Time Series Analysis Course Work

Matěj Adámek

January 15, 2017

University of Plymouth

Module: ECN2005

Student: 10531306

Question 1

Microeconomic theory behind this hypothesis of dependence of share of beards in men on industrial production asusmes the increase in share of beards to be an opportunity cost of increase in production. Because of this, the elasticity of beard share towards industrial production will be analyzed, and therefore the variables need to be transformed logarithmically. Then, because due to the ADF tests both variables do exhibit a unit root, a first difference of both needs to be taken.

Variable	With constant		Unit root
	(a-1)	p-value	
$l\_PERCENT\_BEARD$	-0.458	0.869	Present
$l\_US\_INDPROD$	-0.012	0.838	Present
$d\_l\_PERCENT\_BEARD$	-2.117	0.000***	Absent
$d\_l\_US\_INDPROD$	-1.334	0.000***	Absent

Table 1: Unit root tests for share of beards and US industrial production

Then, an ADL model is attempted to be constructed with the dependent variable being the share of beards. Because trends in shaving are relatively long-term matter, the maximum number of lags is chosen to be 4, testing down to the last significant lags. As the last column shows, the share of men with beard consistently depends on the past values of itself, rather than on industrial production in any way.

Independent variable: $d\_l\_PERCENT\_BEARD$		
Equation	**1**	**7**
constant	1.307 0.811	3.328 0.317
$d\_l\_PERCENT\_BEARD_{t-1}$	-0.889 0.000	-0.863 0.000
$d\_l\_PERCENT\_BEARD_{t-2}$	-0.732 0.002	-0.542 0.008
$d\_l\_PERCENT\_BEARD_{t-3}$	-0.591 0.012	-0.383 0.025
$d\_l\_PERCENT\_BEARD_{t-4}$	0.1782 0.071	-
$d\_l\_US\_INDPROD_{t}$	-0.343 0.000	-
$d\_l\_US\_INDPROD_{t-1}$	9.248 0.234	-
$d\_l\_US\_INDPROD_{t-2}$	-4.954 0.542	-
$d\_l\_US\_INDPROD_{t-3}$	13.757 0.1171	-
$d\_l\_US\_INDPROD_{t-4}$	4.475 0.712	-
p(F)	0.0004	0.0000
R^2	0.591	0.487
RESET p-value	0.999	0.674
B-P p-value	0.702	0.819
QLR	-	Pass
CUSUM	-	Pass

Table 2: ADL model build up for share of beards

Question 2

(a) Firstly, elasticity defines as the percentage change in endogenous variable resulting from 1% change in the exogenous. In the equation $Y = aX^b$, elasticity of X towards Y is expressed by coefficient b. In order to deal with a multiplicative equation using the Ordinary Least squares analysis, a logarithmic transformation must be applied. $ln(Y) = ln(a) + b \times ln(X)$. Then, in the OLS output, the coefficients express a power to which each of the variable is raised to, i.e. elasticity of Y towards it.

The transformation to be made is therefore taking a natural logarithm of both variables. Two new variables are generated: $l\_CORN\_PR = ln(CORN\_PR)$ and $l\_HOG\_PR = ln(HOG\_PR)$

(b) Inspecting time-series plots of both variables shown in figure 1, there is somewhat apparent over-time increase in the hog price and a bit less apparent over-time increase in the rather volatile corn price.

