Econometrics Case Project

Date: 21/12/2023

Solution for (a):

The likelihood ratio tests can be given by the test statistic

$$LR = -2\big(\log\big(L(b_0)\big) - \log\big(L(b_1)\big)\big) \sim \chi^2(m)$$

where $L(b_0)$ is the maximum likelihood value in the restricted model, $L(b_1)$ is the maximum likelihood in the full model and m is the number of parameter restrictions in the restricted model. The critical value of the test is $\chi^2_{0.05}(2) = 5.991$,

$$H_0 : the\ m\ parameter\ restrictions\ are\ correct$$

```
> # model with constant (m = 2 restrictions) vs full model - c
ritical value=5.991
> LR_full <- -2*(-152.763 - (-134.178)) ; LR_full
[1] 37.17
>
>
> # model with li1 (m =1 restrictions) vs full model - critica
l value = 3.841
> LR_1 <- -2*(-139.747 - (-134.178)) ; LR_1
[1] 11.138
>
>
> # model with li2 (m=1 restrictions) vs full model
> LR_2 <- -2*(-149.521 - (-134.178)) ; LR_2
[1] 30.686
```

For the model with two restrictions (li1=li2=0, m=2) we got $LR = 37.17 > 5.991$ so we reject the null hypothesis (at 5% significance level) and thus the full model is statistically significant.

For the model with restriction li2=0 (m=1), we got LR=11.138 > 3.841 (3.841 is the critical value for m=1) and so we reject the null hypothesis (at 5% significance level).

For the model with restriction li1=0 (m=1), we got LR=30.686>3.841, and so we reject the null hypothesis at 5% significance level.

Solution for (b):

We can use the McFadden R^2 because the dependent variable (predictor) of the model is a nominal variable. The R squared is given by the formula:

$$R^2 = 1 - \frac{\log(L(b))}{\log(L(b_1))}$$

where b1 is the model with only the intercept and b is the model under consideration.

```
> Rsq_li11 <- 1 - (-134.178)/(-152.763) ; Rsq_li11
[1] 0.121659
> Rsq_li12 <- 1 - (-134.126)/(-152.763) ; Rsq_li12
[1] 0.1219994
> Rsq_li21 <- 1 - (-130.346)/(-152.763) ; Rsq_li21
[1] 0.1467436
> Rsq_li22 <- 1 - (-130.461)/(-152.763) ; Rsq_li22
[1] 0.1459908
```

The R squared for li1(-1) and li2(-1) is 0.1216.

The R squared for li1(-1) and li2(-2) is 0.1219.

The R squared for li1(-2) and li2(-1) is 0.1467.

The R squared for li1(-2) and li2(-2) is 0.1459.

We got that the highest R squared is 0.1467 which corresponds to the model with variables li1(-2) and li2(-1).

Solution for (d):

We denote $Y_t = LOGGDP$. The model for the Augmented Dickey-Fuller test is the following:

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} + \gamma \Delta Y_{t-1} + \omega_t$$

```
 Call:
lm(formula = DYt ~ T + Yt1 + DYt_Lag1, data = df_new)

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0113538 -0.0024430 -0.0001446  0.0027148  0.0182796 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.344e-02  3.724e-02   2.509   0.0128 *  
T            6.382e-05  2.573e-05   2.481   0.0138 *  
Yt1         -2.001e-02  8.073e-03  -2.479   0.0139 *  
DYt_Lag1     6.090e-01  5.064e-02  12.027   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004542 on 236 degrees of freedom
Multiple R-squared:  0.3976,	Adjusted R-squared:  0.3899 
F-statistic: 51.91 on 3 and 236 DF,  p-value: < 2.2e-16
```

The estimated model is $\Delta Y_t = 0.093 + 0.000 * t - 0.02 * Y_{t-1} + 0.609\ \Delta Y_{t-1} + \omega_t$

The test statistic is $t_\rho = -2.479 > -3.5$ so we do not reject the null hypothesis (H_0 : the time series Y_t is non-stationary) at 5% significance level.

Solution for (e):

For k1=k2=1 we get the following:

```
Call:
lm(formula = GrowthRate ~ GrowthRate_Lag1 + li1_Lag_k1 + li2_L
ag_k2,
    data = df_new)

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0097644 -0.0028559 -0.0001994  0.0026151  0.0158224 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1.737e-03  3.197e-04   5.433 1.38e-07 ***
GrowthRate_Lag1  4.616e-01  4.830e-02   9.556  < 2e-16 ***
li1_Lag_k1      -1.023e-03  1.298e-04  -7.880 1.20e-13 ***
li2_Lag_k2      -1.494e-04  6.421e-05  -2.326   0.0209 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004106 on 236 degrees of freedom
Multiple R-squared:  0.508,     Adjusted R-squared:  0.5017 
F-statistic: 81.22 on 3 and 236 DF,  p-value: < 2.2e-16
```

