Lab 3
Cointegration & Error Correction Model

Case study of biodiesel fuel and soybean oil

UBC MFRE | Master of Food and Resource Economics

Roadmap

- Review - ADF Tests
- Cointegration
  - Engle Granger 2 step test
  - Engle Granger function to retrieve correct critical values
  - Johansen Procedure
- Error Correction Model

Packages

- Load the following packages

pacman::p_load(here, readxl, dplyr, janitor, Quandl, xts, lubridate, urca, forecast, tidyverse, vars, modelsummary)

Data

- Load *data_lecture2.csv* and do the data cleaning steps – same data as last week
- Last week, we used price levels. Today, we will focus on log prices

```
> head(soydiesel)
           biodiesel soyoil diesel    lnbio    lnsoy lndiesel  lncrude
2007-04-13     3.100 29.900  2.877 1.131402 3.397858 1.056748 4.136446
2007-04-20     3.100 29.310  2.851 1.131402 3.377929 1.047670 4.144087
2007-04-27     3.075 30.190  2.811 1.123305 3.407511 1.033540 4.178379
2007-05-04     3.140 31.130  2.792 1.144223 3.438172 1.026758 4.156067
2007-05-11     3.140 31.060  2.773 1.144223 3.435921 1.019930 4.125520
2007-05-18     3.175 32.865  2.803 1.155308 3.492408 1.030690 4.152771
```

ADF Tests

- General to specific
- Type = c("trend") $\Delta \boldsymbol{y_t} = \boldsymbol{\alpha} + \boldsymbol{\delta t} + \boldsymbol{\tau y_{t-1}} + \boldsymbol{\varepsilon_t}$
  - Tau3 (τ = 0) – If fail to reject, then unit root is present (not stationary)
  - Phi3 (τ = δ = 0) – If fail to reject, then there is a unit root AND there is no trend term.
  - Phi2 (α = τ = δ = 0) – If fail to reject, then there is a unit root AND there is no trend term AND there is no drift term.
- Type = c("drift") $\Delta \boldsymbol{y_t} = \boldsymbol{\alpha} + \boldsymbol{\tau y_{t-1}} + \boldsymbol{\varepsilon_t}$
  - α is the constant or drift term
  - Tau2 (τ = 0) – If fail to reject, then unit root is present (not stationary)
  - Phi1 (α = τ = 0) – If fail to reject, then there is a unit root AND there is no drift term.
- Type = c("none") $\Delta \boldsymbol{y_t} = \boldsymbol{\tau y_{t-1}} + \boldsymbol{\varepsilon_t}$
  - Tau1 (τ = 0) – If fail to reject, then unit root is present (not stationary)

ADF Tests – Lag Length

- VARselect(soydiesel$lnbio, lag.max = 5)**$select**

```
> VARselect(soydiesel$lnbio, lag.max = 5)$select
AIC(n)  HQ(n)  SC(n) FPE(n)
     4      1      1      4
> VARselect(soydiesel$lnbio, lag.max = 5)
$selection
AIC(n)  HQ(n)  SC(r
     4      1

$criteria
                  1                                     4            5
AIC(n) -6.761081153 -6.761597249 -6.760340277 -6.765082550 -6.762501692
HQ(n)  -6.756426556 -6.754615354 -6.751031084 -6.753446059 -6.748537902
SC(n)  -6.748987950 -6.743457445 -6.736153872 -6.734849544 -6.726222085
FPE(n)  0.001157977  0.001157379  0.001158835  0.001153352  0.001156333
```

```
Value of test-statistic is: 0.3569

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

ADF Tests – Log Prices

summary(ur.df(soydiesel$lnbio, type = c("trend"), lags = 4))

summary(ur.df(soydiesel$lnbio, type = c("drift"), lags = 4))

summary(ur.df(soydiesel$lnbio, type = c("none"), lags = 4))

```
Value of test-statistic is: -1.5295 1.0646 1.3833

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

```
Value of test-statistic is: -1.6272 1.5378

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

```
Value of test-statistic is: 0.3569

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

UBC
MFRE

ADF Tests – First Difference

VARselect(diff.xts(soydiesel$lnbio, na.pad = F), lag.max = 5)$select

summary(ur.df(diff.xts(soydiesel$lnbio, na.pad = F), type = c("trend"), lags = 3))

```
Value of test-statistic is: -12.9334 55.7676 83.6491

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Roadmap

- Review - ADF Tests
- Cointegration
  - Engle Granger 2 step test
  - Engle Granger function to retrieve correct critical values
  - Johansen Procedure
- Error Correction Model

Engle Granger Method

- We meet the necessary condition that prices in levels are I(1) and prices in first differences are I(0)
- Step 1 – estimate the longrun relationship between biodiesel and soybean oil
- Collect the residuals
- Step 2 – run an ADF test on the residuals

Engle Granger Method

- Step 1 – estimate the longrun relationship between biodiesel and soybean oil

$$PBioDiesel_t = \alpha + \beta PSoyoil_t + \varepsilon_t$$

reg_lnbiodieselsoy <- lm(lnbio ~ lnsoy, data = soydiesel)

summary(reg_lnbiodieselsoy)

