Research of factors that influence on oil prices

Resource Market Analysis

Kim JaeHun
Hwang Hyeju
Lee YoungHoon

Importance

Energy

Clothes

Fuel

War For Oil

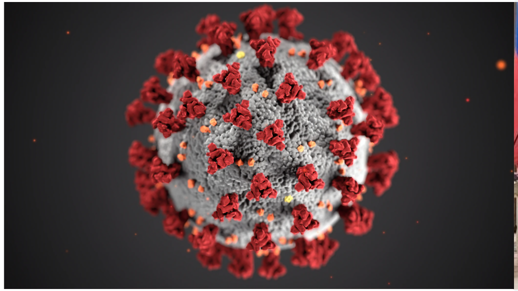
Supply & Demand

Decreasing Demand

- Covid-19 Economic
- Travel restrict

Increasing Supply

- 2020 OPEC–Russia oil price war



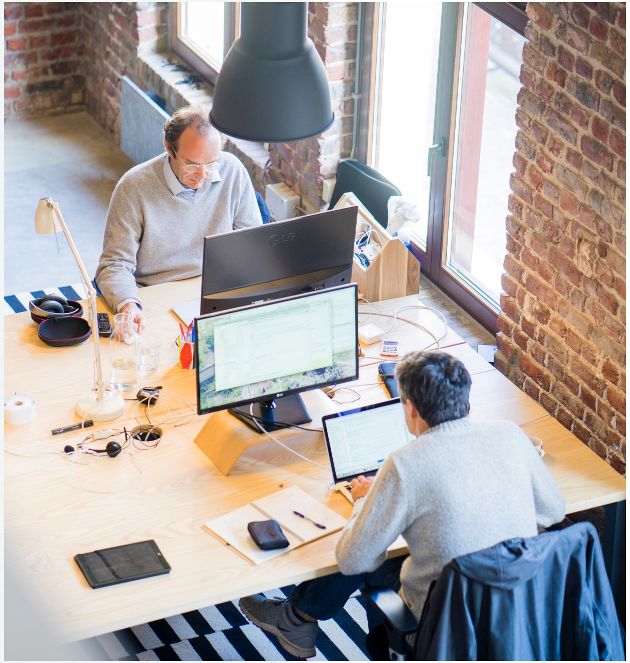
Recently Oil Price

국제유가, 최악 코로나에 3.2% 폭락...3주來 최저치로 '뚝'

입력 2020.10.27 08:17 | 수정 2020.10.27 08:17

[국제유가] 미국-유럽 코로나19 재확산... WTI, 3주 만에 최저

Candidate Factors Influencing Oil Prices



1. Worldwide car sales
2. US trade balance
3. US unemployment rate
4. Tesla stock price
5. US natural gas price
6. Korea’s renewable energy generation
7. Korea’s nuclear energy generation
8. Gold price

