1 ARALIK DERSİ

STATA

webuse lutkepohl2 (common file)(dataset)

plot a graph
graphics -> twoway graph ->create -> select Y variable as ln_consump and X variable as qtr ->For basic plots, select line ->accept -> ok

Grafik önümüze çıkıyor, smooth olduğu için random walk with a drift and trend'I seç

Statistics -> time series->tests->augmented dickey fuller unit root test->variable is ln_consump -> include trend in the regression -> ok

Suppress constant term in the regression = pure random walk
Include trend in the regression = random walk with drift and trend
Include drift term in the regression = random walk with drift

If test statistics is more negative than 5% critical value, you reject the null hypothesis
Null hypothesis = there is unit root

In our example, ln_consump has a unit root bc test statistics is not more negative than 5% critical value

If reported test statistics is not negative, do not continue this test

If you choose pure random walk which is a smaller version, most probably reported test statistics will be positive

Even if we use random walk with drift and trend, we can get a positive test statistics value, the reason for that is seasonality(mevsimsellik)

To solve this problem, you can deseasonalize the data. Or you can use another test. Turksat'ta mevsimsellikten arındırılmış data var, onu kullanabiliriz.

Testing the unit root of ln_consump

$$\Delta ln\_consump(t)= \underbrace{a0}_{\text{Constant}}+\underbrace{a1*t}_{\text{Trend term(must be added)}}+ \theta ln\_consump(t-1)+u(t)$$

For this purpose, we need to create/generate trend term in stata
gen t =_n

data editor(edit)- click on this

D.ln_consump yaparsan delta yapmışsın gibi oluyor, first difference'I alıyor
D1.ln_consump yaparsan da aynı

DD.ln_consump = D2.ln_consump

L.ln_consump yaparsak da ilk lagini alıyor
L2.ln_consump yaparsak ikinci lagini alıyor

HOW TO WRITE REGRESSION IN STATA

regress dependent variable expl.var1 expl.var2

böyle yazarsak constant otomatik olarak ekleniyor

örnek regress Y X Z yazarsak

Y(t)= β(0)+ β(1)X(t)+ β(2)Z(t)+u(t)

Note: If you don’t want to include constant term, regress Y X Z, no constant

Operators are not case sensitive, but variables are

regress d.ln_consump t L.ln_consump

```
      Source |       SS       df       MS              Number of obs =      91
-------------+------------------------------           F(  2,    88) =    1.41
       Model |  .000336465     2  .000168233           Prob > F      =  0.2488
    Residual |  .010475315    88  .000119038           R-squared     =  0.0311
-------------+------------------------------           Adj R-squared =  0.0091
       Total |  .010811781    90  .000120131           Root MSE      =  .01091

------------------------------------------------------------------------------
 D.ln_consump|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -.0000583   .0006614    -0.09   0.930    -.0013727    .0012561
  ln_consump |
         L1. |  -.0007525   .0332704    -0.02   0.982    -.0668705    .0653655
       _cons |   .0266234    .199203     0.13   0.894    -.3692507    .4224976
------------------------------------------------------------------------------
```

tau

same

wrong

```
. dfuller ln_consump, trend lags(0)

Dickey-Fuller test for unit root                   Number of obs   =        91

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.023            -4.060            -3.459            -3.155
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9940
```

Yanlış olmasının nedeni stata unit root testing yaptığımızı bilmiyor, sadece regressionla ilgileniyor

To carry out autocorrelation test, Breush Godfrey (BG)-LM test

estat bgodfrey

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	0.541	1	0.4619

H0: no serial correlation

Hence there is no autocorrelation in the auxiliary regression. So the reported dfuller test can be trusted. This test is an application of DF test beacuse we did not need to add any lags to solve the AC problem.

Phillips Perron Test (PP)

This is a DF variant test

Null hypothesis: there is unit root
Main advantage of this test, you don't need to add any lags to the auxiliary regression to solve autocorrelation. Automatically, PP test uses the AC robust standard errors so that the reported tests values become valid even if there is autocorrelation

```
. pperron ln_consump, trend

Phillips-Perron test for unit root                Number of obs   =        91
                                                  Newey-West lags =         3
```

	Test Statistic	Interpolated Dickey-Fuller: 1% Critical Value	5% Critical Value	10% Critical Value
Z(rho)	-1.065	-27.094	-20.538	-17.374
Z(t)	-0.323	-4.060	-3.459	-3.155

MacKinnon approximate p-value for Z(t) = 0.9890

Since -0.323 is not more negative than reported 5% critical value (-3.459); we do not reject Ho

not rejected ← Ho. There is unit root in ln_consump

DF-GLS UNIT ROOT TEST

Advantage: If your data(variable) shows a clear trend, this trend reported as more powerful in the literature based on some monte carlo studies about its power.

Go to statistics ->time series ->tests->DF GLS test for a unit root->variable is ln_consump
Presentle başlayan seçeneği seçiyor olabiliriz

KPSS TEST FOR UNIT ROOT

This is not a DF variant test. Its null hypothesis is "there is no unit root". Since it is not a DF variant test, you do not need to check if the tau statistics is negative. The test statistics of KPSS is positive.

Generally, it is advised to report this test together with DF test (or DF variant tests) because it is a completely different test. It looks to the situation from a different point of view so drawing the same conclusion with this test and, say, DF test is generally accepted a better result (more trustable)

Unfortunately, this is not included in the core installation of Stata. You need to install it. This is a add-in (ado file) written by a researcher. To install any add-in (ado files) in Stata you need to run following commend: SSC install the name of commend
SSC install KPSS

kpss ln_consump, qs auto ile run edin dedi

Test statistics is far from the 5% critical value, reject the null hypothesis
The null hypothesis was there is no unit root, so there is unit root

THE ORDER OF INTEGRATION

If you need to take first difference of a variable to remove the unit root in this series, then this series is called "series which integrated of first order" and say for any Y(t), it is dusted by

$$Y(t) \sim I(1)$$

Y(t) has a unit root if you take first difference, it will not have unit root. If taking first difference is not enough to remove unit root, you take another difference.

$\Delta\Delta Y(t)$ or $\Delta^2 Y(t)$

Generally, economic variables are expected to be integrated of order 1. In some cases, it may be $I(2)$. But never $I(3)$ or more.

Remark: For any variable, without unit root, we don't need to take any difference.

$$Y(t) \sim I(0)$$

Why do we care about the order of integration?

We care it because to check if there is cointegration (LR relationship in economics)
If we have 3 variables in the regression, all of them must be same order of integration

For stata,

dfuller ln_consump, trend lags(0)
this tests only show ln_consump has a unit root

graphics -> two way graph-> we have plot 1, we can edit it, change the Y variable to d.ln_consump->accept->ok
statistics->time series->augmented dickey fuller unit root test->change the Y variable to d.ln_consum and click only to suppress constant term in the regression bc pure random walk is the best for our situation
test statistics is more negative than 5%critical value so you reject the null hypothesis so there is no unit root in Δln_consump

which means

$$\Delta ln\_consump \sim I(0)$$
$$ln\_consump \sim I(1)$$

PP Test

Statistics-> time series ->tests->Phillips Perron unit root test -> change the Y variable to d.ln_consum and click only to suppress constant term in the regression

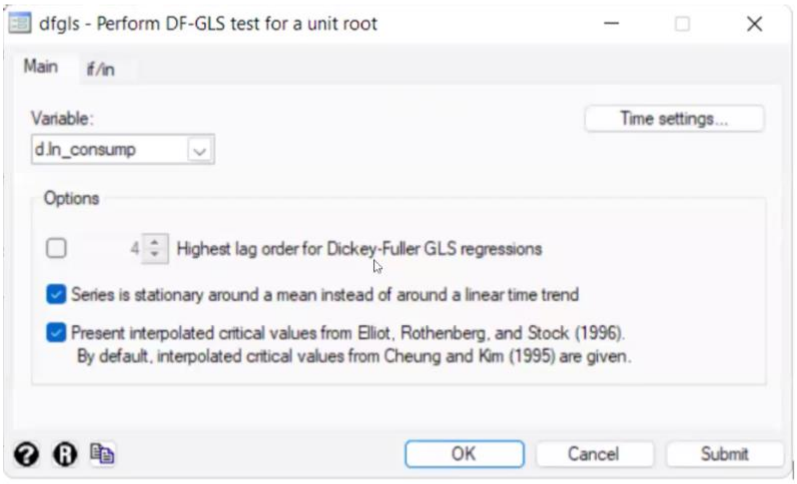
Focus on just the Z(t) column, test statistics is more negative than 5%critical value, reject the null hypothesis which means

$$\Delta ln\_consump \sim I(0)$$
$$ln\_consump \sim I(1)$$

DF-GLS

Go to statistics ->time series ->tests->DF GLS test for a unit root->variable is ln_consump

Trend yoksa buna bas



Focus on the 9th row

-1.412 is not more negative than 5%critical value so we do not reject the null hypothesis so there is still unit root. Then we need to check ΔΔln_consump for unit root

[lags]	DF-GLS mu Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
11	-0.454	-2.604	-1.950	-1.610
10	-0.496	-2.604	-1.950	-1.610
9	-0.539	-2.604	-1.950	-1.610
8	-0.661	-2.604	-1.950	-1.610
7	-0.597	-2.604	-1.950	-1.610
6	-0.660	-2.604	-1.950	-1.610
5	-0.797	-2.604	-1.950	-1.610
4	-0.977	-2.604	-1.950	-1.610
3	-1.449	-2.604	-1.950	-1.610
2	-2.101	-2.604	-1.950	-1.610
1	-4.524	-2.604	-1.950	-1.610

```
Opt Lag (Ng-Perron seq t) =  9 with RMSE  .0104703
Min SC   = -8.610335 at lag  6 with RMSE  .0111017
Min MAIC = -8.859013 at lag  9 with RMSE  .0104703
```

-0.539 is not more negative than 5%critical value so we do not reject the null hypothesis so there is still unit root. Then we need to check ΔΔΔln_consump for unit root. But we do not do that, theoretically not accepted

KPSS

kpss d.ln_consump, qs auto

```
KPSS test for D.ln_consump

Automatic bandwidth selection (maxlag) = 3
Autocovariances weighted by Quadratic Spectral kernel

Critical values for H0: D.ln_consump is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    3             .202
```

Test statistics is far from the 5% critical value, reject the null hypothesis
The null hypothesis was there is no unit root, so there is unit root

```
10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    3             .202
```

falls beyond the 5% critical value (0.146)

So we reject H0.

H0: Δln_consump has no unit root

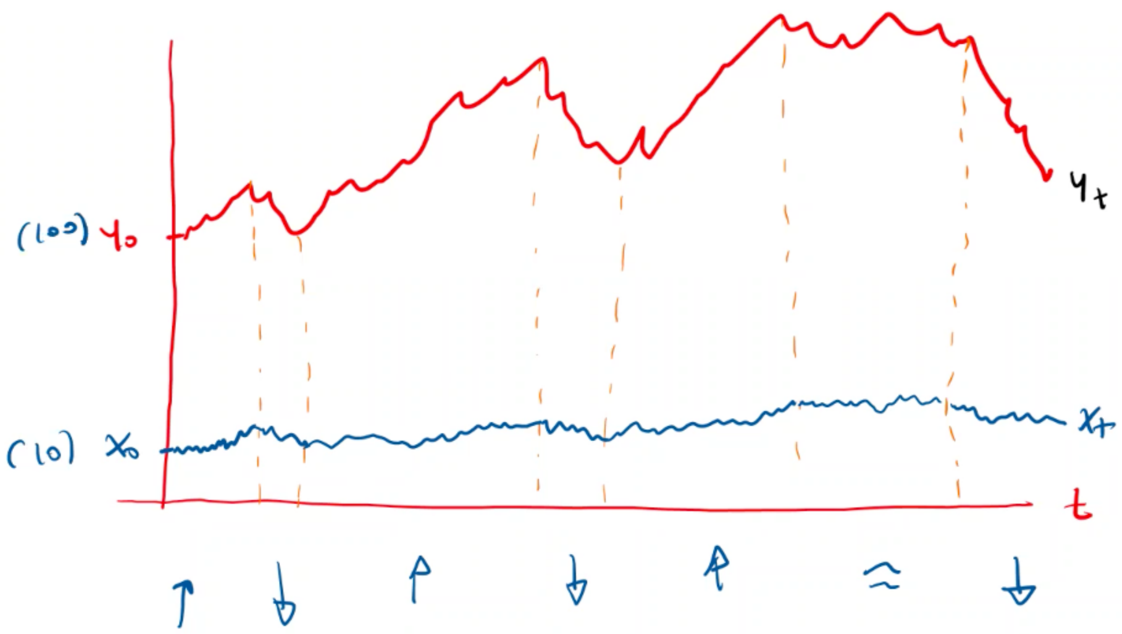
we reject it. Hence Δln_consump has still unit root we need to check ΔΔΔln_consump for unit root

```
. kpss d2.ln_consump, qs auto

KPSS test for D2.ln_consump

Automatic bandwidth selection (maxlag) = 3
Autocovariances weighted by Quadratic Spectral kernel

Critical values for H0: D2.ln_consump is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    3            .067
```