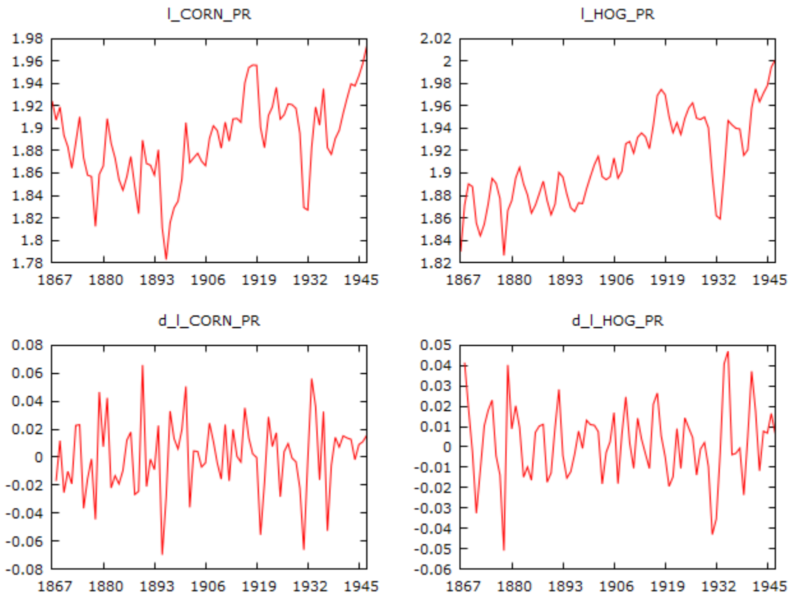


Figure 1: Time series plots

In order to confirm the non-stationarity of variables, the augmented Dickey-Fuller test needs to be conducted. The lag order appropriate for this test is 1 as the data has annual frequency. As shown in table 3, both of the variables are proven to have a unit-root. For this reason, first difference of both variables is taken and another round of ADF test conducted. It can be observed both from the plots in figure 1 and the p-values in the test result that the first differences of both variables are stationary and hence suitable for further analysis.

Variable	Without constant		With constant		Unit root
	(a-1)	p-value	(a-1)	p-value	
l_CORN_PR	0.00024	0.729	0.2173	0.051*	Present
l_HOG_PR	0.00056	0.833	-0.1184	0.1876	Present
$d\_l\_CORN\_PR$	-1.2436	0.000***	-1.2448	0.000***	Absent
$d\_l\_HOG\_PR$	-1.0119	0.000***	-1.0251	0.000***	Absent

Table 3: ADF test output

(c) Then, for the purposes of estimation of autoregressive model estimating the dependence of current price of hogs on its past values , correlogram needs to be produced to establish number of previous values of significant influence on the current value. Shown in figure 2, the first three lags present with partial autocorrelation coefficients outside of the blue confidence intervals, the approppriate number of lags to be included in the autoregressive model is 3.

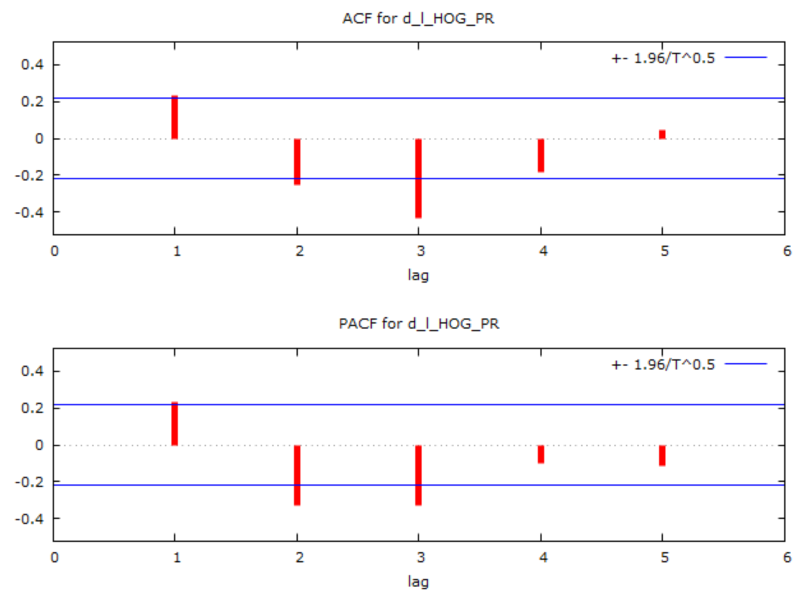


Figure 2: Correlogram for $d\_l\_HOG\_PR$

Independent: $d\_l\_HOG\_PR$	
Variable	**Coefficient**
constant	0.002 0.242
$d\_l\_HOG\_PR_{t-1}$	0.184 0.098
$d\_l\_HOG\_PR_{t-2}$	-0.224 0.042
$d\_l\_HOG\_PR_{t-3}$	-0.335 0.002
p(F)	0.000
R^2	0.258
RESET p-value	0.226
B-P p-value	0.040
QLR	Pass

