統計學一下 期末報告

研究主題:
翡翠水庫未來進水量預測分析

第三組

資管二 B09705010 吳和謙
資管二 B09705014 吳欣紘
資管二 B09705025 徐懷山
資管二 B09705026 詹景棠
資管二 B09705030 柯師為
資管二 B09705031 李奕杰
資管二 B09705039 劉惟恩
資管二 B09705044 王裕勳

研究動機與目的

極端氣候 → 供水不穩

以水庫逕流量之歷史資料，預估未來進流量

保障民生用水

促進糧食生產

穩定工業用水

研究方法

日進流量 v.s. 月進流量

Part 1:
日進流量 → 時間序列分析 → 預測 **111/1/1** 進流量

Part 2:
日進流量 → 季節分析（一天為一區間）
→ 預測未來一年每一天的進流量

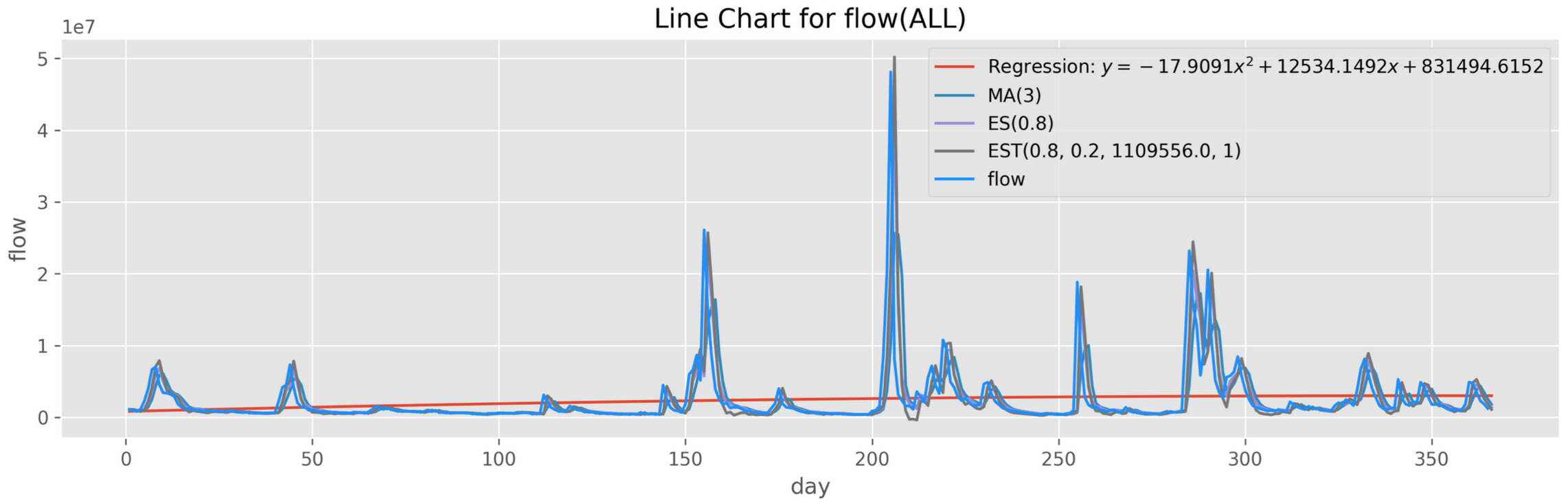
Part 3:
月進流量 → 季節分析 （一年為一區間）
→ 預測未來一年（**111**年）每個月平均進流量

Part 1

110年時間序列分析

MA(3)　　ES(0.8)　　EST(0.8,0.2,1)　　Regression

折線圖


Line Chart for flow(ALL)
1e7
flow
day
Regression: $y = -17.9091x^2 + 12534.1492x + 831494.6152$
MA(3)
ES(0.8)
EST(0.8, 0.2, 1109556.0, 1)
flow
0
1
2
3
4
5
0
50
100
150
200
250
300
350

MA(3)

MAD	1.456729e+06
MSE	1.499028e+13
RMSE	3.871728e+06
MAPE	44.577%

ES(0.8)

MAD	1.210197e+06
MSE	1.237267e+13
RMSE	3.517480e+06
MAPE	31.695%

EST(0.8,0.2,1)	MAD	1.336593e+06
	MSE	1.461905e+13
	RMSE	3.823487e+06
	MAPE	36.426%

Regression

MAD	2.06687e+06
MSE	1.610129e+13
RMSE	4.012641e+06
MAPE	156.321%

誤差比較

(MAPE)

由於進流量的數量級太大，
我們選用以百分比做基準的
MAPE來評估四種方法的誤
差大小

MA **44.577%**

ES **31.695%** 

EST **36.426%**

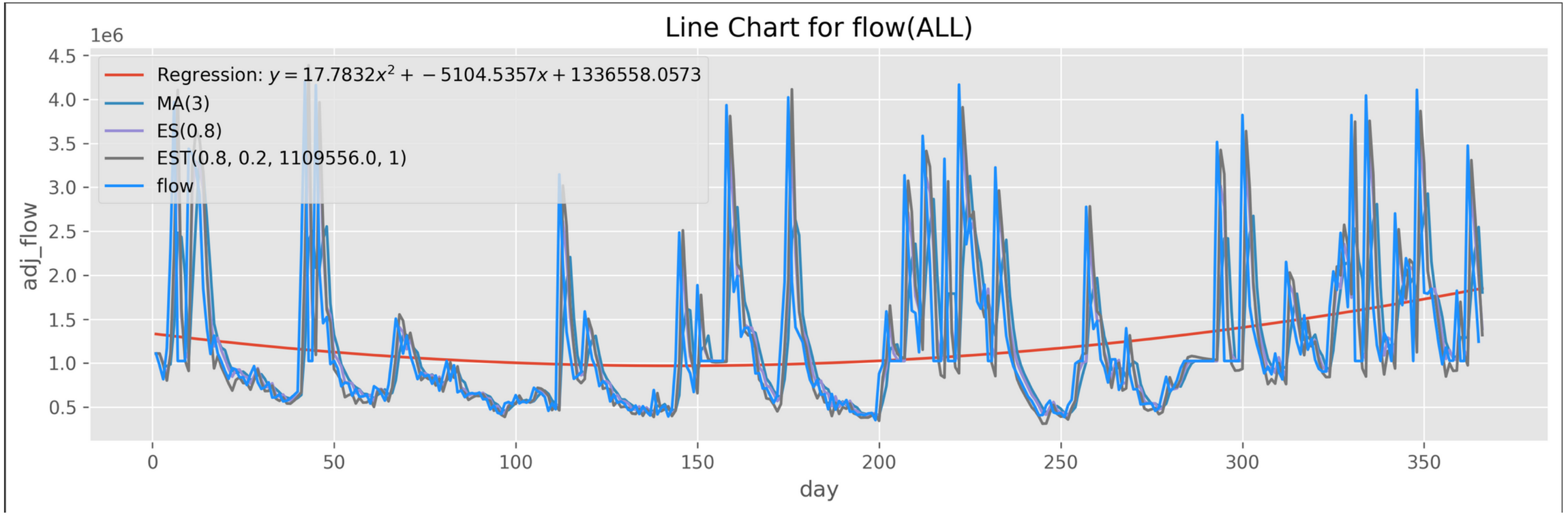
Regression **156.321%**

由於誤差數值太大，我們推測可能是進水量部分極端值的影響，因此我們將極端值剔除後，重新建立一次預測模型。

110年時間序列分析（調整）

MA(3)　　ES(0.8)　　EST(0.8,0.2,1)　　Regression

折線圖


Line Chart for flow(ALL)
1e6
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
adj_flow
Regression: $y = 17.7832x^2 + -5104.5357x + 1336558.0573$
MA(3)
ES(0.8)
EST(0.8, 0.2, 1109556.0, 1)
flow
0
50
100
150
200
250
300
350
day

MA(3)

MAD	4.454673e+05
MSE	5.728891e+11
RMSE	7.568944e+05
MAPE	30.672%

ES(0.8)

MAD	3.943312e+05
MSE	5.503311e+11
RMSE	7.418430e+05
MAPE	25.809%

EST(0.8,0.2,1)

MAD	4.378451e+05
MSE	6.506094e+11
RMSE	8.066036e+05
MAPE	28.445%

Regression

MAD	5.41979e+05
MSE	6.018484e+11
RMSE	7.757889e+05
MAPE	52.028%

誤差比較

(MAPE)

由於進流量的數量級太大，我們選用以百分比做基準的**MAPE**來評估四種方法的誤差大小

MA **30.672%**

ES **25.809%** 

EST **28.445%**

Regression **52.028%**

誤差比較 (調整前後)

	Before		After
MA	44.577%	MA	30.672%
ES	31.695%	ES	25.809%
EST	36.426%	EST	28.445%
Regression	156.321%	Regression	52.028%

預測值對比實際值

根據我們的資料，**111/1/1**日的進水量為**1020824**立方公尺，接下來我們要用模型生成的預測值對比實際的數值。

預測準確度 (Top3)

1st　EST　1110900立方公尺

2nd EST(adj) 1319431立方公尺

3rd　ES(adj) 1390645立方公尺

小結

原本調整前的模型的MAPE大概落在30%-40%。而在調整過後下降到20%-30%，而在與真實數值的比對中兩者都有不錯的準確度，因此我們認為時間序列的模型具有一定程度的準確性。

Part 2

Centered Moving Average (Day)