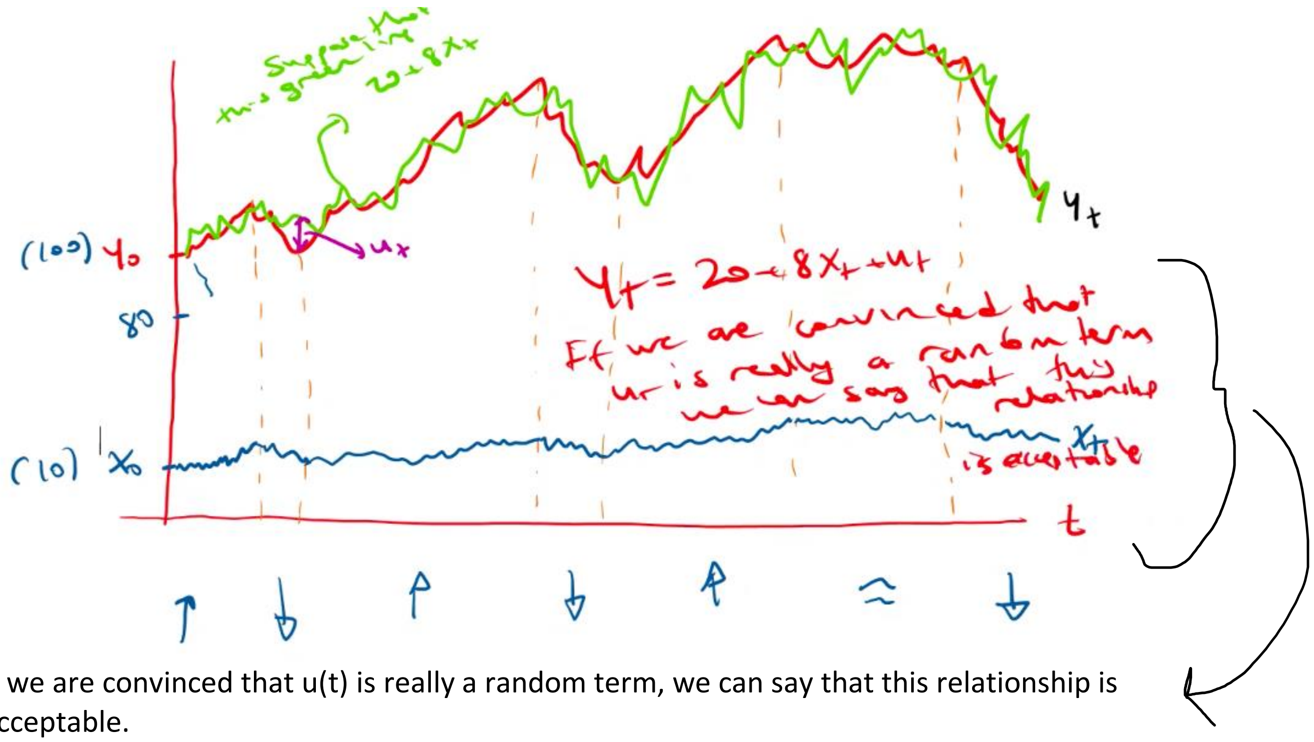
Lag order   Test statistic
    3   .067

does not fall beyond the 5% critical value (0.146). So we do not Reject the null hypothesis that there is no unit root

8 ARALIK DERSİ

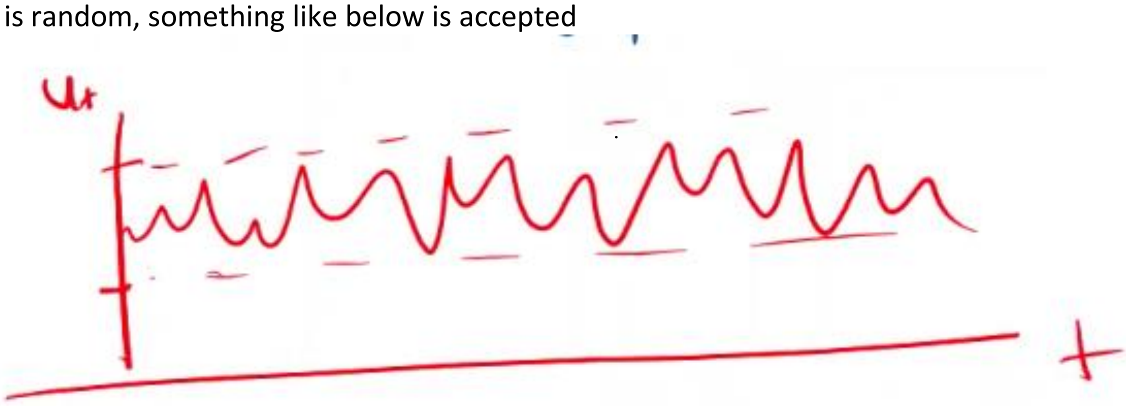


If we multiple X by 8 and add 20, it will be the green line



If we are convinced that u(t) is really a random term, we can say that this relationship is acceptable.

If u(t) is random, something like below is accepted



This is typical I(0) variable plot. I(0) means a variable without unit root.

So test if u(t) has the I(0) pattern, we can use u(t) hat series since u(t) will always be unknown.

Summarizing:

1) You need to find I(1) series. (If you need to take 1 difference of a Y(t) series to remove its unit root then this series is I(1) series)
   Say X(t) and Y(t) (generally check the economic theory to decide them). You regress Y(t) (the variable that the economic theory says that it is dependent) on X(t) variable by usual OLS. You save the residuals of this regression

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

The residual will be u(t) hat

Then you test if the u(t) hat has a unit root by using ADF test. If you conclude that u(t) hat doesn't have a unit root which implies that u(t) hat is a I(0) series.

You say that there is a long run equilibrium relationship between X(t) and Y(t), and the estimated equation which is below

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

is this long run equilibrium relationship (in other words, this is the cointegration equation) and you can use this equation for policy scenarios.

If you get the idea behind this procedure, you get the intuition behind Engle Granger Two Step Cointegration Analysis.

Engle Granger Two Step Cointegration Analysis's first step bu

The second step is checking if an error correction mechanism (going to a new equilibrium after a shock) or not

If this ECM (error correction mechanism) is validated, we trust more on the finding that there is a long run relationship(cointegration)

If ECM is not verified, then do not trust to the finding of first step.

Hoca odtuclass'a iki tane dosya yüklemiş bu derslerin altına, ikisi de aynı
Onlardan birini indir, stata'yı aç. File -> dosyayı seç

Cointegration analysis between short and long run interest rate in Euro Area

1) Firstly, we need to check if they are integrated of order 1.
   For this purpose, we can use ADFD, PP, KPSS, DF-GLS tests etc. We can also use Zivot Andrews test. This test takes into account one possible break in the data. For simplicity, we use this test. We need to install Zivot Andrews test, to do that write "ssc install zandrews" and run it.

Hoca bu şekilde kullanmayı önerdi, trim önemli değil

```
Examples

    . webuse turksales

    . dfuller sales
    . zandrews sales, graph

    . zandrews sales, break(trend)

    . zandrews sales, break(both) trim(0.10)

    . zandrews sales, lagmethod(BIC)

    . zandrews D.sales, graph

    . webuse grunfeld

    . zandrews invest if company==3, break(trend) graph
```

zandrews STN, break(both)
böyle yazınca hata alıyoruz, time variable'ını define etmemiz lazım, elimizde sadece quarter data olduğu için şu şekilde define ediyoruz →tsset quarter
tsset'in anlamı time series set

tekrar commandi yazıyoruz → zandrews STN, break(both) ve çalışıyor

hoca örnek olarak şunu dedi
gen time= _n yazarsak adı time olan 1den başlayan bir variable yaratıyoruz

```
. zandrews STN, break(both)

Zivot-Andrews unit root test for  STN

Allowing for break in both intercept and trend

Lag selection via TTest: lags of D.STN included = 1

Minimum t-statistic -5.500 at 1979q2  (obs 38)

Critical values: 1%: -5.57 5%: -5.08 10%: -4.82
```

To reject the null hypothesis, the t-statistic must be more negative then 5%critical value.

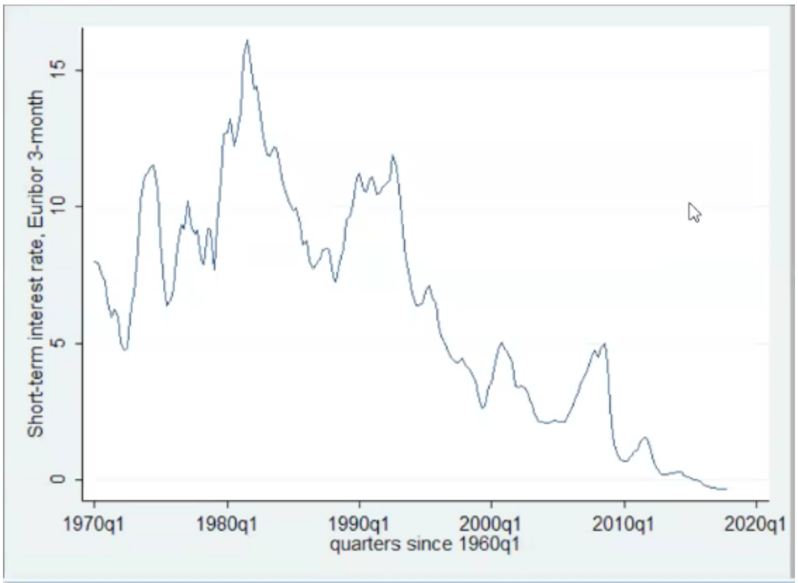
In this case, it is more negative, so we reject the null hypothesis. So, STN series doesn't have a unit root.

Here, we carry out the unit root under structural break (Zivot Andrews). The result of this test that STN variable is I(0). Under this conclusion, we cannot continue to do cointegration analysis since the variables in EG cointegration analysis must be I(1)

In fact, we used zandrews, just for an application of a unit root test under structural break.

Let us first check if there is really structural break over its time plot.

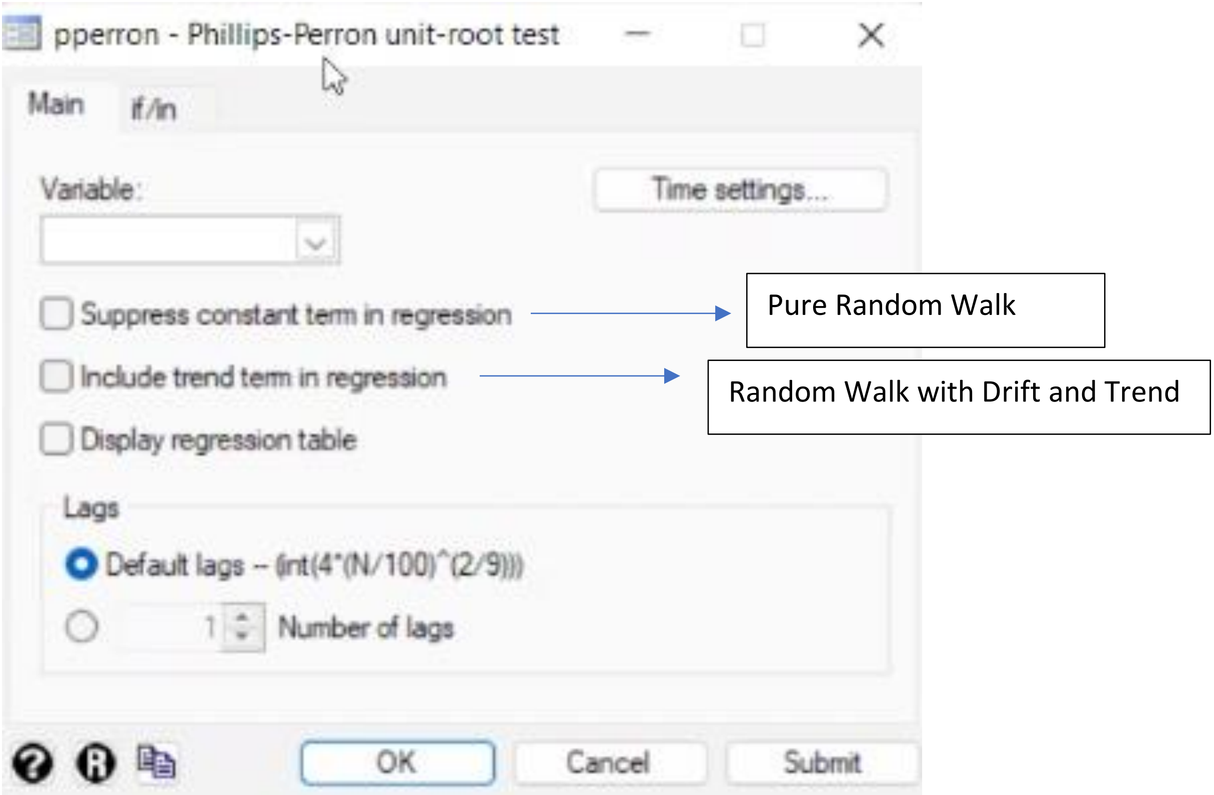
Graphics → Twoway graph → Create → Basic plots: Line → Y variable = STN, X variable= quarter → accept



This is quite obvious there is unit root in this variable because there is upward and downward trends which is stochastic process characteristic and deviations are large which is also result of a unit root. Also there is no point that shows there is a structural break.
Using Zandrews does not seem a good choice here since we do not observe any structural breaks. Hence let us switch to ADF, PP, KPSS and DF-GLS test. But we did the applications of all these test before so to gain time let us only use PP test.

