Lab 2
Stationarity and Cointegration

Case study of biodiesel fuel and soybean oil

UBC MFRE | Master of Food and Resource Economics

Roadmap

- Stationarity Tests - Levels
  - Dickey Fuller test
  - Augmented Dickey Fuller test
    - Lags
    - Flow chart of testing specification
- Stationarity Tests – First differences
- Cointegration
  - Engle Granger 2 step test
  - Engle Granger function to retrieve correct critical values
  - Johansen Procedure

Packages

- Load the following packages

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

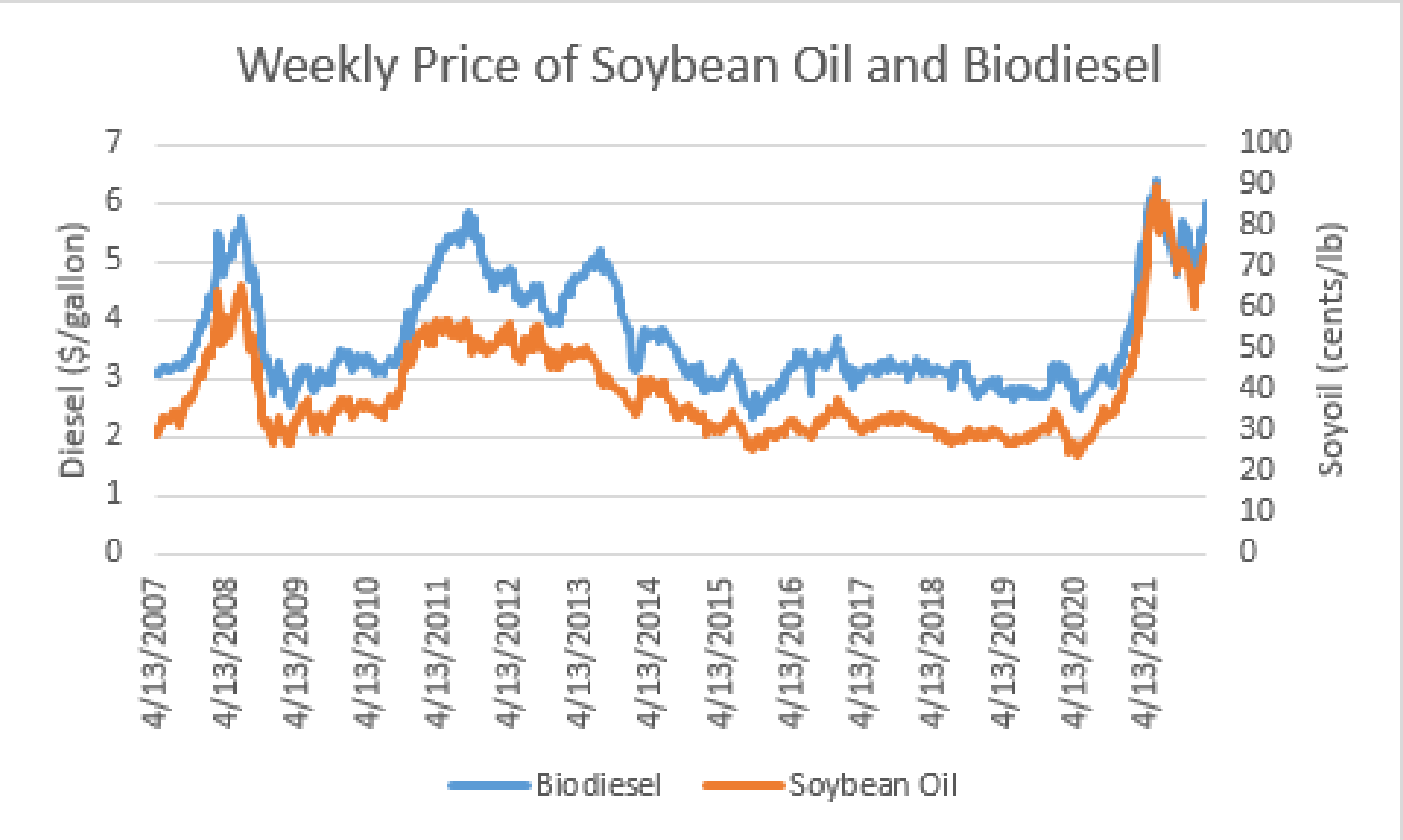
UBC MFRE

Data

- Data sources
  - Biodiesel and soybean oil – [Iowa Stata University]
  - Diesel – [U.S. EIA]
  - Crude – [U.S. EIA]
- Load *data_lecture2.csv* posted on Canvas

```
> head(data)
# A tibble: 6 x 5
  date      biodiesel soyoil diesel crude
  <chr>         <dbl>  <dbl>  <dbl> <dbl>
1 4/13/2007      3.1    29.9   2.88  62.6
2 4/20/2007      3.1    29.3   2.85  63.1
3 4/27/2007      3.08   30.2   2.81  65.3
4 5/4/2007       3.14   31.1   2.79  63.8
5 5/11/2007      3.14   31.1   2.77  61.9