Residual Analysis

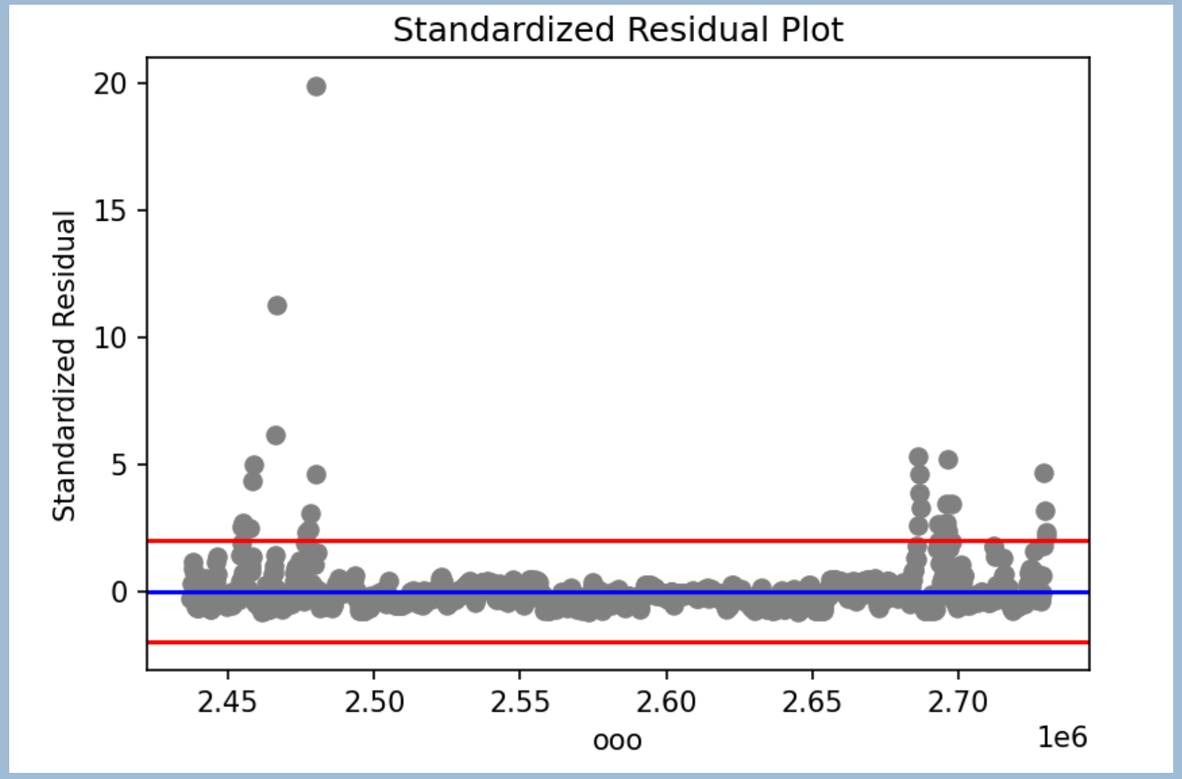
Shapiro Test

Statistics=0.454, p=0.000

```
runs =  159
n1 =  549
n2 =  548
runs_exp =  549.4995442114858
stan_dev =  16.5529246995655
z =  -23.59096965032023
pval_z =  4.771199460866353e-123
p_value for Z-statistic=  4.771199460866353e-123
```

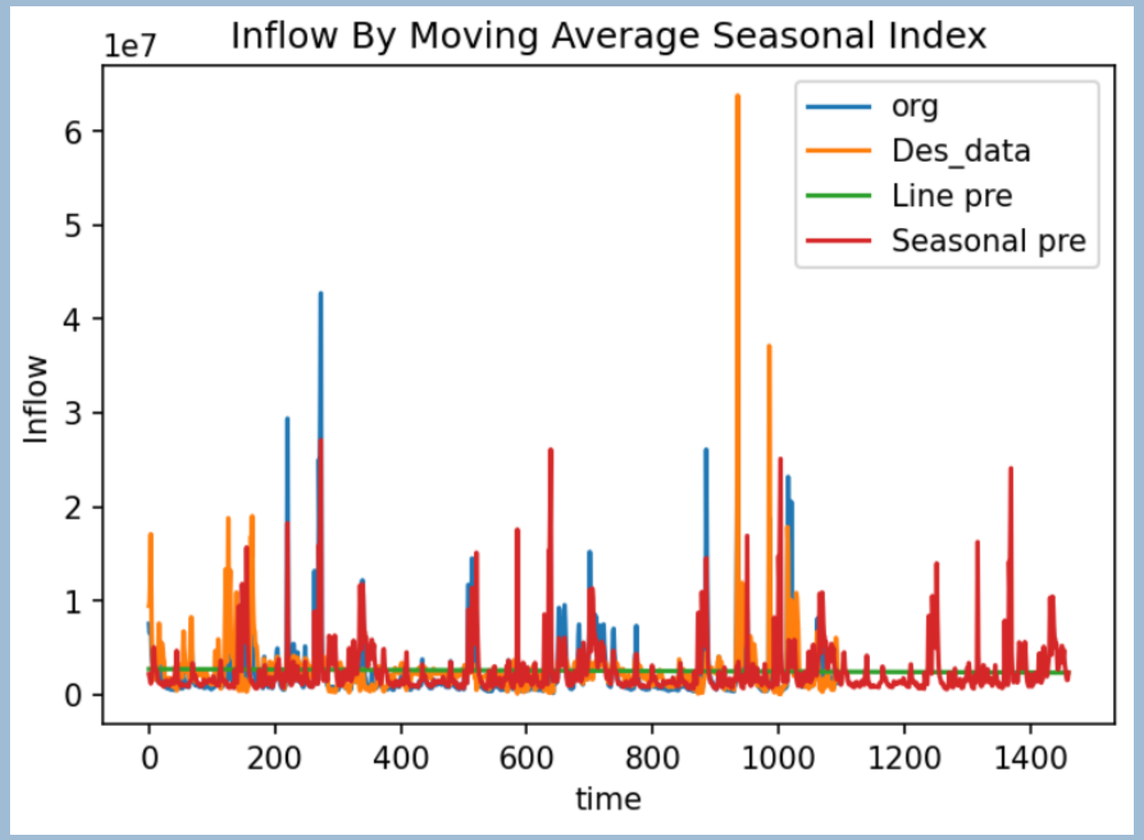
```
x_square_sum =  1096.6565029778735
size =  1096
x_d =  [0. 0. 0. ... 0. 0. 0.]
x_d =  [ 0.          0.12971404  0.87474338 ... -0.59582245 -0.50485788
 -0.09409877]
d =  0.8881689541207388
0.8881689541207388
```

For n = 1100, k = 1, dL = 1.899, dU = 1.903.



未來一年每天的進水量預測

	time	org	Des_data	Line pre	Seasonal pre
0	0.0	7543380.0	9.448437e+06	2.730245e+06	2.179755e+06
1	1.0	6572620.0	9.848032e+06	2.729978e+06	1.821999e+06
2	2.0	7019088.0	1.254411e+07	2.729710e+06	1.527416e+06
3	3.0	7374848.0	1.709533e+07	2.729442e+06	1.177469e+06
4	4.0	4338772.0	8.335019e+06	2.729174e+06	1.420664e+06
...	...	...	...	...	...
1456	1456.0	NaN	NaN	2.340369e+06	2.444691e+06
1457	1457.0	NaN	NaN	2.340102e+06	1.828982e+06
1458	1458.0	NaN	NaN	2.339834e+06	1.550185e+06
1459	1459.0	NaN	NaN	2.339566e+06	1.586998e+06
1460	1460.0	NaN	NaN	2.339298e+06	2.350376e+06



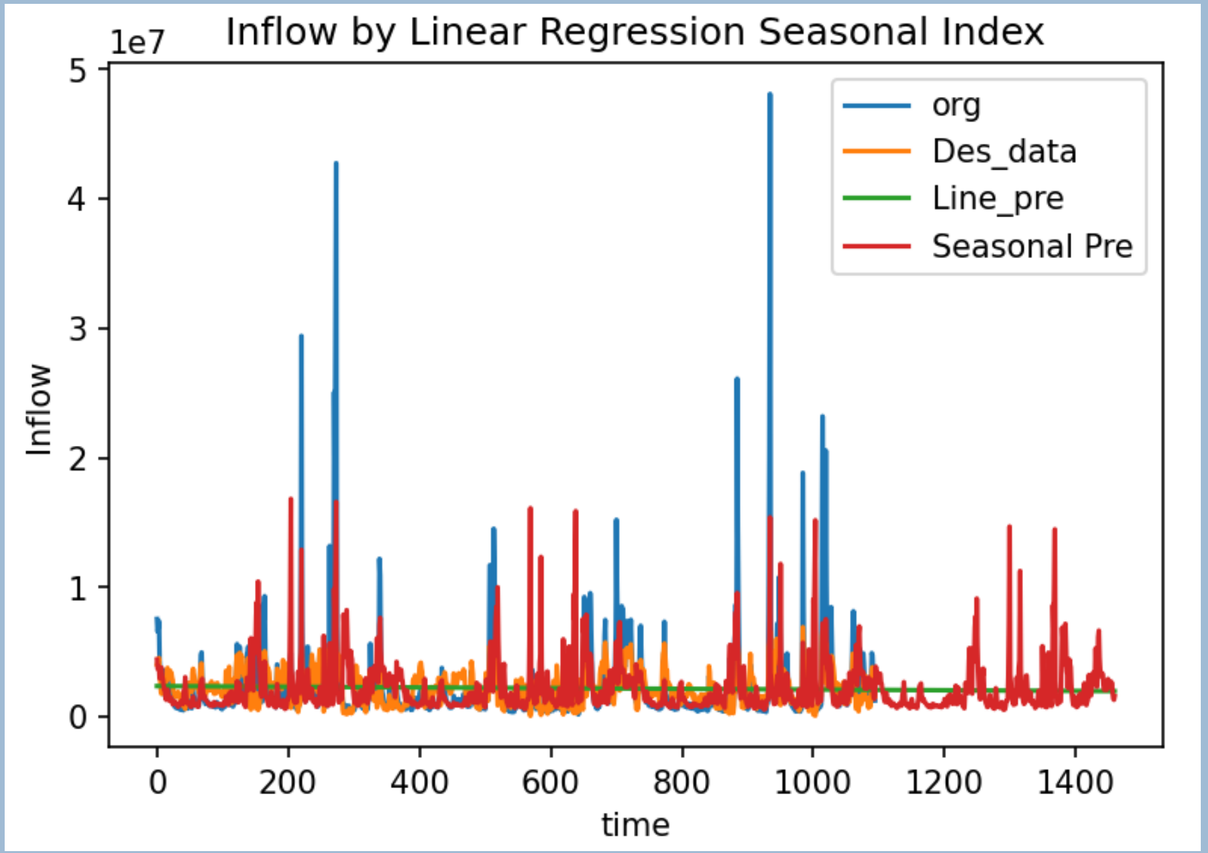
CMA(365)

