Week3Assignment

Pradeepta Das

16 November 2020

This test exercise is of an applied nature and uses data that are available in the data file TestExer3. We consider the so-called Taylor rule for setting the (nominal) interest rate. This model describes the level of the nominal interest rate that the central bank sets as a function of equilibrium real interest rate and inflation, and considers the current level of inflation and production. The model is:

$$i_t = r^* + \pi_t + 0.5(\pi_t - \pi^*) + 0.5g_t$$

with i_t the Federal funds target interest rate at time t, r^* the equilibrium real federal funds rate, π_t a measure of inflation, π^* the target inflation rate and g_t the output gap (how much actual output deviates from potential output).

We simplify the Taylor rule in two manners.

First, we avoid determining r^* and π^* and simply add an intercept to the model to capture these two variables (and any other deviations in the means).

Second, we consider production y_t rather than the output gap.

In this form the Taylor rule is

$$i_t = \beta_1 + \beta_2\pi_t + \beta_3 y_t + \epsilon_t$$

Monthly data are available for the USA over the period 1960 through 2014 for the following variables: • INTRATE: Federal funds interest rate

- INFL: Inflation
- PROD: Production
- UNEMPL: Unemployment
- COMMPRI: Commodity prices
- PCE: Personal consumption expenditure
- PERSINC: Personal income
- HOUST: Housing starts

A sneak-peek into the data:

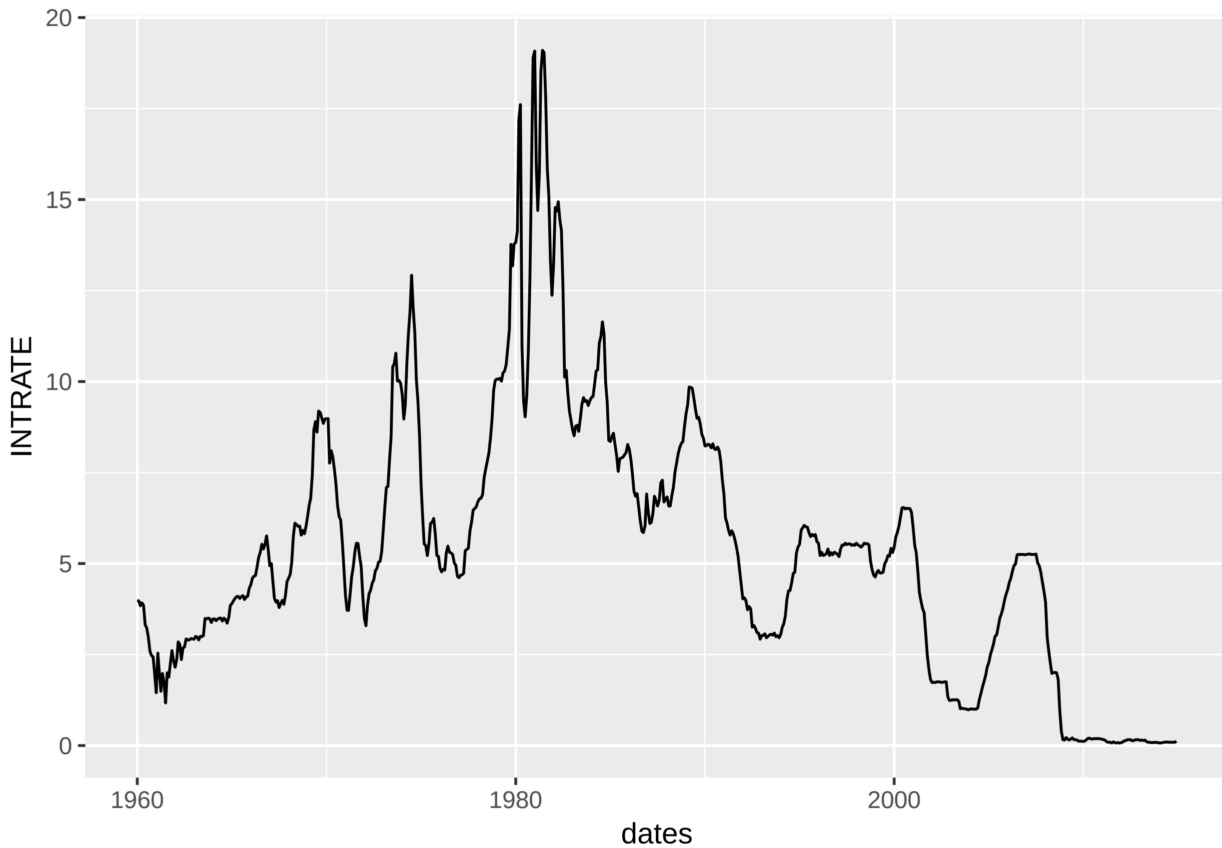
```
tail(data)
```

```
##          OBS INTRATE    INFL    PROD  UNEMPL   COMMPRI     PCE PERSINC
## 655  2014:7    0.09 1.99655 5.05203 2.00632  21.42857 4.00272 1.57378
## 656  2014:8    0.09 1.71141 4.35206 1.97110   7.40741 4.39035 1.51486
## 657  2014:9    0.09 1.66422 4.53604 2.04202   5.30973 4.19565 1.29569
## 658 2014:10    0.09 1.65111 4.40681 2.03575  -3.60360 4.17424 1.88465
## 659 2014:11    0.09 1.28144 5.18284 2.10822 -15.23810 4.05860 2.21122
## 660 2014:12    0.12 0.66285 4.87824 2.26658 -28.03738 3.58454 2.98341
##        HOUST      dates
## 655 22.27171 2014-07-01
## 656  8.81356 2014-08-01
## 657 19.11935 2014-09-01
## 658 16.66667 2014-10-01
## 659 -5.61086 2014-11-01
```

```
## 660  5.31915 2014-12-01
```