6 5/18/2007      3.18   32.9   2.80  63.6
```

Data



Weekly Price of Soybean Oil and Biodiesel
Diesel ($/gallon)
Soyoil (cents/lb)
7
6
5
4
3
2
1
0
100
90
80
70
60
50
40
30
20
10
0
4/13/2007
4/13/2008
4/13/2009
4/13/2010
4/13/2011
4/13/2012
4/13/2013
4/13/2014
4/13/2015
4/13/2016
4/13/2017
4/13/2018
4/13/2019
4/13/2020
4/13/2021
Biodiesel
Soybean Oil

Data Cleaning

```
data <- data %>%
  mutate(date = mdy(date),
         lnbio = log(biodiesel),
         lnsoy = log(soyoil),
         lndiesel = log(diesel),
         lncrude = log(crude))

soydiesel <- xts(data[,c("biodiesel", "soyoil", "diesel", "lnbio", "lnsoy", "lndiesel", "lncrude")], order.by = data$date)
```

```
> 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
```

UBC MFRE

Dickey-Fuller Test

- General model: $\mathbf{y_t} = \boldsymbol{\alpha} + \boldsymbol{\delta t} + \boldsymbol{\rho y_{t-1}} + \boldsymbol{u_t}$
- We wish to test the null hypothesis that S_t is a random walk, which is equivalent to testing that y_t has a unit root ➔ $\rho = 1$

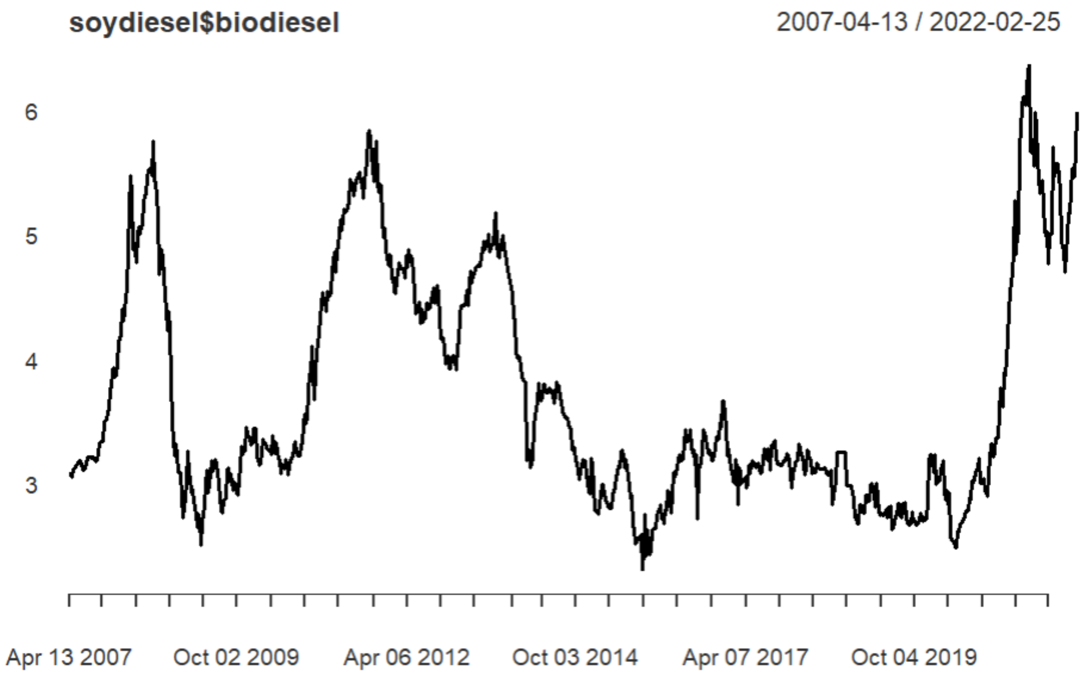
- The unit root test is known as the Dickey-Fuller test