MAD	1522689.8999
MSE	1.064358e+13
RMSE	3262450.6540
MAPE	114.3271 %

Smoothing by Linear Regression Model (Day)

未來一年每天的進水量預測

	time	org	Des_data	Line_pre	Seasonal Pre
0	0.0	7543380.0	4.030803e+06	2.359633e+06	4.415895e+06
1	1.0	6572620.0	4.185165e+06	2.359366e+06	3.705282e+06
2	2.0	7019088.0	4.455707e+06	2.359099e+06	3.716296e+06
3	3.0	7374848.0	4.534867e+06	2.358833e+06	3.836062e+06
4	4.0	4338772.0	3.465800e+06	2.358566e+06	2.952646e+06
...	...	...	...	...	...
1455	1455.0	NaN	NaN	1.971609e+06	2.438915e+06
1456	1456.0	NaN	NaN	1.971343e+06	1.810687e+06
1457	1457.0	NaN	NaN	1.971076e+06	1.384601e+06
1458	1458.0	NaN	NaN	1.970809e+06	1.300637e+06
1459	1459.0	NaN	NaN	1.970543e+06	1.575111e+06



Residual Analysis

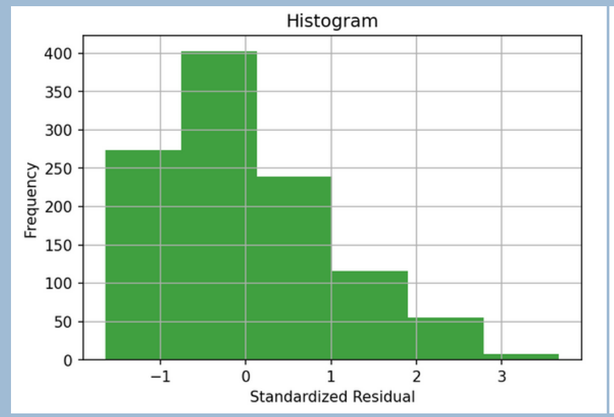
```
Shapiro Test
Statistics=0.949, p=0.000
```

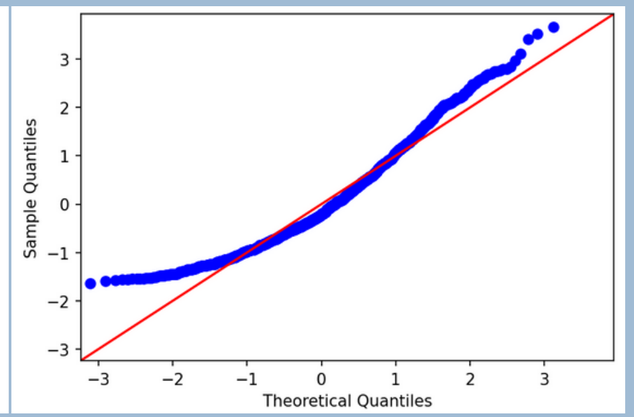
```
Chi-squared test: statistics = 56.1178, p-value = 0.0000
Critical value = 7.8147 (defree of freedom = 3)
3.964652969087883e-12
```

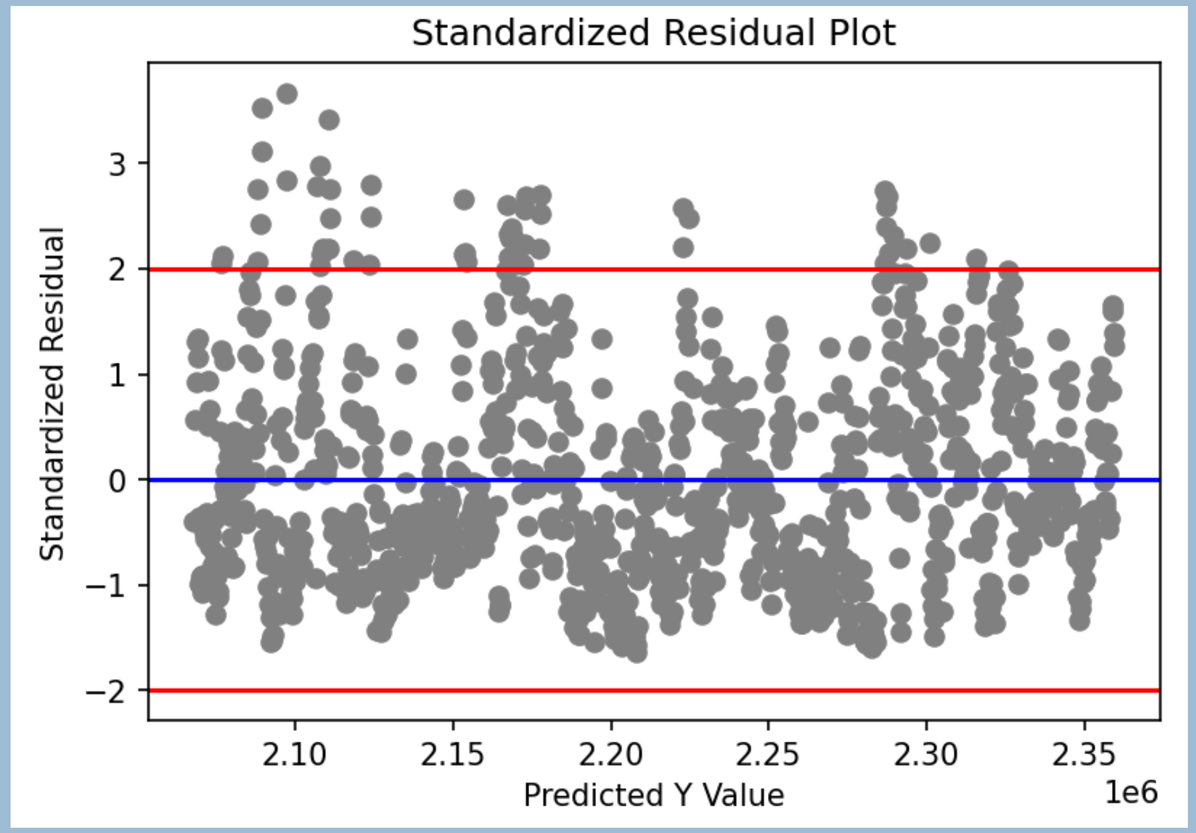
```
runs =  145
n1 =  549
n2 =  547
runs_exp =  548.9981751824818
stan_dev =  16.545330062438l8
z =  -24.417655837501446
pval_z =  1.1106226534461274e-131
p_value for Z-statistic=  1.1106226534461274e-131
```

```
size =  1095
x_d =  [0. 0. 0. ... 0. 0. 0.]
x_d =  [ 0.          0.11793104  0.20654279 ... -0.37565253 -0.36089632
 -0.95944569]
d =  0.3111379338674694
0.3111379338674694
```

For n = 1100, k = 1, dL = 1.899, dU = 1.903.







SI_LR

MAD	1384335.28017
MSE	7892040429938.033
RMSE	2809277.563705
MAPE	95.075719 %

Regression Model by Indicator Variables (Day)

因為要生成 365 個 dummy variables，模型的結果很容易出現偏差，因此我們沒有進行此方法。

Comparing

Comparing

	CMA(365)	SI_LR
MAD	1522689.8999	1384335.28017
MSE	1.064358e+13	7.892040e+12
MAPE	114.3271 %	95.075719 %

預測準確度

111/1/1日的進水量為1020824立方公尺

1st　CMA(365) 1945450.0055立方公尺

2nd　SI_LR 3869403.9367928立方公尺

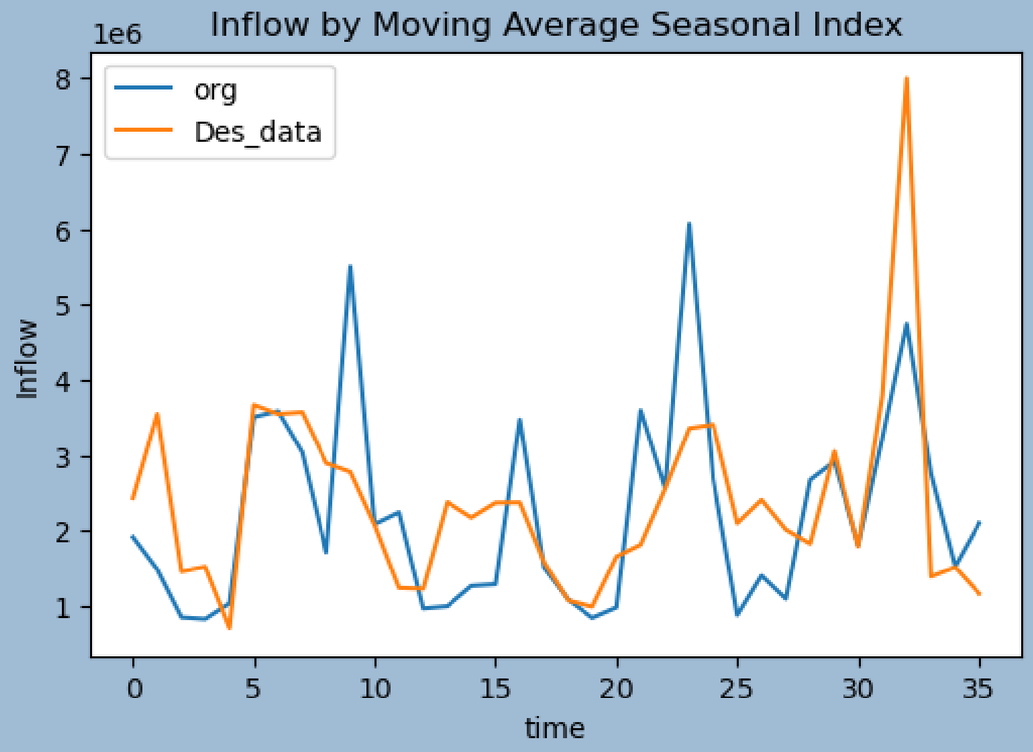
小結

雖然在 error 的比較上 Linear Regression 的模型看起來比較合適，但是其實兩者的殘差分析都沒過，再加上實際預測值的結果皆不符預期，因此 Part 2 的模型可靠性都不高。

