



# Research of factors that influence on oil prices

## Resource Market Analysis

Kim JaeHun  
Hwang Hyeju  
Lee YoungHoon

# 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

# Importance

---



Energy



Clothes



Fuel

# War For Oil

---



# Supply & Demand

## Decreasing Demand

- Covid-19 Economic
- Travel restrict



## Increasing Supply

- 2020 OPEC–Russia oil price war

# WTI

**Min** 45.20

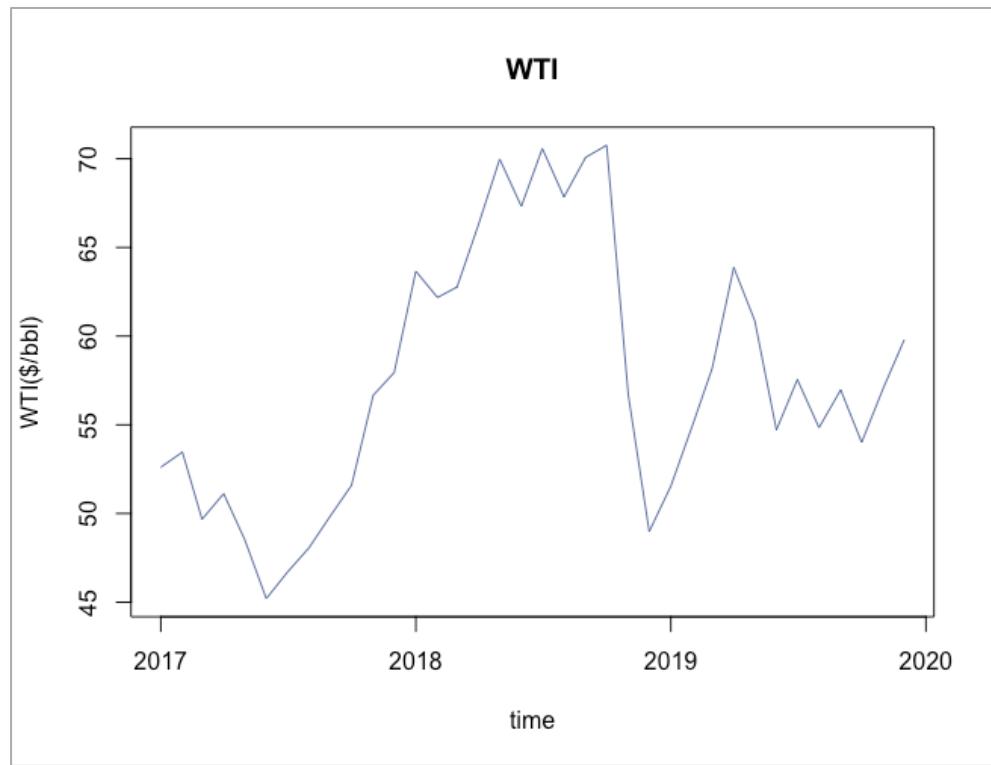
**Max** 70.76

**Mean** 57.58

**Median** 56.83

**Variance** 54.40

**Standard Deviation** 7.38



# Worldwide Car Sale

**Min** 17.01

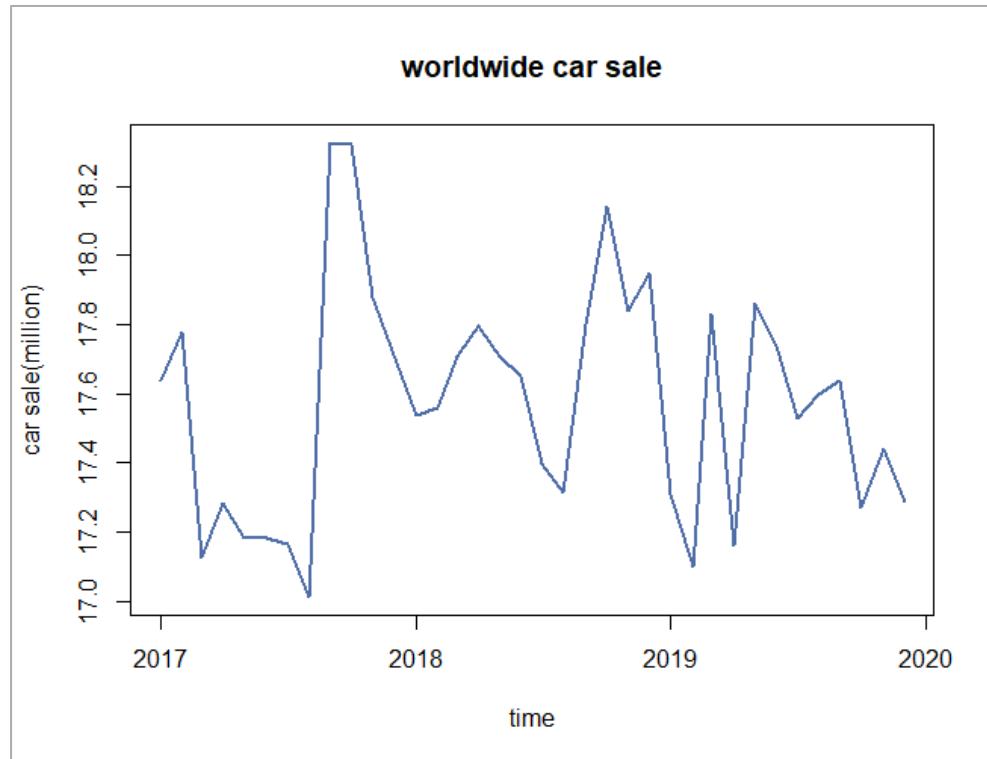
**Max** 18.32

**Mean** 17.58

**Median** 17.62

**Variance** 0.11

**Standard Deviation** 0.34



# US Trade Balance

**Min** -89,362,720

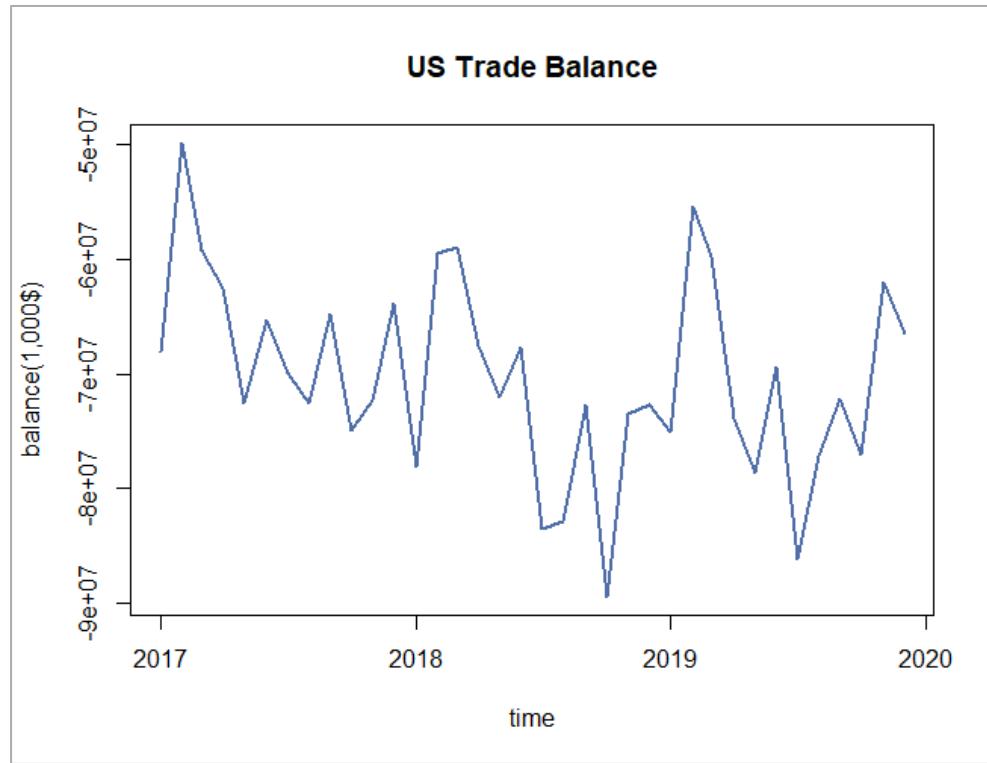