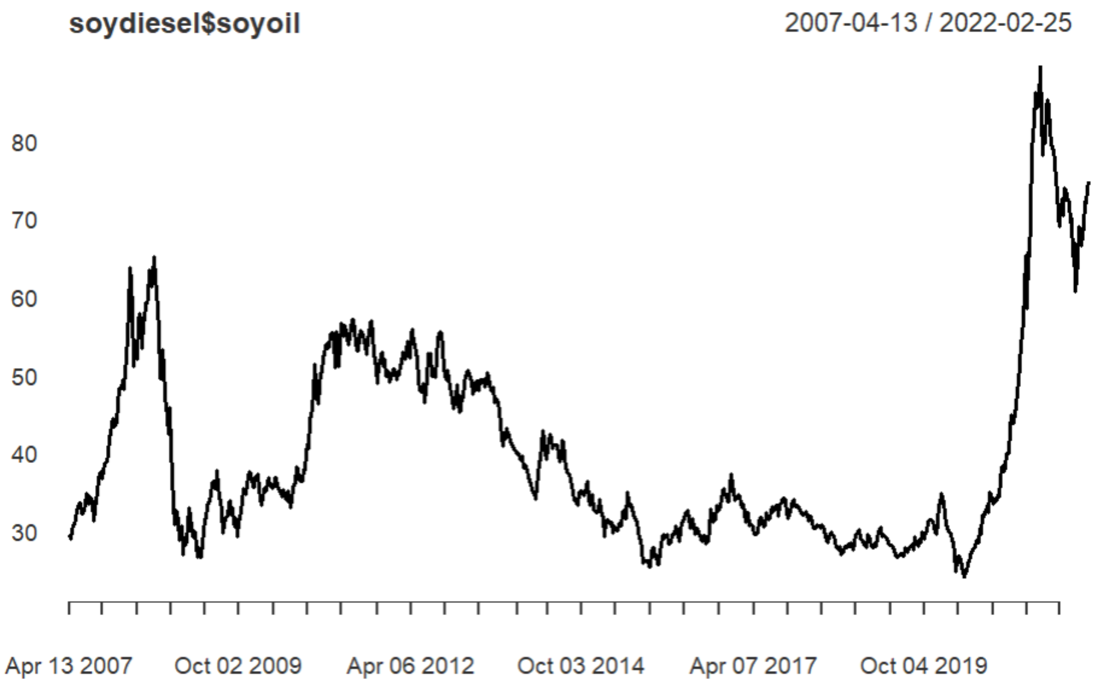
- We will use the ur.df() function of the {urca} package

ur.df(y, type = c("none”), lags = 0)

UBC MFRE

Dickey-Fuller Test





R – Dickey Fuller Test

df_biodiesel <- ur.df(soydiesel$biodiesel, type = c("none"), lags = 0)

summary(df_ biodiesel)

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.68384 -0.05700 -0.00171  0.06738  0.69698

Coefficients:
         Estimate Std. Error t value Pr(>|t|)
z.lag.1 0.0006016  0.0012122   0.496     0.62

Residual standard error: 0.1287 on 771 degrees of freedom
Multiple R-squared:  0.0003194, Adjusted R-squared:  -0.0009772
F-statistic: 0.2463 on 1 and 771 DF,  p-value: 0.6198


Value of test-statistic is: 0.4963

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

Interpretation

- Absolute value of τ_1 t-statistic is smaller than the absolute value of the critical values
- Fail to reject the null hypothesis of a unit root
- Biodiesel price is not stationray

UBC MFRE

Augmented Dickey-Fuller (ADF) test

$$y_t = \alpha + \delta t + \rho y_{t-1} + u_t$$

- In recent years an Augmented Dickey Fuller (ADF) test has been developed to account for potential autocorrelation in the residuals

