Application of Generalized Additive Model in IHSG Data Prediction

Kadek Ardya Novi Diani

09211850096004

Table of content

Data Structure

This study is utilizing 9 variables data, which consist of 1 dependent variable, and 8 independent variables.

The dependent variable is IHSG (Index Harga Saham Gabungan), monthly closing data from January 2010-Dec 2018.

Independent variables are listed below:

- Crude oil price
- Coal price
- Coffee commodity price
- Crude palm oil (CPO) price
- Rubber commodity price
- Gold price
- Exchange rate of IDR to USD
- Indonesia's inflation rate

All independent variables also using the same time frame with the dependent variables.

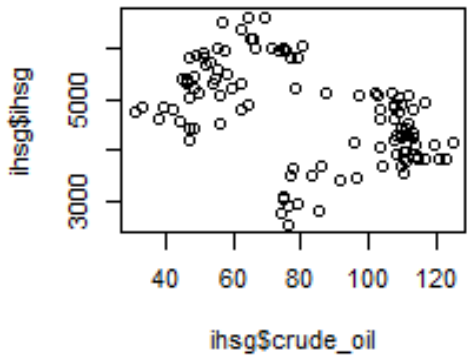
Statistics Descriptive

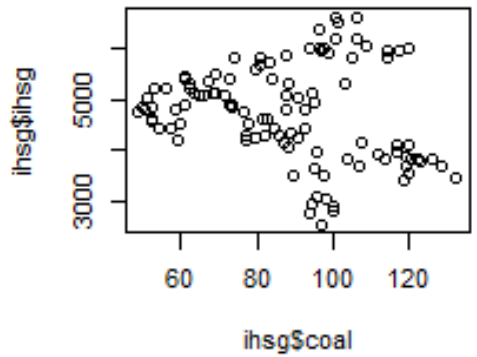
Statistics

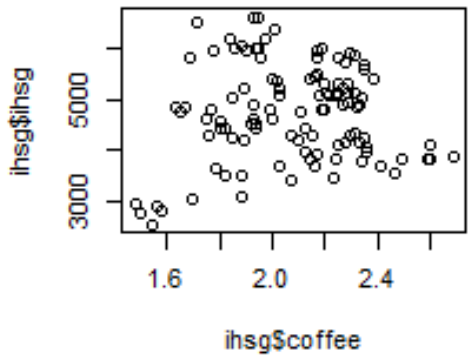
Variable	Total Count	N	N*	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum
IHSG	108	108	0	4738.8	90.1	936.9	877688.5	2549.0	4075.6	4799.7	5386.5	6605.6
Crude Oil	108	108	0	81.36	2.59	26.94	725.52	30.80	55.56	78.12	109.41	124.93
COAL	108	108	0	88.01	2.06	21.46	460.50	49.02	71.79	88.60	102.92	132.48
COFFEE	108	108	0	2.0768	0.0250	0.2600	0.0676	1.4800	1.8825	2.1300	2.2800	2.6900
CPO	108	108	0	832.5	16.5	171.5	29429.1	535.0	709.1	802.5	912.0	1292.0
RUBBER	108	108	0	2.593	0.112	1.167	1.361	1.230	1.652	2.220	3.292	6.260
GOLD	108	108	0	1341.9	17.7	184.4	34019.8	1075.7	1220.9	1281.7	1408.2	1772.1
EX_RATE	108	108	0	11574	201	2085	4348407	8508	9343	12030	13384	15227
Inflasi	108	108	0	5.010	0.159	1.651	2.725	2.790	3.603	4.515	6.320	8.790

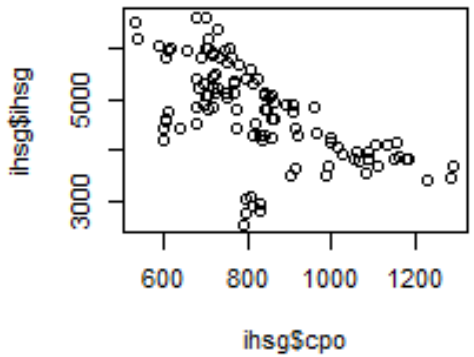
Statistics Descriptive

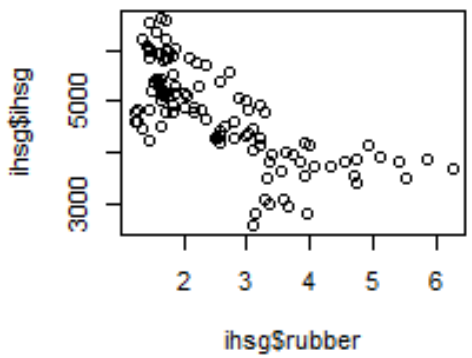
Scatter Plot of IHSG and other variables

