Engle-Granger Method Tutorial

Tyler J. Brough

April 6, 2017

Introduction

Please refer to your notes for the details of the Engle-Granger Method. I will not repeat those details here, but rather show you how to carry out some calculations in R.

We will again use the S&P 500 data.

```
raw.data <- read.csv("SandPHedge.csv", header=T)
head(raw.data)
```

```
##     Date    Spot Futures
## 1 Feb-02 1106.73  1106.9
## 2 Mar-02 1147.39  1149.2
## 3 Apr-02 1076.92  1077.2
## 4 May-02 1067.14  1067.5
## 5 Jun-02  989.82   990.1
## 6 Jul-02  911.62   911.5
```

```
attach(raw.data)
```

```
## The following objects are masked from raw.data (pos = 4):
##
##     Date, Futures, Spot
```

```
## The following objects are masked from raw.data (pos = 5):
##
##     Date, Futures, Spot
```

We will again calculate log-prices, and log-differenced prices:

```
ln.spot <- log(Spot)
ln.futures <- log(Futures)
ln.spot.diff <- diff(ln.spot)
ln.futures.diff <- diff(ln.futures)
```

Testing for Unit Roots

Using the URCA package we will test for Unit Roots.

```
library(urca)

# Test for unit roots in log-prices
adf.s <- ur.df(y=ln.spot, type = "drift", selectlags = "BIC")
summary(adf.s)
```

```
##
## ###############################################
```

```
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.164789 -0.020013  0.007188  0.025378  0.113151
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.25057    0.16695   1.501   0.1358
## z.lag.1     -0.03515    0.02360  -1.489   0.1388
## z.diff.lag   0.21240    0.08657   2.453   0.0155 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04512 on 130 degrees of freedom
## Multiple R-squared:  0.05235,    Adjusted R-squared:  0.03777
## F-statistic: 3.591 on 2 and 130 DF,  p-value: 0.03035
##
##
## Value of test-statistic is: -1.4893 1.238
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

```r
adf.f <- ur.df(y=ln.futures, type = "drift", selectlags = "BIC")
summary(adf.f)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.169865 -0.021061  0.007673  0.023368  0.117402
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.24522    0.16451   1.491   0.1385
## z.lag.1     -0.03441    0.02326  -1.479   0.1415
```

```
## z.diff.lag   0.21671    0.08644   2.507   0.0134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04465 on 130 degrees of freedom
## Multiple R-squared:  0.05398,    Adjusted R-squared:  0.03942
## F-statistic: 3.709 on 2 and 130 DF,  p-value: 0.02714
##
##
## Value of test-statistic is: -1.4792 1.2201
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

From which it is easy to see that we fail to reject the null hypothesis that there is a Unit root in log-prices.