```
Call:
lm(formula = lnbio ~ lnsoy, data = soydiesel)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.17870 -0.04941 -0.01361  0.03892  0.26583 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.542059   0.035108  -43.92   <2e-16 ***
lnsoy        0.773298   0.009579   80.72   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07578 on 771 degrees of freedom
Multiple R-squared:  0.8942,	Adjusted R-squared:  0.8941 
F-statistic:  6517 on 1 and 771 DF,  p-value: < 2.2e-16
```

Engle Granger Method

- Collect the residuals from the previous regression
- In R, (1) save the residuals, (2) convert residuals to ts object, (3) merge to soydiesel

resid_lnsoydiesel <- resid(reg_lnbiodieselsoy)

lnresid_ts <- xts(resid_lnsoydiesel, order.by = index(soydiesel))

Will use this term later in ECM

soydieselr <- merge.xts(soydiesel, lnresid_ts)

```
> head(soydieselr)
           biodiesel soyoil diesel    lnbio    lnsoy lndiesel  lncrude  lnresid_ts
2007-04-13     3.100 29.900  2.877 1.131402 3.397858 1.056748 4.136446  0.045904189
2007-04-20     3.100 29.310  2.851 1.131402 3.377929 1.047670 4.144087  0.061315803
2007-04-27     3.075 30.190  2.811 1.123305 3.407511 1.033540 4.178379  0.030342904
2007-05-04     3.140 31.130  2.792 1.144223 3.438172 1.026758 4.156067  0.027550531
2007-05-11     3.140 31.060  2.773 1.144223 3.435921 1.019930 4.125520  0.029291353
2007-05-18     3.175 32.865  2.803 1.155308 3.492408 1.030690 4.152771 -0.003305434
```

UBC MFRE

Format regression output

```
models <- list(
  "Biodiesel-Soy (Log)" = lm(lnbio ~ lnsoy, data = soydiesel)
)
modelsummary(models, estimate = "{estimate}{stars}")
```

	Biodiesel-Soy (Log)
(Intercept)	-1.542***
	(0.035)
lnsoy	0.773***
	(0.010)
Num.Obs.	773
R2	0.894
R2 Adj.	0.894
AIC	-1790.9
BIC	-1776.9
Log.Lik.	898.437
F	6516.511

Engle Granger Method

- Step 2 – run an ADF test for a unit root on the residuals
    - No need to use time trend or intercept because, by construction, the residuals will have a zero mean
    - If we reject the null hypothesis of a unit root, then we conclude that the two price series are cointegrated

VARselect(soydieselr$lnresid_ts)$select

summary(ur.df(soydieselr$lnresid_ts, type = c("none"), lags = 3))@teststat

```
Value of test-statistic is: -3.6603

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Interpretation
- It **appears** that we can reject the unit root null hypothesis
- But this is **not correct** because we must use different critical values

Engle Granger Method

- Stata gives the correct critical values as part of the **egranger** test, but R does not
- We wrote an R function that will give you the correct critical values

englegranger(var, trend, n)

- var = # of variables, in our case **2** (biodiesel and soybean oil)
- trend = **0** if no trend in step 1 regression, 1 if we included a trend in step 1 regression
- n = number of observations = **773**

```
> englegranger(2, 0, 773)
$crit1
[1] -3.913778

$crit5
[1] -3.345434

$crit10
[1] -3.051473
```

UBC MFRE

Engle Granger Method

- **summary(ur.df(soydieselr$lnresid_ts, type = c("none"), lags = 3))@teststat**
- We must compare the t-statistic of -3.6603 with the critical values of -3.91 (1%), -3.34 (5%), -3.05 (10%). We can reject the null hypothesis of a unit root at the 5% level.
- We have evidence that the biodiesel and soybean oil prices are cointegrated

Roadmap

- Review - ADF Tests
- Cointegration
  - Engle Granger 2 step test
  - Engle Granger function to retrieve correct critical values
  - Johansen Procedure
- Error Correction Model

Testing for cointegration with multiple prices

```
jotest <- soydiesel[,c("lncrude", "lndiesel", "lnbio", "lnsoy")]
VARselect(jotest, lag.max = 10)
summary(ca.jo(jotest, type = c("trace"), ecdet = c("none"), K = 3, spec = c("transitory")))
```

```
######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.041681178 0.031194654 0.014892591 0.006756065

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 3 |  5.22  6.50  8.18 11.65
r <= 2 | 16.77 15.66 17.95 23.52
r <= 1 | 41.18 28.71 31.52 37.22
r = 0  | 73.96 45.23 48.28 55.43
```

Interpretation

- We work from the bottom row to to top, and stop only when it is no longer possible to reject the null
- r = 0 -> reject null of rank 0
- r <=1 -> reject null of rank <=1
- r <= 2 -> reject null at 90% confidence