Figure 3: AR(3) for $d\_l\_HOG\_PR$

The resulting autoregressive model seen in table 3 has an insignificant constant, while the coefficients determining the direction and strength of influence of given lag on the current values with at least 90% confidence have been formed for each of the three lags $d\_l\_HOG\_PR_{t-p}$. As the last rows of table 3 shows, there does not seem to be any functional form misspecification or structural breaks, although heteroskedasticity might be a problem.

(d) Using this information, it is possible to create forecasts of the price of hogs. The formula for the transformed variable at any point in time t is $X_t = 0.184X_{t-1} - 0.224X_{t-2} - 0.335X_{t-3}$ where $X = d\_l\_HOG\_PR$.

In order to construct the dynamic forecast, three known values preceding the predicted period are necessary. In this case, the range of forcast is 1948 - 1951, therefore it requires values for years 1945 - 1947. Then, $X_{1948} = 0.184X_{1947} - 0.224X_{1946} - 0.335X_{1945} \cong 0.184 \times 0.006468 - 0.224 \times 0.016466 - 0.335 \times 0.006802 \cong -0.00478$. This step is then repeated for each consecutive year up until the end of the forecast period.

In order to use the forecast values for creation of predictions of the hog prices rather than the transformed variables, the transformations have to be reversed. The last known value of log of hog price preceding the forecast period is necessary to undo the first difference. Then, by raising the mathematical constant $e \cong 2.718$ to the power of the prediction of l_HOG_PR, the logarithm is canceled-out. Visually, the data gained from the forecast can be plotted as infigure 5. In MS excel, the process is illustrated by figure 4a.

DFC_d_l_HOG_PR	DFC_l_HOG_PR	DFC_HOG_PR
0.006802		
0.016466		
0.006468	**2.001150216**	
=0.184191*I82-0.224771*I81-0.335629*I80		
-0.007863	1.98849468	7.304529827
-0.002542	1.985952953	7.285987282
0.002908	1.988860693	7.307203871

(a) Forecast of first difference in natural log of hog price

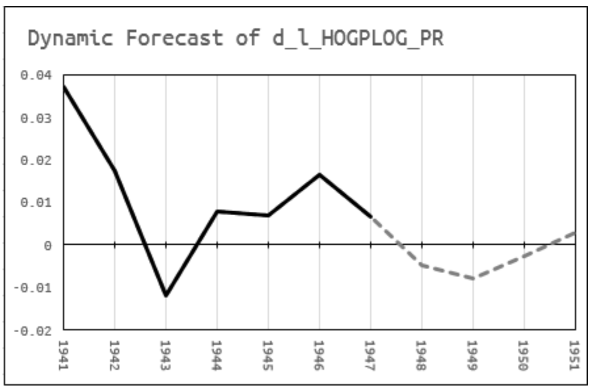
DFC_d_l_HOG_PR	DFC_l_HOG_PR	DFC_HOG_PR
0.006802		
0.016466		
0.006468	**2.001150216**	
-0.004793	=J82+I83	7.362190763
-0.007863	1.98849468	7.304529827
-0.002542	1.985952953	7.285987282
0.002908	1.988860693	7.307203871

