob(F-statistic)	0.000000		

[21]
```
A_data = Baseline_data[['Log(Cost)']].copy()
A_data['Log(Cost)-Log(Pk)'] = A_data['Log(Cost)'] - Baseline_data['Log(Pk)']
A_data['Log(Y(kWh))'] = Baseline_data['Log(Y(kWh))']
A_data['Log(Pl)-Log(Pk)'] = Baseline_data['Log(Pl)'] - Baseline_data['Log(Pk)']
A_data['Log(Pf)-Log(Pk)'] = Baseline_data['Log(Pf)'] - Baseline_data['Log(Pk)']

A_data = A_data.iloc[0:,1:]
A_data.head()
```

	Log(Cost)-Log(Pk)	Log(Y(kWh))	Log(Pl)-Log(Pk)	Log(Pf)-Log(Pk)
0	-7.710522	0.693147	-4.472322	-2.324685
1	-5.573057	1.098612	-4.441216	-1.600854
2	-5.151714	1.386294	-4.423824	-1.583462
3	-6.267170	1.386294	-4.507672	-1.640021
4	-7.075590	1.609438	-4.699622	-2.097632

[22]
```
Model_A = sm.OLS(A_data['Log(Cost)-Log(Pk)'], sm.add_constant(A_data[['Log(Y(kWh))','Log(Pl)-Log(Pk)','Log(Pf)-Log(Pk)']])).fit()
print(Model_A.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Log(Cost)-Log(Pk)   R-squared:                       0.927
Model:                            OLS   Adj. R-squared:                  0.926
Method:                 Least Squares   F-statistic:                     597.6
Date:                Wed, 24 May 2023   Prob (F-statistic):           6.16e-80
Time:                        03:00:51   Log-Likelihood:                -67.838
No. Observations:                 145   AIC:                             143.7
Df Residuals:                     141   BIC:                             155.6
Df Model:                           3
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              -4.6908      0.885     -5.301      0.000      -6.440      -2.941
Log(Y(kWh))         0.7207      0.017     41.334      0.000       0.686       0.755
Log(Pl)-Log(Pk)     0.5929      0.205      2.898      0.004       0.188       0.997
Log(Pf)-Log(Pk)     0.4145      0.099      4.189      0.000       0.219       0.610
==============================================================================
Omnibus:                       49.416   Durbin-Watson:                   1.015
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              166.633
Skew:                           1.252   Prob(JB):                     6.55e-37
Kurtosis:                       7.617   Cond. No.                         234.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Conducting Econometric Analysis 1 in Python

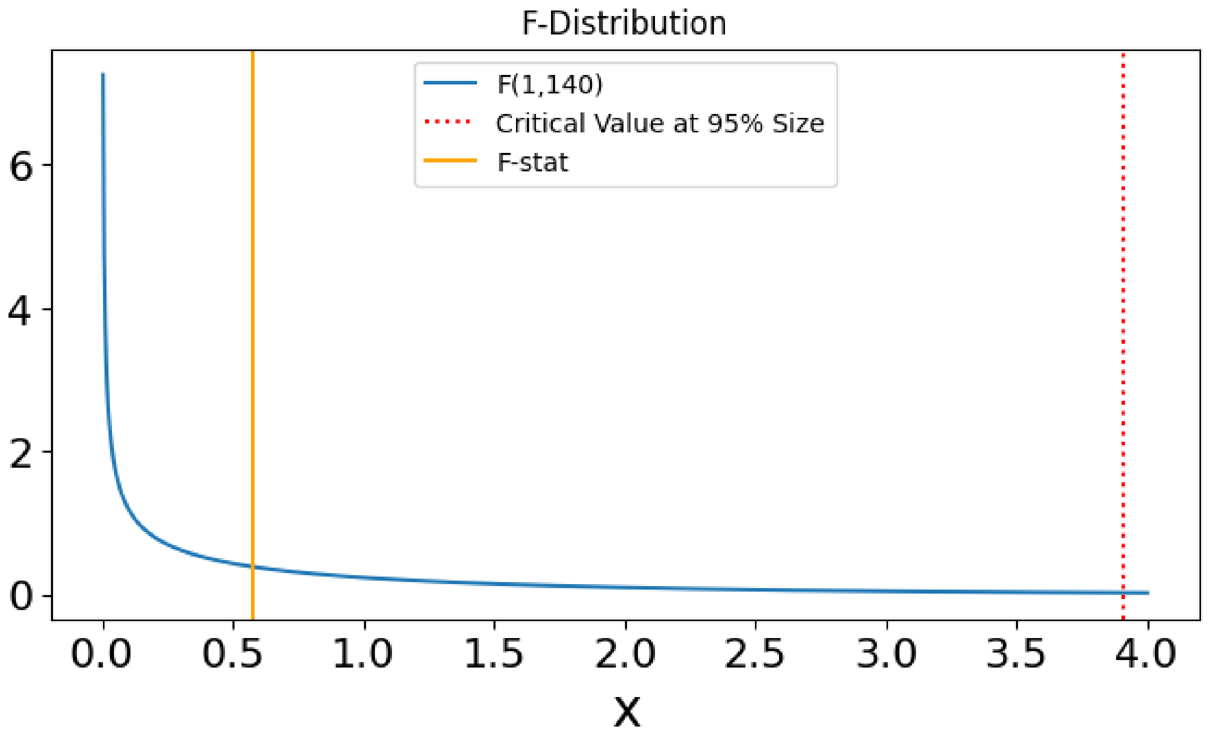
Indirect F–test