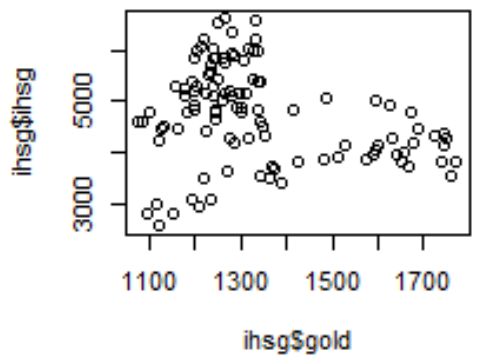


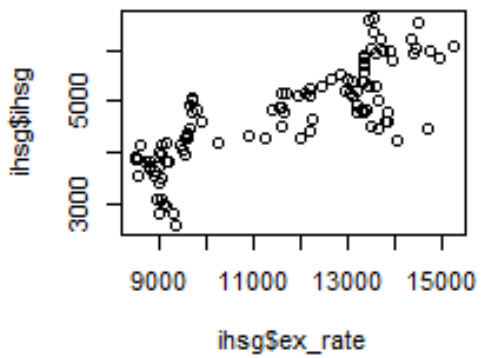


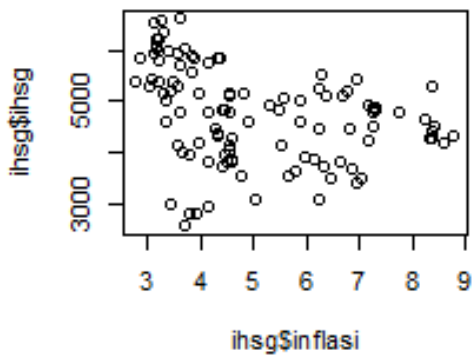






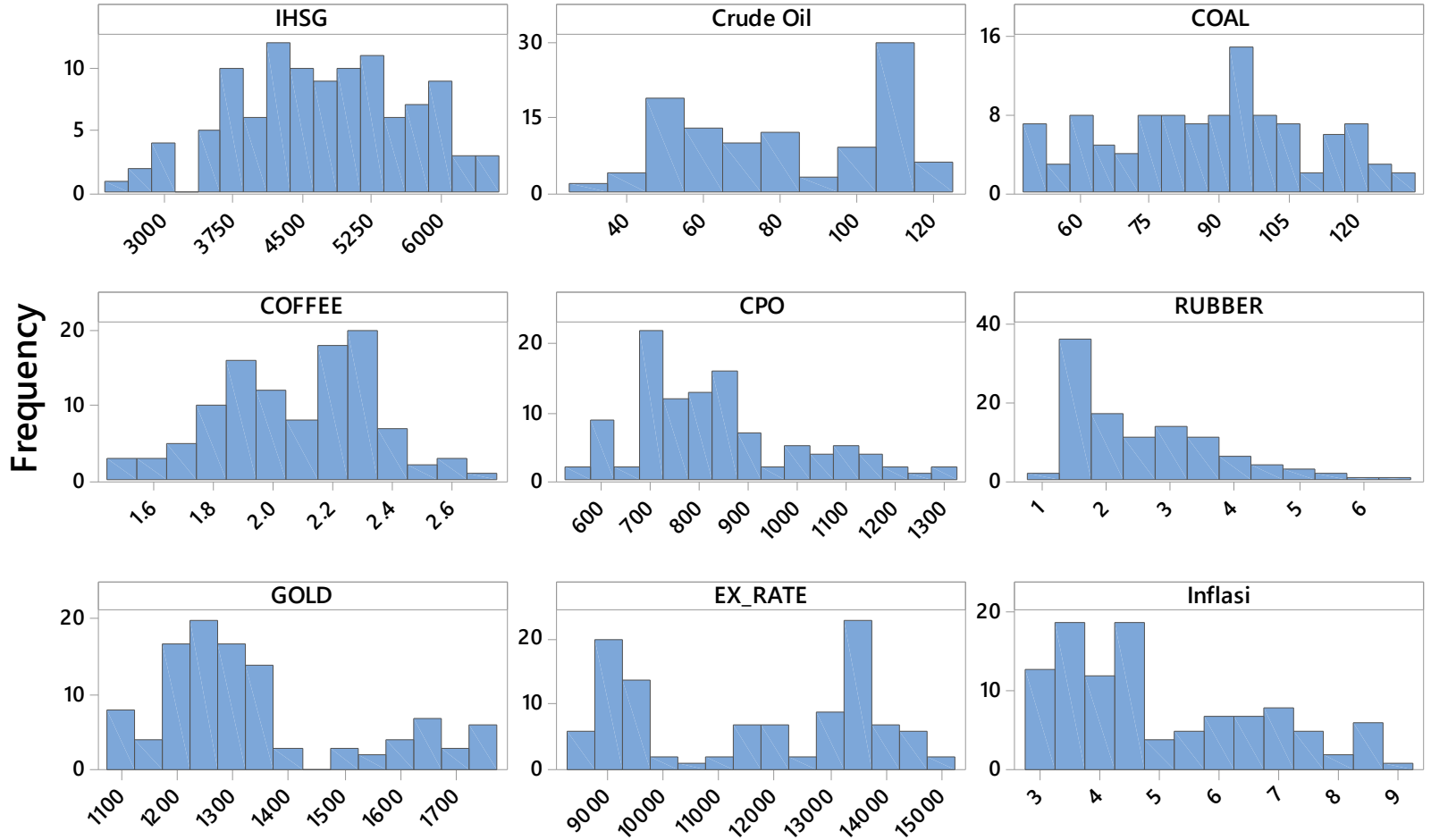






Statistics Descriptive

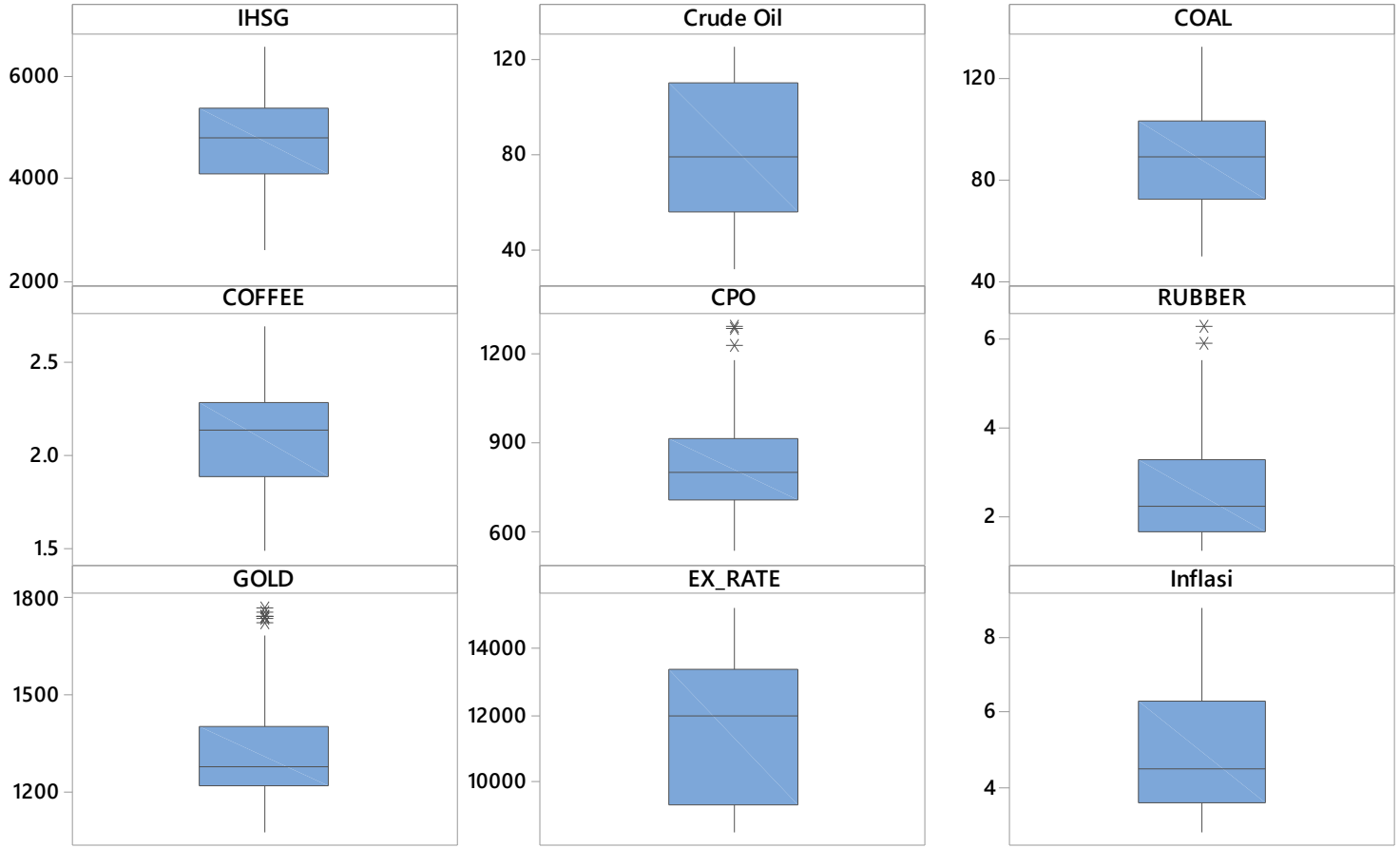
Histoaram of each variables



Histogram graph shows the distribution of each variables

Statistics Descriptive

Boxplot of each variables



Boxplot shows that there is some outlier data in variable CPO, rubber, and gold price.

GAM Modeling

Linearity test

To check the linearity pattern between IHSG and other variables

```
        RESET test

data:  ihsg$ihsg ~ ihsg$crude_oil
RESET = 0.21461, df1 = 1, df2 = 105, p-value = 0.6441
```
→ **linear**

```
data:  ihsg$ihsg ~ ihsg$coffee
RESET = 26.955, df1 = 1, df2 = 105, p-value = 1.025e-06
```
→ **nonlinear**

```