Visualize the interest rate evolution with time:

```r
data%>% ggplot(aes(dates,INTRATE)) + geom_line()
```



(a) Use general-to-specific to come to a model. Start by regressing the federal funds rate on the other 7 variables and eliminate 1 variable at a time.

you start with the most general model, including as many variables as are at hand. Then, check whether one or more variables can be removed from the model. This can be based on individual t-tests, or a joint F-test in case of multiple variables. In case you remove one variable at a time, the variable with the lowest absolute t-value is removed from the model. The model is estimated again without that variable, and the procedure is repeated. The procedure continues until all remaining variables are significant.

So next we need to fit the model:

```r
model <- lm(INTRATE ~ INFL + PROD + UNEMPL + COMMPRI + PCE + PERSINC + HOUST, data = data)
print(summary(model))
```

```
##
## Call:
## lm(formula = INTRATE ~ INFL + PROD + UNEMPL + COMMPRI + PCE +
##     PERSINC + HOUST, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4066 -1.4340 -0.1175  1.3555  7.7386
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.221161   0.244995  -0.903   0.3670
## INFL         0.696059   0.062229  11.185  < 2e-16 ***
## PROD        -0.057743   0.039900  -1.447   0.1483
## UNEMPL       0.102481   0.096757   1.059   0.2899
## COMMPRI     -0.005521   0.002974  -1.857   0.0638 .
## PCE          0.344380   0.069455   4.958 9.08e-07 ***
## PERSINC      0.246999   0.060590   4.077 5.13e-05 ***
## HOUST       -0.019411   0.004672  -4.155 3.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.188 on 652 degrees of freedom
## Multiple R-squared:  0.6385, Adjusted R-squared:  0.6346
## F-statistic: 164.5 on 7 and 652 DF,  p-value: < 2.2e-16
```

```r
print(paste("AIC:" ,AIC(model)," and BIC :", BIC(model)))
```

```
## [1] "AIC: 2916.30140865829  and BIC : 2956.73156717348"
```