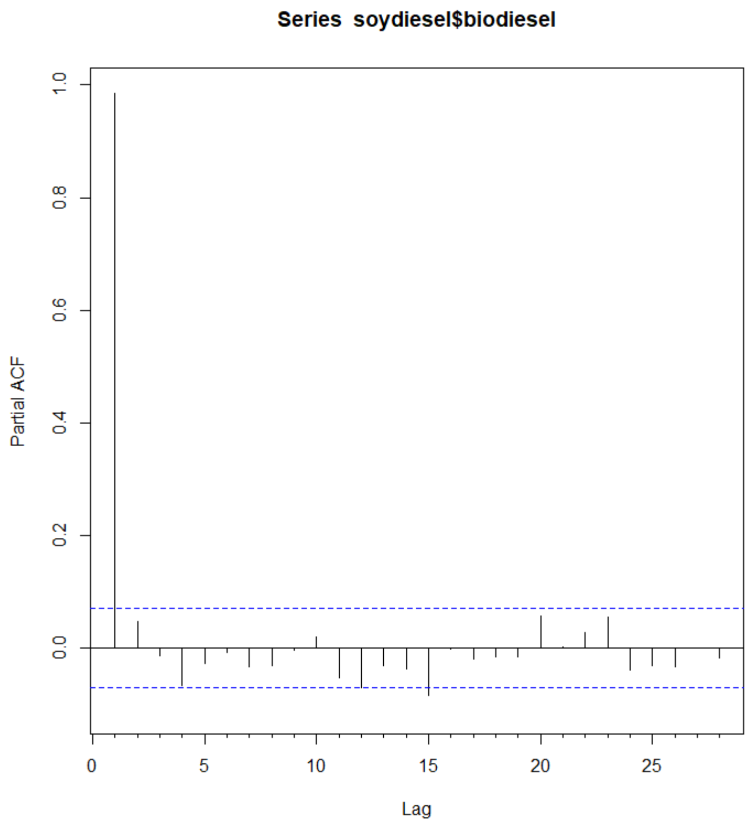
$$\Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \tau_i \Delta y_{t-i} + \varepsilon_t$$

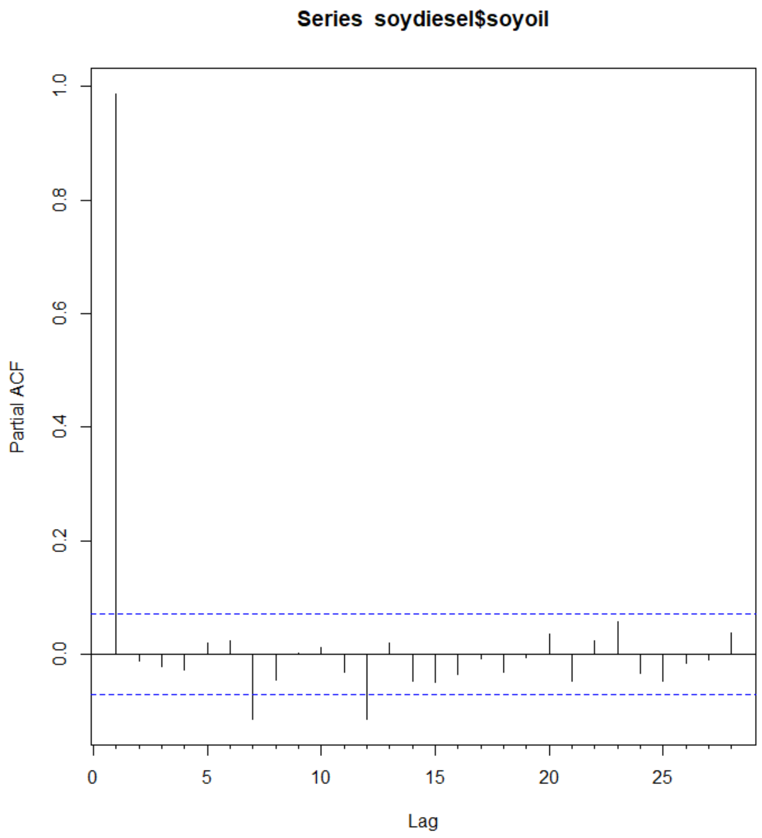
- We will use the ur.df() function of the {urca} package

```
ur.df(y, type = c("trend", "drift"), lags = 5, selectlags = c("AIC")
```