**Max** -49,946,423

**Mean** -70,224,383

**Median** -72,121,613

**Variance**  $7.54 \times 10^{13}$

**Standard Deviation** 8,683,593



# US Unemployment People

**Min** 5,753

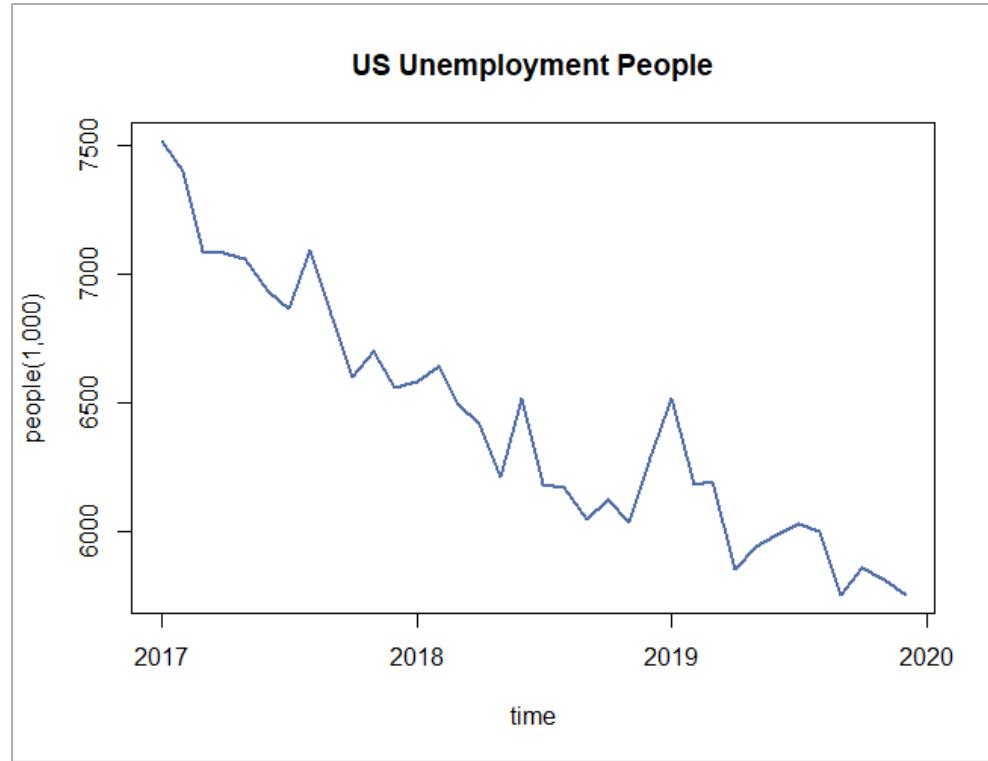
**Max** 7,518

**Mean** 6,425

**Median** 6,352

**Variance** 231,782

**Standard Deviation** 481



# Tesla Stock Price

**Min** 37.03

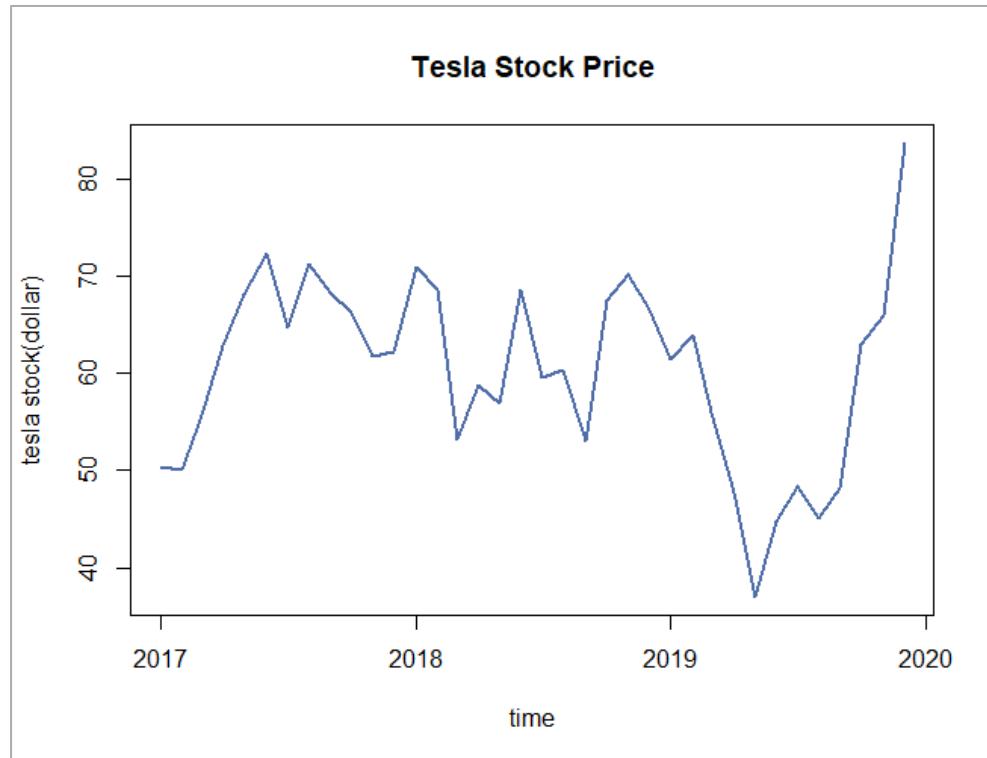
