10.03.2023
1st Lecture

Johansen Procedure (VECM Estimation)

VAR Models (Vector Autoregressive Models)

A VAR(2) model for 3 variables (X, Y, Z) is written as follows

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + a_3 Y_{t-1} + a_4 Y_{t-2} + a_5 Z_{t-1} + a_6 Z_{t-2} + u_{1t}$$
$$Y_t = b_0 + b_1 X_{t-1} + b_2 X_{t-2} + b_3 Y_{t-1} + b_4 Y_{t-2} + b_5 Z_{t-1} + b_6 Z_{t-2} + u_{2t}$$
$$Z_t = c_0 + c_1 X_{t-1} + c_2 X_{t-2} + c_3 Y_{t-1} + c_4 Y_{t-2} + c_5 Z_{t-1} + c_6 Z_{t-2} + u_{3t}$$

<u>Remark</u>: All the variables entering to the VAR system (X, Y, Z) must be I(0). [There is no unit root]

- If any variable among X, Y and Z has unit root, we cannot use VAR modeling.

- For this purpose, we need to check if the variables have unit root or not: ADF test, PP test, KPSS test, DF-GLS test, Zivot Andrews and Kapetanous Test can be used for this purpose.

- On the other hand, if all the variables have unit root, in other words if all the variables are I(1) then we can check if there is cointegration (existing of long term equilibrium relationship between variables) between these variables. This test is done by Johansen Test of Cointegration. For this purpose, we can use Trace Test (most powerful) and Maximum Eigenvalue Test (less powerful)

- If cointegration is detected, then we can estimate a Vector Error Correction Model (VECM). All these applications can be done in Stata.

- After the estimation of VECM, we will check if any error correction mechanism is working → if it is not, then we conclude that there is no trustable cointegration. In this case, VECM cannot be reported. Instead, we can convert I(1) series into I(0) series by taking their 1st differences (i.e using ΔX, ΔY, ΔZ instead of X, Y and Z) and then with these series (ΔX, ΔY, ΔZ) we can run a VAR system estimation.

VECM form of the VAR(2) system given above

<u>Remark</u>: If the VAR system has a m lag as optimum, the corresponding VECM system will have m-1 lags as optimum.

Suppose that there is following cointegration relationship between X, Y and Z

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

Error term of cointegration equation

VAR(2)

$$\begin{cases} x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + a_3 y_{t-1} + a_4 y_{t-2} + a_5 z_{t-1} + a_6 z_{t-2} + u_{t1} \\ y_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + b_3 y_{t-1} + b_4 y_{t-2} + b_5 z_{t-1} + b_6 z_{t-2} + u_{t2} \\ z_t = c_0 + c_1 x_{t-1} + c_2 x_{t-2} + c_3 y_{t-1} + c_4 y_{t-2} + c_5 z_{t-1} + c_6 z_{t-2} + u_{t3} \end{cases}$$

Intercept term is generally dropped, but the addition of the intercept term facilitates a good estimation.

corresponding VECM form

$$\begin{cases} \Delta x_t = d_0 + d_1 \Delta x_{t-1} + d_2 \Delta y_{t-1} + d_3 \Delta z_{t-1} + \lambda_x \hat{e}_{t-1} + v_{t1} \\ \Delta y_t = f_0 + f_1 \Delta x_{t-1} + f_2 \Delta y_{t-1} + f_3 \Delta z_{t-1} + \lambda_y \hat{e}_{t-1} + v_{t2} \\ \Delta z_t = g_0 + g_1 \Delta x_{t-1} + g_2 \Delta y_{t-1} + g_3 \Delta z_{t-1} + \lambda_z \hat{e}_{t-1} + v_{t3} \end{cases}$$

$\lambda_x \hat{e}_{t-1}$: ecm term; λ_x: adjustment coefficient / or ecm coefficient; v_{t1}: error term of VECM eq.

$\hat{e}_{t-1} \rightarrow$ residuals of the cointegration equation (Set with 1 lag)

Here the signs of the error correction terms are important because they must be in the same with the cointegration. In other words, the coefficients of cointegration equation imposes some conditions on these adjustment coefficients (error correction terms)

Remark: If there is really cointegration at least one of the error correction terms must be significant and must have the expected sign. Otherwise, we cannot talk about or trust any cointegration relationship.

Therefore, the required sign of the coefficients must be understood after the estimation of cointegration equation. How will we do that?

Example: Suppose that the cointegration equation is as follows

$$\hat{Y}_t = 10 + 0.5X_t - 2.1Z_t \quad \Big] \text{ Coin. eq.}$$

How can we obtain the requirements of the error correction coefficients?

We can also write in terms of Y_t:

$$Y_t = \underbrace{10 + 0.5X_t - 2.1Z_t}_{\hat{Y}_t} + \hat{e}_t$$

In equilibrium situation there will not be any deviation from equilibrium so $\hat{e}_t$ will be zero. Hence, for LR equilibrium form we can write as

$$Y_t - 10 - 0.5X_t + 2.1Z_t = 0$$

} This is the equation that we will use to determine the sign of ec terms

Suppose that there is a deviation from equilibrium in which Y_t is now above its equilibrium value Y^*

To go back to equilibrium, Y_t must decline → $\Delta Y_t < 0$

Ec term for ΔY_t equation must be negative.

Y_t is the dependent variable of the cointegration equation. That is why always its coefficient is 1. Therefore, it is always positive and must go back equilibrium. ΔY_t must be negative. This says that λ_y must be negative.

2nd Lecture
17.03.2023

We will carry out a Johansen VECM Analysis in Stata
Before starting Johansen Procedure, we be sure that all the variables entering the Johansen Analysis are I(1), which means that the variables must have unit root. For this purpose we can use the Zivot Andrews Test.

This is a variant of ADF test which is developed to take into the account one possible structural chance in the series.

Steps:

- ssc install zandrews

- zandrews LTN, break(both)

```
. zandrews LTN, break(both)

Zivot-Andrews unit root test for  LTN

Allowing for break in both intercept and trend

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

Minimum t-statistic -4.240 at 1983q4  (obs 56)

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

- t-statistic is -4.240 and critical value is -5.08
- t-statistic is not more negative than the 5% critical value, so we do not reject null hypothesis
- Null hypothesis: There is unit root in LTN series
- So, we conclude that there is unit root in LTN.

Now we will check if ΔLTN has unit root. If it does not have unit root, this means that taking the first difference remove the unit root which implies that LTN is I(1).

- zandrews D.LTN, break(both)

```
. zandrews D.LTN, break(both)

Zivot-Andrews unit root test for  D.LTN

Allowing for break in both intercept and trend

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

Minimum t-statistic -8.036 at 1981q4  (obs 48)

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

  - t-statistic is more negative than the 5% critical value. ΔLTN doen not have unit root. We reject the null hypothesis. LTN is I(1).

- zandrews STN, break(both)

```
. 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
```

- t-statistic is more negative than 5% critical value. We reject the null hypothesis. STN does not have unit root. STN is I(0). If STN is I(0), we cannot use Johansen VECM procedure. Therefore now we assume that STN is I(1) not I(0). Already, the other test may point out that it is I(1). Also Zivot Andrews test for say 10% is also saying that there is unit root. To be able to continue the application let us assume it is I(1).

- zandrews URX, break(both)

```
. zandrews URX, break(both)

Zivot-Andrews unit root test for  URX

Allowing for break in both intercept and trend

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

Minimum t-statistic -4.165 at 1997q3  (obs 111)

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

t-statistics are not more negative than critical value, so we do not reject the null hypothesis. So URX has unit root.

- zandrews D.URX, break(both)

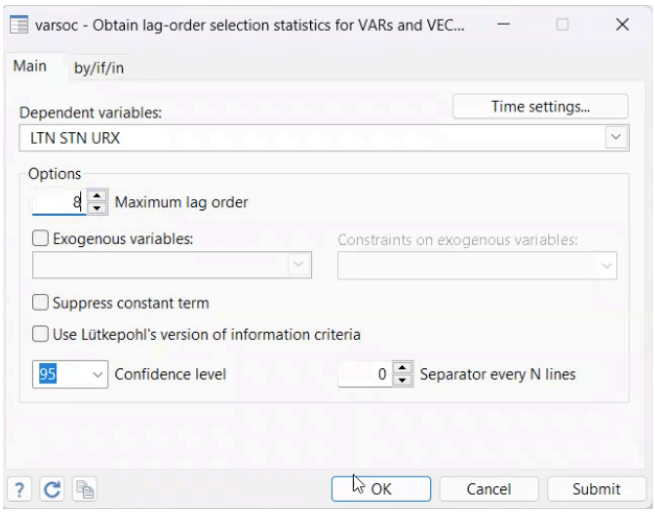
```
. zandrews D.URX, break(both)

Zivot-Andrews unit root test for  D.URX

Allowing for break in both intercept and trend

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

Minimum t-statistic -5.600 at 2008q2  (obs 154)

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

t-statistic is more negative than critical value, so URX is I(1)

Remark: All the variables entering into the VECM must be I(1)

Now let us determine the optimum lag length of the corresponding VAR system.