The variable with the least explanatory power, based on p-value, is Unemployment, so is necessary to eliminate it and create a second model that excludes it:

```r
model <- lm(INTRATE ~ INFL + PROD + COMMPRI + PCE + PERSINC + HOUST, data = data)
print(summary(model))
```

```
##
## Call:
## lm(formula = INTRATE ~ INFL + PROD + COMMPRI + PCE + PERSINC +
##     HOUST, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5322 -1.4982 -0.1005  1.3882  7.6954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.290851   0.236016  -1.232   0.2183
## INFL         0.693309   0.062180  11.150  < 2e-16 ***
## PROD        -0.025460   0.025752  -0.989   0.3232
## COMMPRI     -0.006514   0.002822  -2.308   0.0213 *
## PCE          0.368561   0.065602   5.618 2.86e-08 ***
## PERSINC      0.251581   0.060441   4.162 3.57e-05 ***
## HOUST       -0.021023   0.004417  -4.760 2.39e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.188 on 653 degrees of freedom
## Multiple R-squared:  0.6379, Adjusted R-squared:  0.6346
## F-statistic: 191.7 on 6 and 653 DF,  p-value: < 2.2e-16
```

```r
print(paste("AIC:" ,AIC(model)," and BIC :", BIC(model)))
```

```
## [1] "AIC: 2915.43601815001  and BIC : 2951.37393683018"
```

Also Production has high p-value so we remove it for our third and final round. In fact all remaining variables has absolute t-values above 2, with p-values below 0.05 so they are significant:

```
model <- lm(INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST, data = data)
print(summary(model))
```

```
##
## Call:
## lm(formula = INTRATE ~ INFL + COMMPRI + PCE + PERSINC + HOUST,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1631 -1.5244 -0.1125  1.3715  7.6725
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.240119   0.230366  -1.042  0.29764
## INFL         0.717527   0.057152  12.555  < 2e-16 ***
## COMMPRI     -0.007501   0.002640  -2.841  0.00464 **
## PCE          0.340525   0.059156   5.756 1.32e-08 ***
## PERSINC      0.240242   0.059342   4.048 5.77e-05 ***
## HOUST       -0.020530   0.004389  -4.678 3.52e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.188 on 654 degrees of freedom
## Multiple R-squared:  0.6374, Adjusted R-squared:  0.6346
## F-statistic: 229.9 on 5 and 654 DF,  p-value: < 2.2e-16
```

```
print(paste("AIC:" ,AIC(model)," and BIC :", BIC(model)))
```

```
## [1] "AIC: 2914.42324696844  and BIC : 2945.86892581358"
```

AIC, BIC values seem different (may be different scaling). So, by implementing them:

```
s <- sqrt(deviance(model)/df.residual(model))
k <- length(model$coefficients) - 1
n <- nrow(data)
AIC <- log(s^2) + 2 * k / n
BIC <- log(s^2) + k * log(n) / n
print(round(c(AIC, BIC), 4))
```

```
## [1] 1.581 1.615
```

R^2 seems good at 63.7%.

Conclusion:

Variables elimination after regression, one at a time, produced the following results: After regression round #1: UNEMPL variable was chosen to be removed, After regression round #2: PROD variable was chosen to be removed, After regression round #3: all remaining variables were found to be significant; variables removal stops here.

(b) Use specific-to-general to come to a model. Start by regressing the federal funds rate on only a constant and add 1 variable at a time. Is the model the same as in (a)?

Definition: The specific to general approach follows the same logic [as the general to specific approach], but starts with a very small model, sometimes even only consisting of the constant term. Variables get added one at a time, choosing the one that has the largest absolute t-statistic. This procedure is repeated until no significant variables can be added anymore.

First we start with a model containing only an intercept(the mean average Interest Rate over all time periods), then, thanks to AIC y BIC we added Inflation and continuing with this process is necessary to add Personal Income, then Personal Expenditure, Housing Starts and Commodity Prices.

Also as you know, Unemployment and Production get the AIC higher so are excluded.So the final model shows us:

```
lm(formula = INTRATE ~ INFL + PCE + HOUST + PERSINC + COMMPRI, data = data)
```

```
##
## Call:
## lm(formula = INTRATE ~ INFL + PCE + HOUST + PERSINC + COMMPRI,
##     data = data)
##
## Coefficients:
## (Intercept)         INFL          PCE        HOUST      PERSINC
##   -0.240119     0.717527     0.340525    -0.020530     0.240242
##     COMMPRI
##   -0.007501
```

```
s <- sqrt(deviance(model)/df.residual(model))
k <- length(model$coefficients) - 1
n <- nrow(data)
AIC <- log(s^2) + 2 * k / n
BIC <- log(s^2) + k * log(n) / n
print(round(c(AIC, BIC), 4))
```

```
## [1] 1.581 1.615
```