Lag length selection

- Partial autocorrelation function – **Pacf()**





Lag length selection

- {urca}'s automatic lag selection functionality
- **summary(ur.df(soydiesel$biodiesel, type = c("none"), lags = 4, selectlags = c("AIC")))**

```
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.69096 -0.05993  0.00111  0.06862  0.72149 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)  
z.lag.1      0.0004404  0.0012178   0.362   0.7178  
z.diff.lag1 -0.0268090  0.0362382  -0.740   0.4596  
z.diff.lag2  0.0530633  0.0362628   1.463   0.1438  
z.diff.lag3  0.0767265  0.0363048   2.113   0.0349 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1287 on 764 degrees of freedom
Multiple R-squared:  0.009252,  Adjusted R-squared:  0.004065 
F-statistic: 1.784 on 4 and 764 DF,  p-value: 0.1302


Value of test-statistic is: 0.3616 

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

Lag length selection

- VARselect() function of the {var} package
  - Select lag length with lowest AIC
- **VARselect(soydiesel$biodiesel, lag.max = 5)**

$criteria

	1	2	3	4	5
AIC(n)	-4.09451821	-4.09233331	-4.09269044	-4.09662299	-4.09465117
HQ(n)	-4.08986361	-4.08535141	-4.08338124	-4.08498650	-4.08068738
SC(n)	-4.08242501	-4.07419350	-4.06850403	-4.06638998	-4.05837156
FPE(n)	0.01666377	0.01670022	0.01669426	0.01662874	0.01666156

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

$criteria

	1	2	3	4	5
AIC(n)	0.9274930	0.9295556	0.9277867	0.9293476	0.9318206
HQ(n)	0.9321476	0.9365375	0.9370959	0.9409841	0.9457843
SC(n)	0.9395862	0.9476954	0.9519731	0.9595806	0.9681002
FPE(n)	2.5281630	2.5333832	2.5289060	2.5328567	2.5391284

1. Test β = 0 in full model with intercept and time trend →

$$\Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \tau_i \Delta y_{t-i} + \varepsilon_t$$

(use “trend” option and τ_3 test in R)

Reject (no unit root)

No Reject

2. Test significance of time trend (δ = 0) in full model. → Use “trend” option and Φ_3 test in R.

Reject (no unit root)

No Reject (remove and re-estimate) → Use the “drift” option to re-estimate without a trend.

3. Test β = 0 in model with intercept and not trend. → Use the τ_2 test in R.

Reject (no unit root)

No Reject

4. Test significance of constant (α = 0) → Use “drift” option and Φ_1 test in R.

Reject (no unit root)

No Reject (remove and re-estimate) → Use the “no constant” option to re-estimate without an intercept.

5. Test β = 0 Use τ_1 test in R.

Reject: no unit root

No Reject: unit root

ADF test – Step 1

$$\Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \tau_i \Delta y_{t-i} + \varepsilon_t$$

1. **Test β = 0 in full model with intercept and time trend**

(use “trend” option and τ_3 test in R)

Reject (no unit root)

No Reject

2. **Test significance of time trend (δ = 0) in full model.**

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

```
Value of test-statistic is: -1.419 1.0367 1.2969

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
```

Interpretation

- Absolute value of τ_3 t-statistic is smaller than the absolute value of the critical values
- Fail to reject β = 0 null hypothesis
- Proceed with Step 2

ADF test – Step 2

$$\Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \tau_i \Delta y_{t-i} + \varepsilon_t$$

2. Test significance of time trend (δ = 0) in full model. → Use “trend” option and Φ_3 test in R.

Reject (no unit root)

No Reject (remove and re-estimate) → Use the “drift” option to re-estimate without a trend.

3. Test β = 0 in model with intercept and not trend.

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