**Max** 83.67

**Mean** 60.36

**Median** 62.02

**Variance** 94.81

**Standard Deviation** 9.74



# US Natural Gas Price

**Min** 2.22

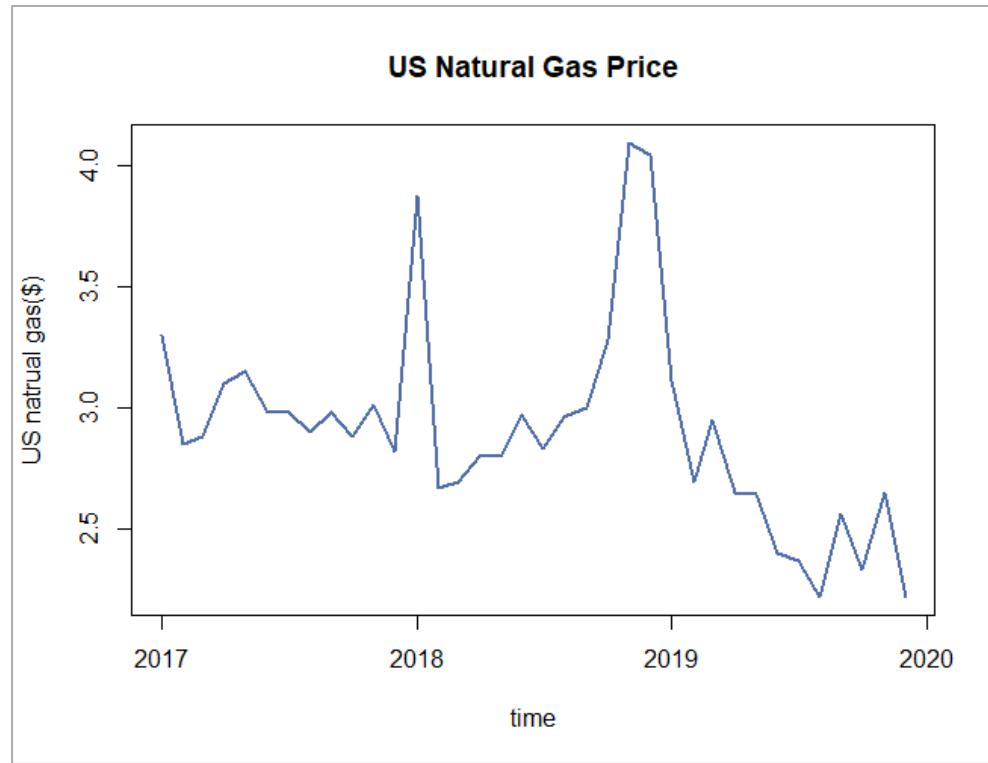
**Max** 4.09

**Mean** 2.91

**Median** 2.88

**Variance** 0.18

**Standard Deviation** 0.43



# Domestic Renewable Energy Generation

**Min** 3,534,270

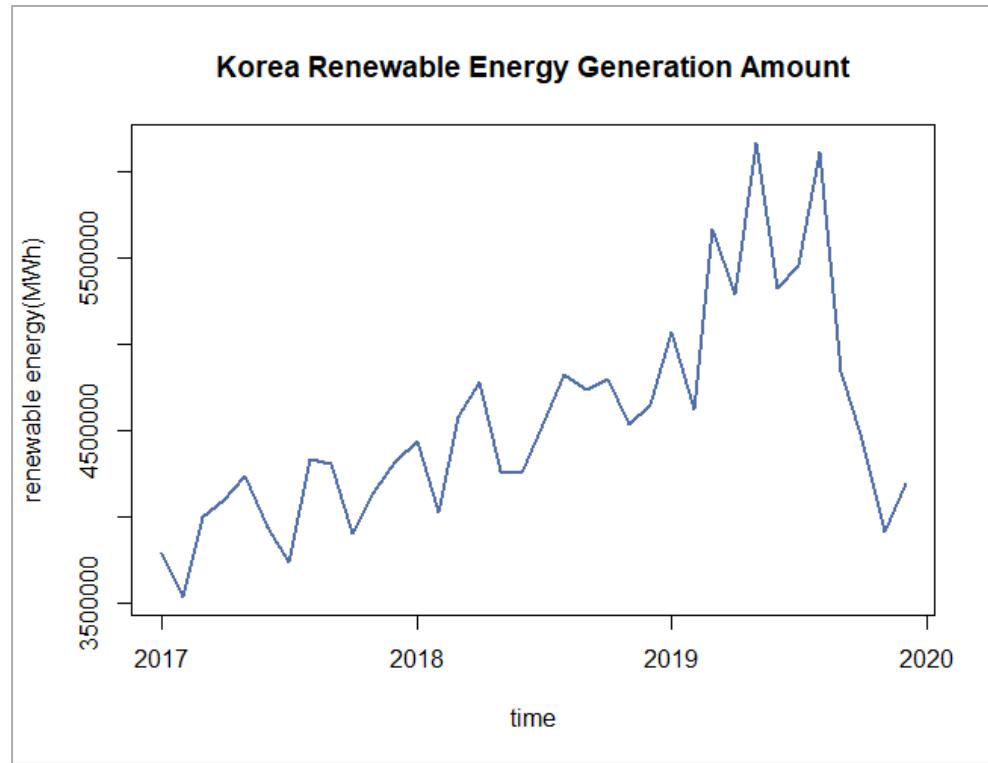
**Max** 6,160,923

**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

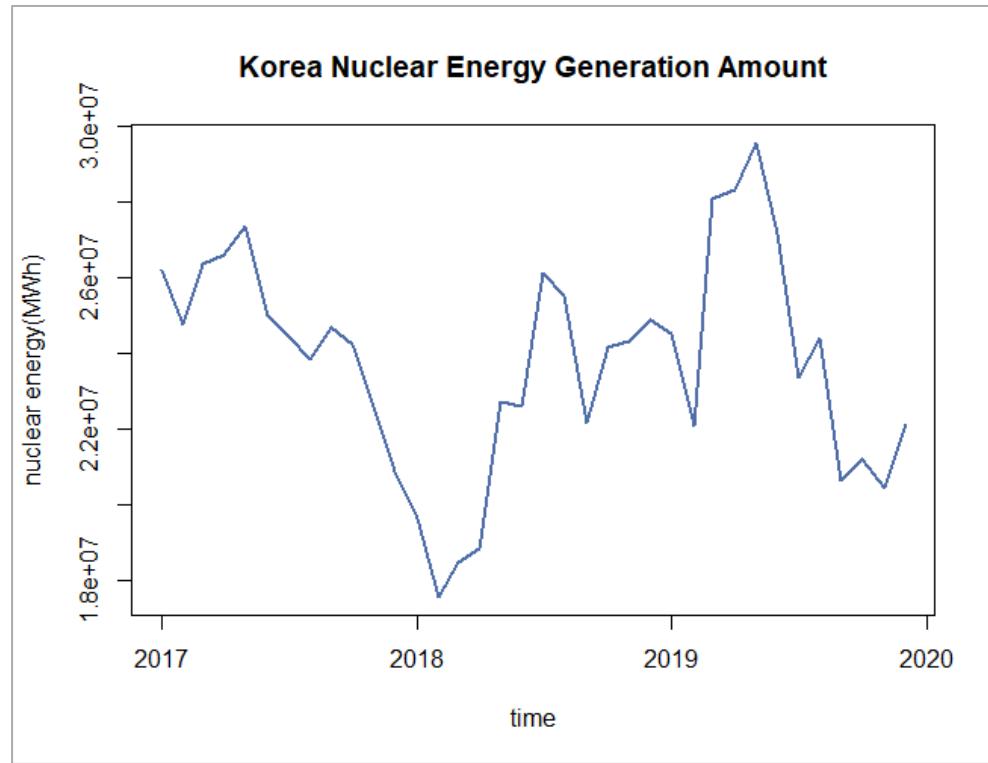