Statistics → Multivariate time series → VAR diagnostics and tests → Lag-order selection statistics (preestimation)



Lag-order selection criteria

Sample: **1972q1** thru **2017q4** Number of obs = **184**

Lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-334.32				.007851	3.66652	3.68776	3.71893
1	705.266	2079.2	9	0.000	1.1e-07	-7.5355	-7.45052	-7.32583
2	825.768	241	9	0.000	3.2e-08	-8.74748	-8.59876*	-8.38056*
3	839.539	27.542*	9	0.001	3.0e-08*	-8.79934*	-8.58688	-8.27516
4	845.599	12.12	9	0.207	3.1e-08	-8.76738	-8.49119	-8.0859
5	852.66	14.123	9	0.118	3.2e-08	-8.74631	-8.40638	-7.90763
6	856.472	7.6236	9	0.572	3.4e-08	-8.68991	-8.28625	-7.69398
7	862.838	12.732	9	0.175	3.5e-08	-8.66128	-8.19389	-7.5081
8	866.567	7.458	9	0.590	3.7e-08	-8.60399	-8.07286	-7.29356

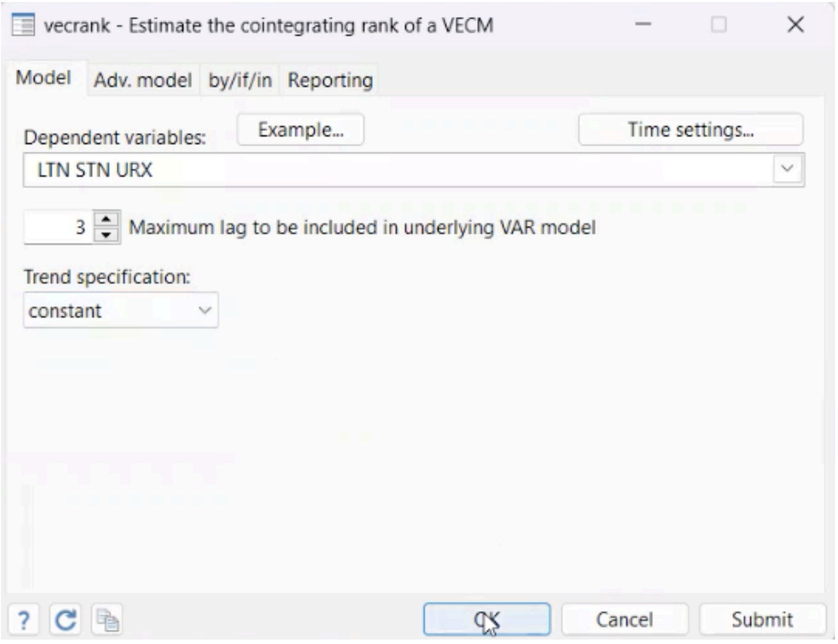
* optimal lag
Endogenous: **LTN STN URX**
Exogenous: **_cons**

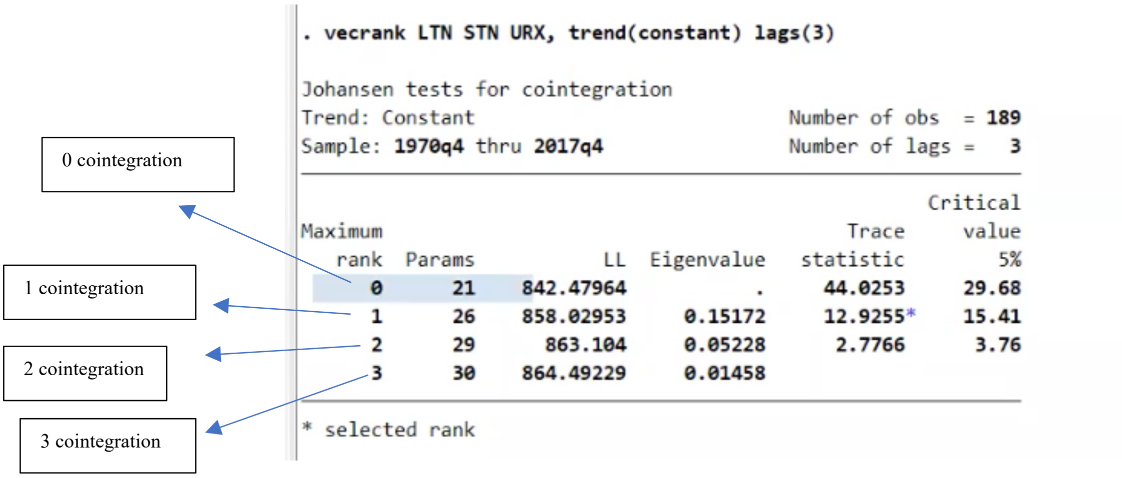
Which one has more stars, choose that one.

- Thus, we select optimal lag length as 3 for the corresponding VAR model.

Statistics → Multivariate time series → Cointegrating rank of a VECM

If the rank is 0, there is no cointegration. If the rank is 1, there is 1 cointegration…






```
. vecrank LTN STN URX, trend(constant) lags(3)

Johansen tests for cointegration
Trend: Constant                          Number of obs  = 189
Sample: 1970q4 thru 2017q4               Number of lags =   3
-------------------------------------------------------------
                                                     Critical
Maximum                                     Trace       value
   rank  Params        LL  Eigenvalue   statistic          5%
      0      21 842.47964           .     44.0253       29.68
      1      26 858.02953     0.15172     12.9255*      15.41
      2      29   863.104     0.05228      2.7766        3.76
      3      30 864.49229     0.01458
-------------------------------------------------------------
* selected rank
```

If you find 3 cointegration in a model with 3 variables, it means all of the variables are I(0).

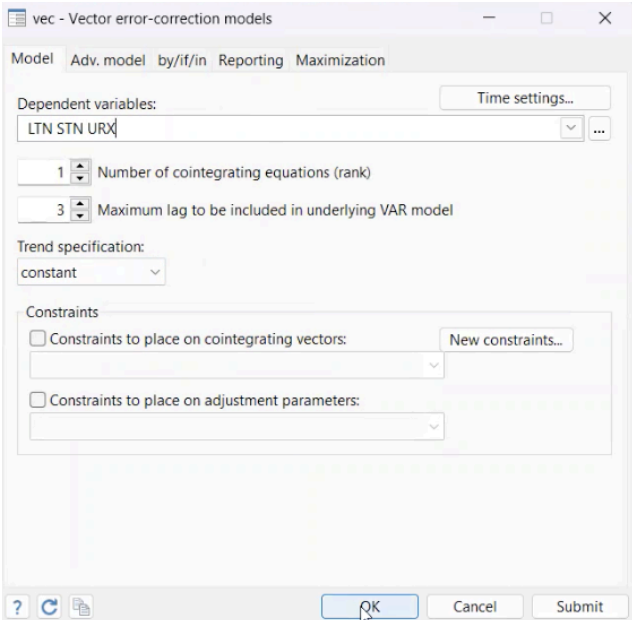
For rank 0: Trace statistic is far beyond critical value so reject the null hypothesis.
For rank 1: Trace statistic is less than critical so do not reject the null hypothesis.

Conclusion: There is 1 cointegration relationship

<u>If you find a cointegration, stop there.</u>

Now we can estimate the VECM.

Statistics → Multivariate time series → Vector error correction model (VECM)



Error term of cointegration

The VECM here is as follows

Suppose that the cointegration equation is like:

$$LTN_t = \beta_0 + \beta_1 STN_t + \beta_2 URX + u_t$$

This coefficient is expected to be negative between -1 and 0 and must be significant

<0.95 → reject H_0 that this is zero (so it is significant)

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
D_LTN						
_ce1						
L1.	-.0572417 ($\hat{\lambda}_{LTN}$)	.0287044	-1.99	0.046	-.1135012	-.0009821
LTN						
LD.	.5209223 ($\hat{a}_1$)	.0797625	6.53	0.000	.3645908	.6772539
L2D.	-.2011449 ($\hat{a}_2$)	.080161	-2.51	0.012	-.3582576	-.0440322
STN						
LD.	.0380691 ($\hat{a}_3$)	.0494745	0.77	0.442	-.0588991	.1350374
L2D.	.0341427 ($\hat{a}_4$)	.0514509	0.66	0.507	-.0666992	.1349847
URX						
LD.	-15.23477 ($\hat{a}_5$)	21.10725	-0.72	0.470	-56.60423	26.13468
L2D.	6.368898 ($\hat{a}_6$)	21.42245	0.30	0.766	-35.61833	48.35613
_cons	-.0151106 ($\hat{a}_0$)	.0241628	-0.63	0.532	-.0624689	.0322477

Since it is significant, LTN can restore the distorted equilibrium

VECM is 1 lag less.