```
[25] RSSu = Baseline.ssr
     RSSr = Model_A.ssr
     q = 1
     n = len(Baseline_data)
     r = len(Baseline.params)
     F_stat = ((RSSr-RSSu)/(q))/(RSSu/(n-r))
     Pval = 1-sp.stats.f(q,n-r).cdf(F_stat)

     print(f"RSSu: {RSSu}\nRSSr: {RSSr}]\nq: {q}\nn: {n}\nr: {r}")
     print(f"\nTest Result: \nF-stat: {F_stat}\nP-value:{Pval}")
```

```
RSSu: 21.552008164185782
RSSr: 21.640319116721381]
q: 1
n: 145
r: 5

Test Result:
F-stat: 0.5736603875052707
P-value:0.45008093212710076
```



Conducting Econometric Analysis 1 in Python

Comparison with Eviews and Python

Eviews	VS	Python
Function as a **single software**	VS	Have to Use **Add-ons (Libraries)**
Strong at **Statistical** Analysis	VS	More **Flexible** Projects
Proper to **Econometrics**	VS	Proper to **Engineering Projects**

Conducting Econometric Analysis 1 in Python

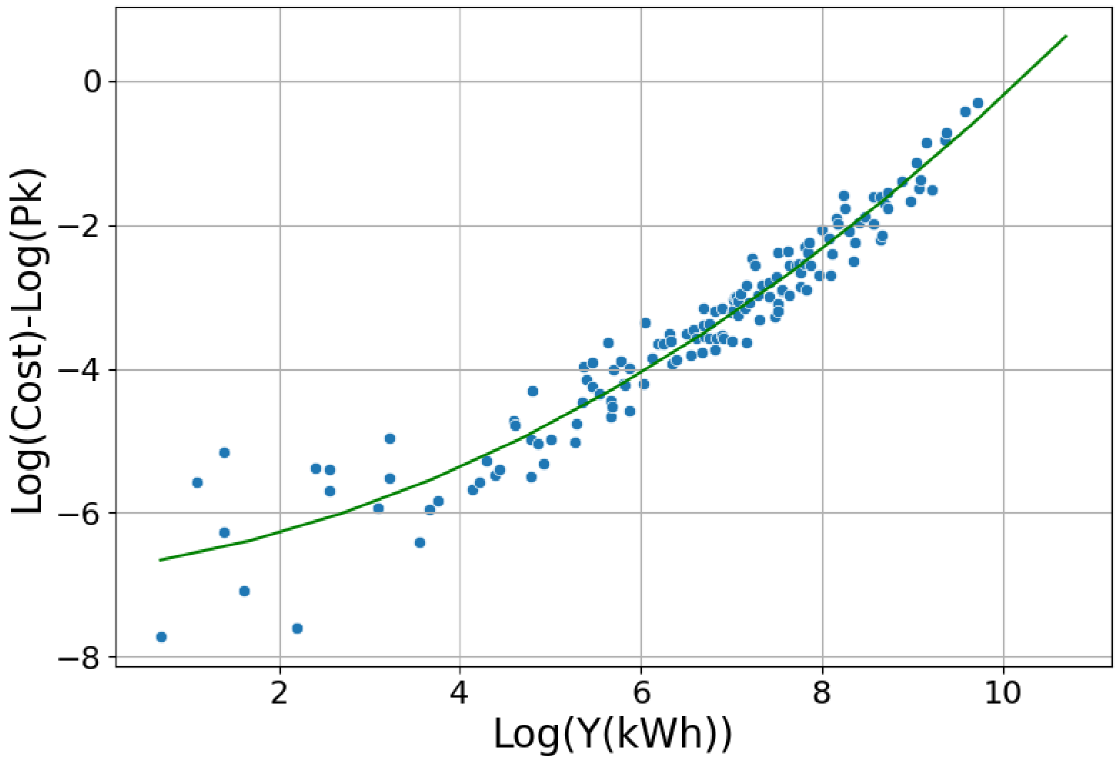
Comparison with Eviews and Python in a Regression Project

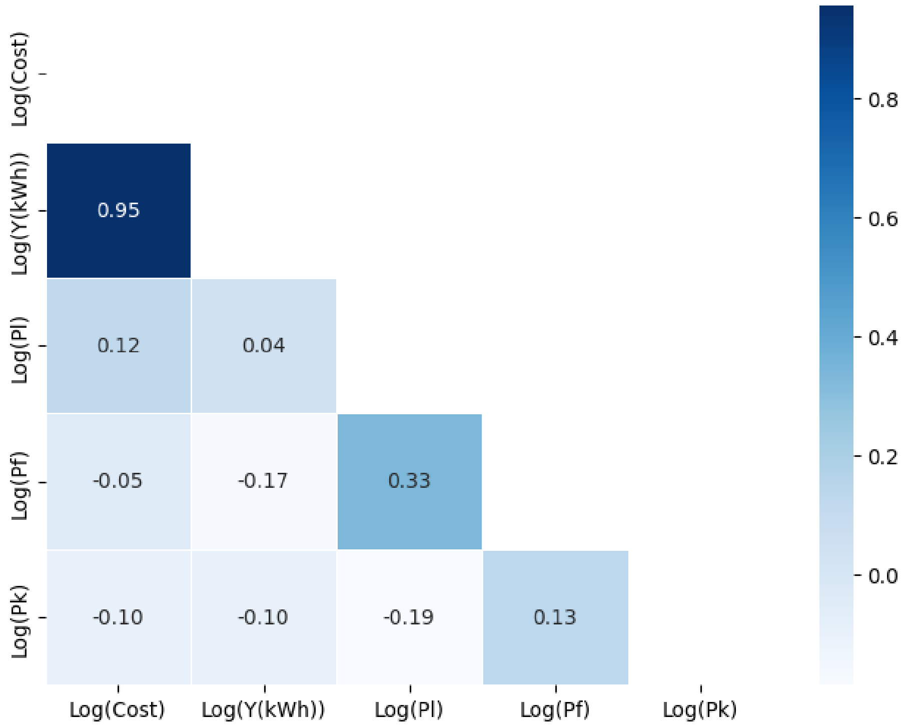
Eviews		Python
Function as a **single software**	VS	Have to Use **Add-ons (Libraries)**
Strong at **Statistical** Analysis	VS	More **Flexible** Projects
Proper to **Econometrics**	VS	Proper to **Engineering Projects**

What is the Strength of Python in a Regression Project?

1. Customized Visualization:

Scatterplot of Model C & Correlation Matrix for Baseline





What is the Strength of Python in a Regression Project?

2. Service Prototyping: Cost Predictor with Model C

https://colab.research.google.com/drive/1MBbHMZe6ffvIEkyqcQoUdbATAXRjm_9K?usp=sharing

What is the Strength of Python in a Regression Project?

2. Service Prototyping:

Cost Predictor with Model C