with 2017 ~ 2019 (3 years) monthly data

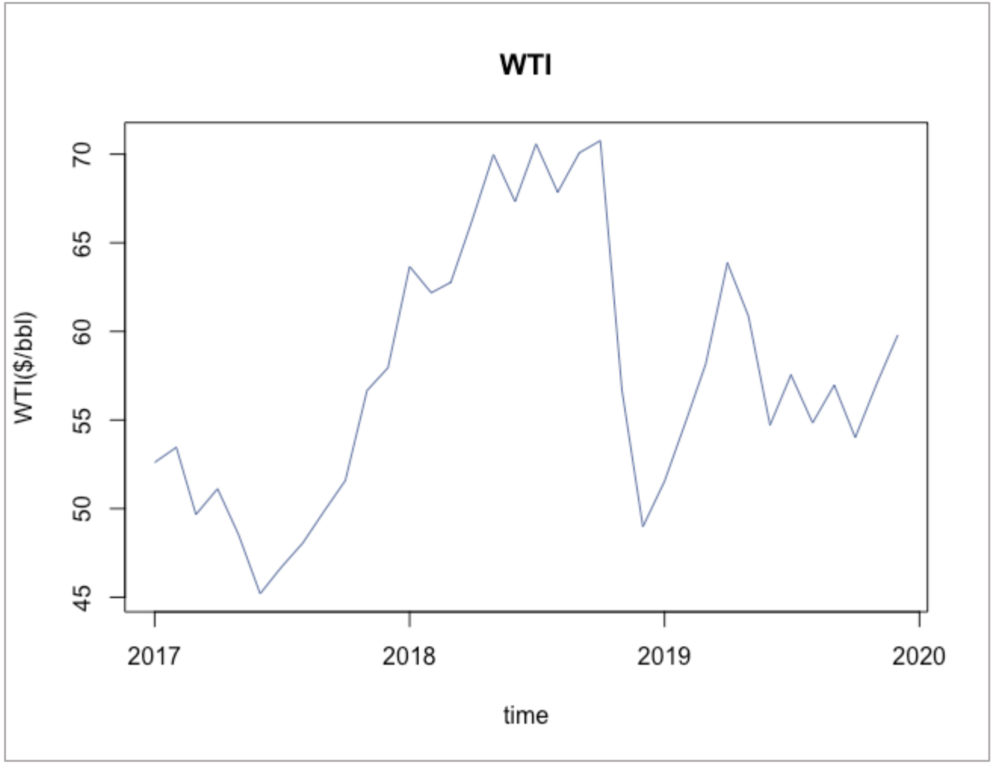
Period

2017.01 ~ 2019.12 Monthly

1. Difficulties of checking availability of 2020 data
2. Difficulties of finding weekly data of some factors

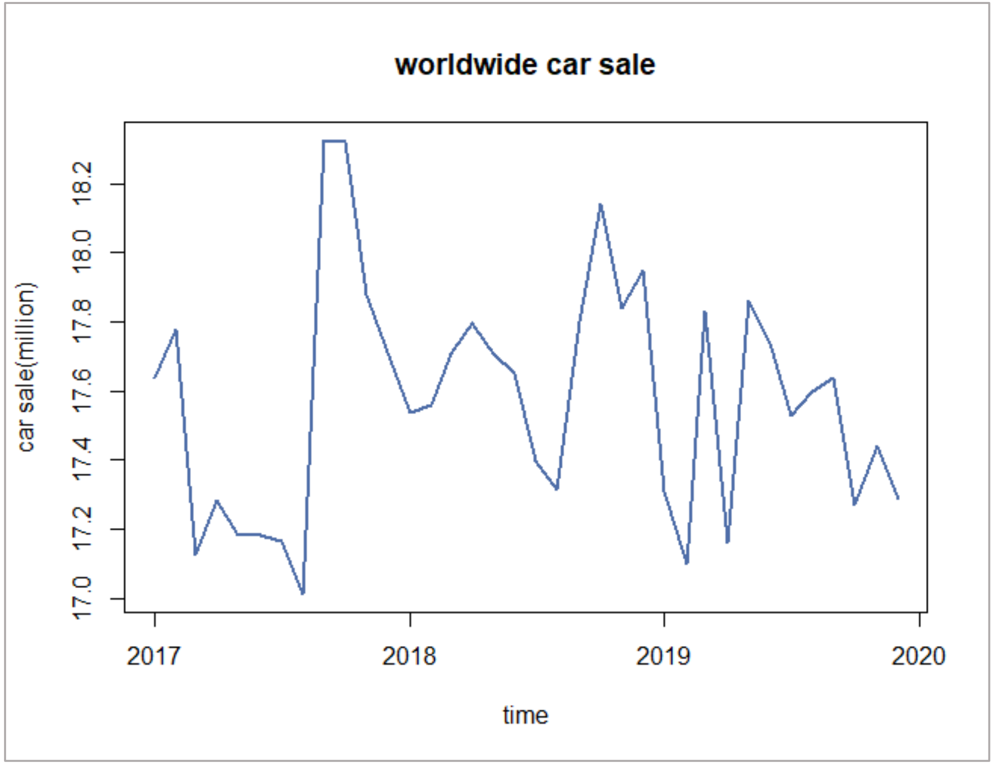
WTI

Min 45.20
Max 70.76
Mean 57.58
Median 56.83
Variance 54.40
Standard Deviation 7.38



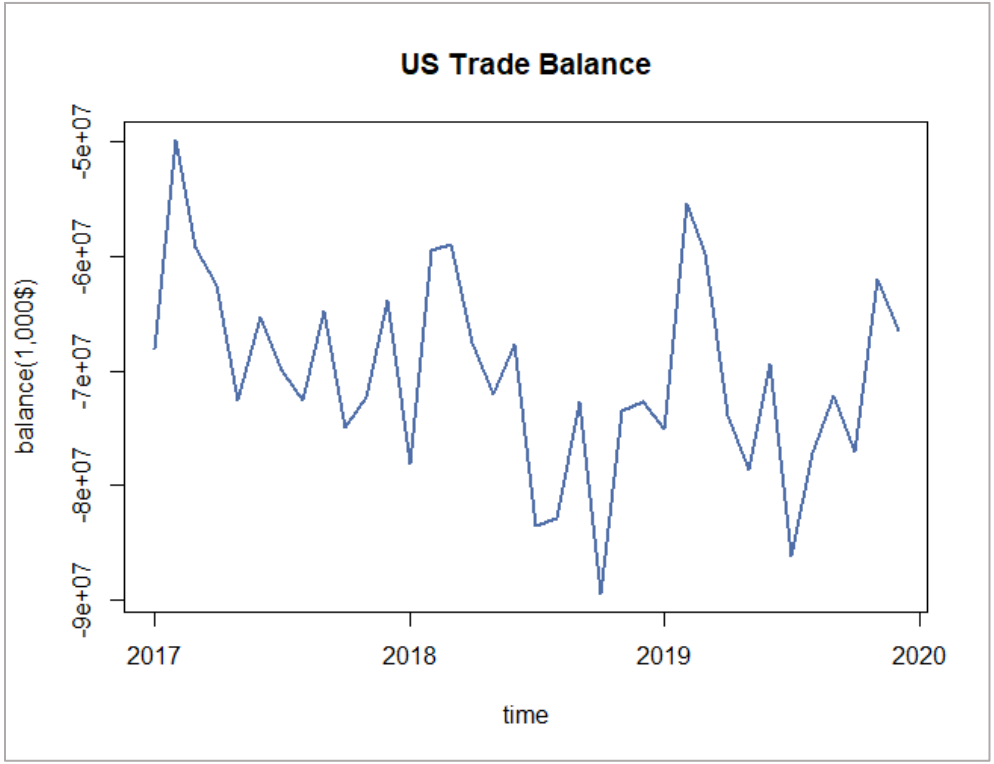
Worldwide Car Sale

Min 17.01
Max 18.32
Mean 17.58
Median 17.62
Variance 0.11
Standard Deviation 0.34



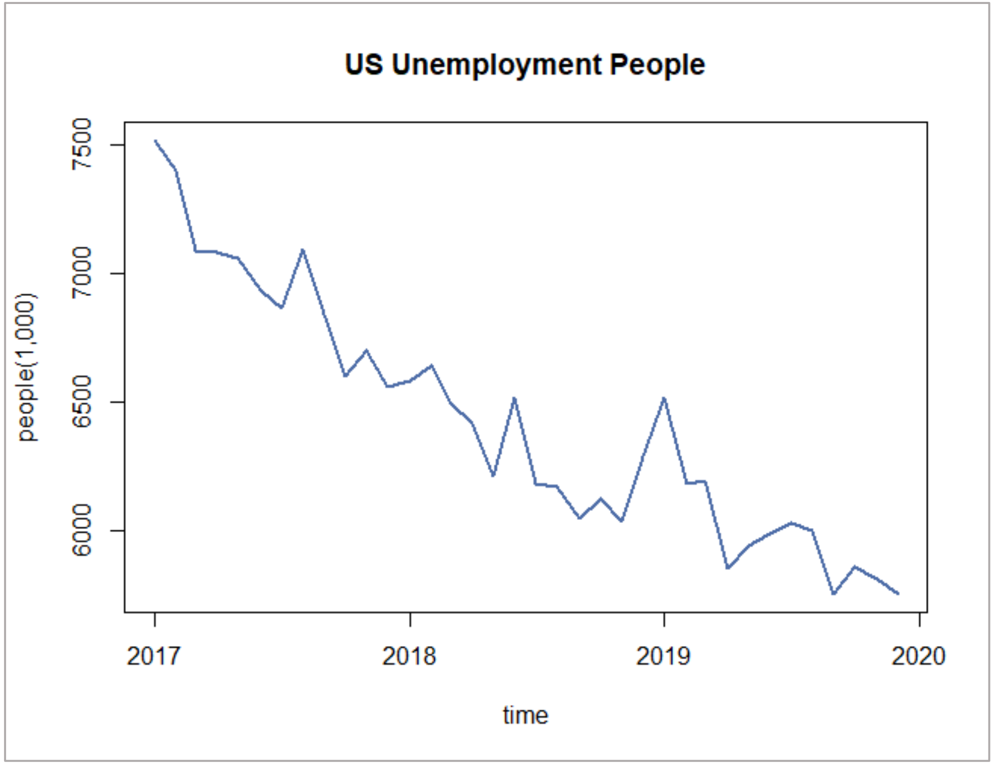
US Trade Balance

Min -89,362,720
Max -49,946,423
Mean -70,224,383
Median -72,121,613
Variance 7.54×10^{13}
Standard Deviation 8,683,593



