Assignment 7

Eric Tu

9/29/2019

Load GPD Data

```
data <- read.xlsx("Case_GDP.xls", 1)
head(data)
```

```
##     Date     GDP GDPIMPR   LOGGDP            GrowthRate li1 li2 T
## 1 1950Q1  94.300      NA 4.546481                    NA   0   0 0
## 2 1950Q2  95.200       1 4.555980   0.00949875215790819   0   0 1
## 3 1950Q3  97.663       1 4.581523     0.025542835184778   3   1 2
## 4 1950Q4  99.728       1 4.602446    0.0209237030843967   4   2 3
## 5 1951Q1 100.445       1 4.609610   0.00716383394761788   2   1 4
## 6 1951Q2 100.406       0 4.609222 -0.000388347585922766   1   3 5
```

Part a

Likelihood Ratio tests

b_0 is parameters with restriction b_1 is parameters w/o restriction

Chi-square distribution LR = $ -2 (L(b_{0}) - L(b_{1}))$

```
# Storing given log likelihood values
## Restricted Models
ll_b0_c <- -152.763
ll_b0_c_li1 <- -139.747
ll_b0_c_li2 <- -149.521

## Unrestriced model
ll_b1 <- -134.178

# Storing number of restrictions
m_b0_c <- 2
m_b0_c_li1 <- 1
m_b0_c_li2 <- 1

# Test log likelihood ratio
llr <- function(ll0, ll1){
  -2 * (ll0 - ll1)
}

llr_b0_c <- llr(ll_b0_c, ll_b1)
llr_b0_c_li1 <- llr(ll_b0_c_li1, ll_b1)
llr_b0_c_li2 <- llr(ll_b0_c_li2, ll_b1)

# Critical values chi-square
prob <- 0.95
crit_b0_c <- qchisq(prob, df = m_b0_c)
crit_b0_c_li1 <- qchisq(prob, df=m_b0_c_li1)
crit_b0_c_li2 <- qchisq(prob, df = m_b0_c_li2)

# Significant Bool
sig_b0_c <- (llr_b0_c > crit_b0_c)
sig_b0_c_li1 <- (llr_b0_c_li1 > crit_b0_c_li1)
sig_b0_c_li2 <- (llr_b0_c_li2 > crit_b0_c_li2)
```

For Joint Li1, Li2. Log Likelihood ratio: 37.17 Crit Val: 5.9914645 Significant: TRUE

For Li1. Log Likelihood ratio: 11.138 Crit Val: 3.8414588 Significant: TRUE

For Li2. Log Likelihood ratio: 30.686 Crit Val: 3.8414588 Significant: TRUE

Part b

We can use McFadden R^2 to select the best lag structure because all the log likelihood ratios are significant. The McFadden R^2 calculation is given by $R^2 = 1 - \dfrac{\log L(b)}{\log L(b_1)}$ where b_1 is the model with only an intercept. The largest Mcfadden R^2 value should be chosen as the preferred logit model.

```
calcMcfadden <- function(ll_test, ll_const){
  1 - ll_test/ll_const
}

ll_logit1 <- -134.178
ll_logit2 <- -134.126
ll_logit3 <- -130.346
ll_logit4 <- -130.461
ll_const <- -152.763

mc_r2_logit1 <- calcMcfadden(ll_logit1, ll_const)
mc_r2_logit2 <- calcMcfadden(ll_logit2, ll_const)
mc_r2_logit3 <- calcMcfadden(ll_logit3, ll_const)
mc_r2_logit4 <- calcMcfadden(ll_logit4, ll_const)
```

The Mcfadden R^2 values are:

- Logit 1 [li1(-1), li2(-1)]: 0.121659
- Logit 2 [li1(-1), li2(-2)]: 0.1219994
- Logit 3 [li1(-2), li2(-1)]: 0.1467436
- Logit 4 [li1(-2), li2(-2)]: 0.1459908

The best model is Logit 3 [li1(-2), li2(-1): with the highest McFadden R^2 value of 0.1467436

Part c

Logit Model 3: $\beta_c = 0.746$ $\beta_{li1(-2)} = -0.429$ $\beta_{li2(-1)} = -0.131$

Pr[$y_i = 1$] = $\dfrac{exp(x_i'b)}{1 + exp(x_i'b)}$ Pr[$y_i = 0$] = $\dfrac{1}{1 + exp(x_i'b)}$

```
beta_c <- 0.746
beta_1 <- -0.429
beta_2 <- -0.131
cutoff <- 0.5
range <- 245:264
x_1 <- data$li1[range]
x_2 <- data$li2[range]

Z <- exp(beta_c + beta_1 * x_1 + beta_2 * x_2)
logit_vals <- Z/(1 + Z)

predic_num = as.integer(logit_vals > cutoff)

logit_eval <- data$GDPIMPR[range]
logit_num <- as.numeric(levels(logit_eval))[logit_eval]
```

```
## Warning: NAs introduced by coercion
```

```
t0_p0 <- sum((predic_num == 0 & logit_num == 0))/ length(range)
t0_p1 <- sum((predic_num == 1 & logit_num == 0))/ length(range)
t1_p0 <- sum((predic_num == 0 & logit_num == 1))/ length(range)
t1_p1 <- sum((predic_num == 1 & logit_num == 1))/ length(range)
```