Let's now check log-price differences:

```r
adf.s.diff <- ur.df(y=ln.spot.diff, type = "drift", selectlags = "BIC")
summary(adf.s.diff)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.164723 -0.021056  0.007511  0.024427  0.110682
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.002613   0.003940   0.663    0.508
## z.lag.1     -0.842050   0.110909  -7.592 5.58e-12 ***
## z.diff.lag   0.050156   0.087072   0.576    0.566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04517 on 129 degrees of freedom
## Multiple R-squared:  0.4059, Adjusted R-squared:  0.3967
## F-statistic: 44.08 on 2 and 129 DF,  p-value: 2.58e-15
##
##
## Value of test-statistic is: -7.5923 28.8334
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
```

```
## phi1  6.52  4.63  3.81
```

```
adf.f.diff <- ur.df(y=ln.futures.diff, type = "drift", selectlags = "BIC")
summary(adf.f.diff)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.170200 -0.020380  0.006299  0.023271  0.115766
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.002558   0.003898   0.656    0.513
## z.lag.1     -0.826602   0.110530  -7.478 1.02e-11 ***
## z.diff.lag   0.038257   0.087013   0.440    0.661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04468 on 129 degrees of freedom
## Multiple R-squared:  0.4027, Adjusted R-squared:  0.3935
## F-statistic: 43.49 on 2 and 129 DF,  p-value: 3.659e-15
##
##
## Value of test-statistic is: -7.4785 27.9765
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

From which we can easily see that the null hypothesis of a Unit root for log-price differences is strongly rejected.

We conclude that log-prices are Unit root nonstationary, but log-price differences are stationary!

The Engle-Granger Two Step Procedure

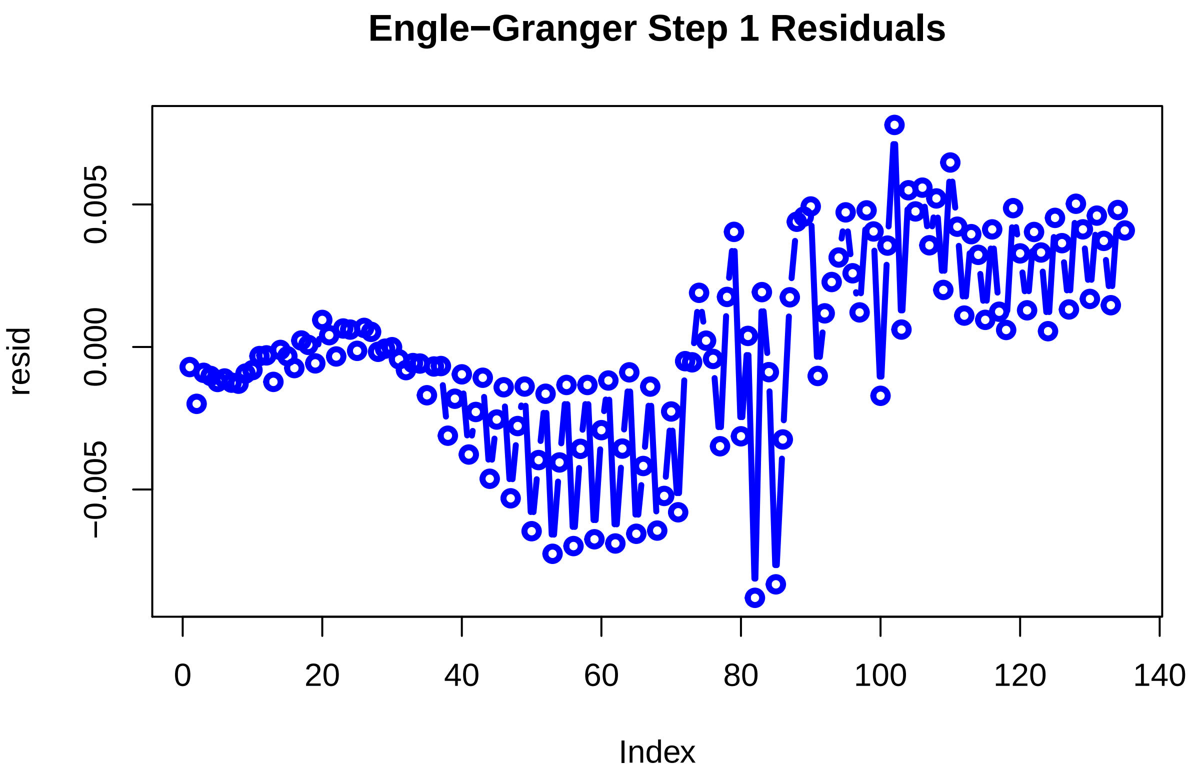
Now let's run the first cointegrating regression and get the residuals and plot them:

```
fit <- lm(ln.spot ~ ln.futures)
summary(fit)
```

```
##
## Call:
## lm(formula = ln.spot ~ ln.futures)
##
## Residuals:
```

```
##        Min         1Q     Median         3Q        Max
## -0.0088011 -0.0014023 -0.0003197  0.0021394  0.0077900
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.025310   0.012060   2.099   0.0377 *
## ln.futures  0.996468   0.001704 584.666   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003362 on 133 degrees of freedom
## Multiple R-squared:  0.9996, Adjusted R-squared:  0.9996
## F-statistic: 3.418e+05 on 1 and 133 DF,  p-value: < 2.2e-16
```

```r
resid <- fit$residuals
plot(resid, type = "b", lwd = 3, col = "blue", main = "Engle-Granger Step 1 Residuals")
```



Now let's test the residuals for the presence of a Unit root using the ur.df (Augmented Dickey-Fuller) test as above:

```r
adf.resid <- ur.df(resid, type = "drift", selectlags = "BIC")
summary(adf.resid)
```

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

```
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0086379 -0.0010582  0.0000525  0.0017935  0.0068720
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.153e-05  2.295e-04   0.225 0.822692
## z.lag.1     -2.876e-01  7.790e-02  -3.692 0.000327 ***
## z.diff.lag  -3.164e-01  8.389e-02  -3.771 0.000245 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002646 on 130 degrees of freedom
## Multiple R-squared:  0.2889, Adjusted R-squared:  0.278
## F-statistic: 26.41 on 2 and 130 DF,  p-value: 2.375e-10
##
##
## Value of test-statistic is: -3.6917 6.8545
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

You can see that there is fairly strong evidence of stationarity between the S&P 500 and the Futures contract! That is, at all levels of significance we reject the null hypothesis of a Unit root.

We tentatively conclude that there is a cointegrating relationship between the S&P 500 and the S&P 500 Futures contract.

The cointegrating vector is obtained as $[1, -\hat{\beta}]$ from the first-step cointegrating regression above. So in this case: $\hat{\beta} = 0.99646$. So for all practical purposes the cointegrating vector is $[1, -1]$.

Question: how is this estimate related to the minimum-variance hedge obtained from the standard procedure using OLS regression?

NB: Step two in the Engle-Granger Procedure is to now use the residuals as a plug-in to the error-correction model. We will not do that here, and you need not do so for your midterm!

Good luck!