R squared = 0.508 (k1=k2=1)

For k1=1 and k2=2 :

```
Call:
lm(formula = GrowthRate ~ GrowthRate_Lag1 + li1_Lag_k1 + li2_L
ag_k2,
    data = df_new)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.009785 -0.002792 -0.000221  0.002526  0.015915 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1.754e-03  3.188e-04   5.501 9.79e-08 ***
GrowthRate_Lag1  4.578e-01  4.845e-02   9.449  < 2e-16 ***
li1_Lag_k1      -1.018e-03  1.296e-04  -7.860 1.37e-13 ***
li2_Lag_k2      -1.471e-04  6.416e-05  -2.293   0.0227 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004107 on 236 degrees of freedom
Multiple R-squared:  0.5077,    Adjusted R-squared:  0.5014 
F-statistic: 81.12 on 3 and 236 DF,  p-value: < 2.2e-16
```

R squared = 0.5077 (for k1=1 k2=2)

For k1=2, and k2=1 we get the following results:

```
Call:
lm(formula = GrowthRate ~ GrowthRate_Lag1 + li1_Lag_k1 + li2_L
ag_k2,
    data = df_new)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.009053 -0.002661 -0.000229  0.002342  0.016491 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1.894e-03  3.376e-04   5.611 5.60e-08 ***
GrowthRate_Lag1  4.109e-01  5.445e-02   7.546 9.70e-13 ***
li1_Lag_k1      -9.871e-04  1.479e-04  -6.674 1.76e-10 ***
li2_Lag_k2      -1.585e-04  6.675e-05  -2.374   0.0184 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004233 on 236 degrees of freedom
Multiple R-squared:  0.4772,	Adjusted R-squared:  0.4705 
F-statistic:  71.8 on 3 and 236 DF,  p-value: < 2.2e-16
```

The R squared is 0.4772 (for k1=2, k2=1).

And finally for k1=k2=2 :

```
Call:
lm(formula = GrowthRate ~ GrowthRate_Lag1 + li1_Lag_k1 + li2_L
ag_k2,
    data = df_new)

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0091209 -0.0026155 -0.0002767  0.0023941  0.0161471 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1.911e-03  3.369e-04   5.673 4.10e-08 ***
GrowthRate_Lag1  4.069e-01  5.466e-02   7.444 1.82e-12 ***
li1_Lag_k1      -9.833e-04  1.476e-04  -6.662 1.88e-10 ***
li2_Lag_k2      -1.579e-04  6.667e-05  -2.368   0.0187 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004233 on 236 degrees of freedom
Multiple R-squared:  0.4771,	Adjusted R-squared:  0.4705 
F-statistic: 71.78 on 3 and 236 DF,  p-value: < 2.2e-16
```

R squared = 0.4771 (k1=k2=2)

From the above we conclude that the model with largest R squared (equal to 0.508) is the model for k1=k2=1 . The model can be written as:

$$GrowthRate_t = \alpha + \rho GrowthRate_{t-1} + \beta_1 li1_{t-k_1} + \beta_2 li2_{t-k2} + \varepsilon_t$$

where $\alpha = 0.001737$, $\rho = 0.4616$, $\beta_1 = -0.001023$ and $\beta_2 = 0.000149$.

Solution for (f):

We take the residuals from the model (with k1=k2=1)

$$GrowthRate_t = \alpha + \rho GrowthRate_{t-1} + \beta_1 li1_{t-1} + \beta_2 li2_{t-1} + \varepsilon_t$$

and regress them (OLS) on the following model (first order serial correlation):

$$e_t = \alpha + \rho GrowthRate_{t-1} + \beta_1 li1_{t-1} + \beta_2 li2_{t-1} + \gamma_1 e_{t-1} + \omega_t$$

```
Call:
lm(formula = e ~ GrowthRate_Lag1 + li1_Lag_k1 + li2_Lag_k2 +
    e_Lag1, data = df_new)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0096882 -0.0028198 -0.0002075  0.0026143  0.0158580

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1.004e-04  3.838e-04  -0.261    0.794
GrowthRate_Lag1  2.914e-02  7.915e-02   0.368    0.713
li1_Lag_k1       3.136e-05  1.492e-04   0.210    0.834
li2_Lag_k2       2.768e-06  6.496e-05   0.043    0.966
e_Lag1          -4.999e-02  1.032e-01  -0.484    0.629

Residual standard error: 0.004121 on 234 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.001002,	Adjusted R-squared:  -0.01607
F-statistic: 0.05868 on 4 and 234 DF,  p-value: 0.9936
```

We got that the R squared is equal to 0.001002. Now we calculate the test statistic of the Breusch-Godfrey test, that is $BG = nR^2$ and compare it with the critical value $\chi^2_{0.05}(1) = 3.841$ (at 5% significance level)

$$BG = nR^2 = 239 * 0.001 = 0.239 < 3.841$$

and so we do not need to adjust the model further.