Cut-off value: 0.5

observed	$\hat{y} = 0$	$\hat{y} = 1$	**sum**
$y = 0$	0.2	0.15	0.35
$y = 1$	0.25	0.4	0.65
sum	0.45	0.55	1

Hit rate: 0.2 + 0.4 = 0.6

Part d

Augmented Dicky-Fuller test on LOG GDP. Includes constant, trend, and single lag.

```
est_range <- 4:242
log_gdp <- data$LOGGDP[est_range]
log_gdp_lag1 <- data$LOGGDP[est_range-1]
log_gdp_diff1 <- diff(data$LOGGDP)[est_range]
trend <- data$T[est_range]

df_fit <- lm(log_gdp~log_gdp_lag1+log_gdp_diff1+trend)


crit_val_trend <- -3.5

ro_coef <- summary(df_fit)$coef[3,1]
ro_tval <- summary(df_fit)$coef[3, 3]

reject_non_stationarity <- ro_tval < crit_val_trend
```

- Dicky-Fuller Test: 0.6286648
- Dicky-Fuller t-val: 11.8248886
- Reject Non-Stationary: FALSE Confiremd Log gdp is not stationary

Part e

```
fact2num <- function(fact){
  as.numeric(levels(fact))[fact]
}
growth_rate <- fact2num(data$GrowthRate[est_range])
```

```
## Warning in fact2num(data$GrowthRate[est_range]): NAs introduced by coercion
```

```
growth_rate_lag1 <- fact2num(data$GrowthRate[est_range-1])
```

```
## Warning in fact2num(data$GrowthRate[est_range - 1]): NAs introduced by
## coercion
```

```
li1_lag1 <- data$li1[est_range-1]
li1_lag2 <- data$li1[est_range-2]
li2_lag1 <- data$li2[est_range-1]
li2_lag2 <- data$li2[est_range-2]

lm_logit1 <- lm(growth_rate~growth_rate_lag1+li1_lag1+li2_lag1)
lm_logit2 <- lm(growth_rate~growth_rate_lag1+li1_lag1+li2_lag2)
lm_logit3 <- lm(growth_rate~growth_rate_lag1+li1_lag2+li2_lag1)
lm_logit4 <- lm(growth_rate~growth_rate_lag1+li1_lag2+li2_lag2)

r2_logit1 <- summary(lm_logit1)$r.squared
r2_logit2 <- summary(lm_logit2)$r.squared
r2_logit3 <- summary(lm_logit3)$r.squared
r2_logit4 <- summary(lm_logit4)$r.squared
```

R^2 values are:

- Logit 1: [li1(-1), li2(-1)] 0.5142939
- Logit 2: [li1(-1), li2(-2)] 0.5142321
- Logit 3: [li1(-2), li2(-1)] 0.4873591
- Logit 4: [li1(-2), li2(-2)] 0.4874944

Coefficients for best model Logit 1 are

```
summary(lm_logit1)$coef
```

```
##                       Estimate   Std. Error   t value     Pr(>|t|)
## (Intercept)       0.0016338741 3.235051e-04  5.050536 8.841614e-07
## growth_rate_lag1  0.4968932839 4.698275e-02 10.576080 1.233780e-21
## li1_lag1         -0.0009850297 1.308068e-04 -7.530416 1.080186e-12
## li2_lag1         -0.0001387017 6.491116e-05 -2.136793 3.364804e-02
```

Part f

Breusch-Godfrey test on first-order residuals

```
BG <- bgtest(growth_rate~growth_rate_lag1+li1_lag1+li2_lag1, type="Chisq")
#BG <- length(resid_logit2)*r2_resid
crit_BC <- qchisq(prob, df = 1)
BC_choice <- (BG$statistic < crit_BC)
```

Residuals look like white noise allowing characterization of BG test as chi-square distribution with # lags degrees of freedom.