Variables inclusion (one at a time) after regressions (all possible combinations), produced the following results:

After regressions set round #1 (with each candidate variable, individually): INFL variable was chosen to be included,

After regressions set round #2 (with INFL and each other candidate variable combination): PERSINC variable was chosen to be included as well,

After regressions set round #3 (with INFL, PERSINC and each other candidate variable combination): PCE variable was chosen to be included as well,

After regressions set round #4 (with INFL, PERSINC, PCE and each other candidate variable combination): HOUST variable was chosen to be included as well,

After regressions set round #5 (with INFL, PERSINC, PCE, HOUST and each other candidate variable combination): COMMPRI variable was chosen to be included as well,

After regressions set round #6 (with INFL, PERSINC, PCE, HOUST, COMMPRI and each other candidate variable combination): all remaining variables (PROD and UNEMPL) were found to be insignificant; variables inclusion stops here.

A final single regression round #7 with all selected variables (INFL, PERSINC, PCE, HOUST, COMMPRI) was run in order to determine the concluded reduced model parameters.

Conclusion

Model characteristics: $R^2 = 0.637$, AIC = 1.581, BIC = 1.615.

This is the same reduced model produced with the general-to-specific process, in (a).

(c) Compare your model from (a) and the Taylor rule of equation (1). Consider R^2, AIC and BIC. Which of the models do you prefer?

```
taylor <-lm(formula = INTRATE ~ INFL + PROD, data = data)
summary(taylor)
```

```
##
## Call:
## lm(formula = INTRATE ~ INFL + PROD, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1592 -1.6762  0.0141  1.3730  7.9203
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.24890    0.17619   7.088 3.51e-12 ***
## INFL         0.97498    0.03273  29.785  < 2e-16 ***
## PROD         0.09472    0.01971   4.805 1.92e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.364 on 657 degrees of freedom
## Multiple R-squared:  0.5747, Adjusted R-squared:  0.5734
## F-statistic: 443.9 on 2 and 657 DF,  p-value: < 2.2e-16
```

```
s <- sqrt(deviance(taylor)/df.residual(taylor))
k <- length(taylor$coefficients) - 1
n <- nrow(data)
AIC <- log(s^2) + 2 * k / n
BIC <- log(s^2) + k * log(n) / n
print(round(c(AIC, BIC), 4))
```

```
## [1] 1.7267 1.7403
```

Conclusion:

Model (a) characteristics: $R^2 = 0.637$, AIC $= 1.581$, BIC $= 1.615$. Model taylor characteristics: $R^2 = 0.5747$, AIC $= 1.7267$, BIC $= 1.7403$.

Based on the lower AIC and BIC of the reduced model (a), it is considered better than the Taylor rule model.

The reduced model (a) has also a greater R^2 value than the Taylor rule model, which confirms the conclusion.

(d) Test the Taylor rule of equation (1) using the RESET test, Chow break and forecast test (with in both tests as break date January 1980) and a Jarque-Bera test. What do you conclude?

RESET Test

```
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 3.6.3

## Loading required package: zoo

##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
resettest(taylor)
```

```
##
##  RESET test
##
## data:  taylor
## RESET = 2.2578, df1 = 2, df2 = 655, p-value = 0.1054
```

The RESET test on fitted values is not significant: it does not reject the Null hypothesis that additional variables would not improve the explanatory power of the model. So, additional variables would improve the model.

Chow Test

to check structural breaks

```
# Chow break testing
grp <- data[data$dates < "1980-01-01", ]
x1 <- grp[, c("INFL", "PROD")]; y1 <- data.frame( INTRATE = grp["INTRATE"] )
data1 <- cbind(x1, y1)
grp <- data[data$dates >= "1980-01-01", ]
x2 <- grp[, c("INFL", "PROD")]; y2 <- data.frame( INTRATE = grp["INTRATE"] )
#chow.test
chow.test(y1, x1, y2, x2)
```

```
##      F value        d.f.1        d.f.2      P value
## 2.873501e+01 3.000000e+00 6.540000e+02 1.836802e-17
```

```
data1_new <- data[,c("INFL", "PROD", "INTRATE")]
```

We can see that the Chow test is significant, implying a structural break in 1980.