(b) Forecast of natural log of hog price

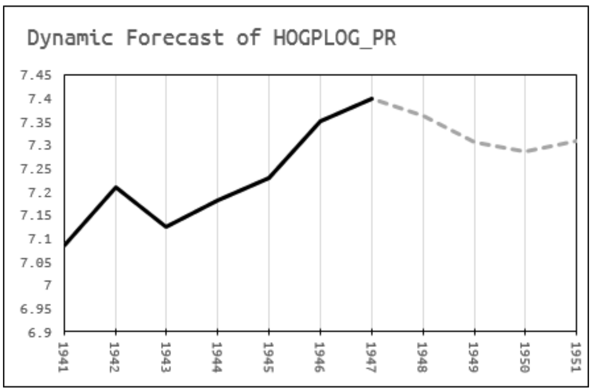
DFC_d_l_HOG_PR	DFC_l_HOG_PR	DFC_HOG_PR
0.006802		
0.016466		
0.006468	**2.001150216**	
-0.004793	1.996358	=EXP(J83)
-0.007863	1.98849468	7.304529827
-0.002542	1.985952953	7.285987282
0.002908	1.988860693	7.307203871

(c) Forecast of hog price

Figure 4: Excel procedure for



(a) of $d\_l\_HOG\_PR$



(b) of HOG_PR

Figure 5: Time series plots for manual dynamic forecasts

(e) To test dependence of hog price on lags of itself as well as another variable, Autocorrelation Distributed Lag model is used. The estimation starts with four lags of both variables, progressively removing the last lags op to the point where these are significant.

Independent variable: $d\_l\_HOG\_PR$		
Equation	**1**	**2**
constant	0.0021 (0.134)	0.0023 (0.094)
$d\_l\_HOG\_PR_{t-1}$	-0.2341 (0.059)	-0.2932 (0.001)
$d\_l\_HOG\_PR_{t-2}$	-0.3466 (0.003)	-0.4202 (0.000)
$d\_l\_HOG\_PR_{t-3}$	-0.2013 (0.068)	-0.2785 (0.003)
$d\_l\_HOG\_PR_{t-4}$	-0.8665 (0.244)	-
$d\_l\_CORN\_PR_{t}$	0.2444 (0.000)	0.2420 (0.000)
$d\_l\_CORN\_PR_{t-1}$	0.3626 (0.000)	0.3886 (0.000)
$d\_l\_CORN\_PR_{t-2}$	0.3891 (0.000)	0.4363 (0.000)
$d\_l\_CORN\_PR_{t-3}$	0.1782 (0.039)	0.2421 (0.001)
$d\_l\_CORN\_PR_{t-4}$	-0.0867 (0.821)	-
p(F)	1.84e-11	4.83e-13
R^2	0.629	0.638
RESET p-value	0.538	0.646
B-P p-value	0.166	0.140

Table 4: ADL model for $d\_l\_HOG\_PR$

The final model, i.e. equation 2 in table 4, presents with high significance of the F statistic being different from 0, which means the error term is independent from any variable in the model. Furthermore, Ramsay's RESET test for functional misspecification has been passed with p-value of 0.646, which suggests that all necessary transformation of variables in the model were done.

(f) In ADL models, the long run effect is calculated as $\frac{\sum(X-coefficients)}{1-\sum(Y-coefficients)}$, where X is the independent variable ($d\_l\_CORN\_PR$) and Y is the dependent variable ($d\_l\_HOG\_PR$). For this specific model, it can be computed as follows.

$$\frac{\sum(X-coefficients)}{1-\sum(Y-coefficients)} = \frac{0.2420+0.3886+0.4363+0.2421}{1-(-0.2932-0.4202-0.2785)} \cong \frac{1.30887}{1-(-0.991795)} \cong 0.657$$

This suggests that 1% increase in price of corn will cause 0.657% increase in the price of hogs.

(g) Chow test is a procedure, during which a dummy and slope dummy variables are added to the model. The splitdum variable is equal to 0 before a specified point in time and 1 after it. For the slope dummies, they are equal to 0 before the suspected point and to a value of specified variable after the point. The Chow test re-runs the regression with the dummies and presents their joint significance. In case their coefficients are significant, the existence of a structural break at the analyzed time point is proven.