```
Value of test-statistic is: -1.419 1.0367 1.2969

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
```

Interpretation

- Absolute value of the Φ_3 t-statistic is smaller than the critical values
- Fail to reject the null hypothesis that time trend is not significant
- Proceed with Step 3

UBC MFRE

ADF test – Step 3

$$\Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \tau_i \Delta y_{t-i} + \varepsilon_t$$

Use the "drift" option to re-estimate without a trend.

3. Test β = 0 in model with intercept and not trend. → Use the τ_2 test in R.

Reject (no unit root)

No Reject

↓

4. Test significance of constant (α = 0)

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

```
Value of test-statistic is: -1.5242 1.4199

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

Interpretation
- We removed the time trend and retest the β = 0 null hypothesis.
- Absolute value of the τ_2 t-statistic is smaller than the critical values
- Fail to reject the β = 0 null hypothesis
- Proceed with Step 4

ADF test – Step 4

$$\Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \tau_i \Delta y_{t-i} + \varepsilon_t$$

4. Test significance of constant (α = 0) → Use “drift” option and Φ_1 test in R.

Reject (no unit root)

No Reject (remove and re-estimate) → Use the “no constant” option to re-estimate without an intercept.

5. Test β = 0 Use τ_1 test in R.

Reject: no unit root

No Reject: unit root

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

```
Value of test-statistic is: -1.5242 1.4199

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

Interpretation

- Absolute value of the Φ_2 t-statistic is smaller than the critical values
- Fail to reject the null hypothesis that the intercept term is not significant
- Proceed with Step 5

ADF test – Step 5

Use the “no constant” option to re-estimate without an intercept.

5. Test β = 0 Use τ_1 test in R.

- Reject: no unit root
- No Reject: unit root

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

```
Value of test-statistic is: 0.3367

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

Interpretation

- We removed the intercept and retest the β = 0 null hypothesis
- Absolute value of the τ_1 t-statistic is smaller than the critical values
- We fail to reject the null hypothesis
- We conclude that biodiesel has a unit root and is there non-stationary

ADF test – output in lecture notes

$Pbiodiesel_t$

Test Type	Test Statistic	1% Critical	5% Critical	10% Critical
No Constant	0.337	-2.580	-1.950	-1.620
Drift	-1.524	-3.430	-2.860	-2.570
Trend	-1.419	-3.960	-3.410	-3.120

Fail to reject unit root with all testing types and all three levels of significance.

$PSoyoil_t$

Test Type	Test Statistic	1% Critical	5% Critical	10% Critical
No Constant	0.722	-2.580	-1.950	-1.620
Drift	-0.815	-3.430	-2.860	--2.570
Trend	-0.783	-3.960	-3.410	-3.120

Fail to reject unit root with all testing types and all three levels of significance.

Roadmap

- Stationarity Tests - Levels
  - Dickey Fuller test
  - Augmented Dickey Fuller test
    - Lags
    - Flow chart of testing specification
- Stationarity Tests – First differences
- Cointegration
  - Engle Granger 2 step test
  - Engle Granger function to retrieve correct critical values
  - Johansen Procedure

First difference

- To take the difference, we use **diff.xts()**

biodiesel	d_biodiesel
<dbl>	<dbl>
3.1	NA
3.1	0
3.08	-0.02
3.14	0.06
3.14	0
3.18	0.03

ADF test – Step 1 (biodiesel)

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

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

```
Value of test-statistic is: -12.4258 51.488 77.2258

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
```