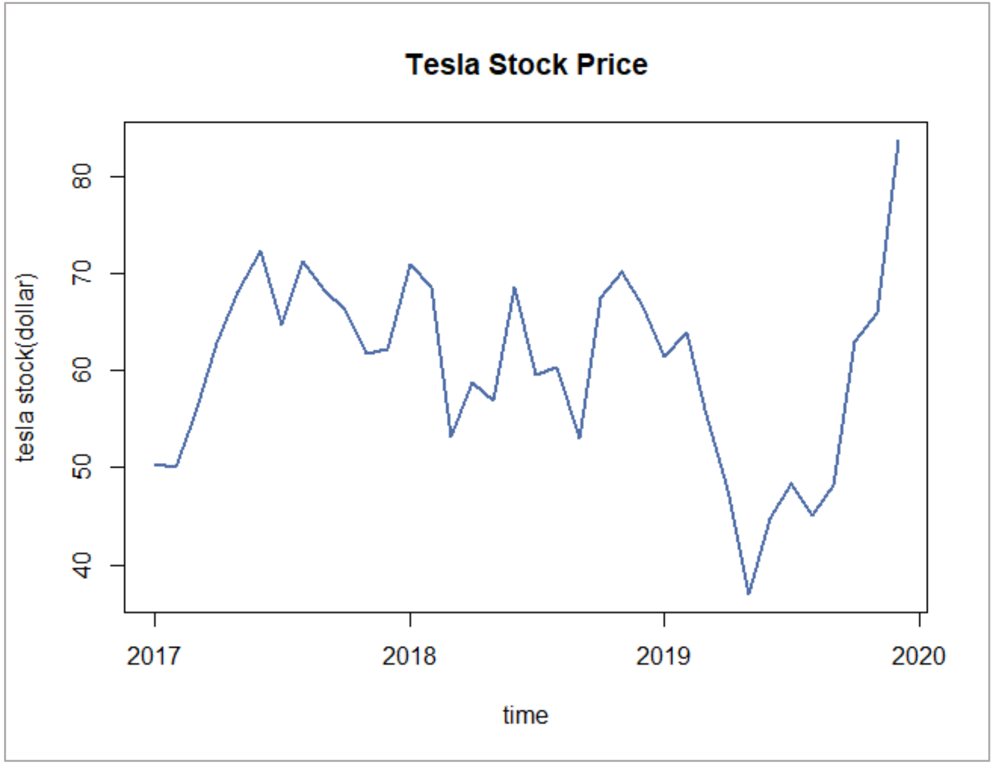
US Unemployment People

Min 5,753
Max 7,518
Mean 6,425
Median 6,352
Variance 231,782
Standard Deviation 481



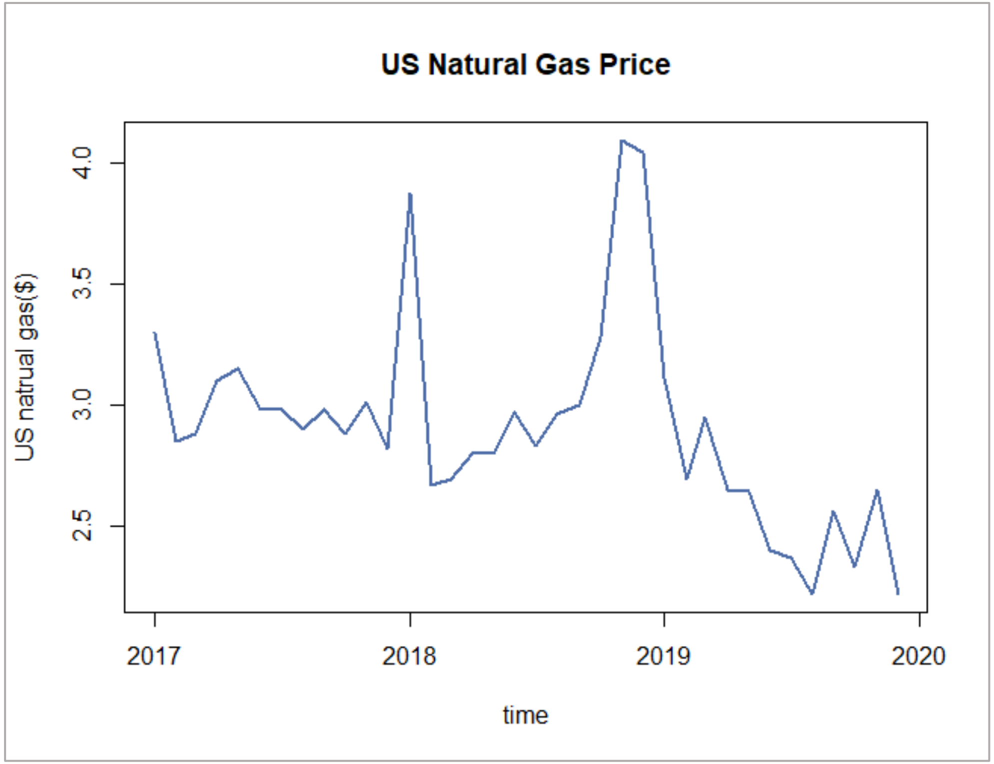
Tesla Stock Price

Min 37.03
Max 83.67
Mean 60.36
Median 62.02
Variance 94.81
Standard Deviation 9.74