To use PP test, we need to decide which form of RW series we use. Best choice for RW seems to be random walk with drift pattern. (Çünkü plota baktığımızda başladığı yerden uzakta bittiğini, aşağı doğru gittiğini ve doğrusal olmayan bir şey olmadığını görüyoruz.)

Statistics → Time Series → Tests → Phillips Perron Unit Root Test



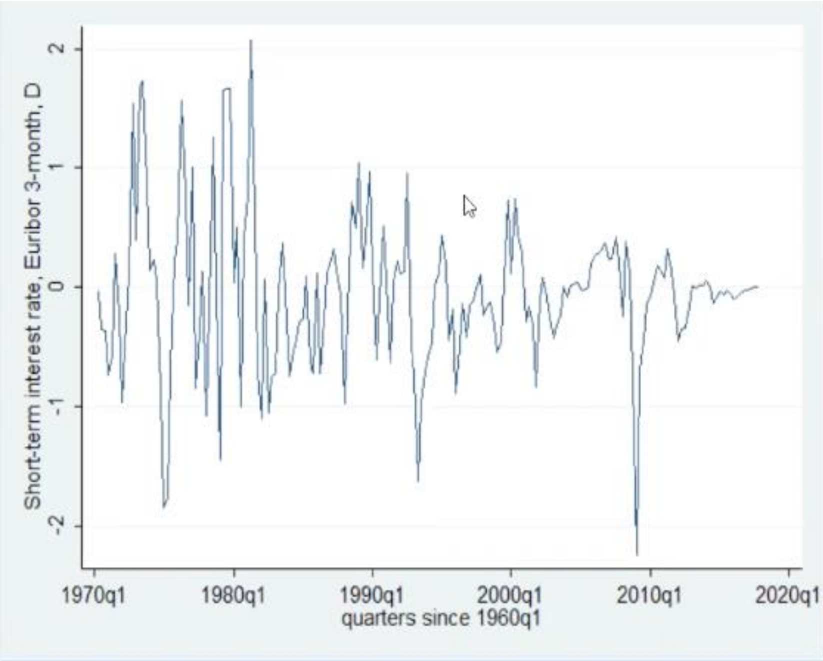
Biz random walk with drift'i seçtiğimiz için herhangi bir şeye basmıyoruz. Sadece Variable olan STN'i seçiyoruz. → Submit

```
Phillips-Perron test for unit root                  Number of obs    =      191
                                                    Newey-West lags  =        4

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(rho)          -3.098          -20.103           -13.882           -11.121
 Z(t)            -1.062           -3.480            -2.884            -2.574
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7298
```

-1.062 is not more negative than -2.884 so we do not reject the null hypothesis. There is unit root.

To decide the order of integration of STN, we need to take its first difference and test it

Graphics → Twoway graph → Edit → as Y variable write d.STN



There is no clear trend behavior so to test for ΔSTN, pure random walk seems better.

Statistics → Time Series → Tests → Phillips Perron Unit Root Test → variable d.STN and click suppress constant term in regression

```
Phillips-Perron test for unit root                 Number of obs   =       190
                                                   Newey-West lags =         4
```

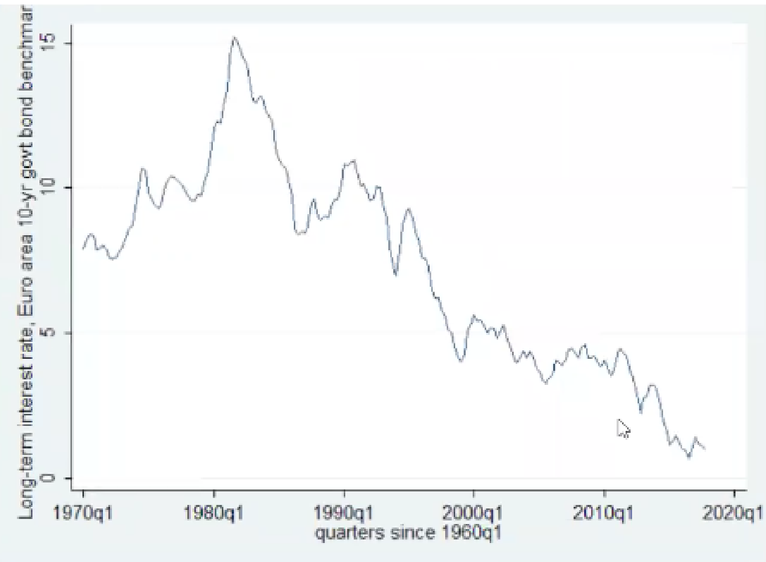
	Test Statistic	Interpolated Dickey-Fuller: 1% Critical Value	5% Critical Value	10% Critical Value
Z(rho)	-100.160	-13.480	-7.960	-5.660
Z(t)	-8.245	-2.588	-1.950	-1.616

-8.245 is more negative than -1.950 so we reject the null hypothesis. There is no unit root. STN is I(1) variable

If a series doesn't have a unit root, we call it I(0) series
STN is I(1) series
ΔSTN is I(0)

Let us check if LTN is I(1) as well

Graphics → Twoway graph → Edit → as Y variable write LTN



Best choice for RW seems to be random walk with drift pattern again.

Statistics → Time Series → Tests → Phillips Perron Unit Root Test → variable LTN and don't click anything for random walk types → submit

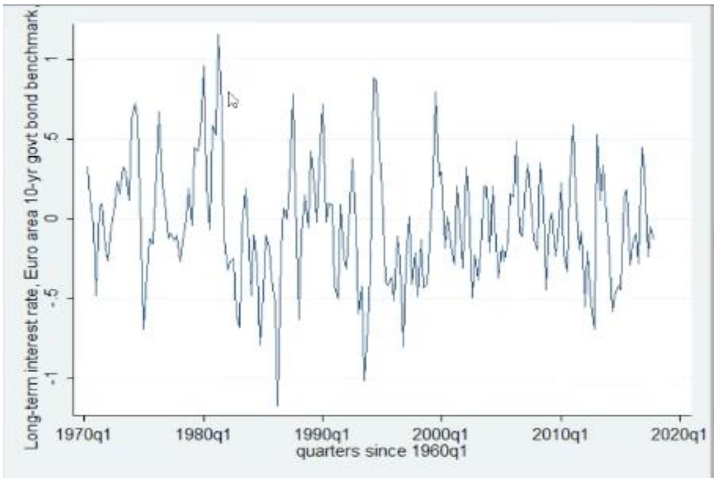
Phillips-Perron test for unit root   Number of obs   =   191
                                            Newey-West lags =   4

		Interpolated Dickey-Fuller		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(rho)	-0.610	-20.103	-13.882	-11.121
Z(t)	-0.292	-3.480	-2.884	-2.574

MacKinnon approximate p-value for Z(t) = 0.9266

-0.292 is not more negative than -2.884 so we do not reject the null hypothesis so there is unit root

Graphics → Twoway graph → Edit → as Y variable write d.LTN



There is no clear trend behavior so to test for ΔSTN, pure random walk seems better.

Statistics → Time Series → Tests → Phillips Perron Unit Root Test → variable d.LTN and click suppress constant term in regression

```
Phillips-Perron test for unit root                 Number of obs   =       190
                                                   Newey-West lags =         4

                         ———— Interpolated Dickey-Fuller ————
                  Test      1% Critical      5% Critical     10% Critical
               Statistic       Value            Value            Value
-----------------------------------------------------------------------------
 Z(rho)         -91.936         -13.480          -7.960           -5.660
 Z(t)            -7.842          -2.588          -1.950           -1.616
```

-7.842 is more negative than -1.950 so we reject the null hypothesis so there is no unit root

ΔLTN doesn't have a unit root. LTN I(1) variable

Now we validated that both variables are I(1) variables. Now regress LTN on STN (bc theory generally says this direction) and save the residuals.

Command:
regress LTN STN

to save the residuals we write

predict uhat, resid (after predict you can name anything you want)

before testing for unit root you need to look at its form but since this is OLS residual, always you can use pure random walk

Statistics → Time Series → Tests → Augmented Dickey Fuller Unit Root Test

For variable choose uhat and click suppress constant term in regression

```
Dickey-Fuller test for unit root                   Number of obs   =       191

                         ———— Interpolated Dickey-Fuller ————
                  Test      1% Critical      5% Critical     10% Critical
               Statistic       Value            Value            Value
-----------------------------------------------------------------------------
 Z(t)            -3.473          -2.588          -1.950           -1.616
```

-3.473 is more negative than -1.950 so we reject the null hypothesis so there is no unit root

So we have two I(1) series, we regress one on the other one. We estimated and got the residuals. Residuals that obtained from the estimation is I(0). So there is a cointegration. However, this values are not correct. We need to look MacKinnon table.

Critical Values for Two-Variable Cointegration ADF Test (Based on MacKinnon,1991)

	Level of Significance		
Sample size	*0.01*	*0.05*	*0.1*
25	-4.37	-3.59	-3.22
50	-4.12	-3.46	-3.13
100	-4.01	-3.39	-3.09
∞	-3.90	-3.33	-3.05

This is the best version

-3.473 is more negative than -3.39 so we reject the null hypothesis so there is no unit root

We conclude that there is cointegration between STN and LTN and we estimated regression was the cointegration equation.

There is an easier way to do these = egranger ado file

We need to install it, command = ssc install egranger

egranger dependent variable independent variable = egranger LTN STN

```
Engle-Granger test for cointegration                  N (1st step)  =       192
                                                      N (test)      =       191
-------------------------------------------------------------------------------
               Test          1% Critical       5% Critical      10% Critical
             Statistic          Value             Value             Value
-------------------------------------------------------------------------------
 Z(t)          -3.473           -3.954            -3.368            -3.067

Critical values from MacKinnon (1990, 2010)
```

regress LTN STN

```
      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) = 2711.36
       Model |  2227.30521     1  2227.30521           Prob > F      =  0.0000
    Residual |  156.079851   190   .8214729            R-squared     =  0.9345
-------------+------------------------------           Adj R-squared =  0.9342
       Total |  2383.38506   191  12.4784558           Root MSE      =  .90635

------------------------------------------------------------------------------
         LTN |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         STN |   .8152444   .0156565    52.07   0.000     .7843616    .8461273
       _cons |   2.202493   .1169786    18.83   0.000     1.971749    2.433236
------------------------------------------------------------------------------
```

LTN = 2.2 + 0.81*STN

15 ARALIK

COINTEGRATION TEST UNDER STRUCTURAL BREAK

GREGORY HANSEN COINTEGRATION TEST

Recall that we have seen unit root testing under structural break in the last lecture: Zivot-Andrews

ssc install zandrews

webuse lutkepohl2 for data

to test if ln_consump has unit root, we write → zandrews ln_consump

```
. zandrews ln_consump

Zivot-Andrews unit root test for  ln_consump

Allowing for break in intercept

Lag selection via TTest: lags of D.ln_consump included = 3

Minimum t-statistic -2.245 at 1970q3  (obs 43)

Critical values: 1%: -5.34 5%: -4.80 10%: -4.58
```

Since -2.245 is not more negative than -4.8, we do not reject the null hypothesis. There is unit root in ln_consump. This implies that ln_consump is not I(0).

To decide level of integration, we need to take the first difference and test if taking the first difference remove the unit root.

zandrews d.ln_consump

```
. zandrews d.ln_consump

Zivot-Andrews unit root test for  D.ln_consump

Allowing for break in intercept

Lag selection via TTest: lags of D.D.ln_consump included = 2

Minimum t-statistic -5.220 at 1968q2  (obs 34)

Critical values: 1%: -5.34 5%: -4.80 10%: -4.58
```

Since -5.220 is more negative than -4.8, we reject the null hypothesis. There is no unit root in ln_consump. This implies that ln_consump is I(1).