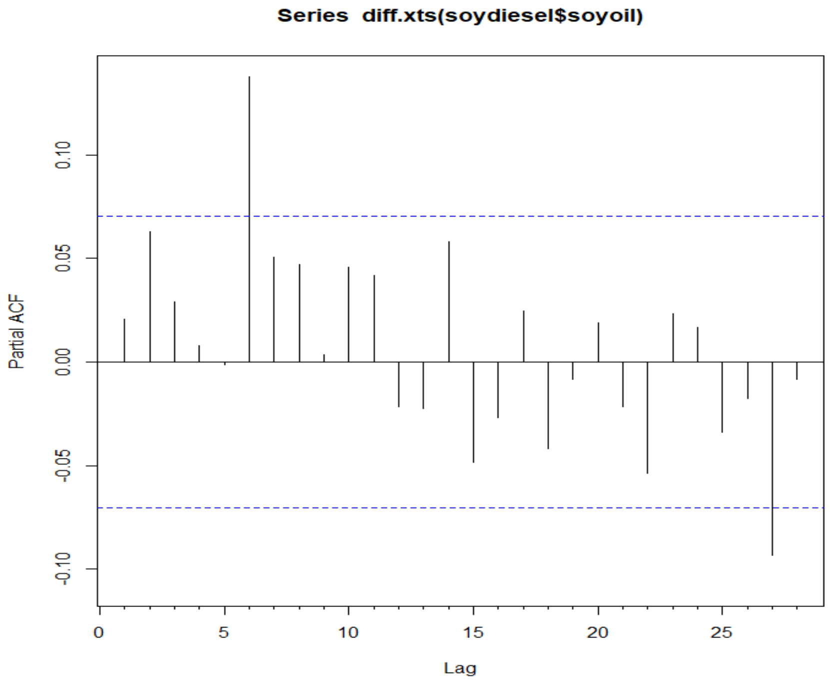
Interpretation

- Absolute value of τ_3 t-statistic is bigger than the absolute value of the critical values
- Reject the unit root null hypothesis
- Differenced price series is I(0) stationary
- Not necessary to proceed with next steps
- If you do, you will still arrive at the same conclusion

UBC MFRE

ADF test – Step 1 (soybean oil)

- R's VARselect() function does not allow for lags = 0. Output suggests lags = 2
- Stata's varsoc function does, and the output suggests lags = 0
- Partial autocorrelation function suggests lags = 0




```
Value of test-statistic is: -27.1734 246.1313 369.1964

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
```

Interpretation
- Differenced price series is I(0) stationary

Roadmap

- Stationarity Tests - Levels
    - Dickey Fuller test
    - Augmented Dickey Fuller test
        - Lags
        - Flow chart of testing specification
- Stationarity Tests – First differences
- Cointegration
    - Engle Granger 2 step test
    - Engle Granger function to retrieve correct critical values
    - Johansen Procedure

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_biodieselsoy <- lm(biodiesel ~ soyoil, data = soydiesel)

summary(reg_biodieselsoy)

```
Residuals:
     Min       1Q   Median       3Q      Max 
-1.05258 -0.18850 -0.06624  0.14424  1.30709 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.0332423  0.0400385   25.81   <2e-16 ***
soyoil      0.0664150  0.0009456   70.23   <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_soydiesel <- resid(reg_biodieselsoy)
resid_ts <- xts(resid_soydiesel, order.by = index(soydiesel))
soydieselr <- merge.xts(soydiesel, resid_ts)
```

	biodiesel	soyoil	diesel	lnbio	lnsoy	lndiesel	lncrude	resid_ts
2007-04-13	3.100	29.900	2.877	1.131402	3.397858	1.056748	4.136446	0.08094869
2007-04-20	3.100	29.310	2.851	1.131402	3.377929	1.047670	4.144087	0.12013355
2007-04-27	3.075	30.190	2.811	1.123305	3.407511	1.033540	4.178379	0.03668834
2007-05-04	3.140	31.130	2.792	1.144223	3.438172	1.026758	4.156067	0.03925822
2007-05-11	3.140	31.060	2.773	1.144223	3.435921	1.019930	4.125520	0.04390728
2007-05-18	3.175	32.865	2.803	1.155308	3.492408	1.030690	4.152771	-0.04097183

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$resid_ts)

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

```
Value of test-statistic is: -3.0262

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

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

$crit5
[1] -3.345434

$crit10
[1] -3.051473
```

Engle Granger Method

- We must compare the t-statistic of -3.0262 with the critical values of -3.91 (1%), -3.34 (5%), -3.05 (10%). The absolute value of the test statistic is smaller than the absolute values of the critical values, so we fail to reject the null hypothesis of a unit root.
- We find no evidence that biodiesel fuel and soybean oil prices are cointegrated

```
Value of test-statistic is: -3.0262

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

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

$crit5
[1] -3.345434

$crit10
[1] -3.051473
```

Retest for cointegration using log prices

- We will now work with log of prices – address skewed nature of price data

- Step 1
  - **reg_lnbiodieselsoy <- lm(lnbio ~ lnsoy, data = soydiesel)**
  - **summary(reg_lnbiodieselsoy)**

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

  - **soydieselr <- merge.xts(soydiesel, lnresid_ts)**

UBC MFRE

Retest for cointegration using log prices

- Step 2
  - **VARselect(soydieselr$lnresid_ts)**
  - **summary(ur.df(soydieselr$lnresid_ts, type = c("none"), lags = 3))**
  - **englegranger(2, 0, 773)**

- We can reject the null hypothesis of a unit root at the 95% confidence level. We have evidence that the pair of prices are cointegrated.