VECM Equations:

$$\Delta LTN_t = a_0 + a_1 \Delta LTN_{t-1} + a_2 \Delta LTN_{t-2} + a_3 \Delta STN_{t-1} + a_4 \Delta STN_{t-2} + a_5 \Delta URX_{t-1} + a_6 \Delta URX_{t-2} + \lambda_{LTN} \hat{u}_{t-1} + e_{t1}$$

Adjustment coefficient of LTN

$$\Delta STN = b_0 + b_1 \Delta LTN_{t-1} + b_2 \Delta LTN_{t-2} + b_3 \Delta STN_{t-1} + b_4 \Delta STN_{t-2} + b_5 \Delta URX_{t-1} + b_6 \Delta URX_{t-2} + \lambda_{STN} \hat{u}_{t-1} + e_{t2}$$

Adjustment coefficient of STN

<u>In the cointegration relationship, which variable is in the left-hand side, the adjustment coefficient of that variable should be negative</u>

Identification: beta is exactly identified

Johansen normalization restriction imposed

beta	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_ce1						
LTN	1	.	.	.	.	.
STN	-.8611155	.0437127	-19.70	0.000	-.9467907	-.7754402
URX	3.459075	5.965053	0.58	0.562	-8.232215	15.15036
_cons	-2.117078	.	.	.	.	.

To determine λ_{STN} we need to go and check the cointegration regression.

$$LTN_t - 0.86 STN_t + 3.45 URX_t - 2.11 = 0$$

Or

$$LTN_t = 2.11 + 0.86 STN_t - 3.45 URX_t$$

<u>If LTN_t is increasing, STN_t must increase because the coefficient is negative (-0.86) to neutralize the equation. That's why λ_{STN} must be positive.</u>

As we determined above, its sign is positive and between 1 and 0 and its prob<0.05 so it is significant

D_STN						
_ce1						
L1.	.1511061	.0466773	3.24	0.001	.0596203	.2425918
LTN						
LD.	.4146899	.1297047	3.20	0.001	.1604733	.6689064
L2D.	-.3284814	.1303528	-2.52	0.012	-.5839682	-.0729946
STN						
LD.	.3921514	.0804523	4.87	0.000	.2344678	.5498351
L2D.	.0325483	.0836663	0.39	0.697	-.1314345	.1965312
URX						
LD.	-37.78027	34.32329	-1.10	0.271	-105.0527	29.49214
L2D.	-29.37253	34.83584	-0.84	0.399	-97.64952	38.90446
_cons	-.005724	.0392921	-0.15	0.884	-.0827351	.0712871

- If it is significant STN can correct the system.

$$\Delta URX_t = C_0 + C_1 LTN_{t-1} + C_2 LTN_{t-2} + C_3 STN_{t-1} + C_4 STN_{t-2} + C_5 URX_{t-1} + C_6 URX_{t-2} + \lambda_{URX} \hat{u}_{t-1} + e_{t3}$$

Adjustment coefficient of URX

To determine the adjustment coefficient of URX

```
Identification:  beta is exactly identified

                 Johansen normalization restriction imposed
```

beta	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_ce1						
LTN	1	.	.	.	.	.
STN	-.8611155	.0437127	-19.70	0.000	-.9467907	-.7754402
URX	3.459075	5.965053	0.58	0.562	-8.232215	15.15036
_cons	-2.117078	.	.	.	.	.

To determine λ_{URX} we need to go and check the cointegration regression.

$$LTN_t - 0.86 STN_t + 3.45 URX_t - 2.11 = 0$$

Or

$$LTN_t = 2.11 + 0.86\, STN_t - 3.45\, URX_t$$

<u>If LTN_t is increasing, URX_t must decline because the coefficient is negative (3.45) to neutralize the equation. That's why λ_{URX} must be negative.</u>

As we determined above, its sign is negative and between -1 and 0 and it is significant

λ_{URX}

D_URX						
_ce1						
L1.	-.0002642	.0000965	-2.74	0.006	-.0004533	-.000075
LTN						
LD.	-.0003596	.0002682	-1.34	0.180	-.0008853	.000166
L2D.	.0008554	.0002695	3.17	0.002	.0003271	.0013836
STN						
LD.	-.0001647	.0001663	-0.99	0.322	-.0004907	.0001614
L2D.	.0001481	.000173	0.86	0.392	-.0001909	.0004872
URX						
LD.	.6638319	.0709657	9.35	0.000	.5247417	.8029221
L2D.	.2022372	.0720254	2.81	0.005	.06107	.3434045
_cons	.0000745	.0000812	0.92	0.359	-.0000847	.0002337

URX is not significant in cointegration equation but the adjustment of URX is significant→ we can think in a Keynesian way such as short tun Phillips curve: it was saying that unemployment is negatively related with inflation.

URX is increasing → INF is declining so LRT is declining

We will adopt this explanation but this analysis in fact (This negative relationship) is a short analysis. Cointegration is a long run analysis that is why some can have criticism about this explanation. But this can be done.

The other statistical/econometric solution is to correct the inconsistence by chancing model structure.

There is a theoretical problem in the model. Possible solutions:

1) Check its significance if it is not significant → do not bother
2) If it is significant, →try other models

Check significance

Johansen normalization restriction imposed

beta	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_ce1						
LTN	1	.	.	.	.	.
STN	-.8611155	.0437127	-19.70	0.000	-.9467907	-.7754402
URX	3.459075	5.965053	0.58	0.562	-8.232215	15.15036
_cons	-2.117078	.	.	.	.	.

The estimated coefficient of URX is not significant. That is why we can ignore it. But the adjustment coefficient u hat not theory consistent so we need to drop it.

To do that, we need to impose a restriction on the adjustment coefficient of URX.

Now we will see how we can impose some restrictions within the VECM estimation. Basically, two types of restrictions case imposed:

1) Constraint on cointegration equation (command: bconstaints)
2) Constraint on adjustment coefficients (command: aconstaints)

Now let us restrict that λ_{URX} zero. Aşağıdakiler sadece tanımlamak için:

```
* IMPOSE RESTRICTIONS ON BETA
constraint 1 [_ce1]LTN = 1
constraint 2 [_ce1]STN = -1


* IMPOSE CONSTRAINTS ON ALPHA
constraint 3 [D_LTN]L._ce1 = 0
constraint 4 [D_STN]L._ce1 = 0
constraint 5 [D_URX]L._ce1 = 0
```

D_LTN'de adjustment coefficient 0 olsun demek

aconstraints

URX'in adjusted coefficienti istediğimiz işarette değil, 0 olsun istiyoruz. Bunu yazmak için:

- vec LTN STN URX, trend(constant) lag(3) aconstraints(5/5)

We will run this command, but VECM system must be identified this is a mathematical matter. In some cases, the restrictions may cause the unidentification of the system, then the model may not be estimated as we wanted. In this case, the other options of cointegration can be tried.

Steps:

constraint 1 [_ce1]LTN = 1
constraint 2 [_ce1]STN = -1

* IMPOSE CONSTRAINTS ON ALPHA
constraint 3 [D_LTN]L._ce1 = 0
constraint 4 [D_STN]L._ce1 = 0
constraint 5 [D_URX]L._ce1 = 0

vec LTN STN URX, trend(constant) lag(3) aconstraints(5/5)

```
          L1.  |   .0426181   .0246054     1.73   0.083    -.0056077    .0908438
               |
          URX  |
          L1.  |   .0324235   .0187196     1.73   0.083    -.0042663    .0691133
---------------+----------------------------------------------------------------
D_STN          |
          LTN  |
          L1.  |   .1632555   .0475506     3.43   0.001      .070058    .2564529
               |
          STN  |
          L1.  |  -.1355136   .0394704    -3.43   0.001    -.2128742   -.0581531
               |
          URX  |
          L1.  |  -.1030978   .0300288    -3.43   0.001    -.1619531   -.0442424
---------------+----------------------------------------------------------------
D_URX          |
          LTN  |
          L1.  |          0  (omitted)
               |
          STN  |
          L1.  |          0  (omitted)
               |
          URX  |
          L1.  |          0  (omitted)
--------------------------------------------------------------------------------

Identification:  beta is underidentified

LR test of identifying restrictions: chi2(1) = 6.457       Prob > chi2 = 0.011
```

The system became underidentified. Therefore, we did not get what we wanted. We can try other functional forms.

We can change trend(constant) command and try again

vec LTN STN URX, trend(trend) lag(3) aconstraints(5/5) → most elastic version

```
--------------------------------------------------------------------------------

Identification:  beta is underidentified

LR test of identifying restrictions: chi2(1) = 3.971       Prob > chi2 = 0.046
```

`trend(trend)` allow
the undifferenced data. A
equations are assumed to

`trend(rtrend)` defin
but allows for linear tren
a cointegrating equation

`trend(constant)` de
in the undifferenced data
is the default.

`trend(rconstant)`
quadratic trend in the un
stationary around nonzer
indicators are not allowe

`trend(none)` defines

So, we cannot convert the model into an identified one, hence we need to interpret the equation as it is.

Due to mathematical reasons we need to put more restrictions on cointegration equation as well. We need also use bconstaints command as well together with aconstraints.