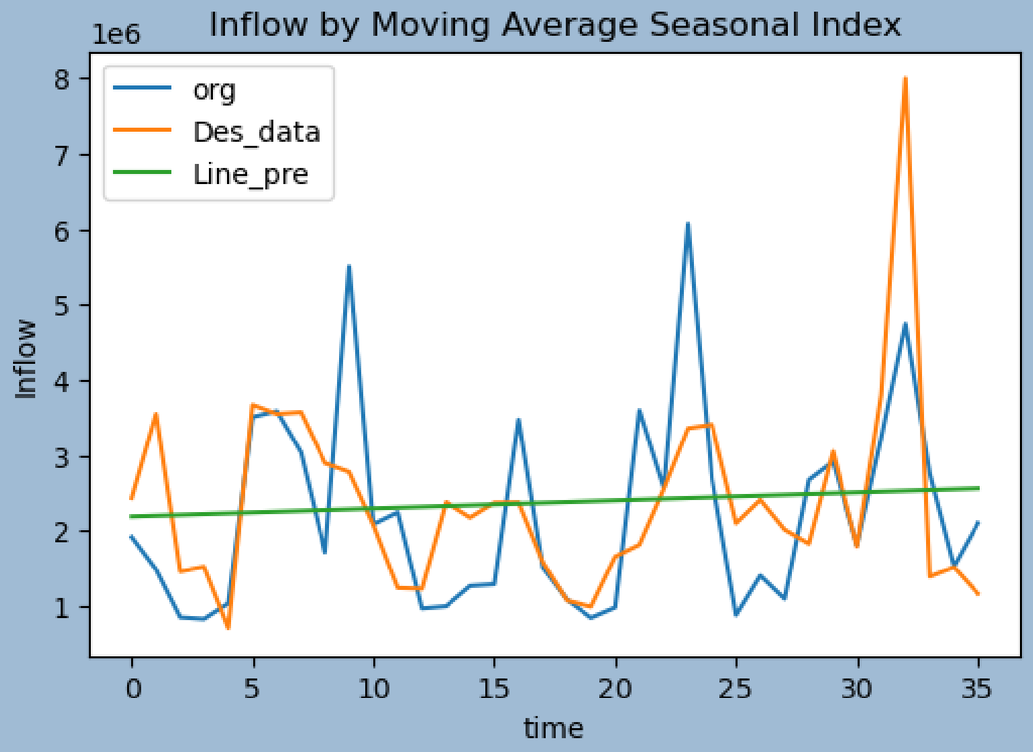
Part 3 (Month Data)

Moving Average

De-Seasonalized Data



New Regression



Regression Results

OLS Regression Results

Dep. Variable:	Des_D	R-squared:	0.007
Model:	OLS	Adj. R-squared:	-0.022
Method:	Least Squares	F-statistic:	0.2551
Date:	Thu, 19 May 2022	Prob (F-statistic):	0.617
Time:	21:34:10	Log-Likelihood:	-557.04
No. Observations:	36	AIC:	1118.
Df Residuals:	34	BIC:	1121.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.189e+06	4.27e+05	5.130	0.000	1.32e+06	3.06e+06
t	1.059e+04	2.1e+04	0.505	0.617	-3.2e+04	5.32e+04

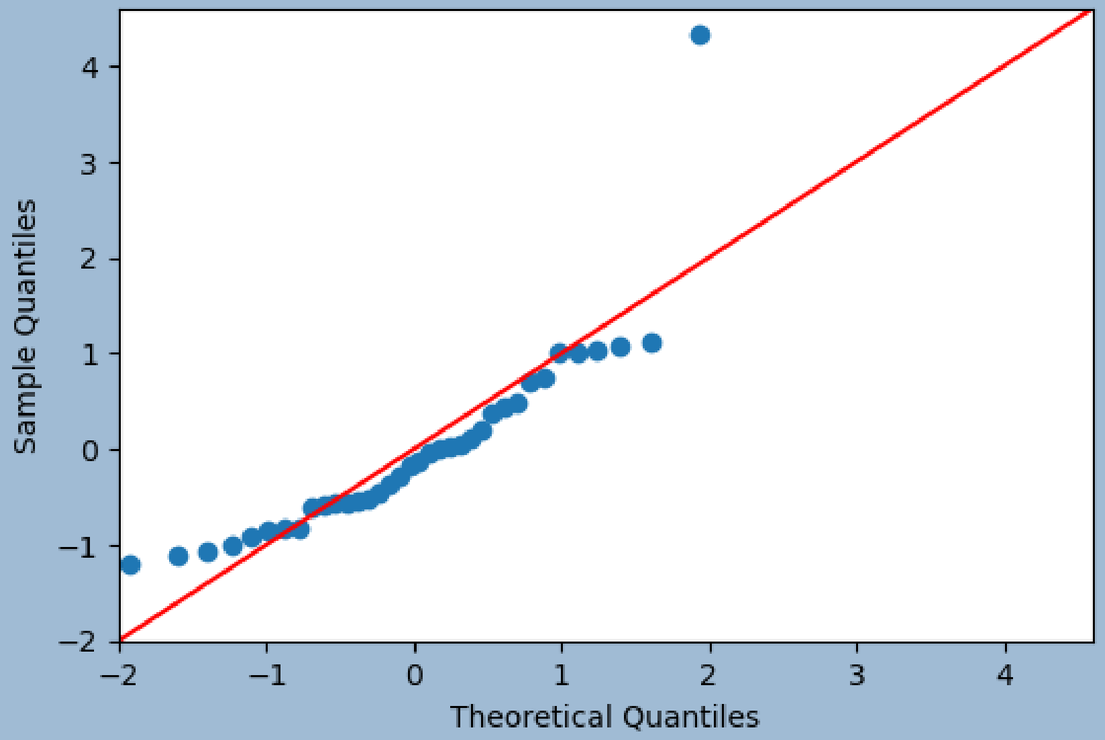
Omnibus:	34.047	Durbin-Watson:	1.568
Prob(Omnibus):	0.000	Jarque-Bera (JB):	100.065
Skew:	2.145	Prob(JB):	1.87e-22
Kurtosis:	9.950	Cond. No.	39.9

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

Average inflow per month = 2.189e+06 + 1.059e+04 t

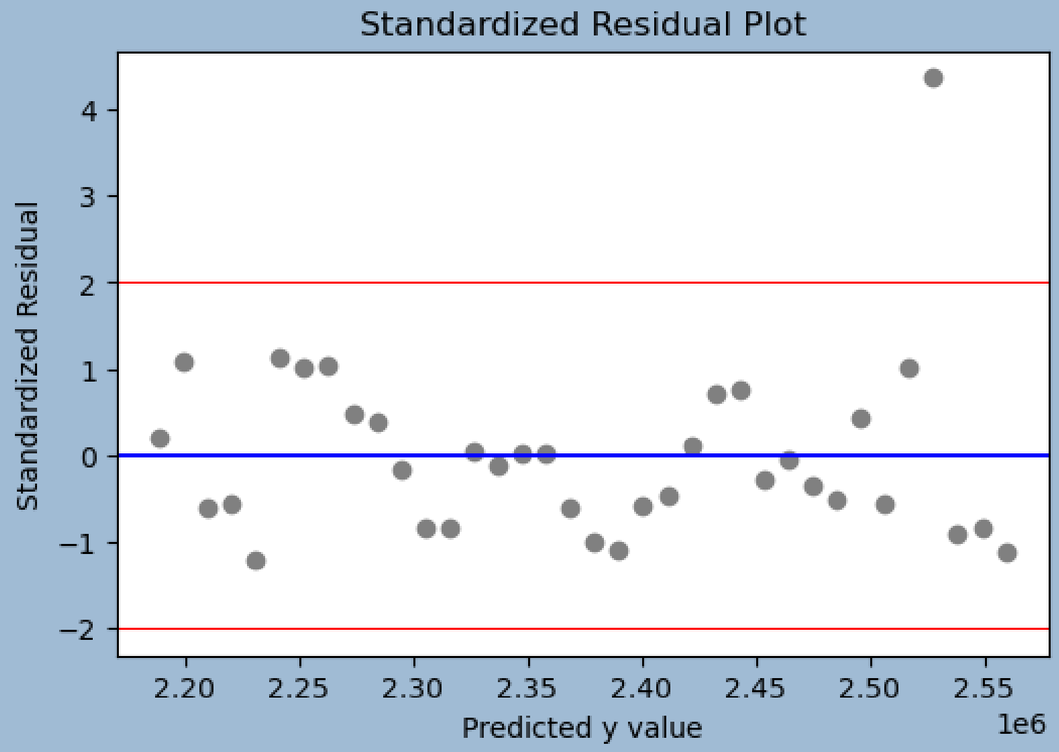
Residual Analysis

Normality Test

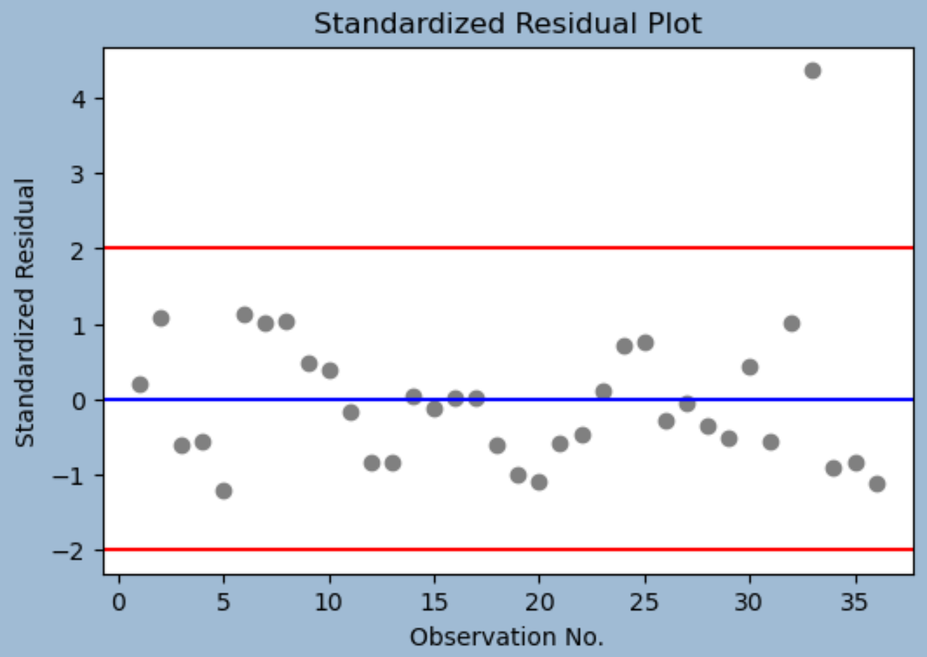



```
Shapiro Test
p_value for Z-statistic=  0.09083702152945593
```

Homoscedasticity and Heteroscedasticity Test



Dependence of the Error Variable Test




```
runs =  14
n1 =  18
n2 =  18
runs_exp =  19.0
stan_dev =  2.9568322818274866
z =  -1.6909988539863081
pval_z =  0.09083702152945593
p_value for Z-statistic=  0.09083702152945593
```

Because n1, n2 < 20, Lr = 12 < R = 14 < Ur = 26, do not reject H_0. There is no evidence to infer that the sample is not random.