**Max** 29,542,098

**Mean** 23,768,981

**Median** 24,267,115

**Variance**  $8.22 \times 10^{12}$

**Standard Deviation** 2,866,422



# US Interest Rate

**Min** 0.75

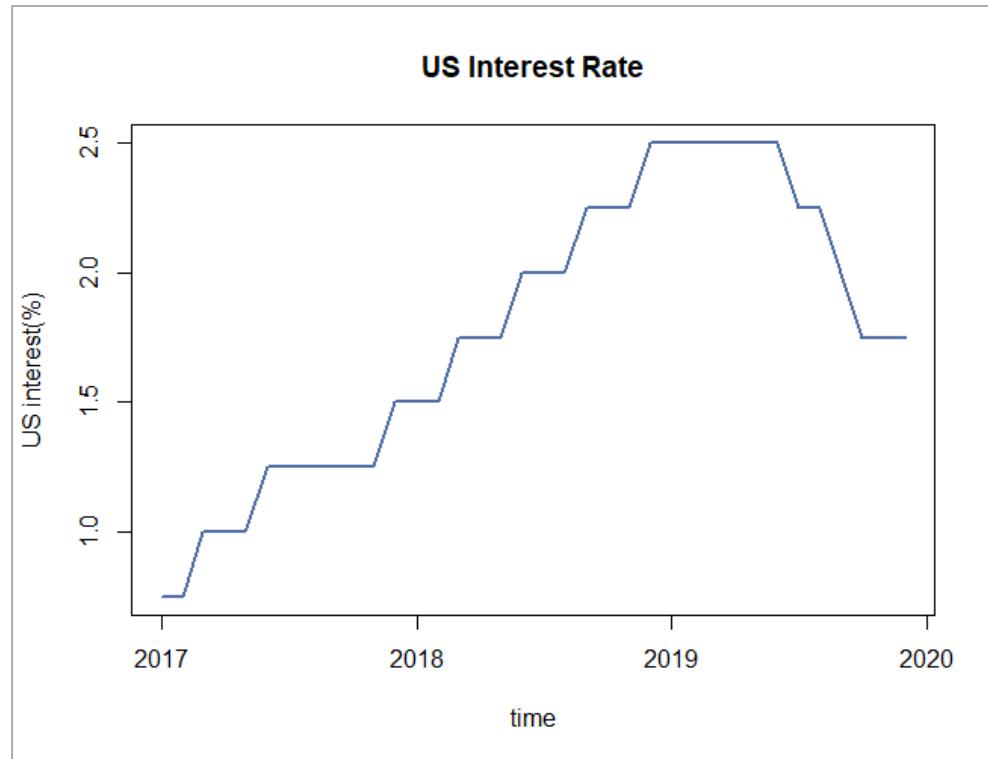
**Max** 2.50

**Mean** 1.77

**Median** 1.75

**Variance** 0.31

**Standard Deviation** 0.56



# Gold Price

**Min** 1,193

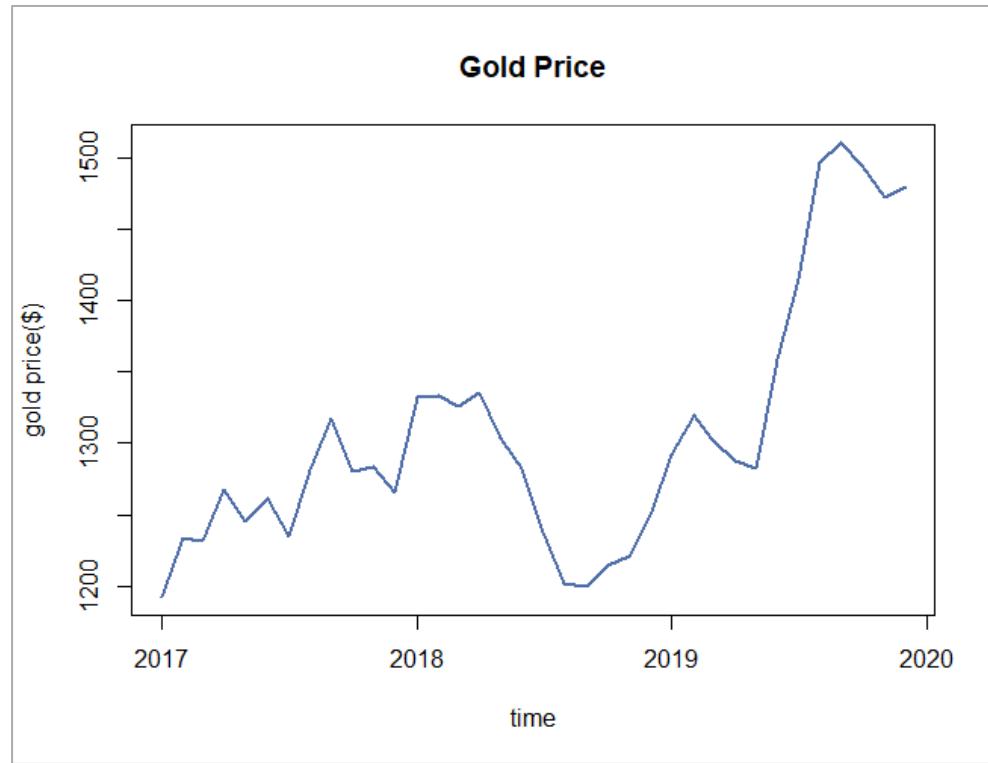
**Max** 1,510

**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_natural_gas_price + domestic_renewable_energy + domestic_nuclearar_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_nuclearar_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_nuclearar_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_nuclearar_energy+US_unemployment_people, data=data)
>
> summary(model)