data:  ihsg$ihsg ~ ihsg$cpo
RESET = 2.1448, df1 = 1, df2 = 105, p-value = 0.146
```
→ **linear**

```
data:  ihsg$ihsg ~ ihsg$rubber
RESET = 11.769, df1 = 1, df2 = 105, p-value = 0.000862
```
→ **nonlinear**

```
data:  ihsg$ihsg ~ ihsg$gold
RESET = 10.813, df1 = 1, df2 = 105, p-value = 0.001372
```
→ **nonlinear**

```
data:  ihsg$ihsg ~ ihsg$ex_rate
RESET = 1.5624, df1 = 1, df2 = 105, p-value = 0.2141
```
→ **linear**

```
data:  ihsg$ihsg ~ ihsg$inflasi
RESET = 16.011, df1 = 1, df2 = 105, p-value = 0.0001175
```
→ **nonlinear**

H0: data is linear
H1: data is nonlinear

p-value < alpha 0.05 → reject H0

GAM Modeling

Input all external variables

```
Formula:
ihsg$ihsg ~ (ihsg$crude_oil + ihsg$coal + ihsg$coffee + ihsg$cpo +
    ihsg$rubber + ihsg$gold + ihsg$ex_rate + ihsg$inflasi)

Parametric coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.049e+03  8.091e+02  -2.532 0.012906 *
ihsg$crude_oil  1.872e+00  2.667e+00   0.702 0.484558
ihsg$coal       1.205e+01  3.018e+00   3.993 0.000126 ***
ihsg$coffee     1.341e+03  1.855e+02   7.229 1.03e-10 ***
ihsg$cpo       -8.482e-01  5.366e-01  -1.581 0.117111
ihsg$rubber    -3.318e+02  1.099e+02  -3.019 0.003225 **
ihsg$gold       6.632e-01  3.528e-01   1.880 0.063060 .
ihsg$ex_rate    3.076e-01  4.832e-02   6.365 6.12e-09 ***
ihsg$inflasi   -1.845e+01  3.085e+01  -0.598 0.551077
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


R-sq.(adj) =   0.86   Deviance explained = 87.1%
GCV = 1.3384e+05  Scale est. = 1.2268e+05  n = 108
```

By using alpha 0.05, some variables are not significantly impact to predict IHSG. They are crude oil price, CPO price, gold price, and inflation.