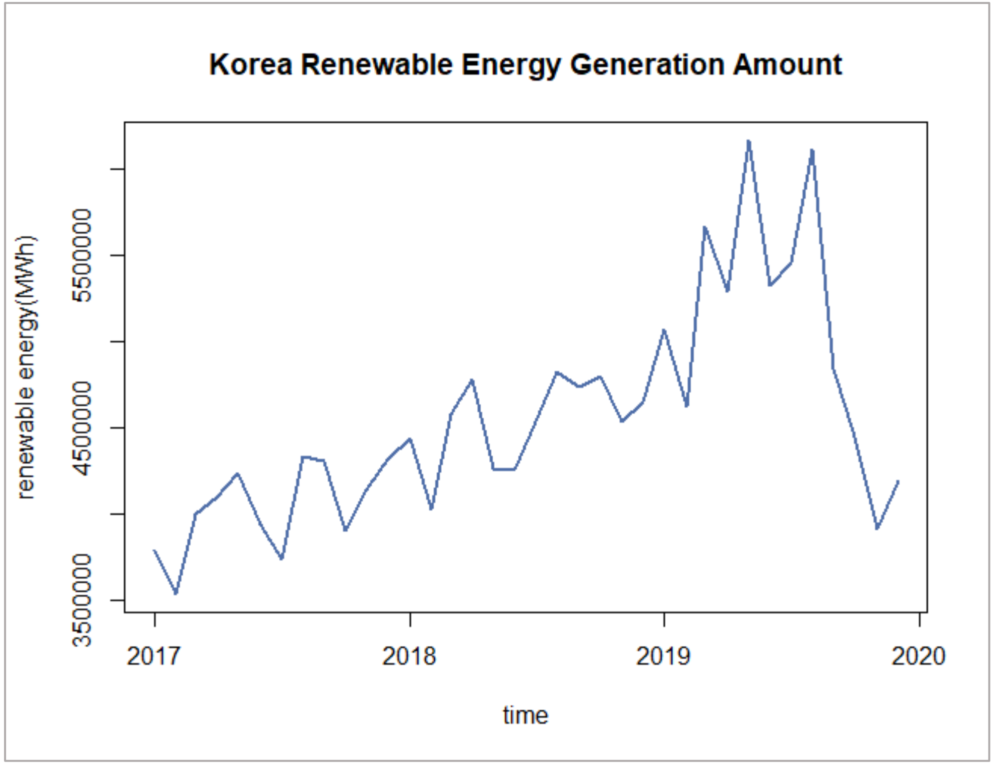
US Natural Gas Price

Min 2.22
Max 4.09
Mean 2.91
Median 2.88
Variance 0.18
Standard Deviation 0.43



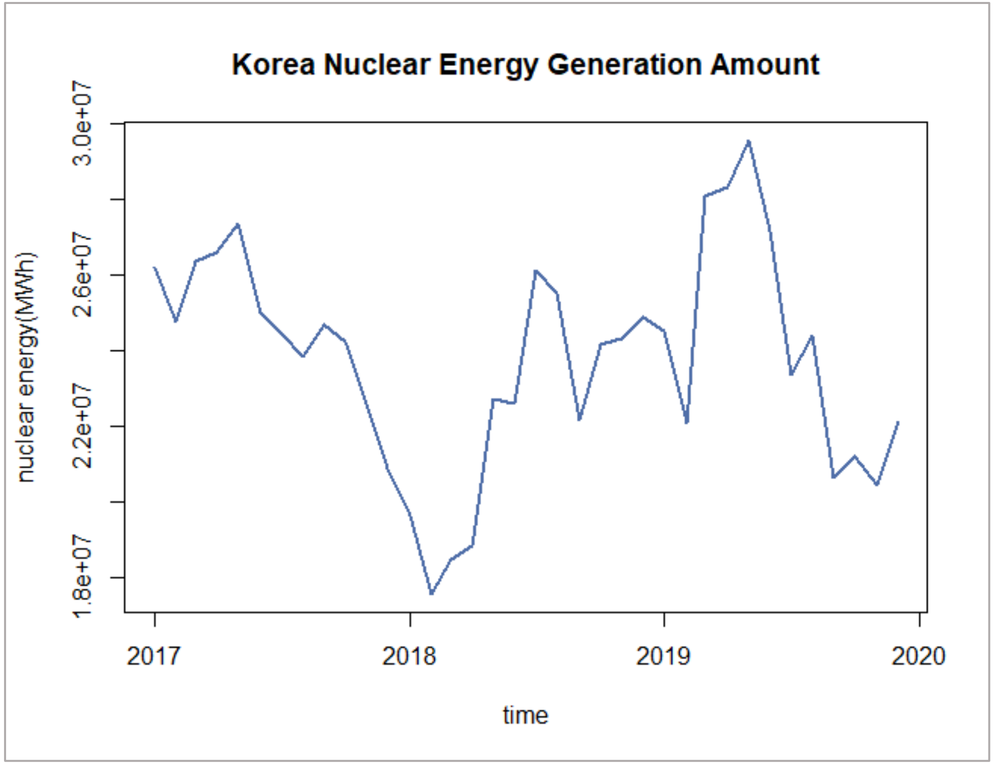
Domestic Renewable Energy Generation

Min 3,534,270
Max 6,160,923
Mean 4,547,127
Median 4,440,754
Variance 395,628,790,264
Standard Deviation 628,990



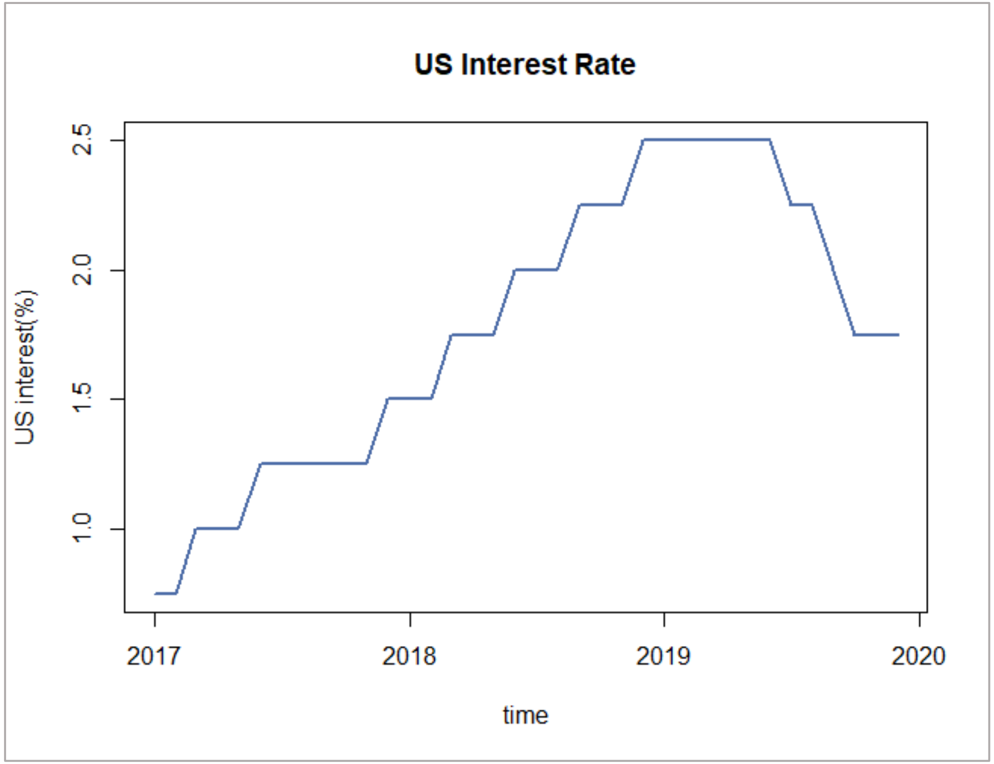
Domestic Nuclear Energy Generation

Min 17,582,516
Max 29,542,098
Mean 23,768,981
Median 24,267,115
Variance 8.22×10^{12}
Standard Deviation 2,866,422