Now, we will see cointegration test under structural break. (Only 1 break is allowed (this is the weakest point of this procedure))

To install Gregory Hansen → ssc install ghansen

To see the syntax of the command, type help ghansen

This test allows us to choose the form of the structural break. To understand it, consider the following general form:

Cointegration equation form, model without any break

$$y_t = \beta_0 + \beta_1 X_t + \beta_2 . t + u_t$$

trend term

After structural break, the comprehensive version (after a point, β0, β1, β2 changed) can be as follows

Dummy variables

slope term

$$y_t = \beta_0 + \beta_1 X_t + \beta_2 t + \underbrace{\alpha_0 D_t} + \underbrace{\alpha_1} D_t X_t + \underbrace{\alpha_2} D_t . t + u_t$$

intercept term

if $\alpha_0 \neq 0$ there is structural break in intercept term

if $\alpha_1 \neq 0$ then there is structural break in slope term of X_t

if $\alpha_2 \neq 0$ then there is structural break in slope term of trend

Suppose there is a structural break at t=2015 and your data runs from 1990 till 2020

$$DT = \begin{cases} 1 & \text{if } t \geq 2015 \\ 0 & \text{if otherwise} \end{cases}$$

If H0: α0 =0
If Ha: α0 ≠0

If H0: β0=β2 =0
If Ha: at least one of them is not zero

```
break(level) specifies a break in the constant term.

break(trend) specifies a break in the constant and the trend.

break(regime) specifies a break in the constant and the slope.

break(regimetrend) specifies a break in the constant, the slope and the trend (Gregory and Hansen 1996a)
```

Hoca bunu kullanmamızı önerdi

H0: β0=β1= β2=0
(There is no structural change in any of them)
Ha: at least one of them is not zero(at least there is a structural change in one of them

If H0: β0=β1 =0
If Ha: at least one of them is not zero

Command:
ghansen ln_consump ln_income, break(regimetrend) lagmethod(aic)

```
. ghansen ln_consump ln_inc, break(regimetrend) lagmethod(aic)

Gregory-Hansen Test for Cointegration with Regime Shifts
Model: Change in Regime and Trend                  Number of obs   =       92
Lags  =  0  chosen by Akaike criterion             Maximum Lags    =        3

           Test        Breakpoint   Date       Asymptotic Critical Values
           Statistic                           1%         5%        10%
------------------------------------------------------------------------
   ADF       -7.52       64         1975q4     -6.02      -5.50     -5.24
   Zt        -7.57       63         1975q3     -6.02      -5.50     -5.24
   Za       -70.10       63         1975q3    -69.37     -58.58    -53.31
------------------------------------------------------------------------
```

Za is asymptotic test in order to use that observations must be higher than or near to 90-100. Otherwise, use Zt.

-70.10 is more negative than -58.58 so we reject the null hypothesis. Since this is a cointegration test, the null hypothesis “there is no cointegration”. Thus, there is cointegration between ln_consump and ln_income.

In the above, we said there is cointegration now it is time to estimate cointegration regression:

$$y_t = \underbrace{\beta_0}_{\text{intercept term}} + \beta_1 X_t + \overset{\text{slope term}}{\beta_2} t + \underbrace{\alpha_0 D_t}_{\text{if } \alpha_0 \neq 0 \text{ there is structural break in intercept term}} + \underbrace{\alpha_1 D_t X_t}_{\text{if } \alpha_1 \neq 0 \text{ then there is structural break in slope term of } X_t} + \underbrace{\alpha_2 D_t . t}_{\text{if } \alpha_2 \neq 0 \text{ then there is structural break in slope term of trend}} + u_t$$

First, we need to create the dummy variable

$$DT = \begin{cases} 0 & \text{if } t < 1975q3 \\ 1 & \text{if otherwise} \end{cases}$$

If H0: $\alpha 0 = 0$
If Ha: $\alpha 0 \neq 0$

command:

gen trend=_n (a trend variable will be added and it would start from 1)

	inv	inc	consump	qtr	ln_inv	dln_inv	ln_inc	dln_inc	ln_consump	dln_consump	trend
1	180	451	415	1960q1	5.192957	.	6.111467	.	6.028278	.	1
2	179	465	421	1960q2	5.187386	-.0055709	6.142037	.03057	6.042633	.0143547	2
3	185	485	434	1960q3	5.220356	.03297	6.184149	.0421114	6.073044	.0304112	3
4	192	493	448	1960q4	5.257495	.0371394	6.200509	.0163603	6.104793	.0317488	4
5	211	509	459	1961q1	5.351858	.0943627	6.232448	.031939	6.12905	.0242572	5
6	202	520	458	1961q2	5.308268	-.0435905	6.253829	.0213809	6.126869	-.0021811	6
7	207	521	479	1961q3	5.332719	.0244513	6.25575	.0019212	6.1717	.0448313	7
8	214	540	487	1961q4	5.365976	.033257	6.291569	.0358191	6.188264	.0165634	8
9	231	548	497	1962q1	5.442418	.0764418	6.306275	.0147061	6.20859	.0203261	9
10	229	558	510	1962q2	5.433722	-.0086956	6.324359	.0180836	6.234411	.0258207	10
11	234	574	516	1962q3	5.455321	.0215993	6.352629	.0282702	6.246107	.0116959	11
12	237	583	525	1962q4	5.46806	.0127387	6.368187	.0155578	6.263398	.0172915	12
13	206	591	529	1963q1	5.327876	-.1401839	6.381816	.013629	6.270988	.0075903	13
14	250	599	538	1963q2	5.521461	.1935849	6.395262	.0134459	6.287858	.01687	14
15	259	610	546	1963q3	5.556828	.035367	6.413459	.0181971	6.302619	.0147605	15
16	263	627	555	1963q4	5.572154	.015326	6.440947	.0274878	6.318968	.0163493	16
17	264	642	574	1964q1	5.575949	.0037951	6.464588	.0236416	6.352629	.0336609	17
18	280	653	574	1964q2	5.634789	.0588403	6.481577	.0169888	6.352629	0	18
19	282	660	586	1964q3	5.641907	.0071177	6.49224	.010663	6.37332	.0206904	19
20	292	694	602	1964q4	5.676754	.0348468	6.542472	.0502319	6.400258	.026938	20
21	286	709	617	1965q1	5.655992	-.020762	6.563856	.0213838	6.424869	.0246115	21
22	302	734	639	1965q2	5.710427	.0544348	6.598509	.0346532	6.459905	.0350356	22
23	304	751	653	1965q3	5.717028	.0066009	6.621406	.0228968	6.481577	.0216722	23
24	307	763	668	1965q4	5.726848	.00982	6.637258	.0158525	6.504288	.0227113	24
25	317	766	679	1966q1	5.758902	.0320539	6.641182	.0039239	6.520621	.0163331	25
26	314	779	686	1966q2	5.749393	-.0095086	6.658011	.016829	6.530878	.0102563	26
27	306	808	697	1966q3	5.723585	-.0258079	6.694562	.036551	6.546785	.0159078	27
28	304	785	688	1966q4	5.717028	-.0065575	6.665684	-.0288782	6.533789	-.0129967	28
29	292	794	704	1967q1	5.676754	-.0402737	6.677083	.0113997	6.556778	.0229897	29
30	275	799	699	1967q2	5.616771	-.0599828	6.683361	.0062776	6.549651	-.0071278	30
31	273	799	709	1967q3	5.609472	-.0072994	6.683361	0	6.563856	.014205	31
32	301	812	715	1967q4	5.70711	.0976386	6.699501	.0161395	6.572282	.0084267	32
33	280	837	724	1968q1	5.634789	-.0723209	6.729824	.0303235	6.584791	.0125089	33

In qtr variable, 1975q3 corresponds to 63 for trend variable.

60	519	1642	1371	1974q4	6.251904	-.0114942	7.40367	.0178194	7.223296	.0117388	60
61	526	1690	1402	1975q1	6.265301	.0133972	7.432484	.0288134	7.245655	.0223594	61
62	510	1759	1452	1975q2	6.234411	-.0308905	7.472501	.0400171	7.280697	.0350423	62
63	519	1756	1485	1975q3	6.251904	.0174932	7.470794	-.0017071	7.30317	.0224729	63
64	538	1780	1516	1975q4	6.287858	.0359545	7.484369	.0135751	7.323831	.0206604	64
65	549	1807	1549	1976q1	6.308098	.0202398	7.499424	.0150547	7.345365	.0215344	65
66	570	1831	1567	1976q2	6.345636	.0375381	7.512618	.0131941	7.356918	.0115533	66

We will create a dummy variable, it will takes a value of 0 until 63. To do that:

gen D=0

replace D=1 if trend>=63

equation :

D_t ln_inc$_t$

DX

D(T)

trend

slope terms

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + \alpha_0 D_t + \alpha_1 D_t X_t + \alpha_2 D_t \cdot t + u_t$$

intercept term

if $\alpha_0 \neq 0$ there is structural break in intercept term

if $\alpha_1 \neq 0$ then there is structural break is slope term of X_t

if $\alpha_2 \neq 0$ then there is structural break in slope term of trend

gen Dln_inc=D*ln_inc

gen Dtrend=D*trend

We generated all the variables that we need to run the cointegration regression. Let us estimate cointegration equation:

equation :

$D_t(\ln\_inc_t)$ DX DT trend

$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + \alpha_0 D_t + \alpha_1 D_t X_t + \alpha_2 D_t . t + u_t$

intercept term

slope term

if $\alpha_0 \neq 0$ there is structural break in intercept term

if $\alpha_1 \neq 0$ then there is structural break in slope term of X_t

if $\alpha_2 \neq 0$ then there is structural break in slope term of trend

regress ln_consump ln_income trend D Dln_income Dtrend

```
. regress ln_consump ln_inc trend D Dln_inc Dtrend
```

Source	SS	df	MS		
Model	25.4538304	5	5.09076608	Number of obs =	92
Residual	.007870551	86	.000091518	F(5, 86) =	55625.82
				Prob > F =	0.0000
Total	25.4617009	91	.279798911	R-squared =	0.9997
				Adj R-squared =	0.9997
				Root MSE =	.00957

ln_consump	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_inc	.9265401	.0364655	25.41	0.000	.8540491	.9990311
trend	.0002122	.0007796	0.27	0.786	-.0013375	.0017619
D	.1210062	.7952853	0.15	0.879	-1.459969	1.701981
Dln_inc	-.0202861	.1231472	-0.16	0.870	-.2650946	.2245223
Dtrend	.001012	.0020019	0.51	0.614	-.0029677	.0049917
_cons	.3531832	.2222986	1.59	0.116	-.0887319	.7950983

significant

Unfortunately, if p value is less than 0.05, it is significant

H0: coefficient is zero

These are all insignificant → don't try to interpret them, do not drop them

ln_inc'in coefficient'ının adı marginal propensity to consume = 0.926 and long run value bc it is cointegration value

If Y has a unit root, most probably lnY also has a unit root.
Taking the difference of lnY is corresponding to growth rate of Y in mathematics

$$\underbrace{\Delta \underbrace{\ln Y_t}_{I(1)}}_{I(0)} \approx g_Y$$

“3 OTHER COINTEGRATION ESTIMATION METHODS”

When there is cointegration, instead of estimating cointegration equation by OLS, we can estimate it by 3 other new estimation methods.

1) FMOLS (Fully Modified OLS) (Hoca bunu kullanmamızı önerdi)
2) DOLS (Dynamic OLS)
3) CCR (Canonical Cointegration Regression)

To install them, use the following command:

ssc install cointreg → we will return to this

ARIMA MODELLING

If the data has AR(1) form, the best way to forecast it is running this following model:

$$y_t = a_0 + \delta y_{t-1} + u_t \quad \text{AR(1)}$$

AR(2)

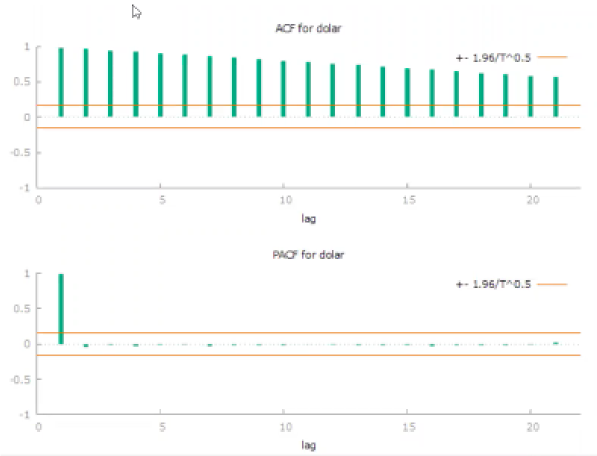
$$y_t = a_0 + \delta_1 y_{t-1} + \delta_2 y_{t-2} + u_t \quad \Big\} \text{ AR(1) or AR(2) is very common}$$

How we will decide that AR form is good for forecasting:

- We look for AR signature → a sudden drop in PACF and a very gradual decline in ACF (If there is sudden drop in ACF and gradually decline in PACF then it is MA)

After deciding that the best form is the AR from, next is determining its level

To decide the level of AR form, we look to the PACF graph, and count the picks
To decide the level of MA form, we look to the ACF graph, and count the picks



There is only one pick so level is AR(1)

We also need to determine order of integration

For our example, the series is I(1)

29 ARALIK DERSİ

So we already determined that the ln_consump, ln_inc and ln_inv series are I(1)

If you have I(1) series, you must check if there is cointegration. If there is no cointegration, then by simply using their first differences we can run a VAR mode. If there is cointegration we need to run a cointegration system estimation: this can be VECM(Vector Error Correction Model- this is also known as Johansen Procedure, or FMOLS (Fully Modified OLS), DOLS (Dynamic OLS), CCR(Canonical Cointegration Regression) and at least Engle Granger.

Today we will not enter to the estimation of cointegration. Today we will assume that there is no cointegration and we procedure by taking their differences and estimating a VAR.