Think like that LTN-STN= 0.5 , this difference is known as spread in finance: in this equation, the spread is 0.5

LTN-STN=0.5 → LTN-STN-0.5=0

Johansen n

beta	Coefficient
_ce1	
LTN	1
STN	-.8611155
URX	3.459075
_cons	-2.117078

İf we impose that this is -1

LTN-STN+µURX-cons=0

LTN-STN=cons-µURX

Spread

Do not confuse here. We want to find an explanation for a constraint that we will impose. All those things are said to explain why we can impose a restriction for the coefficient of STN (as -1)

Bu nedenle aşağıdaki command'i yazmamız lazım

- vec LTN STN URX, trend(constant) lag(3) aconstraints(5/5) bconstraint(2/2)

(1) [_ce1]STN = -1

beta	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
_ce1						
LTN	1.204716	.0675646	17.83	0.000	1.072292	1.33714
STN	-1	.	.	.	.	.
URX	-.7607925	7.790138	-0.10	0.922	-16.02918	14.5076
_cons	-2.392311	.	.	.	.	.

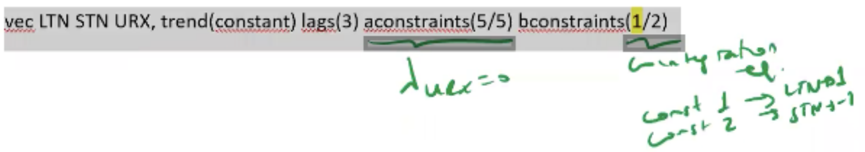
LR test of identifying restrictions: chi2(1) = 6.457 Prob > chi2 = 0.011

Now we got another problem. We want to have LTN-STN
Now new coefficient is 1.204716

Then we need to impose another constraint on LTN as a coefficient of 1.

So, the command will be

- vec LTN STN URX, trend(constant) lag(3) aconstraints(5/5) bconstraint(1/2)



Identification: beta is overidentified

(1) [_ce1]LTN = 1
(2) [_ce1]STN = -1

beta	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
_ce1						
LTN	1	.	.	.	.	.
STN	-1	.	.	.	.	.
URX	-17.11165	9.095518	-1.88	0.060	-34.93854	.7152332
_cons	.3224524	.	.	.	.	.

LR test of identifying restrictions: chi2(2) = 14.08 Prob > chi2 = 0.001

LTN-STN-17.11URX+0.322=0 **OR** LTN-STN=17.11URX-0.32

1 birim işsizlik arttığında 17 birim faiz oranı artacak anlamına geliyor

Now let us check the VECM equation and check their adjustment coefficients

$$-1<\lambda_{LTN}<0$$
$$0<\lambda_{STN}<1$$

At least one of these coefficients must be within the expected range and must be significant. If it is not the case, this implies that we cannot see any equilibrium restoration dynamic within the system. This is not compatible with the idea of stable equilibrium.

(1) [D_URX]L._ce1 = 0

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
D_LTN						
_ce1						
L1.	-.013577	.0231231	-0.59	0.557	-.0588975	.0317435
LTN						
LD.	.5170491	.080754	6.40	0.000	.3587742	.675324
L2D.	-.2124916	.0810251	-2.62	0.009	-.3712979	-.0536853
STN						
LD.	.0502033	.0498628	1.01	0.314	-.0475259	.1479325
L2D.	.0587813	.0516851	1.14	0.255	-.0425196	.1600822
URX						
LD.	-10.38006	21.31211	-0.49	0.626	-52.15102	31.3909

λ_{LTN} is not significant anymore because prob value >0.05

D_STN						
_ce1						
L1.	.127755	.036812	3.47	0.001	.0556048	.1999052
LTN						
LD.	.3981531	.1290214	3.09	0.002	.1452757	.6510305
L2D.	-.3338124	.1294523	-2.58	0.010	-.5875342	-.0800906
STN						
LD.	.3909611	.0796613	4.91	0.000	.2348279	.5470944
L2D.	.0331538	.0825495	0.40	0.688	-.1286404	.1949479
URX						
LD.	-36.99605	34.04826	-1.09	0.277	-103.7294	29.73731
L2D.	-20.85327	34.20982	-0.61	0.542	-87.90329	46.19676
_cons	-.0019045	.03893	-0.05	0.961	-.078206	.0743969

So we conclude that there is cointegration relationship between LTN, STN and URX as

beta	Coefficient
_ce1	
LTN	1
STN	-1
URX	-17.11165
_cons	.3224524

LR test of identifying restr

This relationship implies that LTN-STN=-0.3+17.1URX

$Spread_t$=-0.3+17.1URX →LR relationship

λ_{LTN} is not significant → This means that long run interest rate does not restore the distorted equilibrium.

Λ_{STN} is significant→ This means that shot run interest rate adjust to restore the equilibrium of the form like that.

07.04.2023
3rd Lecture

Pesaran Bounds Test

In the cointegration situation all the variables must be cointegrated in the same order. They all must be cointegrated in the order 1. All variables must be I(1)

This technique allows us to test if there is cointegration when some of the variables are I(0) and I(1).

White root tests are used: ADF Test, Phillips Perron, DF-GLS. These tests are ADF variant tests (H_0: There is unit root in the series)

KPSS (H_0: There is no unit root in the series)

Rule: If the prob value is less than 0.05, reject the null hypothesis.

There is not any preinstalled command in Stata for Bound test, we need to install it. There are several packages but we will focus on one of them: ARDL

Steps:

1. ssc install ardl (adoupdate ardl, update → to reinstall)
2. webuse lutkepohl2

ARDL: Autoregressive Distributed Lag

ARDL(m_1, r_1, r_2) → m: lag length at dependent variable

r_1, r_2 : lag lengths at explanatory variables

ARDL (1,0,2) → 1: soldaki değişken yani dependent variable'ın 1 gecikmesi olduğunu gösteriyor

0,2 yazması 2 tane explanatory variable olduğu, ilk explanatory variable'ın 0 gecikmesi diğerinin de 2 gecikmesi olduğu anlamına geliyor.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \beta_3 y_{t-1} + \beta_4 z_{t-1} + \beta_5 z_{t-2} + u_t$$

Steps:

1. Selection of the ARDL specification
2. ardl ln_inv ln_inc ln_consump

```
. ardl ln_inv ln_inc ln_consump

ARDL(1,0,2) regression
```

Burada bize en iyi specification'un (1,0,2) olduğunu gösterdi

```
Sample: 1961q1 thru 1982q4                        Number of obs =        88
                                                  F(5, 82)      =   1990.82
                                                  Prob > F      =    0.0000
                                                  R-squared     =    0.9918
                                                  Adj R-squared =    0.9913
Log likelihood = 158.83176                        Root MSE      =    0.0412

------------------------------------------------------------------------------
      ln_inv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      ln_inv |
         L1. |   .8432219   .0588646    14.32   0.000     .7261214    .9603224
             |
      ln_inc |  -.4477328   .3143463    -1.42   0.158    -1.073068    .1776022
             |
   ln_consump |
         --. |     1.9247   .5487929     3.51   0.001     .8329761    3.016424
         L1. |  -.3682414   .5622263    -0.65   0.514    -1.486689    .7502058
         L2. |  -.9598887   .4300221    -2.23   0.028     -1.81534   -.1044377
             |
       _cons |  -.0460065   .0706528    -0.65   0.517    -.1865575    .0945445
------------------------------------------------------------------------------
```

Lag length'i bulmuş olduk.

```
ardl ln_inv ln_inc ln_consump, lags(. . 4)
```

Yukardaki gibi kısıtlamalar koyabiliriz. Nokta koymak lag length için programın seçmesini söylemek anlamına geliyor. Eğer sayı koyarsak o sayıdaki leg lengthe baksın diyoruz.

Error correction'ı bulmak için de aşağıdaki komut kullanılıyor:

To estimate the error-correction coefficients, use option ec.

```
ardl ln_inv ln_inc ln_consump, ec
```

ardl ln_inv ln_inc ln_consump, ec

Engle Granger'da adjustment coefficent 0 ile -1 arasında,
Bizim değerimiz de bu aralıkta

```
. ardl ln_inv ln_inc ln_consump, ec

ARDL(1,0,2) regression

Sample: 1961q1 thru 1982q4                      Number of obs =        88
                                                R-squared     =    0.2228
                                                Adj R-squared =    0.1754
Log likelihood = 158.83176                      Root MSE      =    0.0412

------------------------------------------------------------------------------
    D.ln_inv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
ADJ          |
      ln_inv |
         L1. |  -.1567781   .0588646    -2.66   0.009    -.2738786   -.0396776
-------------+----------------------------------------------------------------
LR           |
      ln_inc |  -2.855838   2.522611    -1.13   0.261    -7.874116     2.16244
             |
  ln_consump |
         --. |   3.805188   2.600013     1.46   0.147    -1.367067    8.977442
-------------+----------------------------------------------------------------
SR           |
  ln_consump |
         D1. |    1.32813    .410621     3.23   0.002     .5112741    2.144986
         LD. |   .9598887   .4300221     2.23   0.028     .1044377     1.81534
             |
       _cons |  -.0460065   .0706528    -0.65   0.517    -.1865575    .0945445
------------------------------------------------------------------------------
```

Adjustment coefficent

Cointegration equation

FCM Model

No interpretation

P değeri 0.05'ten küçük yani significant

P değeri 0.05'ten büyük yani significant değil.

Yukardaki kutucuklarda yazdığım conditionlar satisfy olduğu için cointegrated olduğunu bulduk.