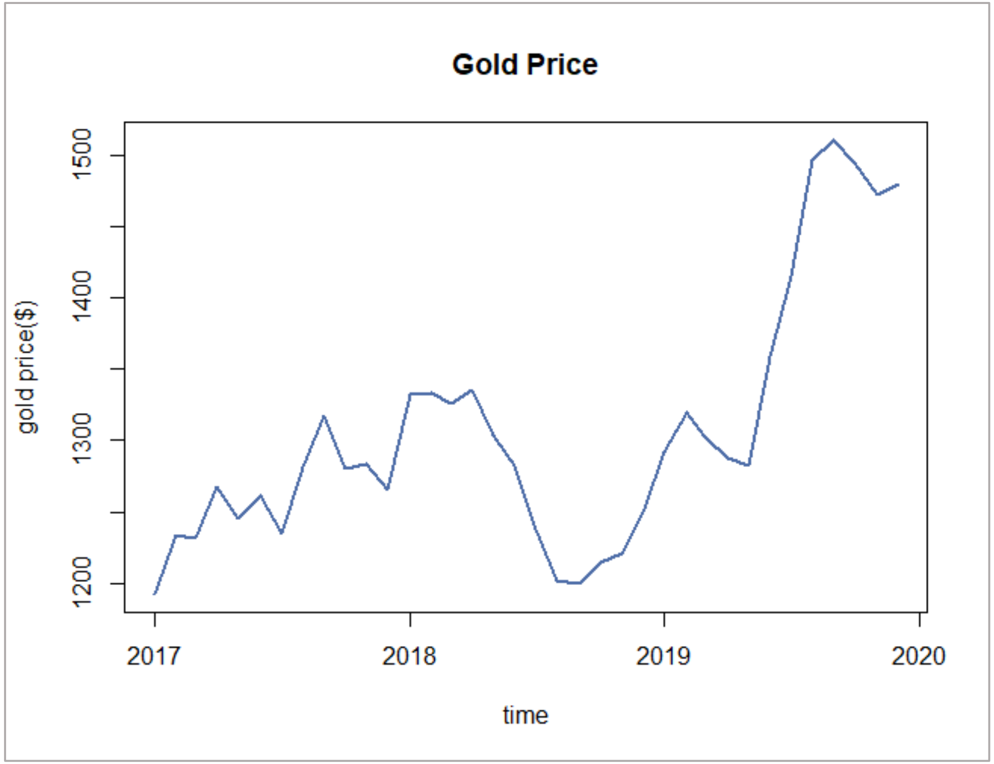
US Interest Rate

Min 0.75
Max 2.50
Mean 1.77
Median 1.75
Variance 0.31
Standard Deviation 0.56



Gold Price

Min 1,193
Max 1,510
Mean 1,307
Median 1,283
Variance 7,881
Standard Deviation 88.78



Ordinary Least Square

```
> summary(lm(WTI~worldwide_car_sale+US_trade_balance+US_unemployment_people+tesla_stock_price+US_
_energy+domestic_nuclerar_energy+US_interest_rate+gold_price ,data=data))

Call:
lm(formula = WTI ~ worldwide_car_sale + US_trade_balance + US_unemployment_people +
    tesla_stock_price + US_natural_gas_price + domestic_renewable_energy +
    domestic_nuclerar_energy + US_interest_rate + gold_price,
    data = data)

Residuals:
   Min      1Q Median      3Q     Max
-8.151 -2.649 -1.102  3.642 10.660

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                2.777e+02  8.164e+01   3.402 0.002174 **
worldwide_car_sale         1.148e+00  2.834e+00   0.405 0.688699
US_trade_balance          -8.919e-08  1.306e-07  -0.683 0.500815
US_unemployment_people    -1.479e-02  5.020e-03  -2.946 0.006710 **
tesla_stock_price         -6.348e-02  1.282e-01  -0.495 0.624785
US_natural_gas_price      -4.415e+00  3.078e+00  -1.434 0.163416
domestic_renewable_energy  3.389e-06  3.537e-06   0.958 0.346931
domestic_nuclerar_energy  -1.530e-06  3.977e-07  -3.846 0.000698 ***
US_interest_rate          -5.690e+00  4.865e+00  -1.170 0.252735
gold_price                -7.950e-02  1.827e-02  -4.353 0.000185 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.057 on 26 degrees of freedom
Multiple R-squared:  0.6507,	Adjusted R-squared:  0.5298
F-statistic: 5.382 on 9 and 26 DF,  p-value: 0.0003471
```

Approach: Add or Remove explanatory variables to get the highest Adjusted R-squared

We started the combination of the lowest p-value ones.

So the starting tentative model is

model = lm(WTI ~ gold_price +domestic_nuclear_energy +US_unemployment_people, data=data)

Ordinary Least Square

```
> model = lm(WTI~gold_price+domestic_nuclerar_energy+US_unemployment_people, data=data)
>
> summary(model)

Call:
lm(formula = WTI ~ gold_price + domestic_nuclerar_energy + US_unemployment_people,
    data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-12.1573  -2.7885   0.4092   3.3066   8.8899

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               2.321e+02  3.143e+01   7.384 2.12e-08 ***
gold_price               -5.401e-02  1.320e-02  -4.093  0.00027 ***
domestic_nuclerar_energy -1.101e-06  3.446e-07  -3.195  0.00314 **
US_unemployment_people   -1.211e-02  2.243e-03  -5.398 6.24e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.3 on 32 degrees of freedom
Multiple R-squared:  0.5279,    Adjusted R-squared:  0.4837
F-statistic: 11.93 on 3 and 32 DF,  p-value: 2.094e-05
```

Starting tentative model's Adjusted R-squared value is 0.4837

Ordinary Least Square

```
> model = lm(WTI~gold_price+domestic_nuclerar_energy+US_unemployment_people+tesla_stock_
> summary(model)

Call:
lm(formula = WTI ~ gold_price + domestic_nuclerar_energy + US_unemployment_people +
    tesla_stock_price + US_natural_gas_price, data = data)

Residuals:
   Min     1Q Median     3Q    Max 
-8.918 -3.034 -0.240  3.146  9.358 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               2.778e+02  3.669e+01   7.570 1.93e-08 ***
gold_price               -7.016e-02  1.544e-02  -4.544 8.41e-05 ***
domestic_nuclerar_energy -1.377e-06  3.529e-07  -3.902   0.0005 ***
US_unemployment_people   -1.192e-02  2.156e-03  -5.526 5.28e-06 ***
tesla_stock_price        -1.167e-01  1.024e-01  -1.140   0.2634    
US_natural_gas_price     -4.180e+00  2.759e+00  -1.515   0.1403    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.026 on 30 degrees of freedom
Multiple R-squared:  0.6019,	Adjusted R-squared:  0.5355 
F-statistic: 9.071 on 5 and 30 DF,  p-value: 2.445e-05
```

The final tentative model is

model = lm(WTI ~ gold_price
+domestic_nuclear_energy
+US_unemployment_people
+tesla_stock_price
+US_natural_gas_price,
data=data)

which has the highest adjusted R-squared value, 0.5355

Ordinary Least Square

$$y = 277.75 - 0.070x_1 - 0.0000014x_2 - 0.012x_3 - 0.12x_4 - 4.18x_5$$

where, y : WTI, x_1: gold price,
x_2 : domestic nuclear energy, x_3: US unemployment people,
x_4 : tesla stock price, x_5 : US natural gas price

Correlation

```
> # Correlation
> cor(data)
                               WTI US_unemployment_people tesla_stock_price US_natural_gas_price domestic_nuclerar_energy  gold_price
WTI                     1.00000000             -0.4887995       -0.14905793          -0.08312462              -0.26621033 -0.04270672
US_unemployment_people -0.48879953              1.0000000        0.17036487           0.32491797               0.12953918 -0.54883582
tesla_stock_price      -0.14905793              0.1703649        1.00000000           0.34827347              -0.32787127 -0.05524263
US_natural_gas_price   -0.08312462              0.3249180        0.34827347           1.00000000               0.09768287 -0.61379706
domestic_nuclerar_energy -0.26621033            0.1295392       -0.32787127           0.09768287               1.00000000 -0.40610590
gold_price             -0.04270672             -0.5488358       -0.05524263          -0.61379706              -0.40610590  1.00000000
```

There is no absolute value over 0.8
We can conclude that there is no correlation between explanatory variables

Heteroscedasticity

```
> # Heteroscedasticity test
> bptest(model)

	studentized Breusch-Pagan test

data:  model
BP = 8.8165, df = 5, p-value = 0.1166

> gqtest(model)

	Goldfeld-Quandt test

data:  model
GQ = 0.83032, df1 = 12, df2 = 12, p-value = 0.6237
alternative hypothesis: variance increases from segment 1 to 2
```

BP Test & GQ Test

Null Hypothesis is that the model is homoscedastic.