Durbin Watson Test

```
The Durbin Watson test
x_square_sum =  36.67482871939605
size =  36
x_d =  [ 0.          0.88771418 -1.68759304  0.04471343 -0.65362713  2.34168332
 -0.10770544  0.00704716 -0.54167356 -0.1013166  -0.55886467 -0.65805598
 -0.01676306  0.88776299 -0.16762035  0.14446894 -0.00568119 -0.62399378
 -0.40138059 -0.07199578  0.50725344  0.11478195  0.56545411  0.61727843
  0.03193954 -1.02818958  0.23348043 -0.31801896 -0.15628584  0.95651488
 -1.00563809  1.58316249  3.34629656 -5.28208926  0.08824531 -0.29887391]
d =  1.5771008430844435
d value = 1.5771008430844435
```

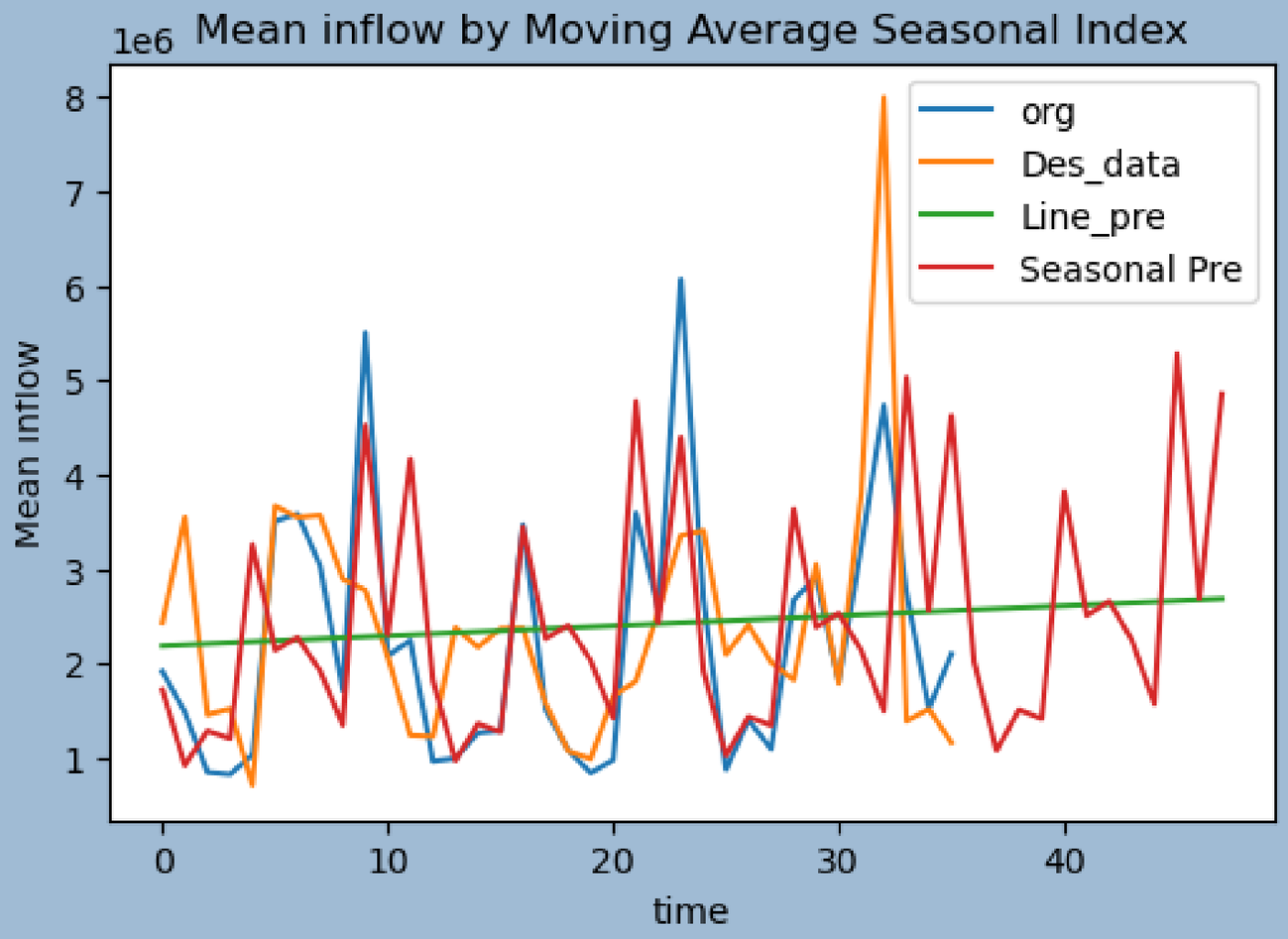
n=36, k=2, α = 0.05

From the Durbin-Watson table we have:

dL = 1.354, dU = 1.587, 4-dU = 2.413, 4-dL = 2.646.

The statistic d is between the dL and dU, so we can not reject the H_0, and conclude that the test is inconclusive.

未來一年每個月的進水量預測

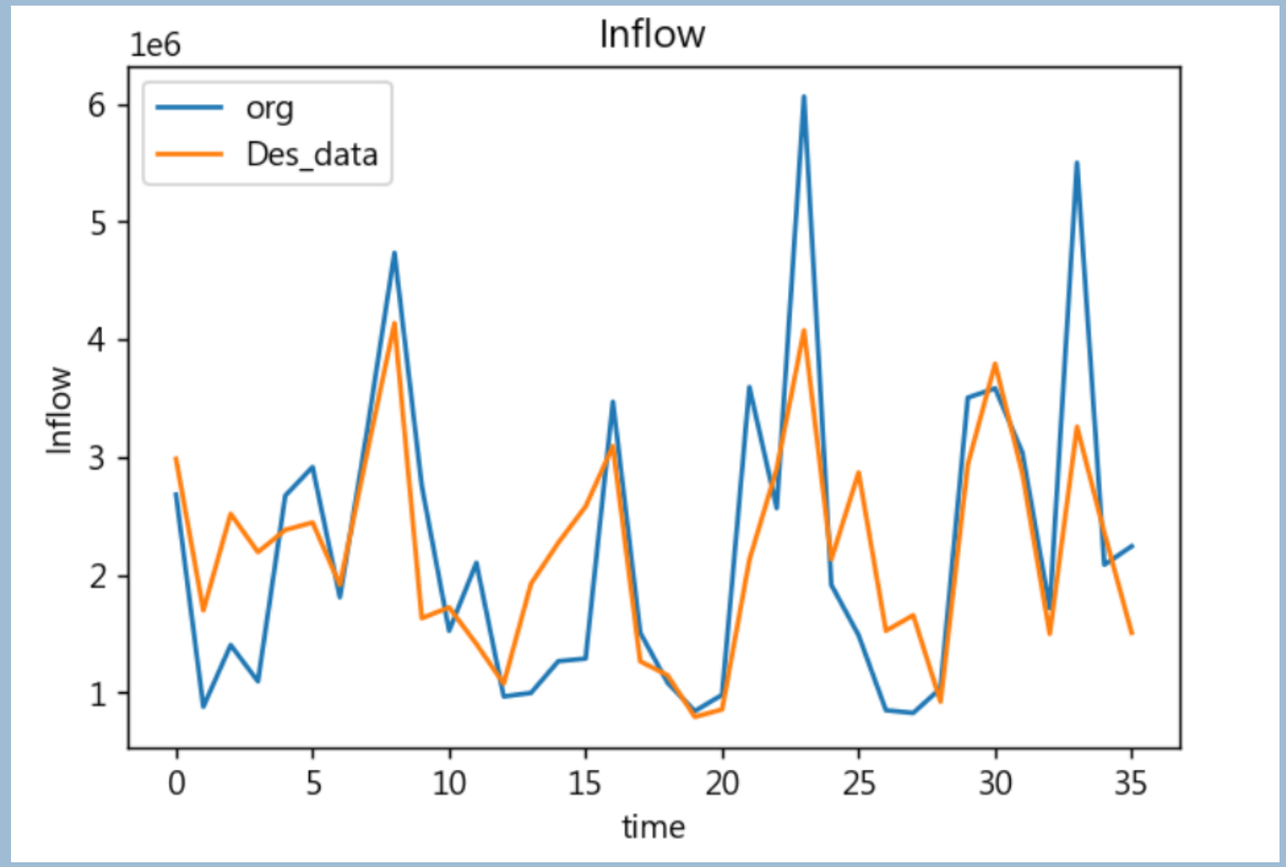


	Year/month	Predict Inflow
36	111/01	2.024442e+06
37	111/02	1.082822e+06
38	111/03	1.511199e+06
39	111/04	1.416061e+06
40	111/05	3.819943e+06
41	111/06	2.507366e+06
42	111/07	2.661643e+06
43	111/08	2.249644e+06
44	111/09	1.573755e+06
45	111/10	5.278628e+06
46	111/11	2.691396e+06
47	111/12	4.853787e+06