Testing for cointegration with multiple prices

- Note that there is a difference in critical values across programs (read [here]() and [here]() for more info); trace statistic is the same though.
- But the important thing is that we reject the null that there is no cointegrating relationship.

```
. vecrank lncrude lndiesel lnbio lnsoy, lags(3)

                    Johansen tests for cointegration
Trend: constant                                    Number of obs =     770
Sample:  4 - 773                                            Lags =       3
---------------------------------------------------------------------------
                                                          5%
maximum                                        trace   critical
  rank    parms       LL        eigenvalue   statistic   value
    0      36     6397.9106         .         73.9584    47.21
    1      43     6414.3019      0.04168      41.1759    29.68
    2      48     6426.5032      0.03119      16.7734    15.41
    3      51     6432.2799      0.01489       5.2198     3.76
    4      52     6434.8899      0.00676
```

```
######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.041681178 0.031194654 0.014892591 0.006756065

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 3 |  5.22  6.50  8.18 11.65
r <= 2 | 16.77 15.66 17.95 23.52
r <= 1 | 41.18 28.71 31.52 37.22
r = 0  | 73.96 45.23 48.28 55.43
```

Roadmap

- Review - ADF Tests
- Cointegration
  - Engle Granger 2 step test
  - Engle Granger function to retrieve correct critical values
  - Johansen Procedure
- Error Correction Model

Error Correction Model

- So far we have established that
  - Biodiesel and soybean are are non-stationary
  - First difference of biodiesel and soybean oil are stationary, i.e. I(1)
  - Esimated $\mathbf{PBioDiesel_t = \alpha + \beta PSoyoil_t + \varepsilon_t}$
  - Collected the residuals and did an ADF test -> reject null hypothesis of a unit root
  - Biodiesel and soybean are cointegrated

- We now want to investigate how quickly prices return to the long run equilibrium if one or both variables experience a shock – error correction model

Error Correction Model

- $Pbiodiesel_t = \ \alpha + \beta Psoyoil_t + \ \varepsilon_t$
  - $\varepsilon_t = resid_t = \ Pbiodiesel_t \ - \alpha \ - \beta Psoyoil_t$
- Enhanced ECM
  - $\Delta\mathrm{Pbiodiesel_t} = \alpha + \ \sum_{i=1}^{k} \beta_i \, \Delta Pbiodiesel_{t-i} + \sum_{j=0}^{l} \gamma_i \, \Delta soyoil_{t-i} + \lambda * resid_{t-1} + \epsilon_t$
- In the lecture, you will learn that λ is the deviation from the LR equilibrium
  - Negative sign
  - Has a value between -1 and 0

Error Correction Model

- Step Pre: Determine the optimal number of lags (i.e., k).
- Step One: Regress y_{t-1} on x_{t-1} to estimate the long run relationship and then collect the residuals.
  - Or regress y_t on x_t, save the residuals, and lag the residuals
- Step Two: Estimate the general ECM with the appropriate number of lags.

Error Correction Model

- Step Pre: Determine the optimal number of lags
  - **VARselect(soydieselr$lnbio, lag.max = 5)** -> 4 lags
  - **VARselect(soydieselr$lnsoy, lag.max = 5)** -> 1 lag

- Step 1: regress y_t on x_t, save the residuals, and lag the residuals
  - **lag.xts(lnresid_ts)**

Error Correction Model

- Step Two: Estimate the general ECM with the appropriate number of lags.

```
ecm <- lm(diff.xts(lnbio, na.pad = T) ~
        lag.xts(diff.xts(lnbio, na.pad = T)) +
        lag.xts(lag.xts(diff.xts(lnbio, na.pad = T))) +
        lag.xts(lag.xts(lag.xts(diff.xts(lnbio, na.pad = T)))) +
        lag.xts(lag.xts(lag.xts(lag.xts(diff.xts(lnbio, na.pad = T))))) +
        diff.xts(lnsoy, na.pad = T) +
        lag.xts(diff.xts(lnsoy, na.pad = T)) +
        lag.xts(lnresid_ts), data = soydieselr)
```

4 lags for biodiesel

Include Δx_t -- see notes for discussion

1 lag for biodiesel

error correction term / speed of adjustment

Error Correction Model

```
Coefficients:
                                                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                                0.000178   0.001011   0.176   0.8603    
lag.xts(diff.xts(lnbio, na.pad = T))                      -0.212774   0.035473  -5.998 3.09e-09 ***
lag.xts(lag.xts(diff.xts(lnbio, na.pad = T)))             -0.025750   0.030394  -0.847   0.3971    
lag.xts(lag.xts(lag.xts(diff.xts(lnbio, na.pad = T))))     0.051046   0.029941   1.705   0.0886 .  
lag.xts(lag.xts(lag.xts(lag.xts(diff.xts(lnbio, na.pad = T))))) 0.022265   0.029963   0.743   0.4577    
diff.xts(lnsoy, na.pad = T)                                0.494722   0.029281  16.896  < 2e-16 ***
lag.xts(diff.xts(lnsoy, na.pad = T))                       0.186336   0.034879   5.342 1.21e-07 ***
lag.xts(lnresid_ts)                                       -0.071474   0.013759  -5.195 2.64e-07 ***
```

Summary

- Review - ADF Tests
- Cointegration
  - Engle Granger 2 step test
  - Engle Granger function to retrieve correct critical values
  - Johansen Procedure
- Error Correction Model