P value is greater than 0.05

→ cannot reject null hypothesis

The model is homoscedastic.

Autocorrelation

```
> # Autocorrelation test
> dwtest(model)

	Durbin-Watson test

data:  model
DW = 1.3635, p-value = 0.003327
alternative hypothesis: true autocorrelation is greater than 0

> bgtest(model)

	Breusch-Godfrey test for serial correlation of order up to 1

data:  model
LM test = 3.9908, df = 1, p-value = 0.04575
```

DW Test & BG Test

Null Hypothesis is that the model doesn't have autocorrelation.

P value is less than 0.05

→ can reject null hypothesis

The model has autocorrelation.

Normality

```
> # Normality test
> jarque.bera.test(resid(model))

        Jarque Bera Test

data:  resid(model)
X-squared = 0.70608, df = 2, p-value = 0.7025
```

Jarque Bera Test

Null hypothesis is the model has normality.

P value is greater than 0.05

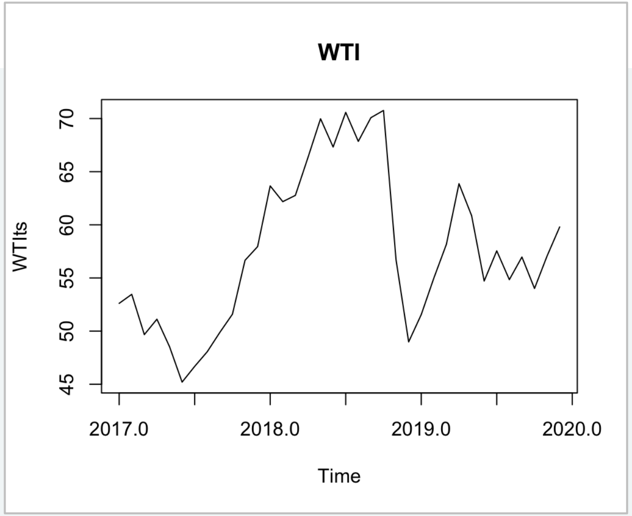
→ cannot reject null hypothesis

The model has normlaity.

WTI in time series

Set the WTI time series variable

```
> # Set time series variable
> WTIts = ts(data=data$WTI, start=c(2017,01), end=c(2019,12), frequency = 12)
> WTIts
       Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
2017 52.61 53.46 49.67 51.12 48.54 45.20 46.68 48.06 49.88 51.59 56.66 57.95
2018 63.66 62.18 62.77 66.33 69.98 67.32 70.58 67.85 70.08 70.76 56.69 48.98
2019 51.55 54.98 58.17 63.87 60.87 54.71 57.55 54.84 56.97 54.01 57.07 59.80
```



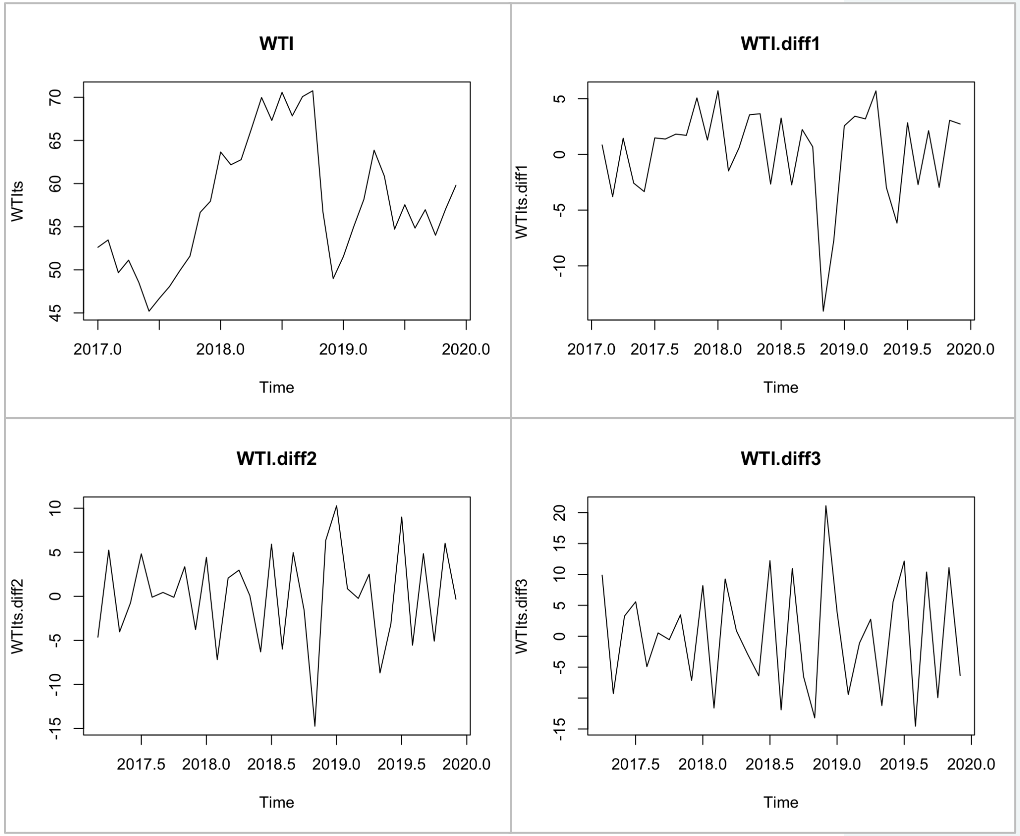
ADF Unit Root Test for stationary check

Calculated tau: -1.9636
Comparing absolute values, it is smaller than critical values' tau

Non-stationary

```
> # Unit Test for WTIts. Test whether WTI is stationary or not
> WTIts.unit = ur.df(WTIts, lags=1, type='drift')
> summary(WTIts.unit)  # non-stationary

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.9030  -2.8834   0.2197   2.7750   5.5202 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 11.05829    5.58839   1.979   0.0568 .
z.lag.1     -0.18904    0.09628  -1.964   0.0586 .
z.diff.lag   0.22261    0.17498   1.272   0.2128  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.021 on 31 degrees of freedom
Multiple R-squared:  0.126,	Adjusted R-squared:  0.0696 
F-statistic: 2.234 on 2 and 31 DF,  p-value: 0.124


Value of test-statistic is: -1.9636 1.9578 

Critical values for test statistics: 
      1pct  5pct 10pct
tau2 -3.58 -2.93 -2.60
phi1  7.06  4.86  3.94
```

Difference to make it stationary




```
# We can see the stationary pattern
# 차분을 통해 정상화(stationary) 패턴을 확인
plot(WTIts, main="WTI")
plot(WTIts.diff1, main="WTI.diff1")
plot(WTIts.diff2, main="WTI.diff2")
plot(WTIts.diff3, main="WTI.diff3")
```