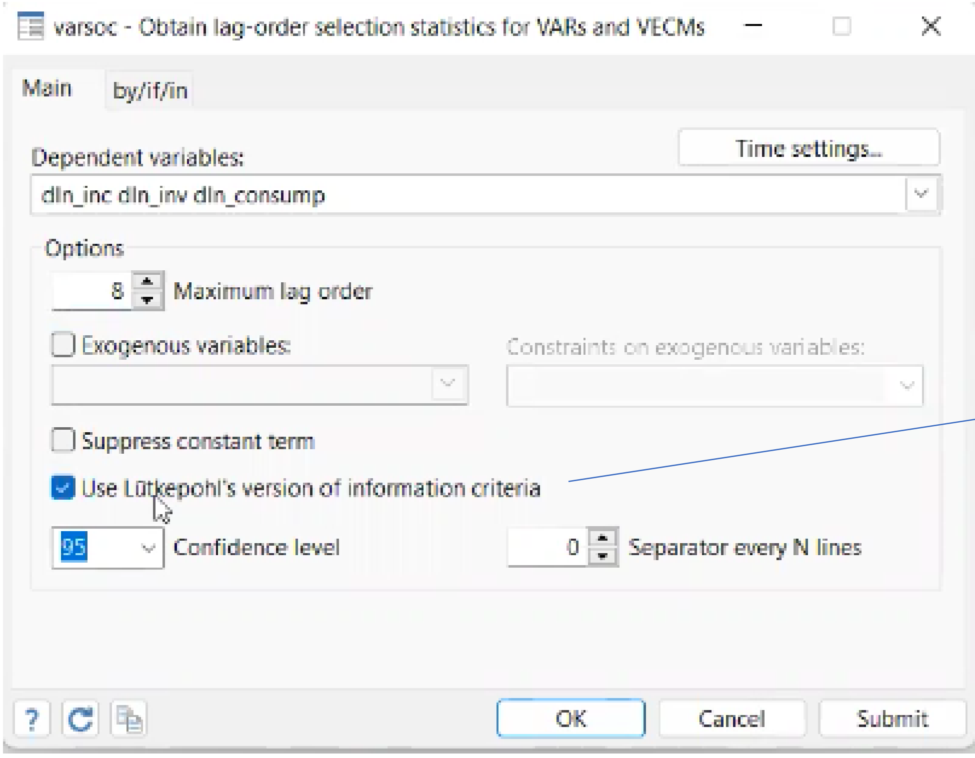
Let us first determine the optimal lag length of a VAR system.

Statistics → multivariate time series → VAR diagnostic and tests → Lag order selection statistics (preestimation)

Order of dependent variables are important (en sondaki her şeyin etkilediği olacak)

I(0) hallerini kullanmamız gerektiği için önlerine d koyduk

1.5-2 seneyi kapsaması için maximum lag order’ı 8 yaptık



Kullanmak zorunda değiliz ama hoca kullanmanız güzel olur dedi.

Sample: 1962q2 thru 1982q4 Number of obs = 83

Lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	673.36				1.9e-11	-24.7392	-24.7392*	-24.7392*
1	684.879	23.037	9	0.006	1.8e-11	-24.7999	-24.6945	-24.5376
2	695.689	21.621	9	0.010	1.7e-11*	-24.8435*	-24.6328	-24.3189
3	701.449	11.519	9	0.242	1.9e-11	-24.7654	-24.4493	-23.9786
4	710.962	19.027	9	0.025	1.9e-11	-24.7778	-24.3563	-23.7286
5	712.493	3.0607	9	0.962	2.3e-11	-24.5978	-24.0709	-23.2864
6	718.143	11.3	9	0.256	2.5e-11	-24.5171	-23.8848	-22.9434
7	720.35	4.4136	9	0.882	3.0e-11	-24.3534	-23.6158	-22.5174
8	731.352	22.005*	9	0.009	2.9e-11	-24.4016	-23.5587	-22.3034

* optimal lag
Endogenous: dln_inc dln_inv dln_consump
Exogenous: _cons

AIC için lag order 2 (yıldızlı olanlara bakıyoruz)
0'dan fazla olan laglere bakıyoruz

If your sample size <50, use SBIC

Tabloda en fazla yıldız olan lag hangisi ise onu seç, VAR için lag 0 anlamsız onu seçme

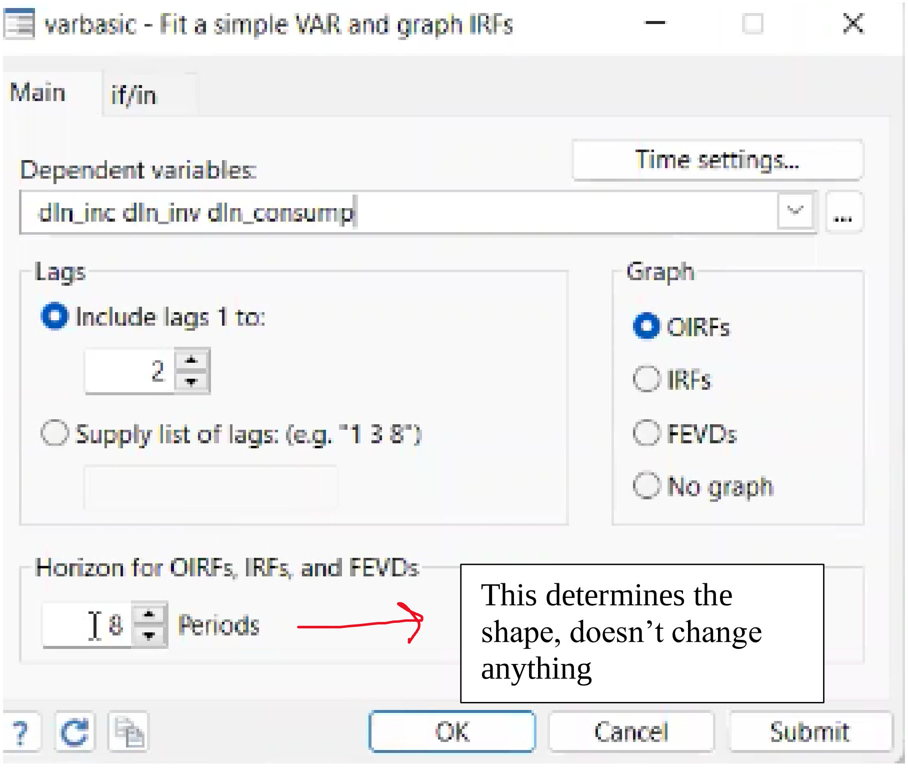
LR aslında daha güçlü, tek başına olsa bile (tek yıldız var) eğer autocorrelation varsa bunu seçeceğiz şimdilik 2 seçiyoruz.

Now er estimate the VAR System

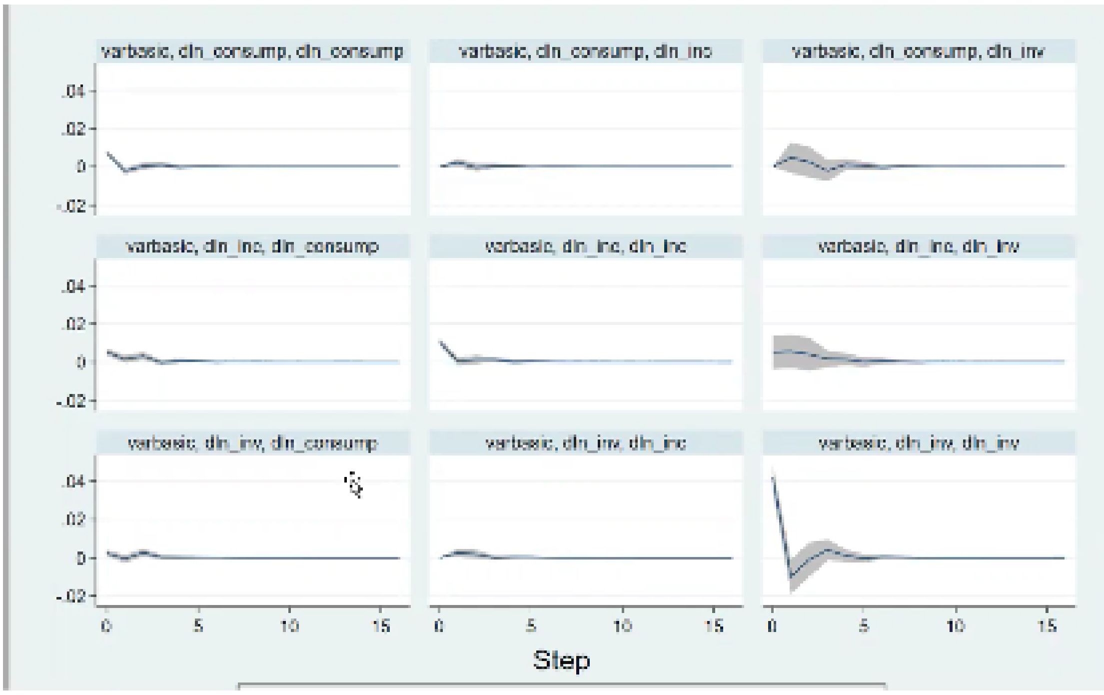
Statistic → multivariate time series → Basic VAR

Enter the same order for dependent variables

Lag length is asked → 2

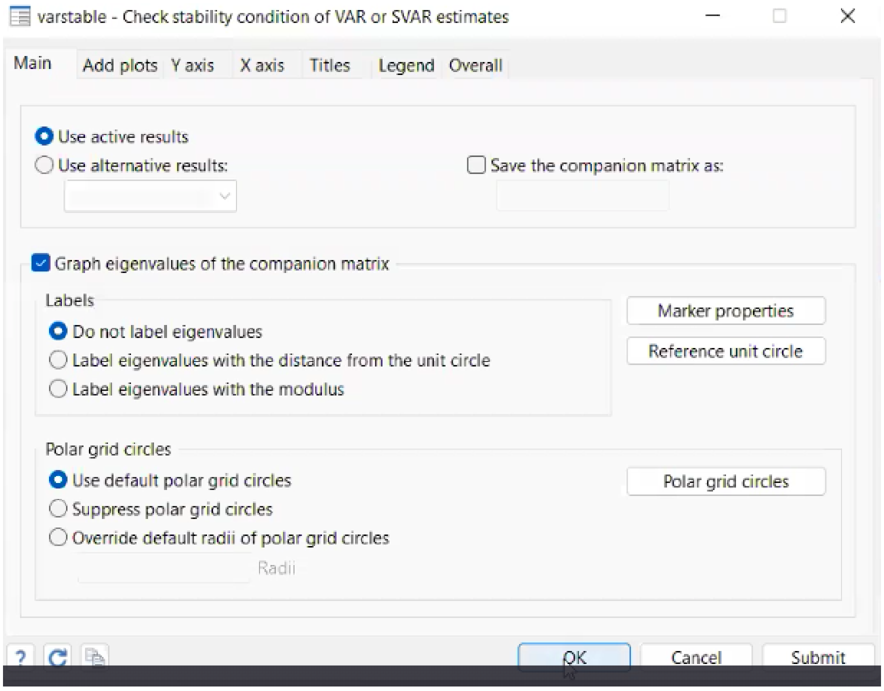


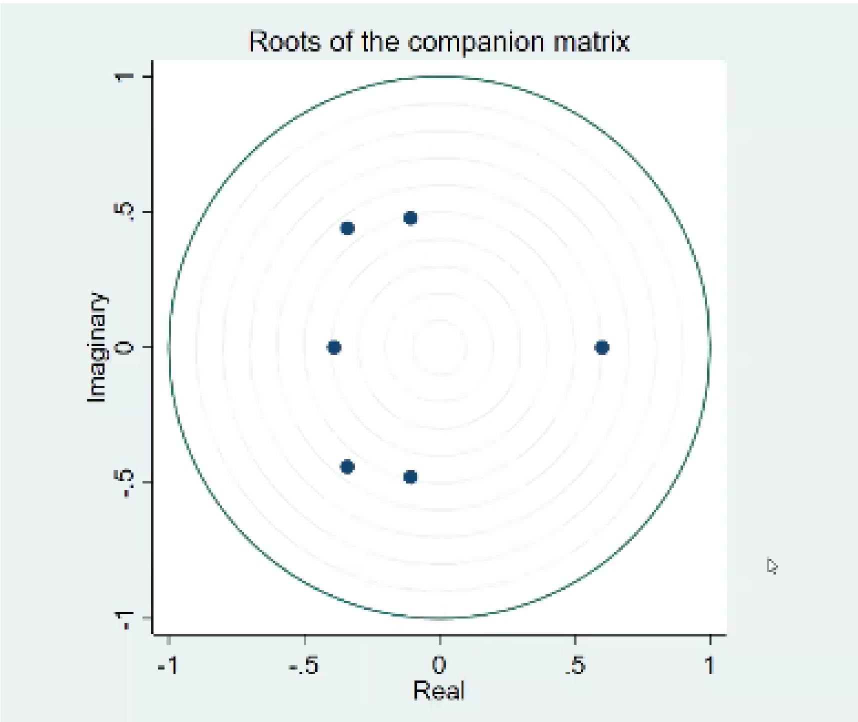
This determines the shape, doesn't change anything



All of them needs to converge to zero.