Call:
lm(formula = WTI ~ gold_price + domestic_nuclearar_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_nuclearar_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_nuclearar_energy+US_unemployment_people+tesla_stock_
> summary(model)

Call:
lm(formula = WTI ~ gold_price + domestic_nuclearar_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_nuclearar_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

$$\text{model} = \text{lm}(\text{WTI} \sim \text{gold\_price} + \text{domestic\_nuclear\_energy} + \text{US\_unemployment\_people} + \text{tesla\_stock\_price} + \text{US\_natural\_gas\_price}, \text{data}=\text{data})$$

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

	WTI	US_unemployment_people	tesla_stock_price	US_natural_gas_price	domestic_nuclerar_energy	gold_price
WTI	1.0000000	-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

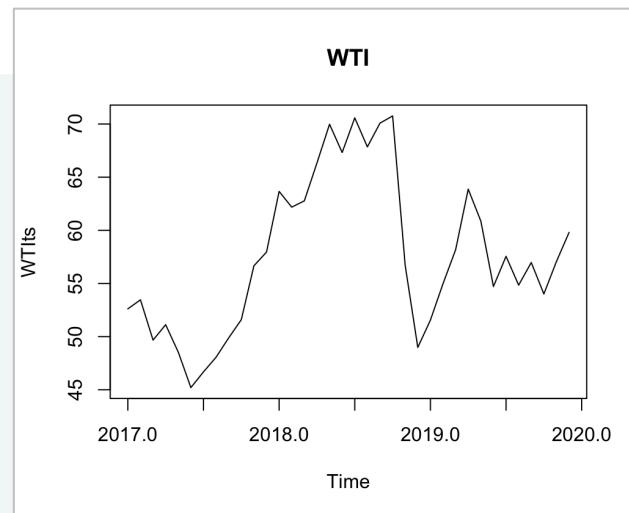
→ cannot reject null hypothesis

The model has normlaity.

# WTI in time series

Set the WTI time series variable

```
> # Set time series variable  
> WTIt = ts(data=data$WTI, start=c(2017,01), end=c(2019,12), frequency = 12)  
> WTIt  
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



```
# Difference to make it stationary
# 차분을 통해 데이터 정상화(stationary)
WTIts.diff1 = diff(WTIts, differences = 1)
WTIts.diff2 = diff(WTIts, differences = 2)
WTIts.diff3 = diff(WTIts, differences = 3)

# We can see the stationary pattern
# 1차 차분을 통해 어느 정도 정상화(stationary) 패턴을 보임
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(WTIt.s.diff2, lags=1, type='drift')) # stationary  
summary(ur.df(WTIt.s.diff3, lags=1, type='drift')) # more stationary
```

### ADF result of the WTIt.s.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
```

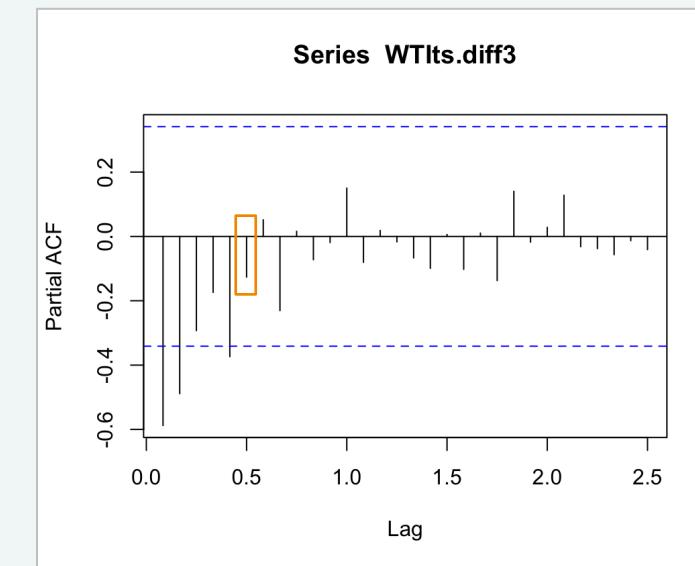
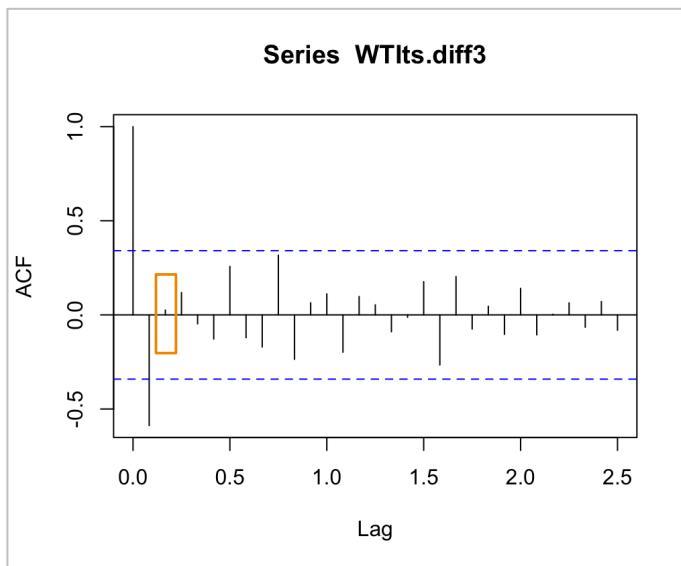
### ADF result of the WTIt.s.diff3

```
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
```