Cointegration equation:

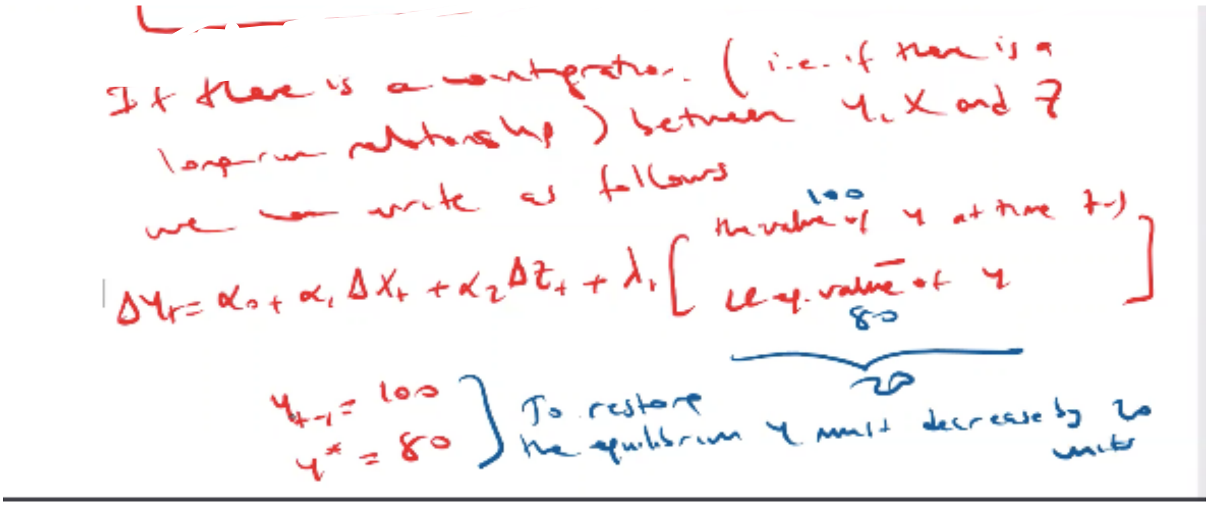
$$ln\_inv_t = \beta_0 + 3.805\, lncons - 2.85\, ln\_inc_t$$

1% increase in consumption increases investment by 3.805 → bu yorumlar aslında doğru değil çünkü

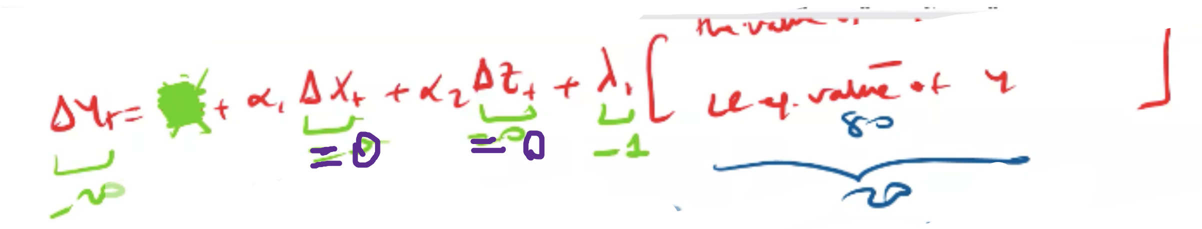
Bunun nedeni de veri setinin küçük olması

Error Correction Model

If there is cointegration (long run relationship) between Y, X and Z we can write as follows



X ve Z’yi sabit tuttuğumuzda ΔX ve ΔZ sıfıra eşit olacak:



If the lambda takes the value of -1, this equation says that if there is a deviation in the long run equilibrium by 20 units, this deviation will be restored.

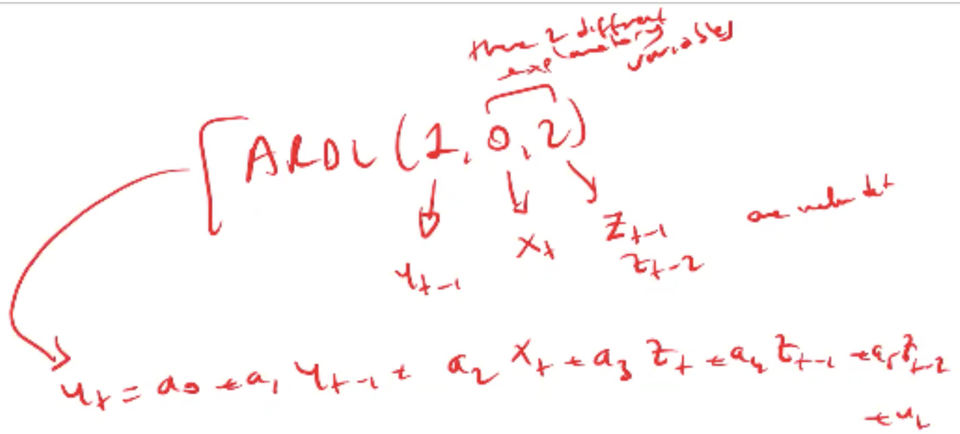
4th Lecture
14.04.2023

net install ardl, from(http://www.kripfganz.de/stata/)

adoupdate ardl, update

This test allows us to test for cointegration for a mixture of I(1) and I(0) series. (Warning: none of the variables can be I(2))

- First, we determine the order of ARDL specification.



To run bounds test, previously we need to determine ARDL specification. Fortunately, ARDL package of stata, does this for us.

- webuse lutkepohl2
- ardl ln_inv ln_inc ln_consump

				R-squared	=	0.9918
				Adj R-squared	=	0.9913
Log likelihood = 158.83176				Root MSE	=	0.0412

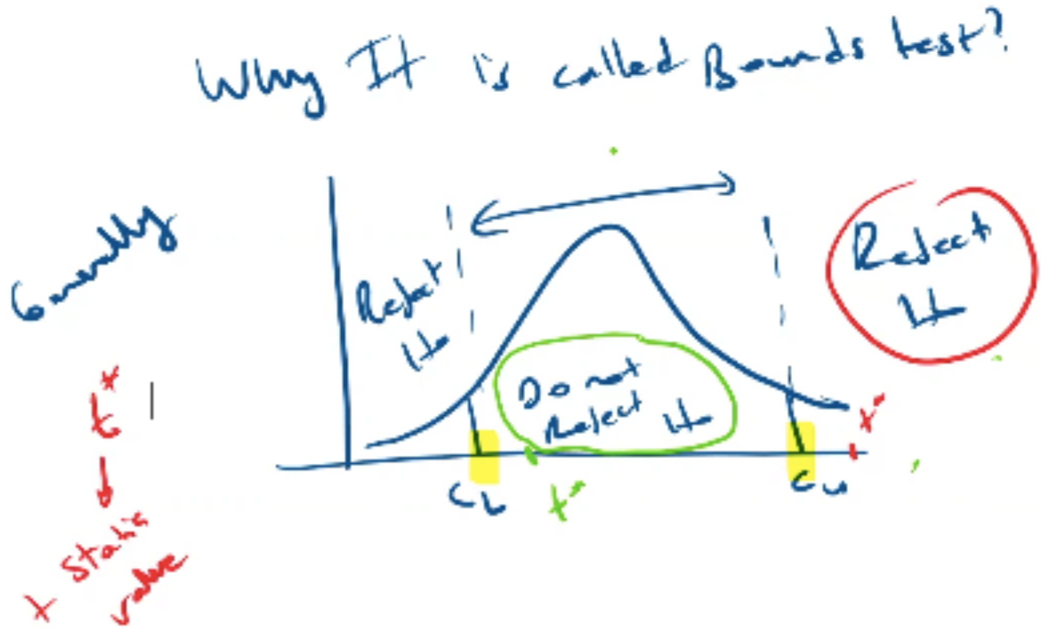
ln_inv	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ln_inv						
L1.	.8432219	.0588646	14.32	0.000	.7261214	.9603224
ln_inc	-.4477328	.3143463	-1.42	0.158	-1.073068	.1776022
ln_consump						
--.	1.9247	.5487929	3.51	0.001	.8329761	3.016424
L1.	-.3682414	.5622263	-0.65	0.514	-1.486689	.7502058
L2.	-.9598887	.4300221	-2.23	0.028	-1.81534	-.1044377
_cons	-.0460065	.0706528	-0.65	0.517	-.1865575	.0945445

Here we must check if there is autocorrelation, heteroskedasticity, ARCH effect, reset test problem etc.