Implement chow test myself:

```
s0 <- sum(residuals(taylor)^2)

data1 <- cbind(x1, y1)
taylor1 <- lm(INTRATE~INFL+PROD, data1)
s1 <- sum(residuals(taylor1)^2)

data2 <- cbind(x2, y2)
taylor2 <- lm(INTRATE~INFL+PROD, data2)
s2 <- sum(residuals(taylor2)^2)

k <- 3  ## WE ARE ONLY USING 2 X variables
n <- length(data$INTRATE)

fcrit <- qf(0.975, df1=taylor1$df, df2=taylor2$df)
print(fcrit)
```

```
## [1] 1.249036
```

```
s12 <- s1 + s2
df2 <- n-2*k
```

```
chow_break_test <- (s0 - s12) * df2 / (k * s12)
print(chow_break_test)
```

```
## [1] 28.73501
```

Palue is

```
pf(chow_break_test, df1=3, df2=654, lower.tail = FALSE)
```

```
## [1] 1.836802e-17
```

if chow test is > F, we reject the null hypothesis.

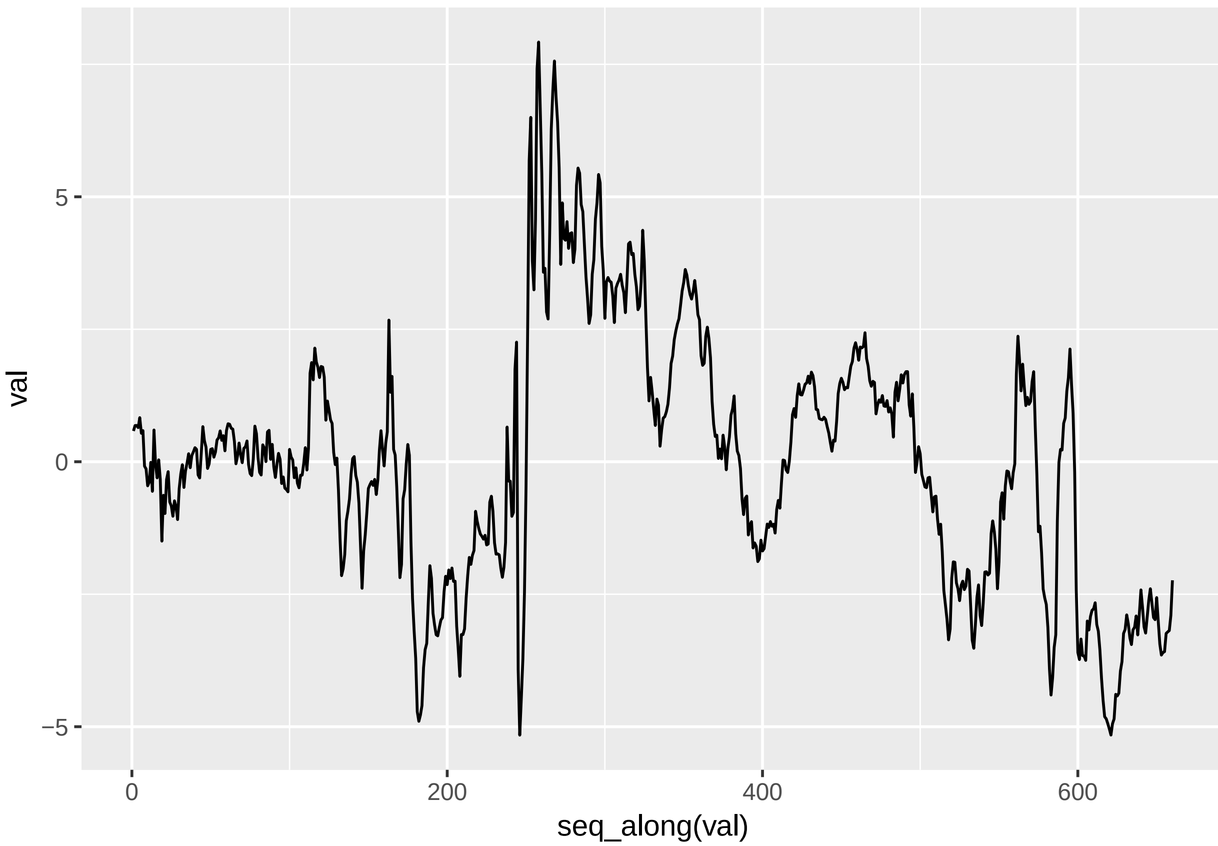
Similarly, chow-forecast test

```
s12 <- s1 + s2
df2 <- length(data1$INTRATE)-k
chow_forecast_test <- (s0 - s1) * df2 / (length(data2$INTRATE) * s1)
print(chow_break_test)
```

```
## [1] 28.73501
```

```
data_frame(val = taylor$residuals) %>% ggplot(aes(seq_along(val), val)) + geom_line()
```

```
## Warning: `data_frame()` is deprecated, use `tibble()`.
## This warning is displayed once per session.
```