Hoca kaydı açmayı unutmuş buradan başlıyor.





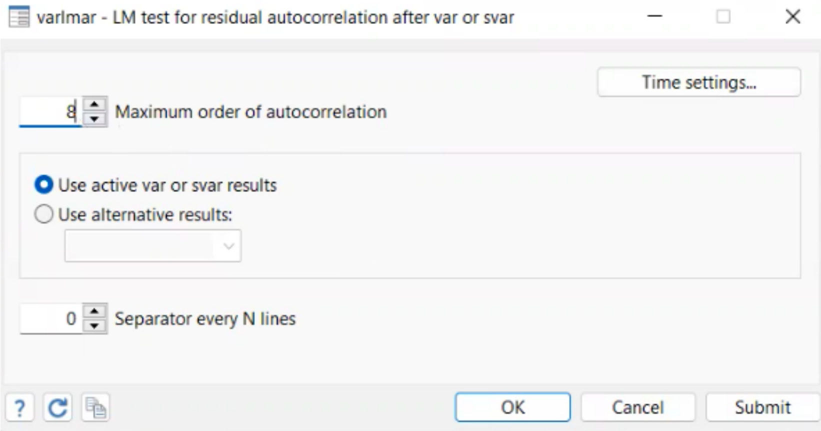
All the roots(dots) lie within the unit circle so there is no problem with stability of the VAR system

To check autocorrelation

Statistics → Multivariate time series → VAR diagnostics and tests → LM test for residual autocorrelation

If there is autocorrelation we need to increase lag length

Maximum order of autocorrelation → 8



Lagrange-multiplier test

lag	chi2	df	Prob > chi2
1	8.8693	9	0.44942
2	10.9722	9	0.27762
3	9.4300	9	0.39856
4	15.5757	9	0.07629
5	2.7613	9	0.97301
6	5.5712	9	0.78195
7	5.3087	9	0.80661
8	10.1816	9	0.33599

H0: no autocorrelation at lag order

Prob > 0.05
You do not reject the null hypothesis

So there is no AR(1) type of autocorrelation

This is the autocorrelation

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Prob > 0.05
You do not reject the null hypothesis

So there is no AR(3) type of autocorrelation

Here none of them needs to be less than 0.005

Null hypothesis: there is no AR(1) type of autocorrelation

Since all of them are higher than 0.05, we do not observe any autocorrelation in the VAR system, so autocorrelation condition is met.

If there is autocorrelation, we increase the lag length as 8 because of this table

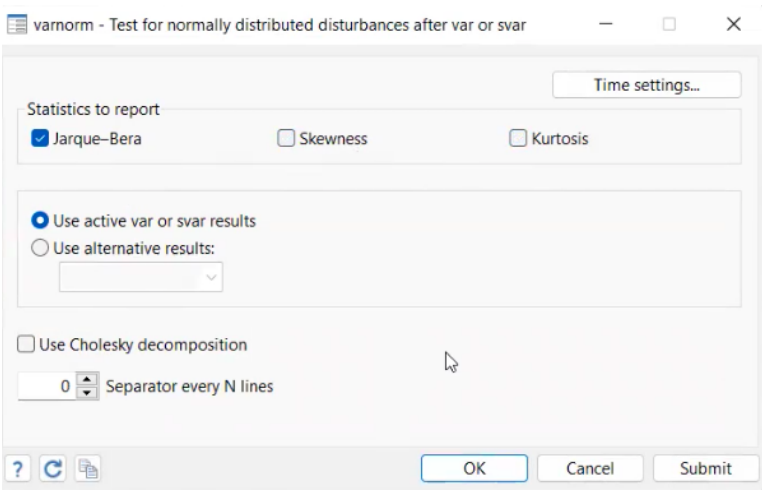
Sample: 1962q2 thru 1982q4 Number of obs = 83

Lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	673.36				1.9e-11	-24.7392	-24.7392*	-24.7392*
1	684.879	23.037	9	0.006	1.8e-11	-24.7999	-24.6945	-24.5376
2	695.689	21.621	9	0.010	1.7e-11*	-24.8435*	-24.6328	-24.3189
3	701.449	11.519	9	0.242	1.9e-11	-24.7654	-24.4493	-23.9786
4	710.962	19.027	9	0.025	1.9e-11	-24.7778	-24.3563	-23.7286
5	712.493	3.0607	9	0.962	2.3e-11	-24.5978	-24.0709	-23.2864
6	718.143	11.3	9	0.256	2.5e-11	-24.5171	-23.8848	-22.9434
7	720.35	4.4136	9	0.882	3.0e-11	-24.3534	-23.6158	-22.5174
8	731.352	22.005*	9	0.009	2.9e-11	-24.4016	-23.5587	-22.3034

* optimal lag
Endogenous: dln_inc dln_inv dln_consump
Exogenous: _cons

Now we test normality, but if the IRFs converge to zero, if there is no autocorrelation and if the stability condition is met (as in this example) even if normality test fails it may not be integrated as a big problem

Statistics → multivariate time series → VAR diagnostics and tests → test for normality distributed disturbances
Sadece Jarque-Bera'yı kullan




```
varnorm, jbera
```

Jarque-Bera test

Equation	chi2	df	Prob > chi2
dln_inc	10.250	2	0.00595
dln_inv	11.669	2	0.00292
dln_consump	0.313	2	0.85499
ALL	22.232	6	0.00110

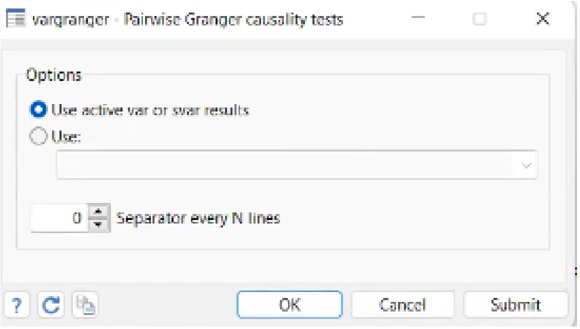
The residual of the 1st equation does not have a normal distribution so there is problem

If any of them is less than 0.05, there is a problem.

But don't bother so much

So according to the result of these tests, we do not observe any big problem with the current VAR estimation. Then we can use it for a Granger causality analysis

Statistics → multivariate time series → VAR diagnostics and tests → Granger Causality Test



0.044 < 0.05 so we reject the null hypothesis so α_3 and α_4 are not zero so they are important

in other words Δln_inv_{t-1} and Δln_inv_{t-2} are important to explain Δln_inc_t

	Equation	Excluded	chi2	df	Prob > chi2
1st	dln_inc	dln_inv	6.2466	2	0.044
	dln_inc	dln_consump	5.1029	2	0.078
	dln_inc	ALL	13.087	4	0.011
2nd	dln_inv	dln_inc	.55668	2	0.757
	dln_inv	dln_consump	1.9443	2	0.378
	dln_inv	ALL	7.3184	4	0.120
3rd	dln_consump	dln_inc	16.275	2	0.000
	dln_consump	dln_inv	4.2446	2	0.120
	dln_consump	ALL	21.717	4	0.000

Prob value of the following test
H0: $\alpha_3 = \alpha_4 = 0$
Ha: at least one of them is nonzero

Prob value of the following test
H0: $\alpha_5 = \alpha_6 = 0$
Ha: at least one of them is nonzero

$$\Delta ln\_inc_t = a_0 + a_1 \Delta ln\_inc_{t-1} + a_2 \Delta ln\_inc_{t-2} + a_3 \Delta ln\_inv_{t-1} + a_4 \Delta ln\_inv_{t-2} + a_5 \Delta ln\_consump_{t-1} + a_6 \Delta ln\_consump_{t-2} + u_t$$

In conclude that, Δln_inv is the granger cause of Δln_inc

Ikisi de 0.05'ten büyük çıkarsa ne Δln_inv ne de Δln_consump, Δln_inc'ın granger cause'u değil ama eğer all 0.05'ten küçükse bu demek oluyor ki Δln_inv ve Δln_consump beraber değiştiğinde Δln_inc'ın granger cause'u oluyor

0.00 < 0.05 so we reject the null hypothesis so c1 and c2 are not zero so they are important

Prob value of the following test
H0: c1= c2=0
Ha: at least one of them is nonzero

dln_consump	dln_inc	16.275	2	0.000
dln_consump	dln_inv	4.2446	2	0.120
dln_consump	ALL	21.717	4	0.000

tyy

$$\Delta ln\_consump_t = c_0 + c_1 . \Delta ln\_inc_{t-1} + c_2 \Delta ln\_inc_{t-2} + c_3 \Delta ln\_inv_{t-1} + c_4 \Delta ln\_inv_{t-2} + c_5 \Delta ln\_consump_{t-1} + c_6 \Delta ln\_consump_{t-2} + u_t$$

Δln_inc is the granger cause of Δln_consump
Δln_inv is not the granger cause of Δln_consump
If Δln_inv and Δln_inc changes together, they are the granger cause of Δln_consump

Error Correction Model

In fact, the basis of Error Correction Model is in Nerlove's Partial Adjustment Model (PAM)

If there is a long run relationship between say X and Y, then the following equation expected to work to restore the distorted equilibrium.

$$\Delta Y_t = \beta_1 \Delta X_t + \lambda \left[\text{the value of } Y \text{ at time } t-1 - \text{Equilibrium value of } Y \right]$$

PAM in its simplest form

$$\boxed{\Delta Y_t = \beta_1 \Delta X_t + \lambda [Y_{t-1} - Y^*]}$$

simple

Suppose $Y^* = 10$
$Y_{t-1} = 15$

At time t-1 the value of y is above its equilibrium value by 5 units

How can the equilibrium be restored in this case
Y_t must decline by 5 units

$$\Delta Y_t = -5$$

Her şeyi aynı anda değiştirirsek ne neden oldu anlayamayacağımız için x'I sabit kabul ediyoruz. ΔX=0 bu yüzden

ΔY'nin -5 olması için λ=-1'e eşit olması lazım. Λ=-1'e eşitse system goes back to equilibrium in one period.

What would happen if λ=-0.2?

ΔY=-0.2*(15-10) = -1 This means that in next period Y will decrease by 1 unit so it will be 14.

ΔY=-0.2*(14-10) = -0.8 So in the second period, Y will decline by 0.8 so now value will be 13.2 ….. so it will never completely return to its equilibrium value. (But you can calculate the time say %99 of the deviation from equilibrium)

So λ is very important for the speed of adjustment,
λ =-1 → The deviation from equilibrium is restored completely in 1st period (100% of the deviation from equilibrium is restored in only 1 period.)

λ= -0.2 → 20% of the deviation from equilibrium is restored in every period.

$\lambda = 2$?

$$\Delta Y_t = \beta_1 \cancel{\Delta X_t} + \lambda [Y_{t-1} - Y^*]$$

$$\Delta Y_t = 2 \cdot \underbrace{(15-10)}_{5} = 10$$

$15 \rightarrow 25 \Rightarrow$ By this way we will observe a divergence.

Conclusion: If there is a stable equilibrium relationship (cointegration) the λ in this regression cannot take a positive value.

For a healthy stable equilibrium relationship, we need that λ lies between -1 and 0.

-1 < λ <0 → range of λ for a healthy estimate of LR relationship

After 1980s with the invention of cointegration analysis the expression within the parenthesis

$$[Y_{t-1} - Y^*]$$

is proved to be written as the residual of the cointegration equation t time t-1

$$[Y_{t-1} - Y^*] \rightarrow \hat{u}_{t-1}$$

you can use the residual serie of the cointegration regression for this expression.

$$\Delta Y_t = \beta_1 \Delta X_t + \lambda \hat{u}_{t-1}$$ In form, the name of equation became Error Correction Model

This equation can be easily estimated, and we can check if λ lies between -1 and 0.

For cointegration, all variables must be I(1) so ΔY~I(0), ΔX~I(0)

In Engle-Granger test, we tested if the residuals are I(0). So if there is cointegration

So - if there is cointegration $\hat{u}_t \sim I(0)$ which implies that $\hat{u}_{t-1} \sim I(0)$

So in the existence of cointegration, this ECM can be estimated by OLS since all the terms are I(0). We will estimate it and we will test if λ is negative or not.

ECM Steps

1)Estimate ECM:

$$\Delta Y_t = \beta_1 \Delta X_t + \lambda \hat{u}_{t-1}$$

2)

$$H_0: \lambda \geq 0 \qquad H_a: \lambda < 0$$

Carry out a t-test to check it. If you confirm that λ<0, then you say that Error Correction Mechanism is working → This implies that the LR relationship is working. In addition, the estimated value of λ, gives us the speed of adjustment.