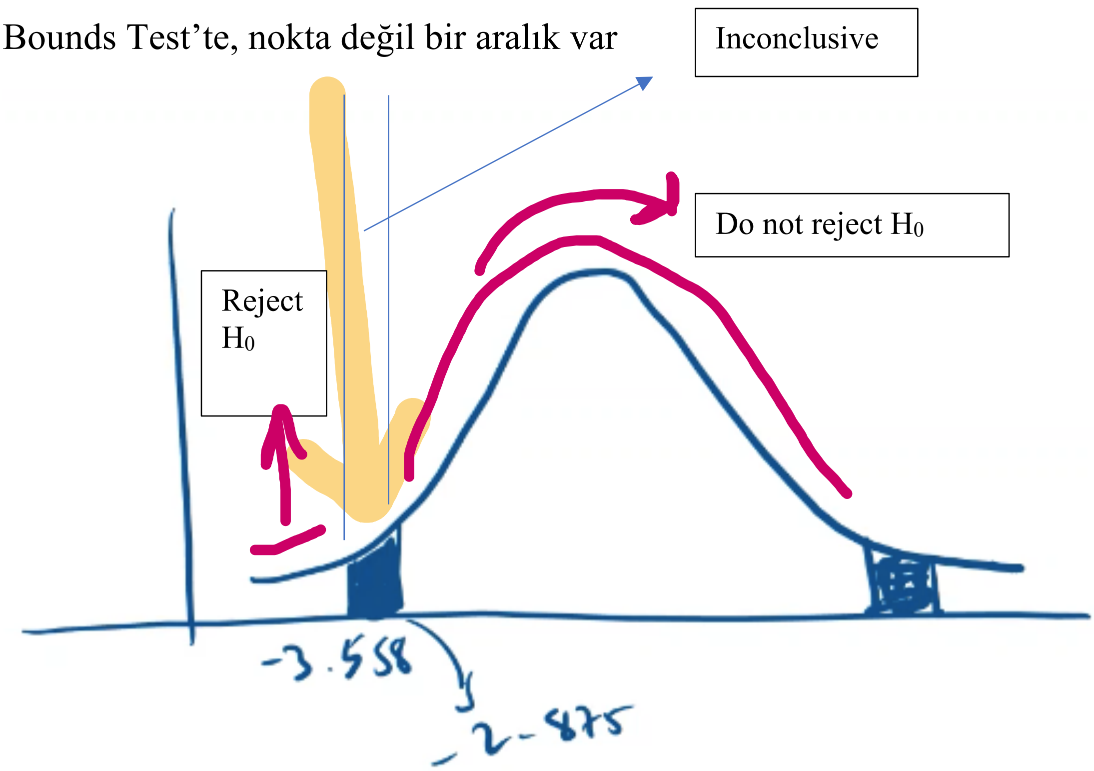
(Otokorelasyon vb varsa bunu kullanma)

This command suggests a specification for ARDL model, but you must keep in mind that this is valid if there is not any major problems in this model. (Major problems: AC, HC, ARCH, reset test (test for omitted variable situation) normality)

However, for the time being let us assume that there is no such a problem and directly show how we can run the cointegration test of Pesaran (Bounds Test)



Bounds Test'te, nokta değil bir aralık var



Let us carry out the test:

- ardl ln_inv ln_inc ln_consump, ec → auxiliary test

```
ARDL(1,0,2) regression

Sample: 1961q1 thru 1982q4                     Number of obs  =        88
                                               R-squared      =    0.2228
                                               Adj R-squared  =    0.1754
Log likelihood = 158.83176                     Root MSE       =    0.0412

------------------------------------------------------------------------------
    D.ln_inv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
ADJ          |
      ln_inv |
         L1. |  -.1567781   .0588646    -2.66   0.009    -.2738786   -.0396776
-------------+----------------------------------------------------------------
LR           |
      ln_inc |  -2.855838   2.522611    -1.13   0.261    -7.874116     2.16244
             |
  ln_consump |
         --. |   3.805188   2.600013     1.46   0.147    -1.367067    8.977442
-------------+----------------------------------------------------------------
SR           |
  ln_consump |
         D1. |    1.32813    .410621     3.23   0.002     .5112741    2.144986
         LD. |   .9598887   .4300221     2.23   0.028     .1044377     1.81534
             |
       _cons |  -.0460065   .0706528    -0.65   0.517    -.1865575    .0945445
------------------------------------------------------------------------------
```

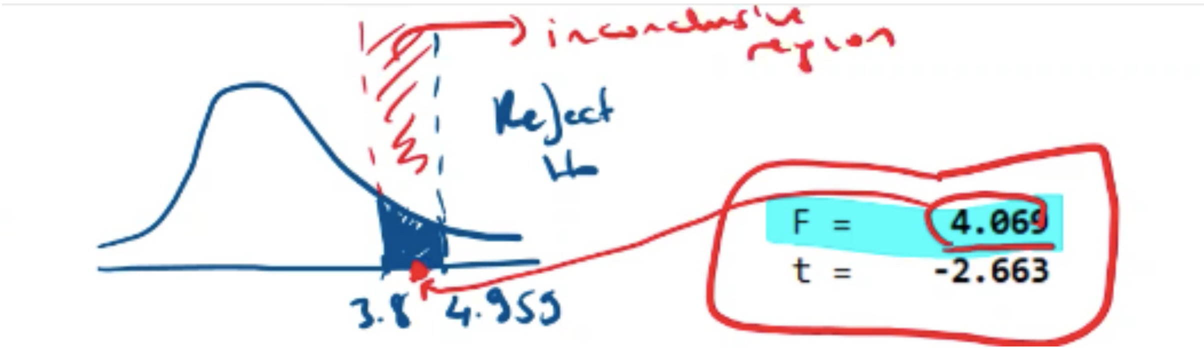
- estat ectest

```
Pesaran, Shin, and Smith (2001) bounds test

H0: no level relationship                                  F =      4.069
Case 3                                                     t =     -2.663

Finite sample (2 variables, 88 observations, 2 short-run coefficients)

Kripfganz and Schneider (2020) critical values and approximate p-values

   | 10%              | 5%               | 1%               | p-value
   |   I(0)     I(1)  |   I(0)     I(1)  |   I(0)     I(1)  |   I(0)     I(1)
---+------------------+------------------+------------------+-----------------
 F |  3.216    4.194  |  3.888    4.959  |  5.398    6.641  |  0.041    0.112
 t | -2.564   -3.224  | -2.875   -3.558  | -3.485   -4.200  |  0.081    0.253

do not reject H0 if
    either F or t are closer to zero than critical values for I(0) variables
      (if either p-value  > desired level for I(0) variables)
reject H0 if
    both F and t are more extreme than critical values for I(1) variables
      (if both   p-values < desired level for I(1) variables)

decision: no rejection (.a), inconclusive (.), or rejection (.r) at levels:

             |     10%        5%        1%
-------------+------------------------------
    decision |      .         .a        .a
```

<u>F-test is the main test. If you conclude that there is cointegration, this is very important finding, but it is not sufficient. You need to check also t-test.</u>

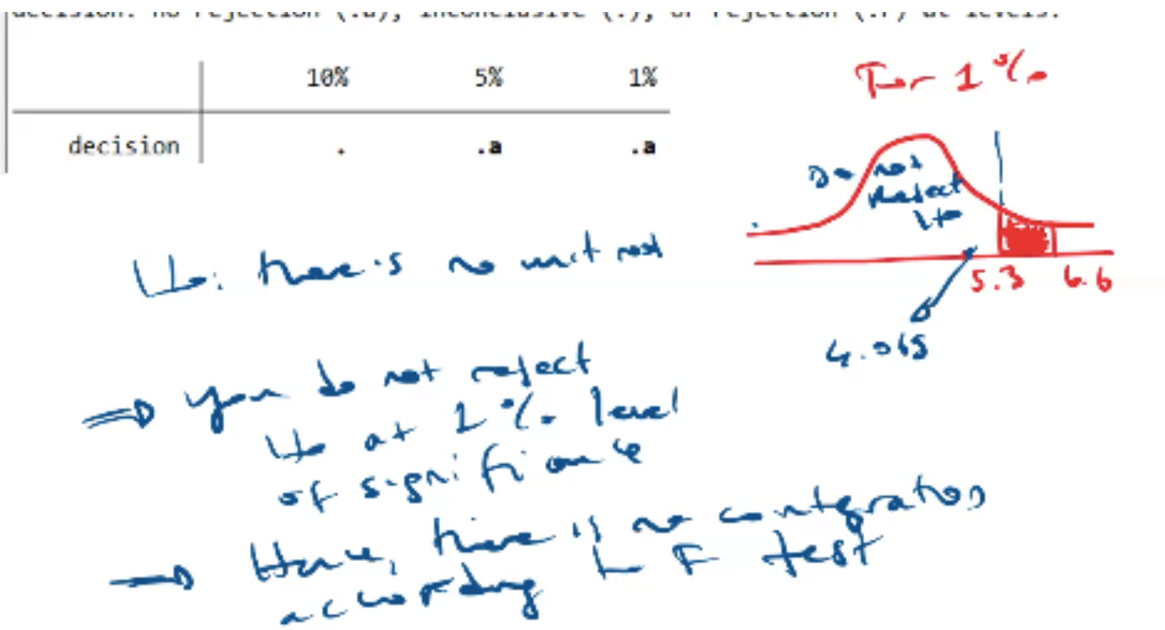
For F-test: (for 5%)

H_0 : There is no cointegration



It is in inconclusive region. F-test is not giving any conclusion.