Here, year 1930 is considered to be a possible break. In figure 6, it can be observed, that elasticities towards the first lag of corn price and the third lag of hog price differ significantly before and after 1930. The joint p-value presented at the bottom of the output then indicates that with 99% confidence there is a structural break at this point.

```
Augmented regression for Chow test
OLS, using observations 1871-1947 (T = 77)
Dependent variable: d_l_HOG_PR

                      coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const                0.00297922   0.00137213    2.171    0.0338   **
  d_l_CORN_PR          0.188090     0.0621069     3.028    0.0036   ***
  d_l_CORN_PR_1        0.336774     0.0618853     5.442    9.89e-07 ***
  d_l_CORN_PR_2        0.361693     0.0659303     5.486    8.37e-07 ***
  d_l_CORN_PR_3        0.272766     0.0649852     4.197    8.92e-05 ***
  d_l_HOG_PR_1        -0.404305     0.110398     -3.662    0.0005   ***
  d_l_HOG_PR_2        -0.442489     0.0957841    -4.620    2.04e-05 ***
  d_l_HOG_PR_3        -0.474714     0.102929     -4.612    2.09e-05 ***
  splitdum            -0.00190873   0.00281488   -0.6781   0.5003
  sd_d_l_CORN_PR       0.0357726    0.121257      0.2950   0.7690
  sd_d_l_CORN_PR_1     0.266599     0.132885      2.006    0.0493   **
  sd_d_l_CORN_PR_2     0.203939     0.190349      1.071    0.2882
  sd_d_l_CORN_PR_3    -0.0941447    0.195798     -0.4808   0.6324
  sd_d_l_HOG_PR_1      0.270205     0.313918      0.8608   0.3927
  sd_d_l_HOG_PR_2     -0.117877     0.203438     -0.5794   0.5644
  sd_d_l_HOG_PR_3      0.597937     0.192296      3.109    0.0028   ***

Mean dependent var   0.001471   S.D. dependent var   0.018001
Sum squared resid    0.006399   S.E. of regression   0.010242
R-squared            0.740162   Adjusted R-squared   0.676268
F(15, 61)            11.58414   P-value(F)           1.07e-12
Log-likelihood       252.4665   Akaike criterion    -472.9330
Schwarz criterion   -435.4321   Hannan-Quinn        -457.9330
rho                 -0.066798   Durbin-Watson        2.129191

Chow test for structural break at observation 1930
  F(8, 61) = 2.99959 with p-value 0.0067
```

Figure 6: GRETL output for the Chow test on ADL model for *d_l_HOG_PR*

(h) In order to choose appropriate lag order for Vector Autoregression model, which tests for Granger Causality, the VAR lag selection is run in GRETL, selecting the highest lag-order to test for to be 4.

As the output in figure 7 shows, based on different criteria, the best lag order is either 2 or 3. Therefore, 3 lags will be chosen for the construction of the Vector Autoregression model.

```
VAR system, maximum lag order 6

The asterisks below indicate the best (that is, minimized) values
of the respective information criteria, AIC = Akaike criterion,
BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

lags        loglik    p(LR)        AIC          BIC          HQC

   1     370.80175            -9.859507    -9.672691    -9.784984
   2     387.66332  0.00000  -10.207117    -9.895757   -10.082911
   3     401.87227  0.00001  -10.483034*  -10.047130*  -10.309147*
   4     402.82302  0.75387  -10.400622    -9.840174   -10.177053
   5     404.28133  0.57187  -10.331928    -9.646936   -10.058676
   6     404.62088  0.95388  -10.232997    -9.423460    -9.910063
```

Figure 7: VAR lag selection output for dependence of $d\_l\_HOG\_PR$ on $d\_l\_CORN\_PR$