Simple Linear Regression

De-Seasonalized Data

	SID	SeaIdx	orig	Des_D
0	1.0	0.898547	2683037.0	2.985972e+06
1	2.0	0.518363	881138.0	1.699847e+06
2	3.0	0.557710	1405088.0	2.519390e+06
3	4.0	0.499758	1096764.0	2.194591e+06
4	5.0	1.122025	2673796.0	2.383009e+06
5	6.0	1.192688	2917327.0	2.446010e+06
6	7.0	0.944780	1811000.0	1.916847e+06
7	8.0	1.061965	3228925.0	3.040520e+06
8	9.0	1.144253	4737023.0	4.139841e+06
9	10.0	1.688389	2757452.0	1.633185e+06
10	11.0	0.884868	1526410.0	1.725014e+06
11	12.0	1.486654	2103162.0	1.414695e+06



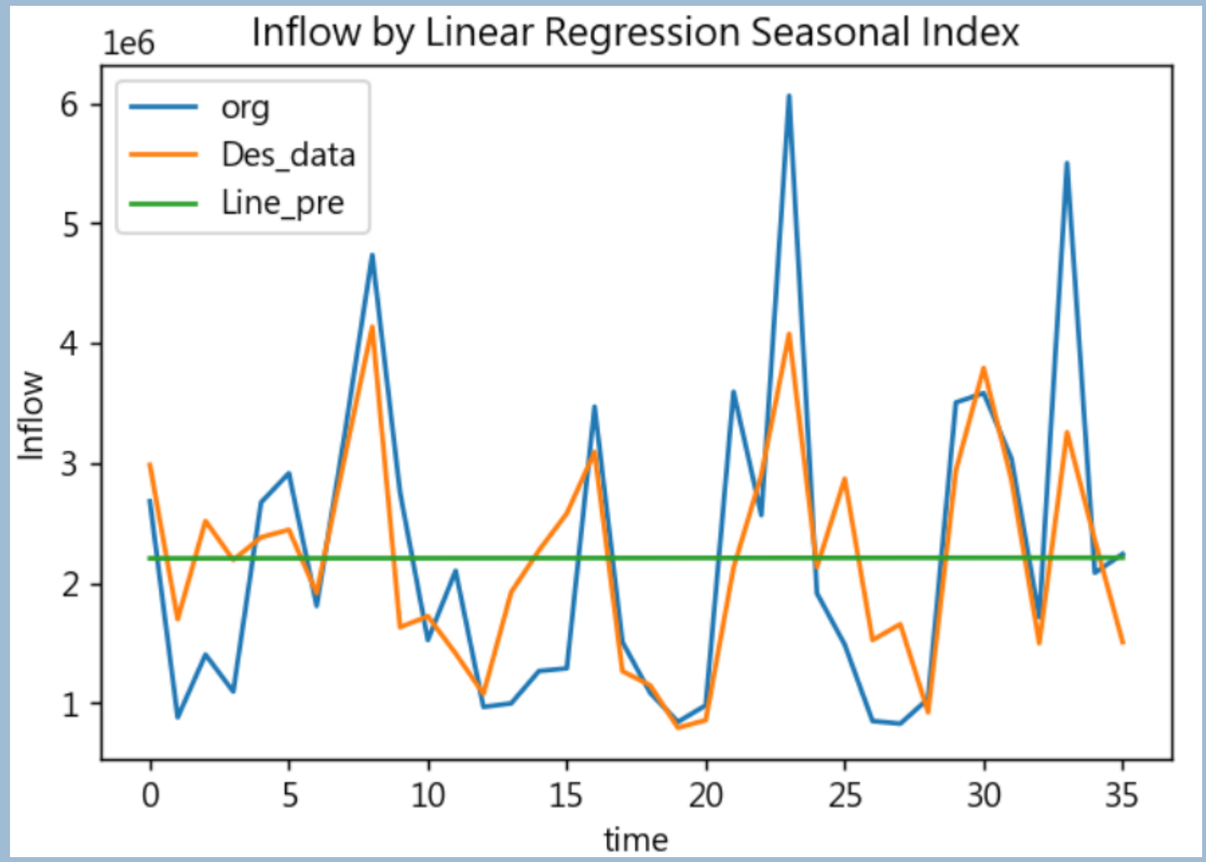
New Regression

OLS Regression Results

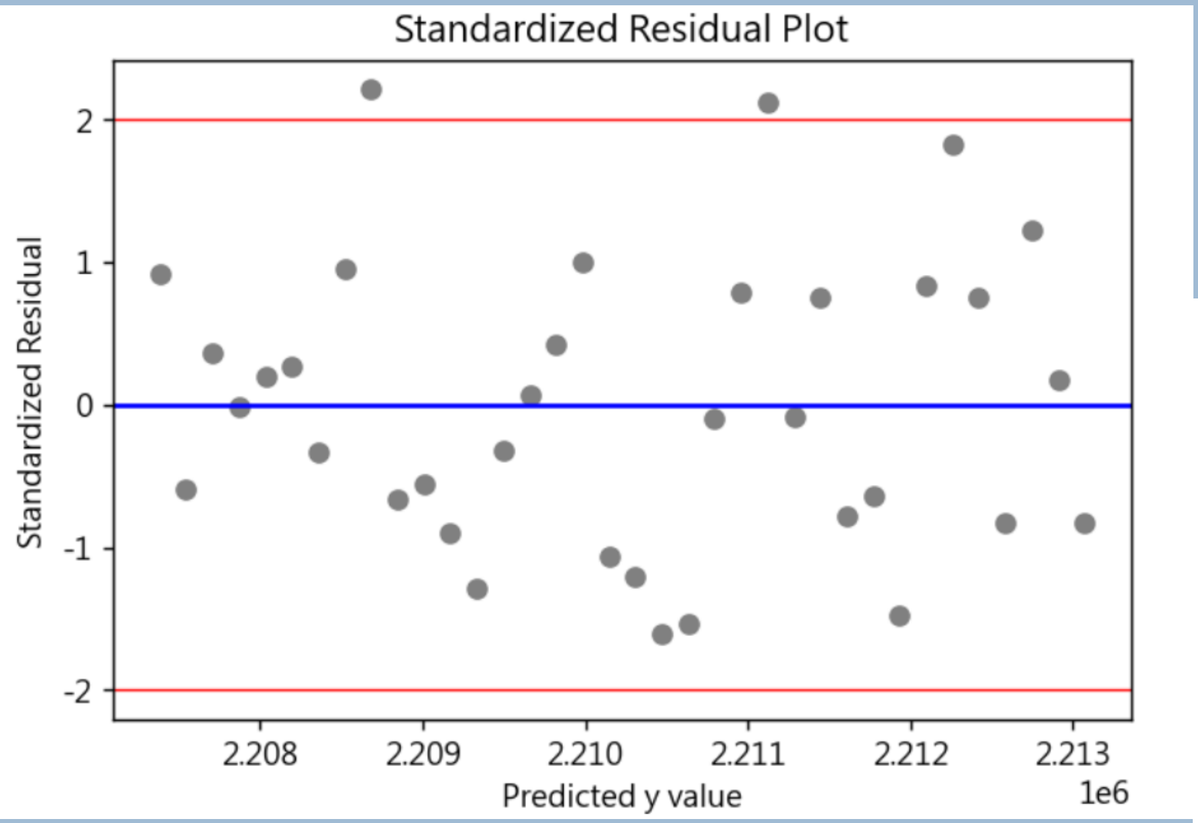
Dep. Variable:	Des_D	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.029
Method:	Least Squares	F-statistic:	0.0001277
Date:	Thu, 19 May 2022	Prob (F-statistic):	0.991
Time:	21:41:26	Log-Likelihood:	-543.46
No. Observations:	36	AIC:	1091.
Df Residuals:	34	BIC:	1094.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.207e+06	2.93e+05	7.544	0.000	1.61e+06	2.8e+06
t	162.4628	1.44e+04	0.011	0.991	-2.91e+04	2.94e+04

Omnibus:	1.266	Durbin-Watson:	1.422
Prob(Omnibus):	0.531	Jarque-Bera (JB):	1.247
Skew:	0.371	Prob(JB):	0.536
Kurtosis:	2.471	Cond. No.	39.9



Residual Analysis




```
Shapiro Test
Statistics=0.966, p=0.334
runs =  16
n1 =  18
n2 =  18
runs_exp =  19.0
stan_dev =  2.9568322818274866
z =  -1.014599312391785
pval_z =  0.3102968672638191
p_value for Z-statistic=  0.3102968672638191
```

because n1, n2 < 20, Lr = 12 < R = 16 < Ur = 26, do not reject H0, There is no evidence to infer that the sample is not random.