ADF Test for differenced WTI

```
# Test whether it is stationary or not by ADF
# ADF를 통해 정상화(stationary) 확인
summary(ur.df(WTIts.diff2, lags=1, type='drift')) # stationary
summary(ur.df(WTIts.diff3, lags=1, type='drift')) # more stationary
```

ADF result of the WTIts.diff2

```
Value of test-statistic is: -6.0808 18.4884

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.58 -2.93 -2.60
phi1  7.06  4.86  3.94
```

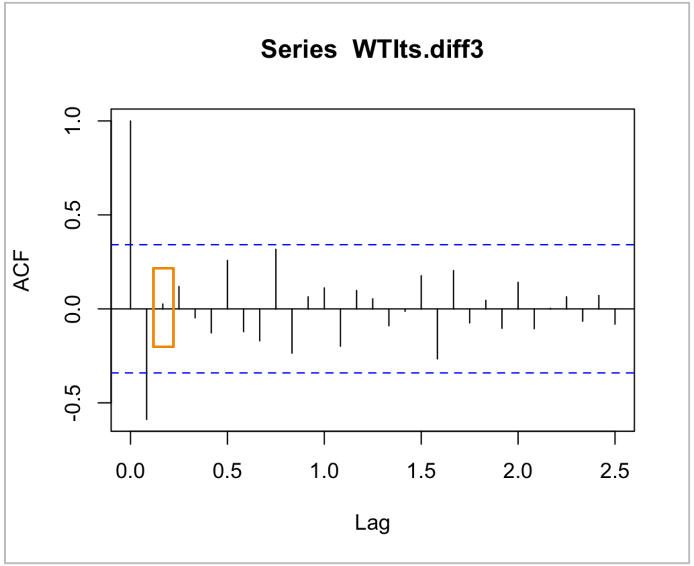
ADF result of the WTIts.diff3

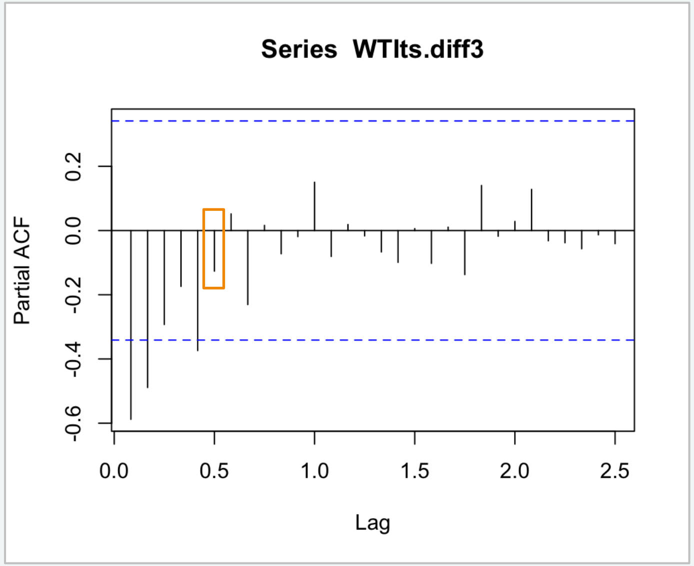
```
Value of test-statistic is: -8.1195 32.9686

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.58 -2.93 -2.60
phi1  7.06  4.86  3.94
```

WTIts.diff2 is staionary.
WTIts.diff3 is more stationary

ACF and PACF for getting q and p

```
# ACF and PACF for getting q and p
# 1차 차분한 데이터로 ARIMA 모형 확인
acf(WTIts.diff3, lag.max = 30)  # lag 2부터 점선 안에 존재. lag 절단값 = 2 --> MA(1)
pacf(WTIts.diff3, lag.max = 30) # lag 6부터 점선 안에 존재. lag 절단값 = 6 --> AR(5)
```





Compare ARIMA models

Calculated Model

```
# My ARIMA model, ARIMA(5,3,1)
# where d=3, p=5, q=1
WTIts.arima = arima(WTIts, order = c(5,3,1))
```

```
Call:
arima(x = WTIts, order = c(5, 3, 1))

Coefficients:
          ar1      ar2      ar3      ar4      ar5      ma1
      -0.6276  -0.5149  -0.3780  -0.3603  -0.2888  -1.0000
s.e.   0.1660   0.1909   0.1991   0.1830   0.1658   0.1074

sigma^2 estimated as 19.08:  log likelihood = -98.84,  aic = 211.68
```

Auto calculated Model

```
# Automatically detect optimal ARIMA parameters
# 자동으로 ARIMA 모형 확인
WTIts.autoArima = auto.arima(WTIts) # (1,0,0)
```

```
Series: WTIts
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1     mean
      0.8328  57.2836
s.e.  0.0834   3.4088

sigma^2 estimated as 15.78:  log likelihood=-100.3
AIC=206.59   AICc=207.34   BIC=211.35
```

AIC is smaller in optimal model. (the lower the better)

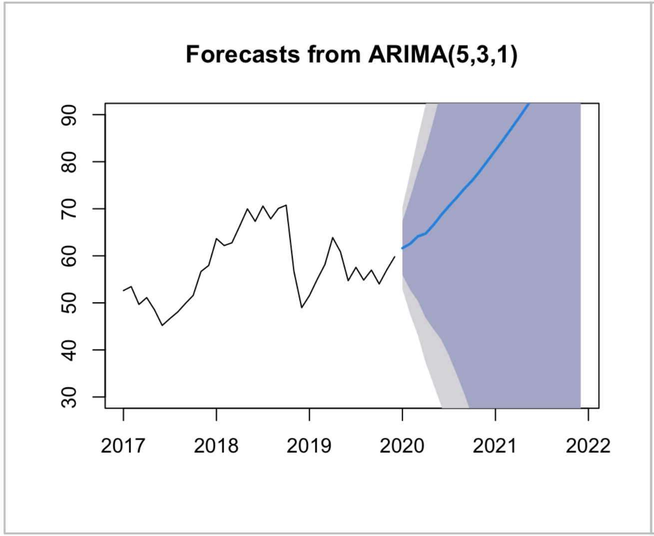
But d in optimal model is 0, which means non-stationary.

The optimal model doesn't make sense theorically.