(i) Running the VAR on Granger causality between $d\_l\_HOG\_PR$ and $d\_l\_CORN\_PR$ suggests that both variables Granger cause the other. While predicting the price of hogs, the p-value for the hypothesis that after removing the $d\_l\_CORN\_PR$ lags R^2 does not change is lower than 0.000, suggesting price of corn Granger causes the price of hogs. Analogically, for the prediction of price of corn, the p-value of a hypothesis that R^2 does not change is 0.005 > 0.05, therefore, the price of hogs Granger causes price of corn.

```
Equation 1: d_l_CORN_PR

                 coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------------
  const           0.00223171   0.00281114    0.7939    0.4299
  d_l_CORN_PR_1  -0.149482     0.124433     -1.201     0.2337
  d_l_CORN_PR_2  -0.215082     0.137228     -1.567     0.1215
  d_l_CORN_PR_3  -0.0517419    0.142098     -0.3641    0.7169
  d_l_HOG_PR_1   -0.00778056   0.225803     -0.03446   0.9726
  d_l_HOG_PR_2    0.0943467    0.186358      0.5063    0.6143
  d_l_HOG_PR_3   -0.608956     0.176021     -3.460     0.0009   ***

Mean dependent var   0.001037   S.D. dependent var   0.026296
Sum squared resid    0.041206   S.E. of regression   0.024262
R-squared            0.215909   Adjusted R-squared   0.148701
F(6, 70)             3.212562   P-value(F)           0.007616
rho                 -0.002594   Durbin-Watson        1.992379

F-tests of zero restrictions:

All lags of d_l_CORN_PR        F(3, 70) =    1.0887 [0.3597]
All lags of d_l_HOG_PR         F(3, 70) =    4.7575 [0.0045]
All vars, lag 3                F(2, 70) =    7.5261 [0.0011]
```

```
Equation 2: d_l_HOG_PR

                 coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------------
  const           0.00279015   0.00147397    1.893    0.0625    *
  d_l_CORN_PR_1   0.352386     0.0652438     5.401    8.61e-07 ***
  d_l_CORN_PR_2   0.384270     0.0719529     5.341    1.09e-06 ***
  d_l_CORN_PR_3   0.229531     0.0745065     3.081    0.0030   ***
  d_l_HOG_PR_1   -0.295042     0.118395     -2.492    0.0151   **
  d_l_HOG_PR_2   -0.397346     0.0977129    -4.066    0.0001   ***
  d_l_HOG_PR_3   -0.425803     0.0922933    -4.614    1.74e-05 ***

Mean dependent var   0.001471   S.D. dependent var   0.018001
Sum squared resid    0.011328   S.E. of regression   0.012721
R-squared            0.539988   Adjusted R-squared   0.500558
F(6, 70)             13.69498   P-value(F)           3.20e-10
rho                  0.025973   Durbin-Watson        1.942615

F-tests of zero restrictions:

All lags of d_l_CORN_PR        F(3, 70) =    14.326 [0.0000]
All lags of d_l_HOG_PR         F(3, 70) =    14.500 [0.0000]
All vars, lag 3                F(2, 70) =    11.755 [0.0000]
```

Figure 8: VAR system output

(j) Assuming that price of corn is equal to the variable cost of production of hogs in a competitive market scenario, it would be reasonable to expect cointegration between the price of corn and price of hogs. This is assessed by the Engle-Granger test.

Firstly, it uses ADF test to check that both $CORN\_PR$ and HOG_PR exhibit a unit root. Because of the annual nature of the data, the appropriate lag order is 1. Output in table 5 confirms that both variables present with a unit-root at 95% significance, which in turn supports the possibility of cointegration.

Then, a cointegrating model with HOG_PR as independent and $CORN\_PR$ as dependent variable is constructed. For presence of cointegration to be proven, it is necessary that the model's residuals are stationary, which appears to be the case, as with a p-value of 0.002 the hypothesis of presence of a unit root is rejected.

Variable	**With constant**		**Unit root**
	(a-1)	**p-value**	
HOG_PR	-0.113	0.212	Present
$CORN\_PR$	-0.220	0.080	Present
$\hat{u}$	-0.557	0.002	Absent
$CORN\_PR$ and HOG_PR are cointegrated.			
Long-run relationship			
$HOG\_PR = 1.528 + 0.791 \times CORN\_PR$			

Table 5: Cointegration output

Finally, the cointegrating relationship shown in table 5 defines the long-run relationship between the price of hogs and the price of corn. The coefficient of $CORN\_PR$, i.e. 0.791 is the long-run effect of change in corn price on the price of a hog. When price of a unit of corn increases by 1, price of hog increases by 0.791 in the long run. (Historically, hog consumes 0.791 units of corn on average before being sold.)