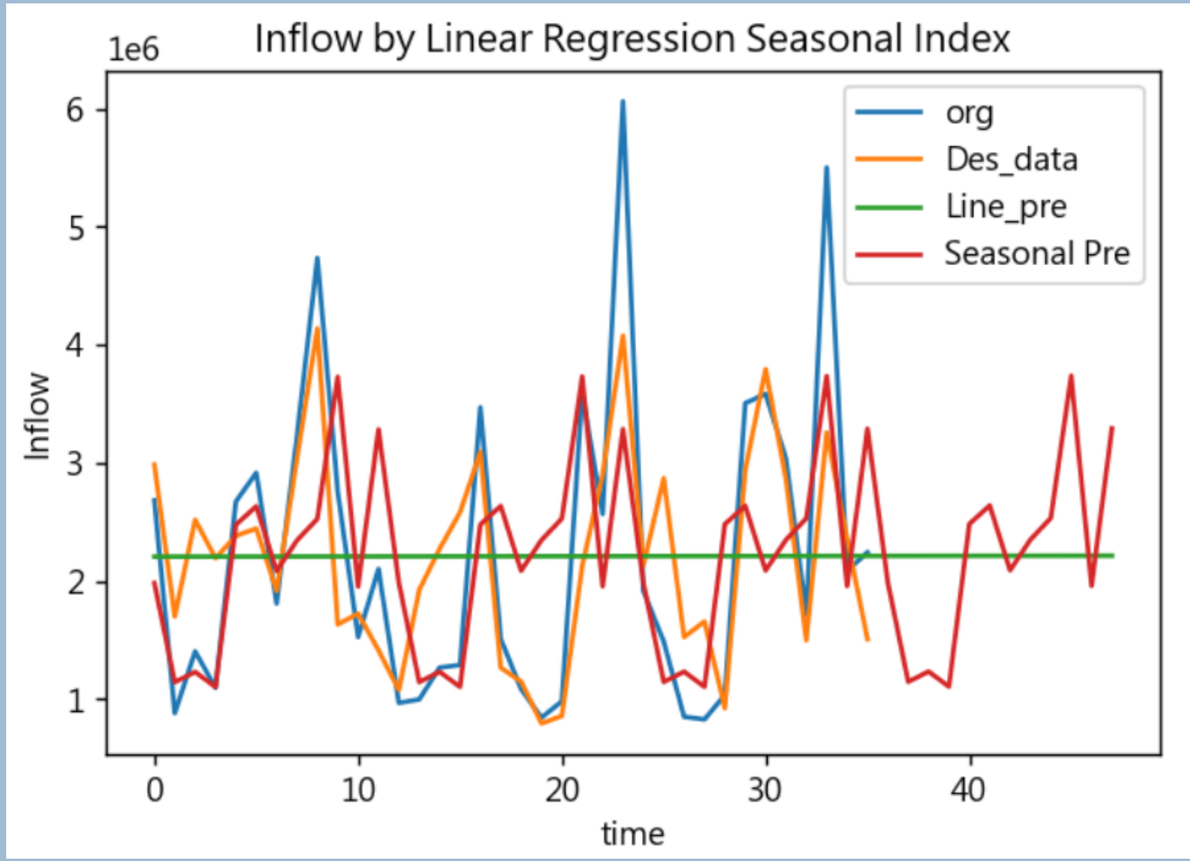
```
The Durbin Watson test
x_square_sum =  35.783926362059596
size =  36
x_d =  [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
x_d =  [ 0.         -1.51563834  0.96097267 -0.37997792  0.2184428   0.07190145
 -0.61071289  1.29145593  1.25645156 -2.86986844  0.10593549 -0.35231829
 -0.38152444  0.96408747  0.39436665  0.34883329  0.57924443 -2.06825063
 -0.13358174 -0.40129046  0.07063841  1.43975943  0.87652652  1.33735354
 -2.21095545  0.84109253 -1.53543001  0.14944828 -0.84544383  2.31803111
  0.99032962 -1.07699237 -1.58191439  2.05317747 -1.05211957 -1.00180468]
d =  1.4313915532378143
d value = 1.4313915532378143

n = 36, k = 1, alpha = 0.05

dL=1.411, dU=1.525 Because dL = 1.411 < d value 1.4313915532378143 < dU=1.525 , it is inconclusive whether the first order correlation exists.
```

未來一年每個月的進水量預測

	time	org	Des_data	Line_pre	Seasonal Pre
36	36.0	NaN	NaN	2.213233e+06	1.988694e+06
37	37.0	NaN	NaN	2.213395e+06	1.147343e+06
38	38.0	NaN	NaN	2.213558e+06	1.234522e+06
39	39.0	NaN	NaN	2.213720e+06	1.106324e+06
40	40.0	NaN	NaN	2.213883e+06	2.484031e+06
41	41.0	NaN	NaN	2.214045e+06	2.640666e+06
42	42.0	NaN	NaN	2.214207e+06	2.091940e+06
43	43.0	NaN	NaN	2.214370e+06	2.351583e+06
44	44.0	NaN	NaN	2.214532e+06	2.533984e+06
45	45.0	NaN	NaN	2.214695e+06	3.739266e+06
46	46.0	NaN	NaN	2.214857e+06	1.959857e+06
47	47.0	NaN	NaN	2.215020e+06	3.292969e+06



Regression Model by Indicator Variables (Dummy)

Regression Results

We can see that 41.5% of variations are explained by this model. However, overfitting exists.

OLS Regression Results

Dep. Variable:	water_in	R-squared:	0.415
Model:	OLS	Adj. R-squared:	0.110
Method:	Least Squares	F-statistic:	1.362
Date:	Wed, 18 May 2022	Prob (F-statistic):	0.253
Time:	17:27:57	Log-Likelihood:	-548.68
No. Observations:	36	AIC:	1123.
Df Residuals:	23	BIC:	1144.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.472e+06	8.78e+05	3.953	0.001	1.65e+06	5.29e+06
t	-69.5788	2.14e+04	-0.003	0.997	-4.44e+04	4.42e+04
month_1	-1.615e+06	1.05e+06	-1.531	0.139	-3.8e+06	5.67e+05
month_2	-2.348e+06	1.05e+06	-2.236	0.035	-4.52e+06	-1.76e+05
month_3	-2.296e+06	1.05e+06	-2.195	0.039	-4.46e+06	-1.32e+05
month_4	-2.399e+06	1.04e+06	-2.301	0.031	-4.56e+06	-2.42e+05
month_5	-1.077e+06	1.04e+06	-1.036	0.311	-3.23e+06	1.07e+06
month_6	-8.257e+05	1.04e+06	-0.797	0.434	-2.97e+06	1.32e+06
month_7	-1.31e+06	1.03e+06	-1.267	0.218	-3.45e+06	8.29e+05
month_8	-1.1e+06	1.03e+06	-1.066	0.297	-3.23e+06	1.03e+06
month_9	-9.92e+05	1.03e+06	-0.963	0.346	-3.12e+06	1.14e+06
month_10	4.814e+05	1.03e+06	0.468	0.644	-1.65e+06	2.61e+06
month_11	-1.41e+06	1.03e+06	-1.371	0.184	-3.54e+06	7.18e+05

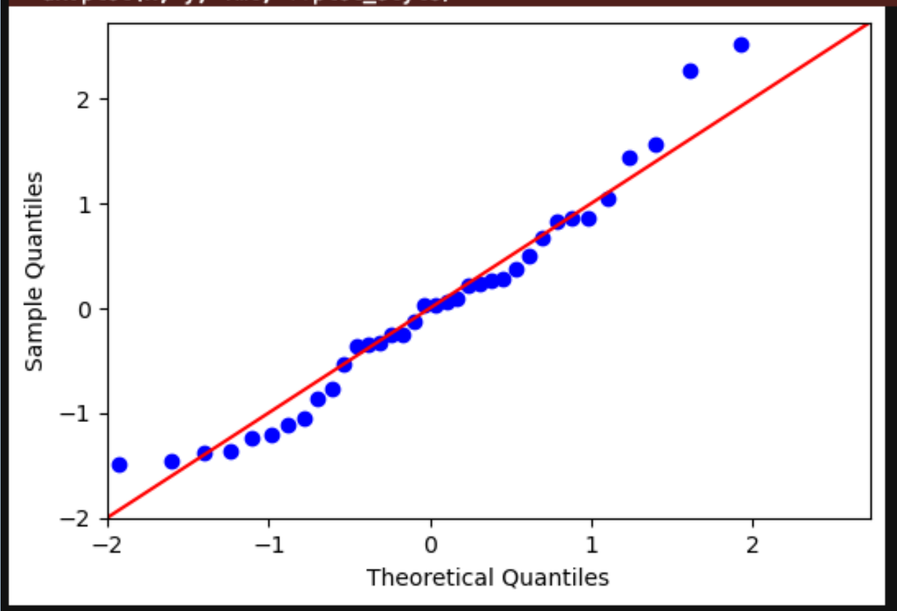
Omnibus:	2.278	Durbin-Watson:	1.556
Prob(Omnibus):	0.320	Jarque-Bera (JB):	1.702
Skew:	0.533	Prob(JB):	0.427
Kurtosis:	2.977	Cond. No.	259.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Fitted Values [1856900.2360215 1123435.46950845 1175619.63387097 1072820.94569893
 2394894.38655914 2645864.1234767 2161432.2360215 2371251.41881721
 2479356.45681004 3952609.91344086 2061418.67903226 3471108.10698925
 1856065.29032258 1122600.52380952 1174784.68817204 1071986.
 2394059.44086021 2645029.17777778 2160597.29032258 2370416.47311828
 2478521.51111111 3951774.96774194 2060583.73333333 3470273.16129032
 1855230.34462366 1121765.5781106 1173949.74247312 1071151.05430108
 2393224.49516129 2644194.23207885 2159762.34462366 2369581.52741936
 2477686.56541219 3950940.02204301 2059748.78763441 3469438.2155914]

Residual Analysis

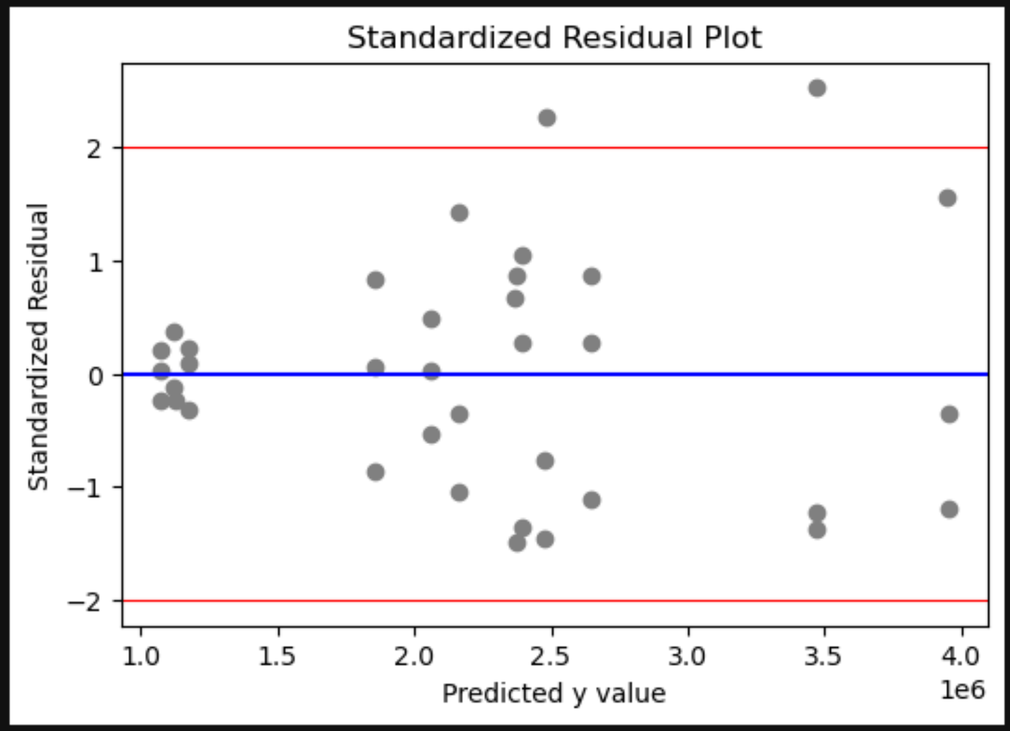
Normality Test

```
Shapiro Test
Statistics=0.957, p=0.180
/Users/liuweien/opt/anaconda3/lib/python3.9/site-packages/statsmodels/graphics/gofplots.py:
g "bo" (-> marker='o'). The keyword argument will take precedence.
  ax.plot(x, y, fmt, **plot_style)
```