- BG value: 1.4268462

- Critical value: 3.8414588
- Model Correctly Specified: TRUE

Part g

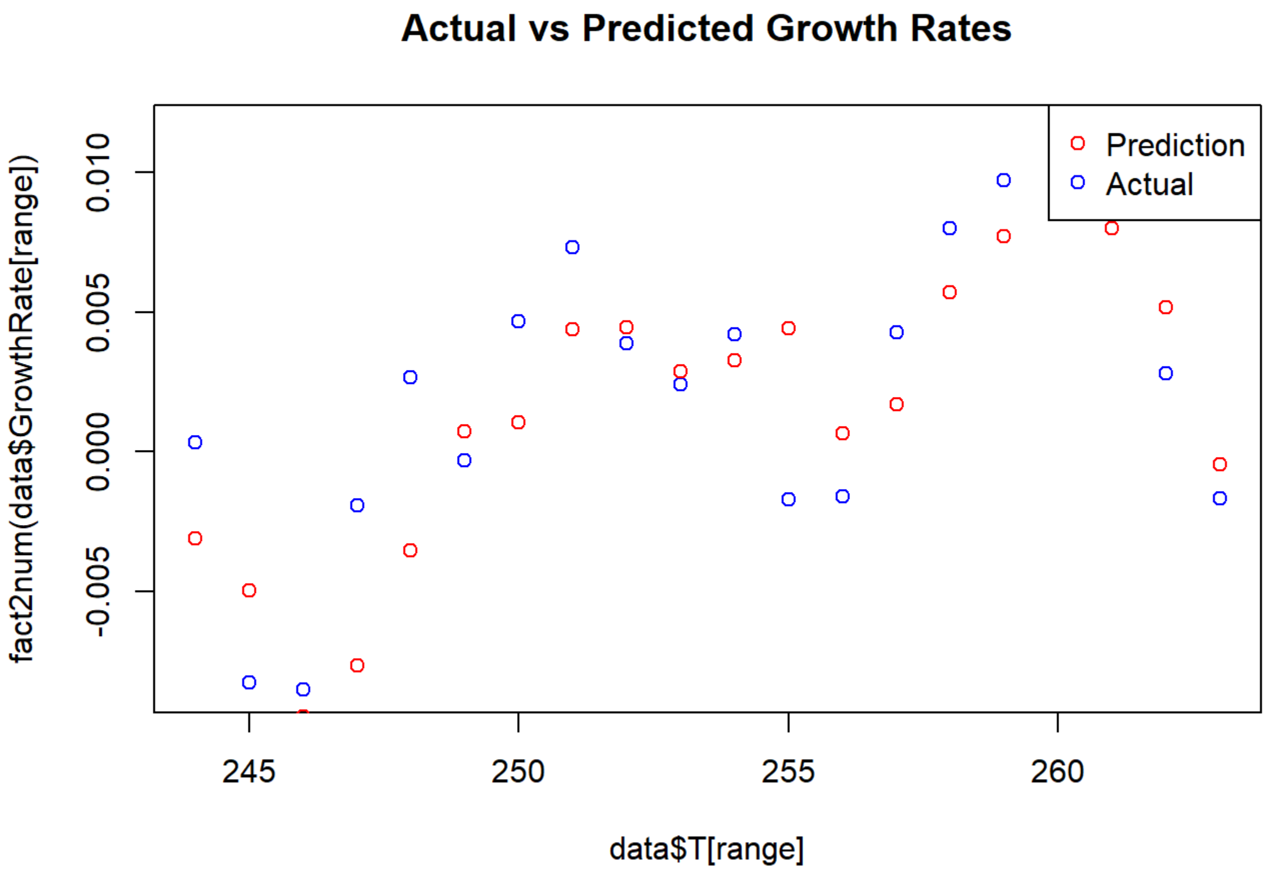
```
ind_gr_lag1 <- fact2num(data$GrowthRate[range-1])
```

```
## Warning in fact2num(data$GrowthRate[range - 1]): NAs introduced by coercion
```

```
ind_l1_lag1 <- data$li1[range-1]
ind_l2_lag1 <- data$li2[range-1]

indicators <- data.frame("growth_rate_lag1" = ind_gr_lag1, "li1_lag1" = ind_l1_lag1, "li2_lag1"
 = ind_l2_lag1)
predictions <- predict(lm_logit1, indicators)

plot(data$T[range], fact2num(data$GrowthRate[range]), col="blue", main="Actual vs Predicted Grow
th Rates")
```

```
## Warning in fact2num(data$GrowthRate[range]): NAs introduced by coercion
```

```
points(data$T[range], predictions, col="red")

legend(x="topright", c("Prediction","Actual"),col=c("red","blue"),pch=c(1, 1))
```




```
MSE <- function(y1, y2){
  sum((y1 - y2) ** 2)/length(y1)
}

mse <- MSE(fact2num(data$GrowthRate[range]), predictions)
```

```
## Warning in fact2num(data$GrowthRate[range]): NAs introduced by coercion
```

Mean squared error of: 1.014324710^{-5}