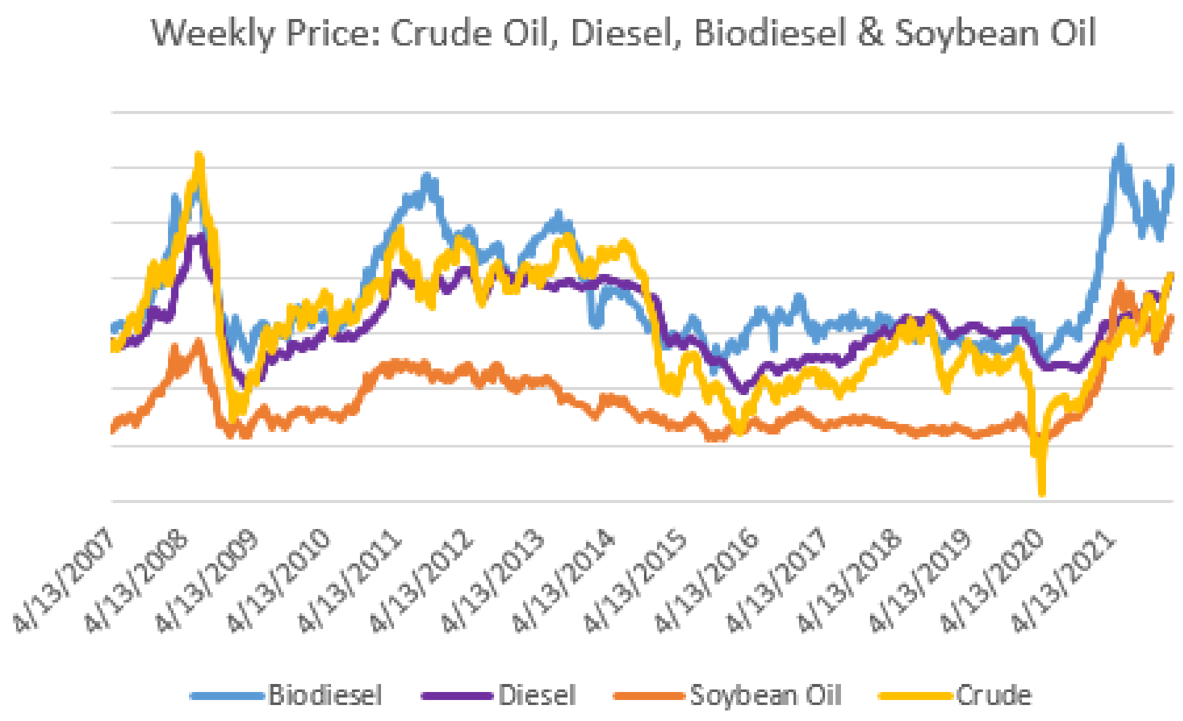
```
Value of test-statistic is: -3.6603

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

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

$crit5
[1] -3.345434

$crit10
[1] -3.051473
```

UBC MFRE

Testing for cointegration with multiple prices

- Let's say we want to test for the cointegration of crude oil, diesel, biodiesel, and soybean oil
- We use the Johansen Procedure to test the cointegration of more than two data series.



UBC MFRE

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

UBC MFRE

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
```

Summary

- Stationarity Tests – DF and ADF tests in R
  - VARselect(y, lag.max = n)
  - ur.df(y, type = c(“none”, “trend”, “drift”), lags = n)
  - ur.df(diff.xts(y, na.pad = F), type = c(“none”, “trend”, “drift”), lags = n)

- Cointegration
  - Step 1 – estimate the long run relationship between prices
  - Collect the residuals
  - Step 2 – conduct an ADF test on residuals and use correct critical values
  - Johansen Procedure
    - Tests for cointegration of more than 2 data series