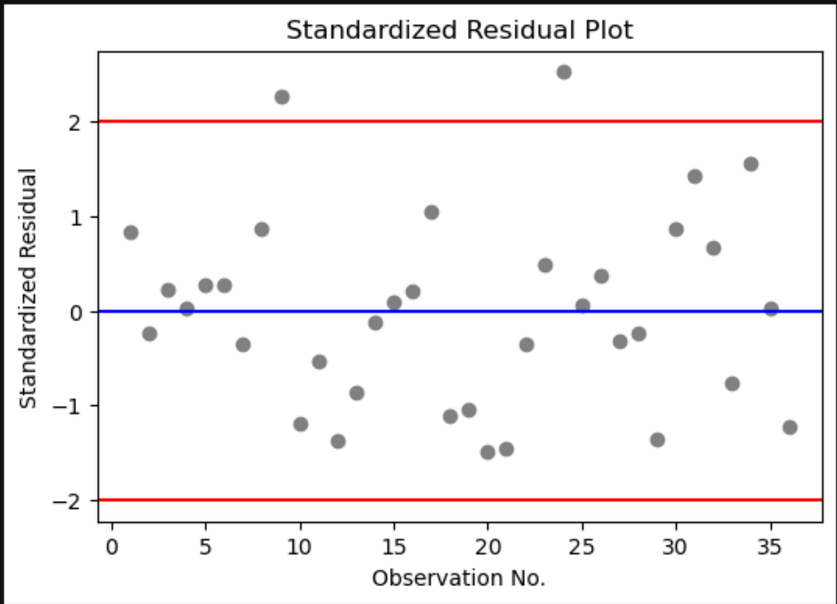
Because p_value = 0.180 > a = 0.05, do not rejected H0. We can assume that errors are normally distributed.

Homoscedasticity and Heteroscedasticity Test



Do not rejected H0. We can assume that the variation is constant and the mean is around 0.

Dependence of the Error Variable Test




```
runs =  16
n1 =  18
n2 =  18
runs_exp =  19.0
stan_dev =  2.9568322818274866
z =  -1.014599312391785
pval_z =  0.3102968672638191
p_value for Z-statistic=  0.3102968672638191
```

Because n1, n2 < 20, Lr = 12 < R = 16 < Ur = 26, do not reject H0. There is no evidence to infer that the sample is not random.

Durbin Watson Test

```
x_square_sum =  35.78236961701012
size =  36
x_d =  [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
x_d =  [ 0.         -1.07313834  0.47384275 -0.20643065  0.25608129 -0.00747103
 -0.62463315  1.21342467  1.40615775 -3.46802759  0.66305591 -0.83660509
  0.51076711  0.74241981  0.21157326  0.12161196  0.83554961 -2.1507051
  0.05735659 -0.43874108  0.02890915  1.10904361  0.83990245  2.02824913
 -2.4602304   0.30637045 -0.69235468  0.08083036 -1.11903322  2.22870971
  0.56539552 -0.76029483 -1.43601499  2.32261227 -1.53050343 -1.25816161]
d =  1.5667996478416744
1.5667996478416744
```

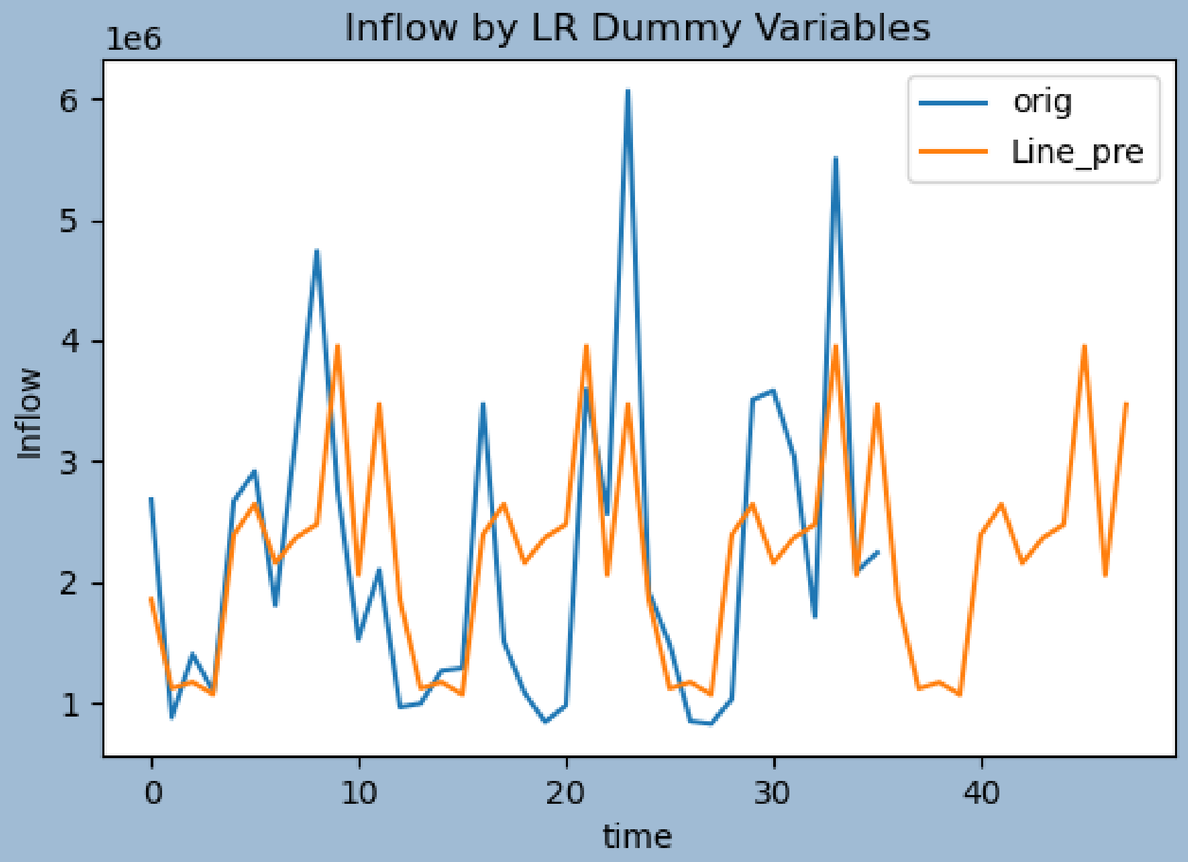
(T = 36, K = 12, alpha = 0.05): dl = 0.748 , du = 2.398

dl < d = 1.5667996478416744 < du. Thus, it is inconcusive whether first order auto-correlation exists.

Regression Functions

Average Inflow of January = -1.615e+06 + -69.5788 t + 3.472e+06
Average Inflow of February = -2.348e+06 + -69.5788 t + 3.472e+06
Average Inflow of March = -2.296e+06 + -69.5788 t + 3.472e+06
Average Inflow of April = -2.399e+06 + -69.5788 t + 3.472e+06
Average Inflow of May = -1.077e+06 + -69.5788 t + 3.472e+06
Average Inflow of June = -8.257e+05 + -69.5788 t + 3.472e+06
Average Inflow of July = -1.31e+06 + -69.5788 t + 3.472e+06
Average Inflow of August = -1.1e+06 + -69.5788 t + 3.472e+06
Average Inflow of September = -9.92e+05 + -69.5788 t + 3.472e+06
Average Inflow of October = 4.814e+05 + -69.5788 t + 3.472e+06
Average Inflow of November = -1.41e+06 + -69.5788 t + 3.472e+06
Average Inflow of December = -69.5788 t + 3.472e+06

Prediction Value and Plot



	Year/month	Predict Inflow
0	111/01	1.854395e+06
1	111/02	1.120931e+06
2	111/03	1.173115e+06
3	111/04	1.070316e+06
4	111/05	2.392390e+06
5	111/06	2.643359e+06
6	111/07	2.158927e+06
7	111/08	2.368747e+06
8	111/09	2.476852e+06
9	111/10	3.950105e+06
10	111/11	2.058914e+06
11	111/12	3.468603e+06

Error Matrix Analysis of Part 3

dataset:

	ErrM	SIMA	SILR	Dummy
0	MAD	897899.420924	777168.888083	788789.261086
1	MSE	1410625116976.3325	1023939149413.8843	1013279558234.3698
2	RMSE	1187697.401267	1011898.784175	1006617.880943
3	MAPE	46.281122	41.716465	42.033599

111/01 Prediction Analysis

	Year/month	Predict Inflow	Seasonal Pre	Predict Inflow	actual value
36	111/01	2.024442e+06	1.988694e+06	1.854395e+06	2.718889e+06

Part 3 conclusion

According to the results in the previous pages,
we can notice that three methods all passed the residual analysis test.
Thus, we can conclude that
the prediction with month average data by seasonal effect
is a quite accurate prediction method compared to the previous methods.

However, by comparing it to the forecast value
with the actual value in 111/01, there is still a gap between it.
We can observe that the influence of seasonal change is not as crucial as we thought it would be.

Total Conclusion

以日為單位：

Part 1 中以時間序列所生成的回歸
相較於 Part 2 中的季節性回歸
所得到的結果較為準確

以月為單位：

seasonal index by simple linear regression 所生成的回歸較佳

以日為單位的預測可以用來進行即時的水庫管理決策

而以月為單位的預測可以用來了解長期的趨勢

資料來源

水庫運轉月報表

來源：https://data.gov.tw/dataset/145840?fbclid=IwAR2jdsH5VSQBdI5jB2YGgZadTnwbM8LZbEyPl2YOoeezFXsH5dvBLsMoU5A

Thank you!