WTIt.s.diff2 is **stationary**.  
WTIt.s.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 절선 안에 존재. lag 절단값 = 2 --> MA(1)
pacf(WTIts.diff3, lag.max = 30) # lag 6부터 절선 안에 존재. lag 절단값 = 6 --> AR(5)
```



# Compare ARIMA models

## Calculated Model

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

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

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1
    -0.3999  -0.0311  -0.0749  -0.1488  -0.1773  0.5230
  s.e.  0.5396   0.1914   0.1821   0.1744   0.1780  0.5274
sigma^2 estimated as 15.33: log likelihood = -97.53,  aic = 209.06
```

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 theoretically.

## Auto calculated Model

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

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

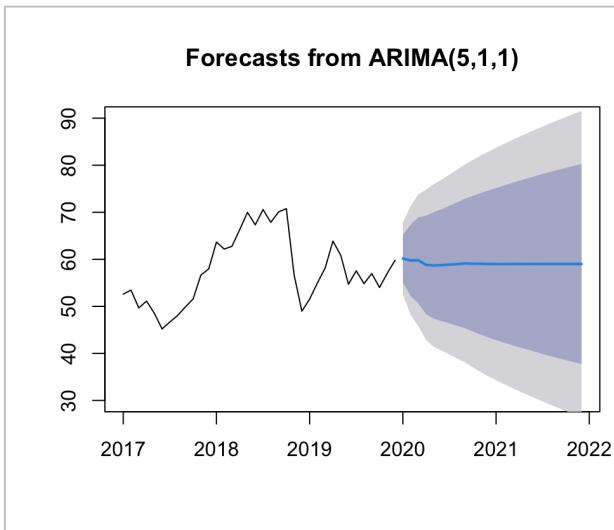
```
Coefficients:
      ar1      mean
    0.8328  57.2836
  s.e.  0.0834   3.4088
```

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

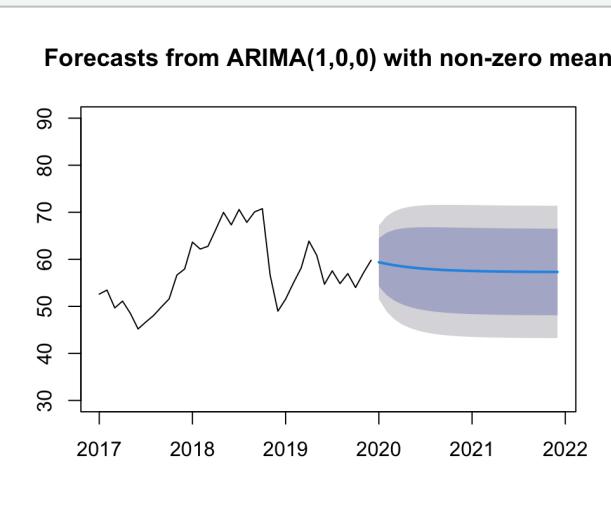
## Compare ARIMA models (cont.)

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

Calculated Model



Auto calculated Model



1134

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](https://github.com/nobel6018/ResourceMarketAnalysis_HYU)



Data from...

- [https://www.opec.org/opec\\_web/en/publications/4054.htm](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](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](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