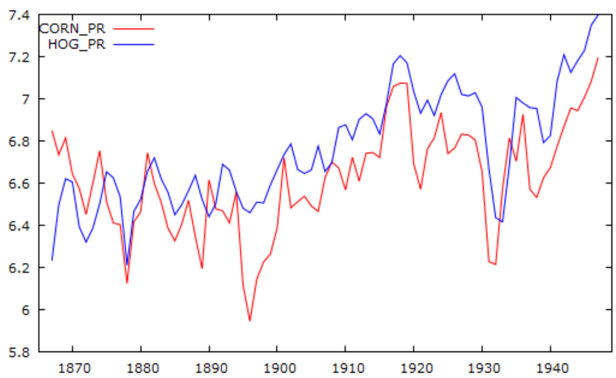


Figure 9: Time-series plot of $CORN\_PR$ and HOG_PR

(k) Error correction model is built by addition of the first lag of the error term from the cointegrating model ($\hat{u}_{t-1}$) to an ADL model of the first differences of the two variables. In order to construct this ADL, the first differences, i.e. l_HOG_PR and l_CORN_PR should be tested for unit root. as these both pass with significance of the null hypothesis (that is the presence of a unit root) lower than 0.000, the ADL can be created. This is sensible, as whenever these variables cointegrate, there must be an error correction model.

Then, the ADL model is estimated with the $\hat{u}_{t-1}$ term added and tested down from four lags to the point where the last lag of both variables is significant, just as in case of exercise (e). After completion of this build-up, results similar to the ADL model appear, with the difference that error term lag is included. The coefficient of $\hat{u}_{t-1}$ represents the speed of adjustment of equilibrium. Inverse value of this represents the time (number of periods, i.e. years) it takes for the equilibrium to be readjusted, that is $\frac{1}{0.125} = 8$ years. Nevertheless, it is important to say that this can only be said with no more than 90% confidence.

Furthermore, assessing this model for functional form misspecification using Ramsay's RESET proves it to have no issues in this respect and based on the result of Breusch-Pagan heteroskedasticity test, there is no problem either. Additionally, testing for structural breaks using QLR results with no evidence for any structural breaks whatsoever, showing that this model is more appropriate, as the equilibrium forces are expected to exist even through the period of The Great Depression. Both CUSUM and CUSUMSQ tests for stability are also passed and so is the test for normality of residuals.

Independent variable: d_HOG_PR		
Equation	**1**	**2**
constant	0.015 0.095	0.016 0.067
$d\_HOG\_PR_{t-1}$	−0.138 0.283	−0.224 0.046
$d\_HOG\_PR_{t-2}$	−0.295 0.012	−0.386 0.000
$d\_HOG\_PR_{t-3}$	−0.156 0.154	−0.245 0.008
$d\_HOG\_PR_{t-4}$	0.043 0.653	-
$d\_CORN\_PR_{t}$	0.273 0.000	0.266 0.000
$d\_CORN\_PR_{t-1}$	0.303 0.000	0.341 0.000
$d\_CORN\_PR_{t-2}$	0.354 0.000	0.412 0.000
$d\_CORN\_PR_{t-3}$	0.167 0.055	0.244 0.001
$d\_CORN\_PR_{t-4}$	−0.082 0.272	-
$\hat{u}_{t-1}$	−0.146 0.050	−.125 0.075
p(F)	2.35e-11	3.92e-13
R^2	0.642	0.656
RESET p-value	0.240	0.802
B-P p-value	0.536	0.257
QLR	Pass ($p = 0.094$)	Pass ($p = 0.076$)
CUSUM	Pass ($p = 0.494$)	Pass ($p = 0.518$)
Normality of residuals	Pass ($p = 0.750$)	Pass ($p = 0.635$)

Table 6: Error correction model build up