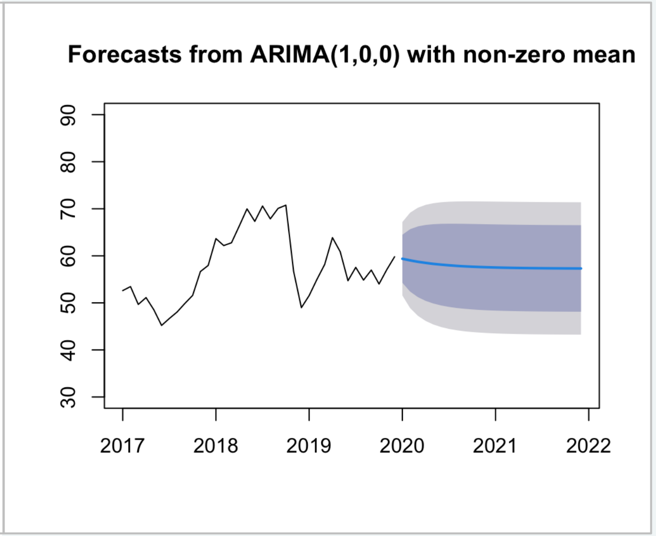
Compare ARIMA models (cont.)

```
# Predict & Plot
WTIts.autoArima.forecast = forecast(WTIts.autoArima)
WTIts.arima.forecast = forecast(WTIts.arima)
plot(WTIts.autoArima.forecast, ylim=range(30, 90))
plot(WTIts.arima.forecast, ylim=range(30, 90))
```

Calculated Model



Auto calculated Model



Conclude

$$y = 277.75 - 0.070x_1 - 0.0000014x_2 - 0.012x_3 - 0.12x_4 - 4.18x_5$$

where, y : WTI, x_1: gold price,
x_2 : domestic nuclear energy, x_3: US unemployment people,
x_4 : tesla stock price, x_5 : US natural gas price

Expect		Calculated
$\beta_1 < 0$	→	$\beta_1 < 0$
$\beta_2 > 0$		$\beta_2 < 0$
$\beta_3 < 0$		$\beta_3 < 0$
$\beta_4 < 0$		$\beta_4 < 0$
$\beta_5 > 0$		$\beta_5 < 0$

Thank you

All questions are welcome

Notice

All stuffs (R code, data and even study logs) are available in
https://github.com/nobel6018/ResourceMarketAnalysis_HYU

Raw Data

	A	B	C	D	E	F	G	H	I	J	K
1	date	WTI	worldwide_car_sale	US_trade_balance	US_unemployment_peopl	**tesla_stock_price**	US_natural_gas_price	domestic_renewable_energy	domestic_nuclerar_energy	US_interest_rate	gold_price
2	2017-01-01	52.61	17.638	-68129645	7518	50.39	3.30	3781496	26208876	0.75	1192.648
3	2017-02-01	53.46	17.776	-49946423	7399	50.00	2.85	3534270	24746976	0.75	1233.39
4	2017-03-01	49.67	17.127	-59216654	7088	55.66	2.88	3996404	26349492	1.00	1231.063
5	2017-04-01	51.12	17.285	-62609344	7085	62.81	3.10	4095016	26603076	1.00	1267.15
6	2017-05-01	48.54	17.184	-72576262	7059	68.20	3.15	4236066	27340064	1.00	1245.25
7	2017-06-01	45.20	17.184	-65319263	6933	72.32	2.98	3935694	24987716	1.25	1261.277
8	2017-07-01	46.68	17.167	-69994712	6867	64.69	2.98	3735339	24446492	1.25	1235.1
9	2017-08-01	48.06	17.014	-72509202	7097	71.18	2.90	4332538	23828874	1.25	1281.72
10	2017-09-01	49.88	18.324	-64854418	6841	68.22	2.98	4301560	24665704	1.25	1317.045
11	2017-10-01	51.59	18.321	-74943885	6599	66.31	2.88	3899918	24217574	1.25	1280.677
12	2017-11-01	56.66	17.878	-72262905	6697	61.77	3.01	4133330	22582938	1.25	1283.189
13	2017-12-01	57.95	17.714	-63809521	6561	62.27	2.82	4309096	20875666	1.50	1265.674
14	2018-01-01	63.66	17.536	-78148971	6582	70.86	3.87	4435313	19651638	1.50	1332.809
15	2018-02-01	62.18	17.557	-59527195	6641	68.61	2.67	4020960	17582516	1.50	1333.775
16	2018-03-01	62.77	17.706	-59046553	6493	53.23	2.69	4572103	18486612	1.75	1325.562
17	2018-04-01	66.33	17.794	-67562581	6418	58.78	2.80	4775906	18844984	1.75	1335.332
18	2018-05-01	69.98	17.709	-72093031	6209	56.95	2.80	4261578	22736846	1.75	1303.618
19	2018-06-01	67.32	17.650	-67656979	6519	68.59	2.97	4246790	22606842	2.00	1282.126
20	2018-07-01	70.58	17.398	-83472820	6180	59.63	2.83	4529512	26104412	2.00	1238.064
21	2018-08-01	67.85	17.314	-82851871	6167	60.33	2.96	4818010	25505100	2.00	1201.859
22	2018-09-01	70.08	17.801	-72753610	6045	52.95	3.00	4730764	22148228	2.25	1199.198
23	2018-10-01	70.76	18.142	-89362720	6123	67.46	3.28	4792956	24165742	2.25	1214.726
24	2018-11-01	56.69	17.84	-73542263	6034	70.10	4.09	4527266	24316656	2.25	1221.275
25	2018-12-01	48.98	17.95	-72659087	6286	66.56	4.04	4642189	24860948	2.50	1249.887
26	2019-01-01	51.55	17.31	-75061394	6516	61.40	3.11	5069776	24512812	2.50	1291.63
27	2019-02-01	54.98	17.10	-55386185	6181	63.98	2.69	4619620	22067436	2.50	1319.755
28	2019-03-01	58.17	17.83	-59802163	6194	55.97	2.95	5660412	28067002	2.50	1302.286
29	2019-04-01	63.87	17.16	-73888656	5850	47.74	2.65	5282161	28299140	2.50	1287.65
30	2019-05-01	60.87	17.86	-78626917	5938	37.03	2.64	6160923	29542098	2.50	1282.46
31	2019-06-01	54.71	17.74	-69384743	5985	44.69	2.40	5322698	27165594	2.50	1358.488
32	2019-07-01	57.55	17.53	-86155115	6027	48.32	2.37	5449488	23355900	2.25	1414.611
33	2019-08-01	54.84	17.60	-77162893	5999	45.12	2.22	6109594	24393832	2.25	1497.102
34	2019-09-01	56.97	17.64	-72150195	5753	48.17	2.56	4837489	20641596	2.00	1510.336
35	2019-10-01	54.01	17.27	-77024605	5857	62.98	2.33	4446194	21199706	1.75	1494.765
36	2019-11-01	57.07	17.44	-62052391	5811	65.99	2.65	3907288	20455888	1.75	1471.921
37	2019-12-01	59.80	17.29	-66532622	5753	83.67	2.22	4186870	22118336	1.75	1480.025

Data from…

https://www.opec.org/opec_web/en/publications/4054.htm - monthly oil market report 2017
https://www.opec.org/opec_web/en/publications/4814.htm - monthly oil market report 2018
https://www.opec.org/opec_web/en/publications/5392.htm - monthly oil market report 2019
https://fred.stlouisfed.org/series/TOTALSA#0 - worldwide car sales
http://stat.kita.net/stat/istat/uts/UsWholeList.screen - us trade balance
https://stats.oecd.org/index.aspx?queryid=36324 - us unemployed people
https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm - us natural gas price
https://kosis.kr/index/index.do - domestic nuclear, renewable energy generation
https://kr.investing.com/economic-calendar/interest-rate-decision-168 - us interest rate
https://datahub.io/core/gold-prices#data - gold price