We can see this in the residual plot, with a lot of variation in the residuals. The residuals are negative before the suggested break (observation 241), and mostly positive for a few years afterwards, going back in range in the 1990's.

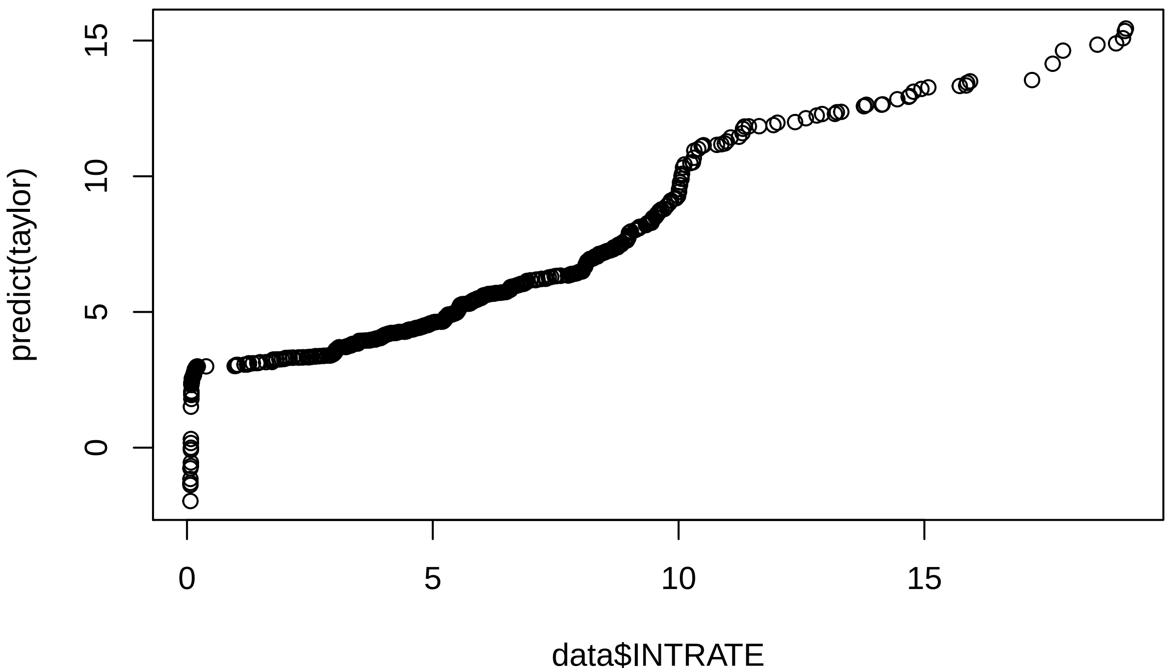
Jarque Bera

normality test is also significant: it signals that data do not have a normal distribution, rejecting the null hypotheses of normality of the residuals.

```
jarque.bera.test(taylor$residuals)
```

```
##
##  Jarque Bera Test
##
## data:  taylor$residuals
## X-squared = 12.444, df = 2, p-value = 0.001985
```

```
qqplot(data$INTRATE,predict(taylor))
```



Conclusion:

The RESET test did not rejected the null hypothesis of correct model specification, however the Chow tests and the Jarque-Bera test rejected the null hypothesis of stability and normality of the residuals.

Conclusively, the Taylor rule model does NOT seem to fit the data very well.