y: 12502

pl: 1.6

pf: 21.5

pk: 189.1

클리어 | 제출하기

output 0

Predicted Cost = 19.96129

output 1

Predicted Log(Cost)-Log(Pk) = -2.24848

output 2

95% Confidence Interval = [-2.98588, -1.51108]

output 3

```
                 OLS Regression Results
==============================================================================
Dep. Variable:    Log(Cost)-Log(Pk)  R-squared:                0.955
Model:                   OLS  Adj. R-squared:           0.954
Method:           Least Squares  F-statistic:             749.2
Date:          Wed, 24 May 2023  Prob (F-statistic):     2.03e-93
Time:                 02:37:33  Log-Likelihood:         -32.259
No. Observations:          145  AIC:                    74.52
Df Residuals:              140  BIC:                    89.40
Df Model:                    4
Covariance Type:     nonrobust
==============================================================================
====
              coef   std err      t    P>|t|   [0.025    0.975]
------------------------------------------------------------------------------
const         -3.7646   0.702   -5.365   0.000   -5.152   -2.377
Log(Y(kWh))    0.1525   0.062    2.466   0.015    0.030    0.275
Log(Y(kWh))^2  0.0505   0.005    9.418   0.000    0.040    0.061
Log(Pl)-Log(Pk)  0.4806   0.161    2.984   0.003    0.162    0.799
Log(Pf)-Log(Pk)  0.4452   0.078    5.726   0.000    0.292    0.599
==============================================================================
Omnibus:              21.061  Durbin-Watson:            1.665
```

What is the Strength of Python in a Regression Project?

3. Expansion to Machine Learning: Decision Tree Example