These variables will be taken out from the model to have a better model.

GAM Modeling

Exclude nonsignificant variables

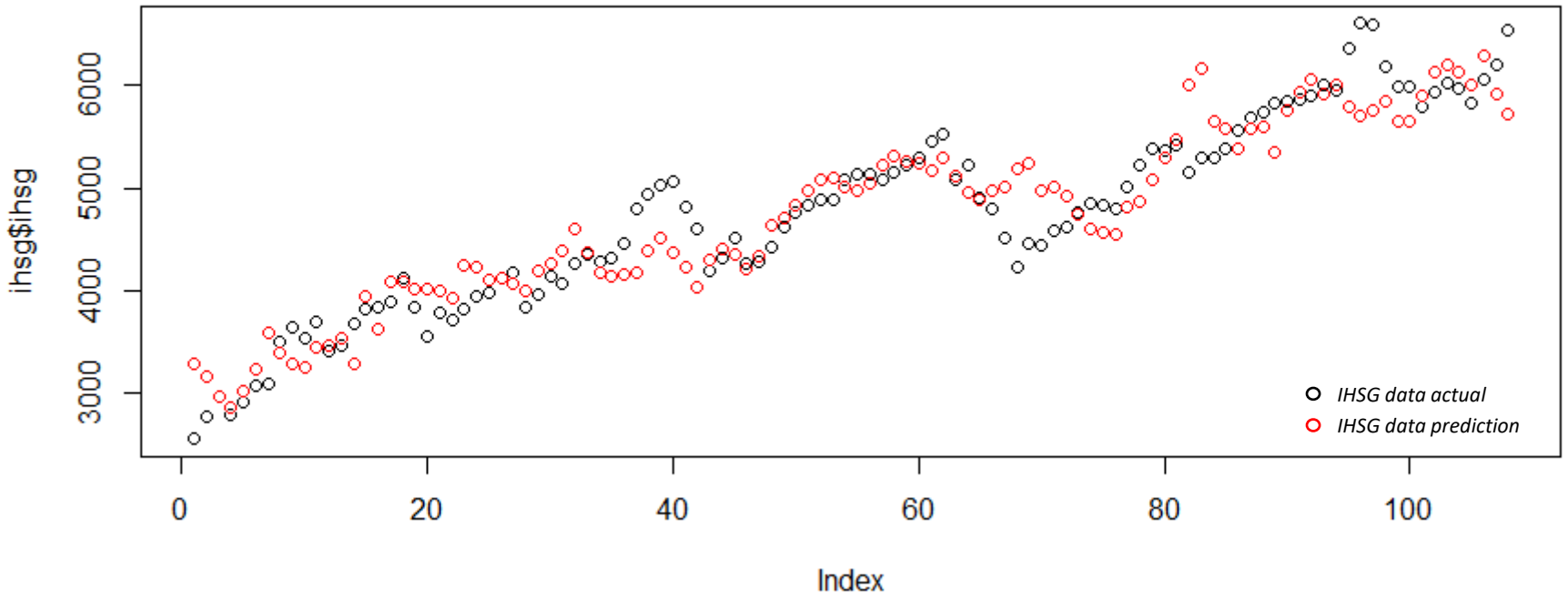
```
Formula:
ihsg$ihsg ~ (ihsg$coal + ihsg$coffee + ihsg$rubber + ihsg$ex_rate)

Parametric coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.231e+03  5.970e+02  -2.063   0.0417 *
ihsg$coal     1.610e+01  2.317e+00   6.947 3.45e-10 ***
ihsg$coffee   1.477e+03  1.473e+02  10.028  < 2e-16 ***
ihsg$rubber  -5.214e+02  8.245e+01  -6.323 6.69e-09 ***
ihsg$ex_rate  2.452e-01  3.766e-02   6.510 2.79e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


R-sq.(adj) =  0.854   Deviance explained = 85.9%
GCV = 1.3437e+05  Scale est. = 1.2815e+05  n = 108
```

All variables that is included in the model are significantly impact to predict IHSG.