For 1%



<u>Warning</u><u>*: If F-test says there is no cointegration, you stop here. You conclude that there is no cointegration. But if the F-test says that there is cointegration, the t-test must also be checked, and it must support there is cointegration. If t-test doesn't support that there is cointegration, the conclusion of F-test must be interpreted as untrustable.*</u>

```
F =      4.069
t =     -2.663
```

The reported t-value is in fact the t value of the adjustment coefficient of the EC model. So basically t-test is checking if the EC mechanism is working or not for the equilibrium. (if any)

If the EC mechanism does not work, we cannot trust the equilibrium finding of the F-test.

Up to now, we implicitly assume that the ARDL specification suggested by ARDL package of stata can be used. (In other words, this specification is assumed that there is no autocorrelation, no heteroscedasticity [at least], no ARCH, no reset test failure, no non-normality)

Now let us go back and check if there is any autocorrelation and or heteroscedasticity problem in the suggested ARDL specification.

ARDL (1,0,2)

For this purpose, let us rerun the model and check the AC, HC and other diagnostic.

- ardl ln_inv ln_inc ln_consump, ec

```
. ardl ln_inv ln_inc ln_consump

ARDL(1,0,2) regression

Sample: 1961q1 thru 1982q4                      Number of obs =         88
                                                F(5, 82)      =    1990.82
                                                Prob > F      =     0.0000
                                                R-squared     =     0.9918
                                                Adj R-squared =     0.9913
Log likelihood = 158.83176                      Root MSE      =     0.0412
```

ln_inv	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ln_inv						
L1.	.8432219	.0588646	14.32	0.000	.7261214	.9603224
ln_inc	-.4477328	.3143463	-1.42	0.158	-1.073068	.1776022
ln_consump						
--.	1.9247	.5487929	3.51	0.001	.8329761	3.016424
L1.	-.3682414	.5622263	-0.65	0.514	-1.486689	.7502058
L2.	-.9598887	.4300221	-2.23	0.028	-1.81534	-.1044377
_cons	-.0460065	.0706528	-0.65	0.517	-.1865575	.0945445

- estat bgodfrey

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          3.874               1                   0.0490
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Prob value < 0.05 but very close

Reject H_0 → There is AC problem at 5% value of significance (but with a very close value)

- estat hettest

```
. estat hettest

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of ln_inv

H0: Constant variance

    chi2(1) =  17.79
Prob > chi2 = 0.0000
```

Prob value < 0.05

H_0 : No heteroscedasticity

Reject H_0 → There is HC problem

This model cannot be used with these problems. However, to see the other test, let us carry out them as well)

- estat archlm

```
. estat archlm
LM test for autoregressive conditional heteroskedasticity (ARCH)
```

lags(p)	chi2	df	Prob > chi2
1	11.100	1	0.0009

```
H0: no ARCH effects      vs.  H1: ARCH(p) disturbance
```

Prob value < 0.05

H_0 : No ARCH effects

Reject H_0 → There is ARCH problem

- estat ovtest → Reset Test (omitted variable test)

```
. estat ovtest

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of ln_inv

H0: Model has no omitted variables

F(3, 79) =   0.84
Prob > F = 0.4767
```

H_0 : Model has no omitted variables

Prob value > 0.05 → do not rejected

But if there is AC and HC, this test is not valid, not trustable

→ **Now let us try to fix the model**

(Teorik olarak gecikme koyarak her zaman AC çözülür. Pratikte bunun tek istisnası mevsimsellik olması)

For this purpose let us try a time span of around 1,5 year (approximately 6 lags)

- ardl ln_inv ln_inc ln_consump, lags(6 6 6)

```
.  ardl ln_inv ln_inc ln_consump, lags(6 6 6)

ARDL(6,6,6) regression

Sample: 1961q3 thru 1982q4                      Number of obs =      86
                                                F(20, 65)     =  488.68
                                                Prob > F      =  0.0000
                                                R-squared     =  0.9934
                                                Adj R-squared =  0.9914
Log likelihood = 166.46136                      Root MSE      =  0.0402

------------------------------------------------------------------------------
      ln_inv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      ln_inv |
         L1. |   .6283761   .1275788     4.93   0.000     .3735836    .8831685
         L2. |    .084336   .1433826     0.59   0.558    -.2020187    .3706908
         L3. |    .255578   .1392123     1.84   0.071    -.0224481    .5336041
         L4. |   .2232205   .1422923     1.57   0.122    -.0609569    .5073978
         L5. |  -.3218368   .1424885    -2.26   0.027     -.606406   -.0372675
         L6. |   -.033245   .1289611    -0.26   0.797    -.2907982    .2243081
             |
      ln_inc |
         --. |  -.3372203   .5243752    -0.64   0.522     -1.38447    .7100293
         L1. |   .3195703   .6399799     0.50   0.619    -.9585577    1.597698
         L2. |  -.7221154    .635151    -1.14   0.260      -1.9906    .5463689
```

```
         L1. |   .3195703   .6399799     0.50   0.619    -.9585577    1.597698
         L2. |  -.7221154    .635151    -1.14   0.260      -1.9906    .5463689
         L3. |   .5240738   .6361705     0.82   0.413    -.7464465    1.794594
         L4. |  -.2902936    .627343    -0.46   0.645    -1.543184    .9625969
         L5. |   .3108682   .6066033     0.51   0.610    -.9006022    1.522339
         L6. |  -.3046237   .5326974    -0.57   0.569    -1.368494    .7592464
             |
  ln_consump |
         --. |   1.822809   .6903529     2.64   0.010     .4440792    3.201539
         L1. |  -.4160103   .7092602    -0.59   0.560    -1.832501     1.00048
         L2. |   -.309473   .7295255    -0.42   0.673    -1.766436     1.14749
         L3. |  -.7158921   .7080093    -1.01   0.316    -2.129884    .6981001
         L4. |  -.3308217   .6907943    -0.48   0.634    -1.710433     1.04879
         L5. |   .4567342   .6448266     0.71   0.481    -.8310734    1.744542
         L6. |   .1514744   .5489302     0.28   0.783     -.944815    1.247764
             |
       _cons |  -.0729444   .0965173    -0.76   0.453    -.2657028     .119814
------------------------------------------------------------------------------
```

- estat bgodfrey

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	**10.217**	**1**	**0.0014**

H0: no serial correlation

Prob value < 0.05

Reject H_0 → There is AC problem (even worse situation)

Let's try with 8 lags

- ardl ln_inv ln_inc ln_consump, lags(8 8 8)
- estat bgodfrey

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	**0.041**	**1**	**0.8401**

H0: no serial correlation

Prob value > 0.05

Do not reject H_0 → There is not AC problem

- estat hettest

```
. estat hettest

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of ln_inv

H0: Constant variance

    chi2(1) =   7.95
Prob > chi2 = 0.0048
```

Prob value < 0.05

So there is a problem, let us try imtest (White test variant)

- estat imtest

```
. estat imtest

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	84.00	83	0.4487
Skewness	35.40	26	0.1032
Kurtosis	1.32	1	0.2501
Total	120.72	110	0.2280

>0.05
No HC is
detected

Let us adopt this conclusion and continue…

➔ Now our aim is to decrease the lag length as much as possible

Lag of dependent variable is much more important usually. The other lags can be easily smaller in most of the cases.

Let us jump to ARDL(8,0,2)

- ardl ln_inv ln_inc ln_consump, lags(8 0 2)

```
. ardl ln_inv ln_inc ln_consump, lags(8 0 2)

ARDL(8,0,2) regression

Sample: 1962q1 thru 1982q4                         Number of obs  =        84
                                                   F(12, 71)      =    844.42
                                                   Prob > F       =    0.0000
                                                   R-squared      =    0.9930
                                                   Adj R-squared  =    0.9919
Log likelihood = 162.66641                         Root MSE       =    0.0380
```

ln_inv	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
ln_inv						
L1.	.5741399	.115102	4.99	0.000	.344633	.8036468
L2.	.1062221	.1321673	0.80	0.424	-.157312	.3697562
L3.	.1818434	.1308368	1.39	0.169	-.0790376	.4427245
L4.	.2849337	.1287399	2.21	0.030	.0282337	.5416338
L5.	-.2831615	.1295011	-2.19	0.032	-.5413793	-.0249437
L6.	.0186785	.1312485	0.14	0.887	-.2430236	.2803806
L7.	.0310161	.1271251	0.24	0.808	-.2224642	.2844965
L8.	-.163815	.1084667	-1.51	0.135	-.3800914	.0524615

- estat bgodfrey → Prob >0.05’ten büyük sorun yok

```
. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
```

lags(*p*)	chi2	df	Prob > chi2
1	0.002	1	0.9634

H0: no serial correlation

- estat imtest → Prob >0.05’ten büyük sorun yok

```
. estat imtest

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	84.00	83	0.4487
Skewness	9.97	12	0.6190
Kurtosis	1.55	1	0.2135
Total	95.51	96	0.4949

- estat hettest → There is still problem (Gecikme koyup çözülecek gibi gözükmüyor)

```
. estat hettest

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of ln_inv

H0: Constant variance

    chi2(1) =  18.43
Prob > chi2 = 0.0000
```

(imtest yapınca sorun yok ama hettest yapınca sorun varsa demek ki önemli bir değişkeni eklemeyi unutuyoruz)

ARDL(7,0,2) → yapınca da problem olmadı, modeli küçültmeye devam edebiliriz

5th Lecture
28.04.2023

PANEL DATA ANALYSIS

We have 2 types of data in econometrics

1.Cross Sectional Data
2.Time Series Data

Cross Sectional Data → Time is fixed

We have several individuals, firms (cross-sections) etc.