Regress ln_consump ln_inc

```
. regress ln_consump ln_inc

      Source |       SS           df       MS      Number of obs   =        92
-------------+----------------------------------   F(1, 90)        >  99999.00
       Model |  25.4403954         1  25.4403954   Prob > F        =    0.0000
    Residual |  .021305495        90  .000236728   R-squared       =    0.9992
-------------+----------------------------------   Adj R-squared   =    0.9992
       Total |  25.4617009        91  .279798911   Root MSE        =    .01539

------------------------------------------------------------------------------
  ln_consump | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      ln_inc |    .967773   .0029521   327.82   0.000     .9619081    .9736379
       _cons |   .0856322    .020934     4.09   0.000     .0440431    .1272214
------------------------------------------------------------------------------
```

We need to save the residuals of this regression predict ourresidual, resid

regress d.ln_consump d.ln_inc L.ourresidual, noconstant

L→ bir gecikmesini al demek

```
. regress d.ln_consump d.ln_inc L.ourresidual, noconstant

      Source |       SS           df       MS      Number of obs   =        91
-------------+----------------------------------   F(2, 89)        =    174.36
       Model |  .033905249         2  .016952625   Prob > F        =    0.0000
    Residual |  .008653454        89  .00009723    R-squared       =    0.7967
-------------+----------------------------------   Adj R-squared   =    0.7921
       Total |  .042558703        91  .000467678   Root MSE        =    .00986

------------------------------------------------------------------------------
D.ln_consump | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      ln_inc |
         D1. |   .8198017   .0453523    18.08   0.000     .7296876    .9099159
             |
 ourresidual |
         L1. |  -.2892177    .067872    -4.26   0.000    -.4240779   -.1543574
------------------------------------------------------------------------------
```

It must be less than 0.05
Null hypothesis: Lambda hat = 0

VECTOR ERROR CORRECTION MODEL (VECM)

This is also known as Johansen Procedure (Cointegration Analysis). This is very similar to the VAR modeling.

Refreshment: A VAR(2) system for 3 variables

All variables must be I(0) here

$$\begin{cases} X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + a_3 Y_{t-1} + a_4 Y_{t-2} + a_5 Z_{t-1} + a_6 Z_{t-2} + u_{t1} \\ Y_t = b_0 + b_1 X_{t-1} + b_2 X_{t-2} + b_3 Y_{t-1} + b_4 Y_{t-2} + b_5 Z_{t-1} + b_6 Z_{t-2} + u_{t2} \\ Z_t = c_0 + c_1 X_{t-1} + c_2 X_{t-2} + c_3 Y_{t-1} + c_4 Y_{t-2} + c_5 Z_{t-1} + c_6 Z_{t-2} + u_{t3} \end{cases}$$

If all of them are not I(0), i.e. if all of them I(1), thus there we can carry out a cointegration analysis in multivariate level.

If there is cointegration between variables of the system (here X, Y and Z), then the following form can be estimated. You take the difference of all equations. Then it will be VAR(1).

If there are only green ones, it is VAR

$$\Delta X_t = d_0 + d_1 \Delta X_{t-1} + d_2 \Delta Y_{t-1} + d_3 \Delta Z_{t-1} + \lambda_x \hat{u}_{t-1} + \varepsilon_{t1}$$
$$\Delta Y_t = e_0 + e_1 \Delta X_{t-1} + e_2 \Delta Y_{t-1} + e_3 \Delta Z_{t-1} + \lambda_y \hat{u}_{t-1} + \varepsilon_{t2}$$
$$\Delta Z_t = f_0 + f_1 \Delta X_{t-1} + f_2 \Delta Y_{t-1} + f_3 \Delta Z_{t-1} + \lambda_z \hat{u}_{t-1} + \varepsilon_{t3}$$

If there is no cointegration, red terms are not necessary

Error Correction Term if there is cointegration we add these terms. And the equation becomes VECM

REMARK:
If there is cointegration

one eq:

Refreshment about ECM

If there is cointegration $Y_t = \beta_0 + \beta_1 X_t + u_t$

$$\boxed{\Delta Y_t = \beta_1 \Delta X_t + \lambda \hat{u}_{t-1}}$$

$$-1 < \lambda < 0$$

If there is cointegration between the variables X, Y and Z such as:

coint 1 $\quad Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t$

→ this form is an appropriate version of the VECM system

But there can be more than one cointegration relationship in a multi equation system. In other words, apart from Coint 1 equation, there can be another cointegration equation such as

$X_t = \alpha_0 + \alpha_1 Z_t + V_t$ Coint 2

In this case, two different error correction mechanism expected to run.

Two LR relationship (Two cointegration relation) → implies → Two Error Correction Mechanism to work

How can we add this addition cointegration relationship (Coint 2) into the multiequation system?

$$\Delta X_t = d_0 + d_1 \Delta X_{t-1} + d_2 \Delta Y_{t-1} + d_3 \Delta Z_{t-1} + \lambda_x \hat{u}_{t-1} + \theta_x \hat{v}_{t-1} + \varepsilon_{t1}$$
$$\Delta Y_t = e_0 + e_1 \Delta X_{t-1} + e_2 \Delta Y_{t-1} + e_3 \Delta Z_{t-1} + \lambda_y \hat{u}_{t-1} + \theta_y \hat{v}_{t-1} + \varepsilon_{t2}$$
$$\Delta Z_t = f_0 + f_1 \Delta X_{t-1} + f_2 \Delta Y_{t-1} + f_3 \Delta Z_{t-1} + \lambda_z \hat{u}_{t-1} + \theta_z \hat{v}_{t-1} + \varepsilon_{t3}$$

Error correction terms for first cointegration relationship

Error correction terms for second cointegration relationship

Let's say Y is inflation, X is current account deficit and Z is exchange rate.

Coint 1

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t$$
$$\Delta Y = g_0 + g_1 \Delta X_t + g_2 \Delta Z_t + (\lambda_y) \hat{u}_{t-1}$$

The signs of λ X and λZ must be analyzed particularly

-1<λY <0

This long run relationship says that the inflation is a function of CAD and exchange rate in the long run.

Coint 2

$$X_t = \alpha_0 + \alpha_1 Z_t + V_t$$

$$\Delta X_t = h_0 + h_1 \Delta Z_t + \theta_x \tilde{v}_{t-1}$$
$$\boxed{-1 < \theta_x < 0}$$

This LR relationship says that there is also another LR relationship for this country in this form: CAD is a function of exchange rate in the long run.

It is possible → Theoretically it is!!

It is main advantage and its main complicity of a VECM estimation.

For simplicity, we continue with only 1 cointegration relationship. Now let us determine the expected signs of the λx and λz.

$\boxed{\text{Coint 1}}$

$$\Delta Y = g_0 + g_1 \Delta X_t + g_2 \Delta Z_t + \lambda_y \hat{u}_{t-1} \qquad -1 < \lambda_y < 0$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t$$

Suppose that Coint 1 relationship is as follows

$$Y_t = 10.1 + 0.1 X_t - 2.1 Z_t + u_t$$

X has a positive and Z has a negative impact on Y in the long run. How can we determine the signs of λx and λz?

In the long run, we can write this equation without error term

$$Y_t = 10.1 + 0.1 X_t - 2.1 Z_t$$

We put everything on the left-hand side and make equal to the zero.

$$Y_t - 10.1 - 0.1 X_t + 2.1 Z_t = 0$$

If Yt increases, left hand size will be positive, so in order to go to the equality Y should decrease delta Y should be negative and λY <0

If Yt increases, we need to increase the Xt to make it equal to 0. (Nötrlüğü korumaya çalışıyoruz. Delta Xt >0, λx>0

If Yt increases, we need to decrease the Zt to make it equal to 0. (Nötrlüğü korumaya çalışıyoruz. Delta Zt <0, λz<0

Üçünün (λy, λx, λz) birden çalışmasına gerek yok, biri çalışırsa eşitliği sağlayabilir.

$$\Delta X_t = d_0 + d_1 \Delta X_{t-1} + d_2 \Delta Y_{t-1} + d_3 \Delta Z_{t-1} + \lambda_x \hat{u}_{t-1} + \varepsilon_{t1}$$
$$\Delta Y_t = e_0 + e_1 \Delta X_{t-1} + e_2 \Delta Y_{t-1} + e_3 \Delta Z_{t-1} + \lambda_y \hat{u}_{t-1} + \varepsilon_{t2}$$
$$\Delta Z_t = f_0 + f_1 \Delta X_{t-1} + f_2 \Delta Y_{t-1} + f_3 \Delta Z_{t-1} + \lambda_z \hat{u}_{t-1} + \varepsilon_{t3}$$

Stata Application

Steps of VECM analysis

1)All the variables must be I(1) [This must be checked]

2)You determine the optimal lag length of VECM system

3)Carry out the Johansen cointegration test and determine of number of cointegration relationships

Since this algorithm allows us to work with more than 1 cointegration, we must check what is the number of cointegration relationship. This is reported with the name of "rank". If rank is zero, there is no cointegration. If rank is 1, there is one cointegration.

To determine the rank of the VECM system, two tests are proposed by Johansen

1)Trace Test → more powerful
2)Max Eigenvalue Test

4)Estimate the VECM system and check if at least one of the error correction terms is in the expected range

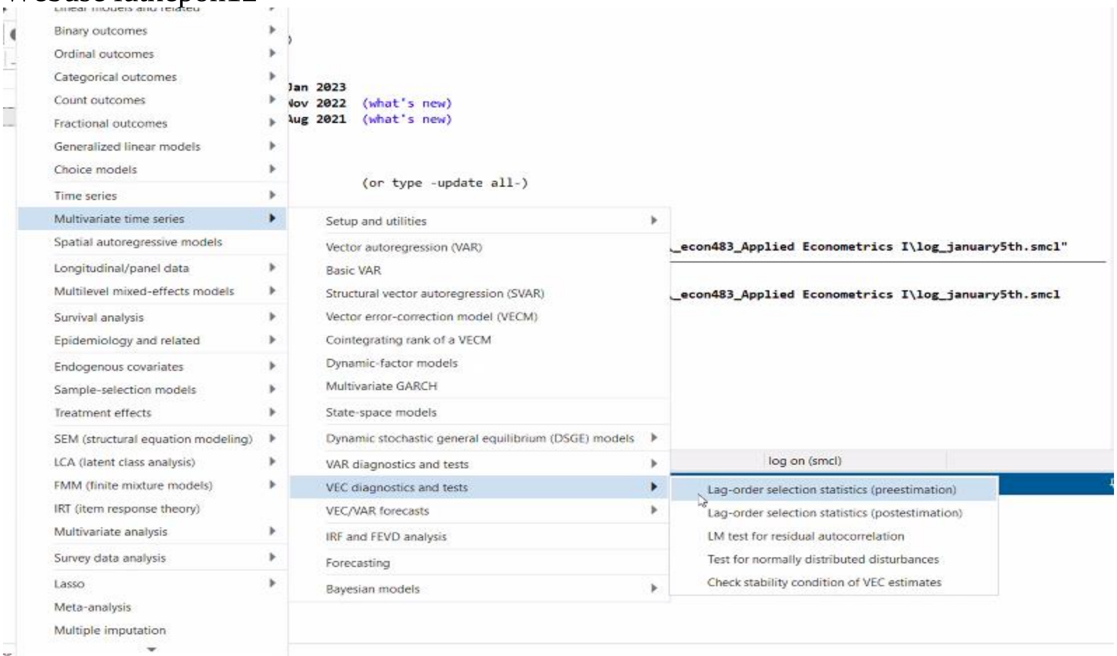
$$\begin{bmatrix} -1 < \lambda_y < 0 \\ \lambda_x > 0 \\ -1 < \lambda_z < 0 \end{bmatrix} \longrightarrow \boxed{\text{For our example}}$$

And statistically significant (its t value is larger than 2 in absolute value or the prob values of the reported lambda terms must be less than 0.05 to be significant)

If at least one of the lambda terms is within the expected range and significant, be happy ☺

You can use this model for economic analysis and even for forecasting.