Compared to previous model, GCV of this model is better.

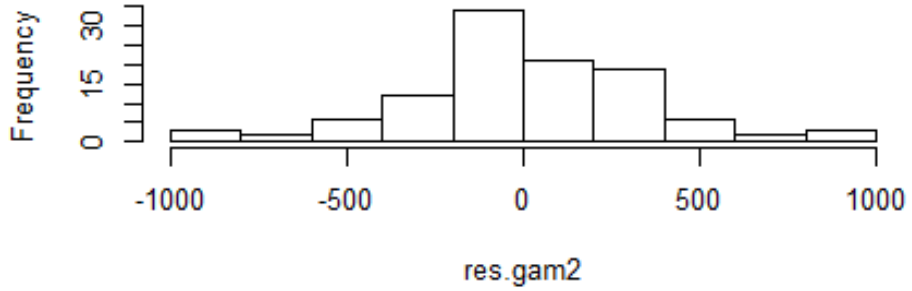
GAM Modeling

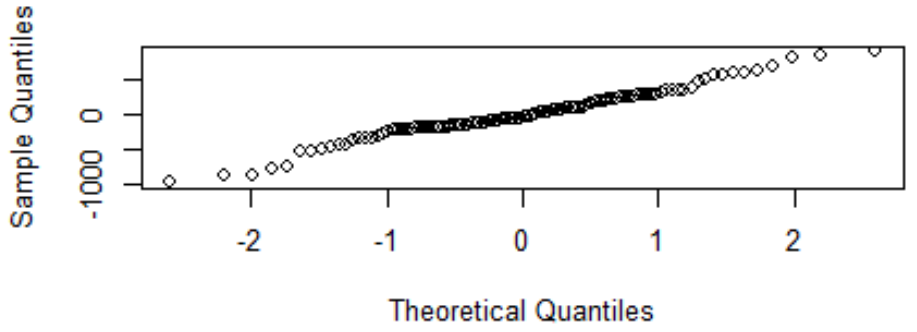
IHSG Prediction



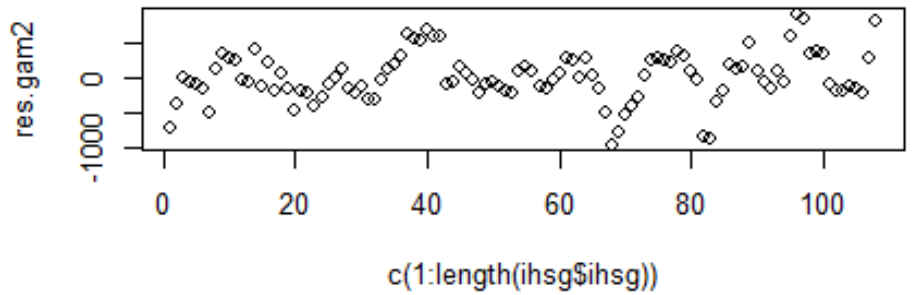
MAPE of this prediction is 5.8%, or around ±275 points

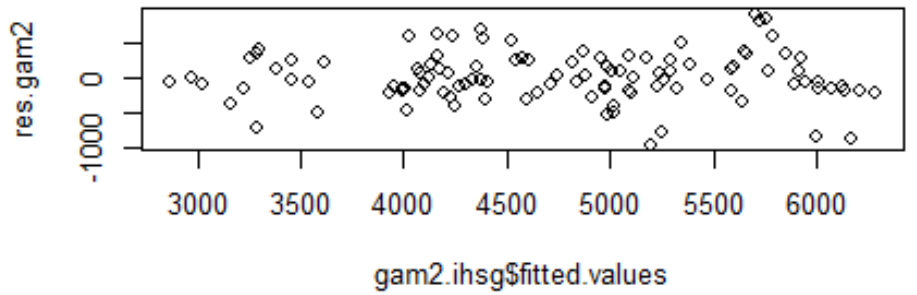
Residuals Checking





These four graphs shows that the residual of this model has fulfilled assumption of identic, independent, & normal distributed (IIDN)





Thank you