Webuse lutkepoh12



Dependent variable → ln_inc ln_inv ln_consump

(en güçlü olan en sona yazılacak)

Maximum lag order →8

Do not suppress constant term

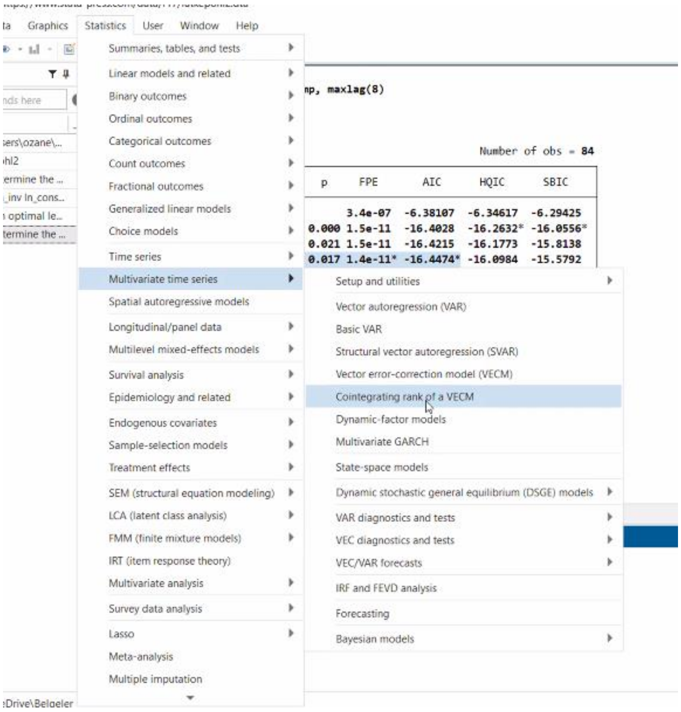
Sample: 1962q1 thru 1982q4 Number of obs = 84

Lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	271.005				3.4e-07	-6.38107	-6.34617	-6.29425
1	700.918	859.83	9	0.000	1.5e-11	-16.4028	-16.2632*	-16.0556*
2	710.705	19.574	9	0.021	1.5e-11	-16.4215	-16.1773	-15.8138
3	720.79	20.17	9	0.017	1.4e-11*	-16.4474*	-16.0984	-15.5792
4	727.16	12.74	9	0.175	1.5e-11	-16.3848	-15.9311	-15.2562
5	733.503	12.686	9	0.177	1.7e-11	-16.3215	-15.7631	-14.9325
6	736.176	5.3449	9	0.803	2.0e-11	-16.1708	-15.5078	-14.5214
7	745.707	19.063*	9	0.025	2.0e-11	-16.1835	-15.4157	-14.2736
8	749.601	7.7872	9	0.556	2.2e-11	-16.0619	-15.1895	-13.8916

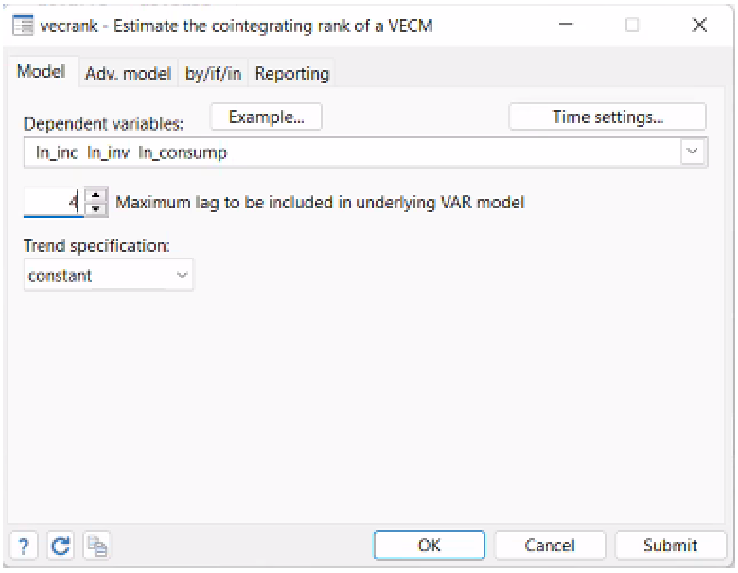
* optimal lag
Endogenous: ln_inc ln_inv ln_consump
Exogenous: _cons

If you have small sample size, use SBIC
If your sample size is large enough, choose AIC

For the system, optimal lag length is 3. Now let us determine the rank (number of cointegration relationship)



Dependent variable → ln_inc ln_inv ln_consump



Lag sayısını yazarken burada VAR için soruyor o yüzden her zaman 1 fazlasını yazacağız çünkü VECM bir lag eksik yazıyor. Optimal lag length 3tü bir fazlasını yazdığımız için 4 yazdık.

Johansen tests for cointegration

Trend: Constant — Number of obs = 88

Sample: 1961q1 thru 1982q4 — Number of lags = 4

Maximum rank	Params	LL	Eigenvalue	Trace statistic	Critical value 5%
0	30	737.55584	.	33.2489	29.68
1	35	748.66973	0.22321	11.0211*	15.41
2	38	753.23176	0.09849	1.8970	3.76
3	39	754.18028	0.02133		

* selected rank

0 rank = there is no cointegration

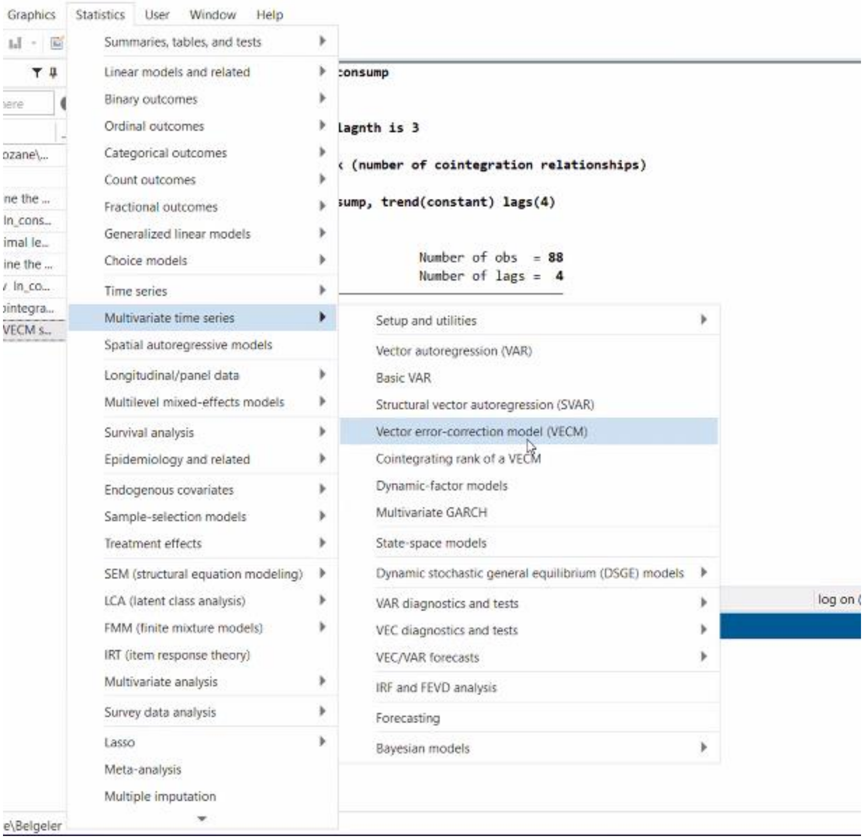
Null hypothesis: There is no cointegration

Trace statistics > critical value → reject the null hypothesis

1 rank= at least 1 cointegration

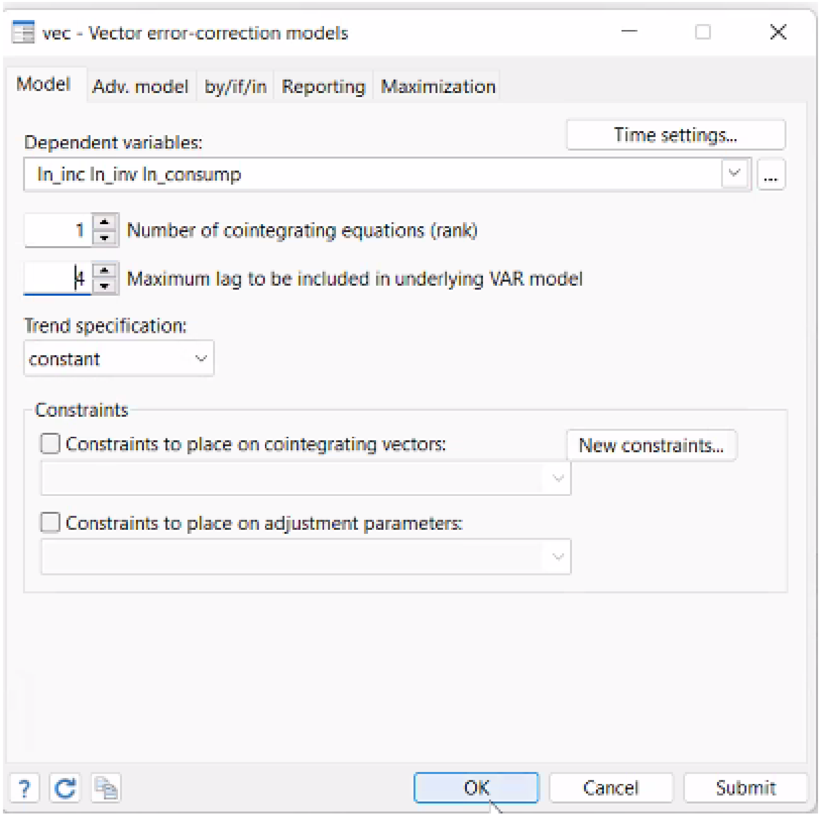
Trace statistics < critical value

Stop here, there is 1 cointegration



Dependent variable → ln_inc ln_inv ln_consump

(asla fark almıyoruz)



D_ln_inc	11	.011312	0.7777	269.3465	0.0000
D_ln_inv	11	.044728	0.2438	24.83139	0.0096
D_ln_consump	11	.009227	0.8378	397.7097	0.0000

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
D_ln_inc						
_ce1						
L1.	.1077006	.1104255	0.98	0.329	-.1087294	.3241306
ln_inc						
LD.	-.1765273	.167732	-1.05	0.293	-.5052761	.1522215
L2D.	.0022134	.1531669	0.01	0.988	-.2979882	.3024149
L3D.	.1707646	.1409287	1.21	0.226	-.1054506	.4469798
ln_inv						
LD.	.0576695	.0315866	1.83	0.068	-.0042391	.119578
L2D.	.0704422	.0326002	2.16	0.031	.006547	.1343375
L3D.	.0255336	.0323427	0.79	0.430	-.037857	.0889242
ln_consump						
LD.	.235813	.1671553	1.41	0.158	-.0918053	.5634313
L2D.	-.0316929	.1675529	-0.19	0.850	-.3600906	.2967048
L3D.	-.067531	.1506836	-0.45	0.654	-.3628654	.2278034

Error correction parameter

ln_inc						
LD.	1.018989	.6632	1.54	0.124	-.2808587	2.318837
L2D.	.6646721	.6056103	1.10	0.272	-.5223023	1.851647
L3D.	.5460865	.5572216	0.98	0.327	-.5460478	1.638221
ln_inv						
LD.	-.328338	.124891	-2.63	0.009	-.5731198	-.0835563
L2D.	-.1973008	.1288988	-1.53	0.126	-.4499379	.0553363
L3D.	-.0216915	.1278807	-0.17	0.865	-.2723331	.2289501
ln_consump						
LD.	.1619585	.6609194	0.25	0.806	-1.13342	1.457337
L2D.	.4291322	.6624918	0.65	0.517	-.8693278	1.727592
L3D.	-.0351118	.5957916	-0.06	0.953	-1.202842	1.132618
_cons	.0029673	.0175806	0.17	0.866	-.03149	.0374246
D_ln_consump						
_ce1				< 0.05		
L1.	.2809805	.0900737	3.12	0.002	.1044392	.4575218
ln_inc						
LD.	.0463056	.1368185	0.34	0.735	-.2218538	.314465
L2D.	.186618	.1249378	1.49	0.135	-.0582555	.4314915
L3D.	.1087082	.1149551	0.95	0.344	-.1165997	.3340161
ln_inv						

Johansen normalization restriction imposed

beta	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_ce1						
ln_inc	1	.	.	.	.	.
ln_inv	-.0035932	.033586	-0.11	0.915	-.0694205	.062234
ln_consump	-1.03087	.0300374	-34.32	0.000	-1.089742	-.9719977
_cons	.1349589	.	.	.	.	.

$$ln\_inc_t + 0.1349 - 0.0035 ln\_inv - 1.03 ln\_consump = 0$$

$$ln\_inc_t + 0.1349 - 0.0035 ln\_inv - 1.03 ln\_consump = 0$$

$$\overline{ln\_inc_t} = -0.1349 + 0.0035\, ln\_inv_t + 1.03\, ln\_consump_t$$

Here we did the opposite form but it is not important, in fact, this form reveals that a positive LR relationship between

$\lambda > 0$ inc. → as expected

D_ln_consump						
_ce1						
L1.	.2809805	.0900737	3.12	0.002	.1044392	.4575218
				sign		
ln_inc						
LD.	.0463056	.1368185	0.34	0.735	-.2218538	.314465
L2D.	.186618	.1249378	1.49	0.135	-.0582555	.4314915
L3D.	.1087082	.1149551	0.95	0.344	-.1165997	.3340161
ln_inv						
LD.	.0296532	.0257651	1.15	0.250	-.0208454	.0801518
L2D.	.0713132	.0265919	2.68	0.007	.0191941	.1234324
L3D.	.0429671	.0263819	1.63	0.103	-.0087404	.0946746
ln_consump						
LD.	-.2796758	.136348	-2.05	0.040	-.5469131	-.0124386
L2D.	-.1222972	.1366724	-0.89	0.371	-.3901703	.1455758
L3D.	.038841	.1229121	0.32	0.752	-.2020624	.2797444
_cons	.0041841	.0036269	1.15	0.249	-.0029245	.0112927

You found that there is error correction mechanism working to restore the LR relationship

So this estimation is not bad, it can be used for economic analysis

To graph ACF and PACF
statada "ac variablename" ve "pac variablename" şeklinde komut girince çiziyordu ACF ve PACF grafiklerini