F	G	H	I	J
ID Number	Household	Incomes	Consumption	
1111	Ahmet	1200	1150	
1112	Leyla	1500	250	
1113	Bahar	1800	900	
5005	Barbara	5000	8000	
6225	Mehmet	10000	15000	

Time Series Data

Wrong: Do not use the data such as 2020, 2022, 2024… etc.

What about finding values estimating for 2021, 2023? This is not a good idea.

Suppose that this value is missing

How can we find the missing value? 6500 ve 6300 arasında bir değer atanacak.

If you estimate 6400 for income, the consumption peak will not match with the value.

But this is a peak value, it will give you lower the t-value.

- ***Panel Data:*** It has both time series and cross-sectional data dimension

	Years	Country	Consumption	Income
N=cross section dimesion	2010	Turkey	7261	5810.09
N=3	2011	Turkey	8891	7111.5
	2012	Turkey	7495	5996.7
	2013	Turkey	7464	5971.86
T=Time series dinesion	2014	Turkey	6594	5275.62
T=11	2015	Turkey	8430	6743.23
	2016	Turkey	7821	6257.48
	2017	Turkey	9523	7618.4
	2018	Turkey	5345	4275.68
	2019	Turkey	5531	4423.43
	2020	Turkey	7741	6191.69
	2010	France	8448	6757.44
	2011	France	7370	5895.43
	2012	France	6688	5349.2
	2013	France	9266	7414.81
	2014	France	9334	7467.96

Cross Sectional | time series | Panel Data

Turkey
Germany
France

Böyle boşluklar olunca buna unbalanced panel diyoruz

	Years	Country	Consumption	Income
N=cross section dimesion	2010	Turkey		-0.1816
N=3	2011	Turkey		0.07998
	2012	Turkey	8490	6791.8
	2013	Turkey	7449	5959.59
T=Time series dinesion	2014	Turkey	9538	7629.82
T=11	2015	Turkey		0.46753
	2016	Turkey		0.1442
	2017	Turkey	8557	6847.13
	2018	Turkey	9223	7379.05
	2019	Turkey	6332	5065.21
	2020	Turkey	7084	5666.43
	2010	France	5491	4392.37
	2011	France	6821	5458.27
	2012	France	9177	7341.28
	2013	France	6925	5540.69
	2014	France	9291	7433.61

Basic Estimation Techniques in Panel Data

First estimation contributions are as follows:

1) Fixed Effects (still can be used)
2) Random Effects (still can be used)
3) Between Effects (not good)
4) Pooled OLS (not good)

<u>STATA ADIMLARI:</u>

- webuse abdata

Değişkenler:

n log of the level of employment for firm 1 at time t
w: log reel wage of firm i at time t
k: log of capital stock for firm i at time t
ys: sector output level for firm i at time t

This is the data of afomsous paper of Arellano and Bond (1991)
This is an unbalanced data

Datanın balance ya da unbalanced olup olmadığını anlamak için "xtset" yazıyoruz.

```
. xtset

Panel variable: id (unbalanced)
 Time variable: year, 1976 to 1984
         Delta: 1 unit
```

Here this is a prepared data for us. The required xtset command has been run. But I we have to create our own panel data; we need to run the "xtsat id year" command

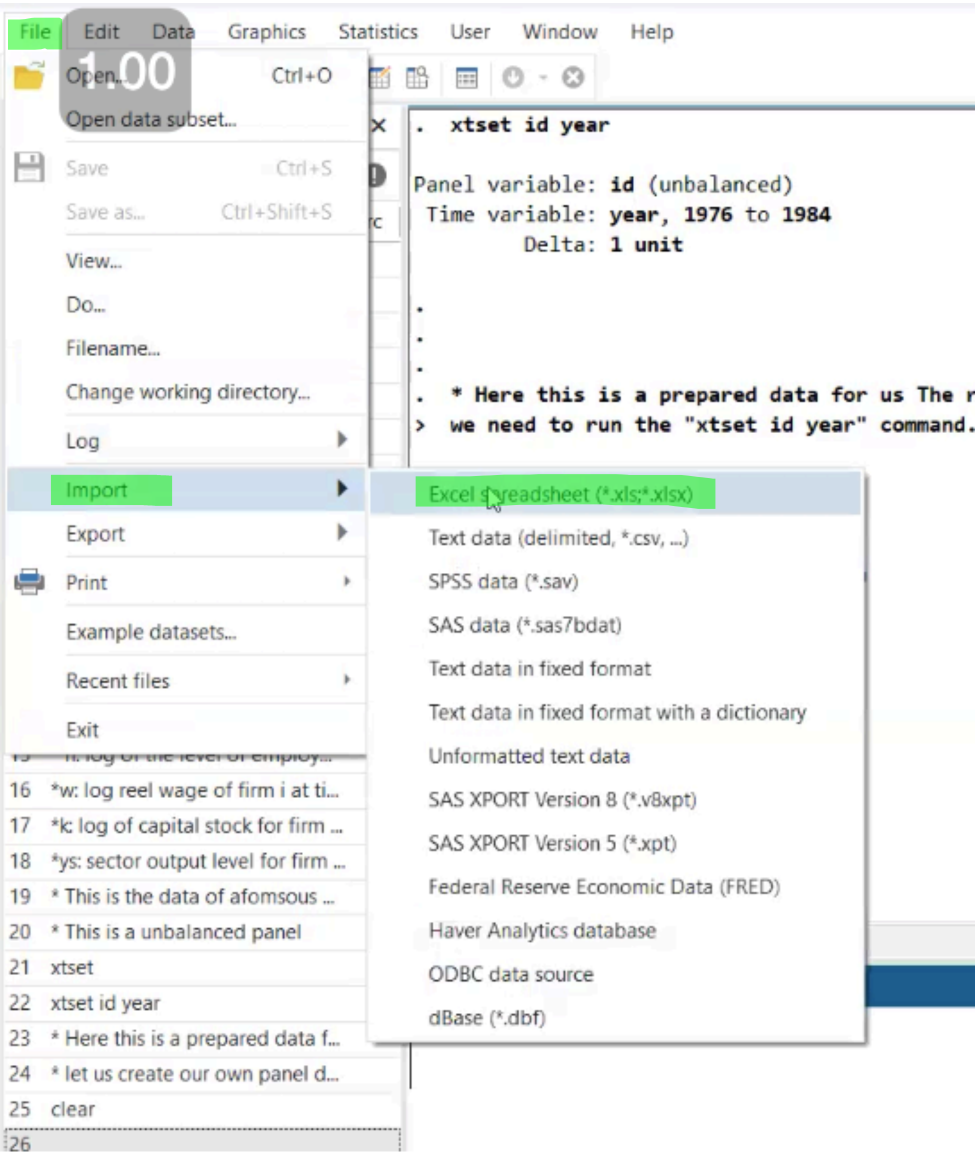
Eğer panel variable ve time variable otomatik gelmezse kendimiz atamak istersek

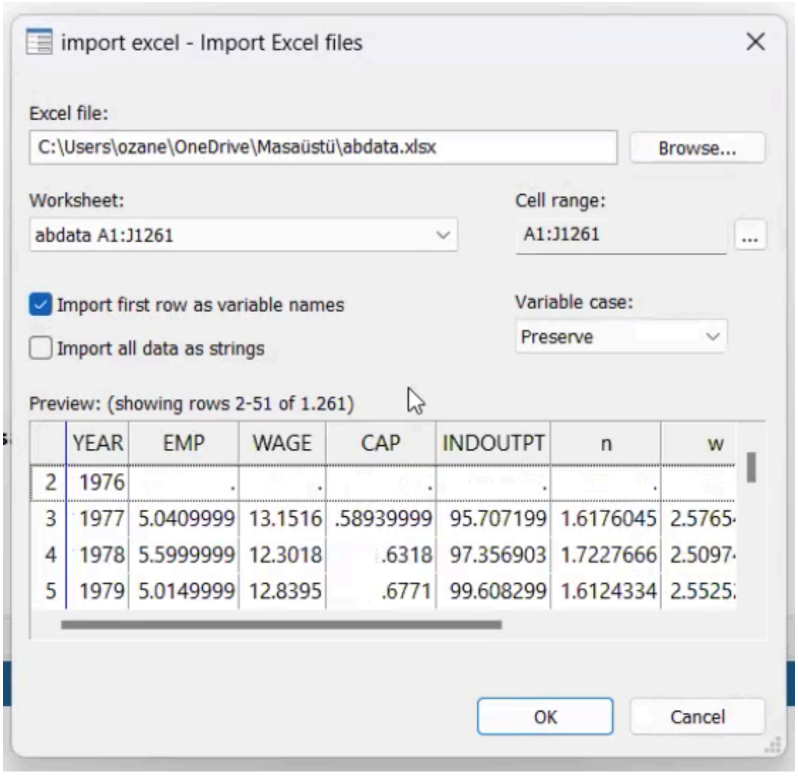
xtset id year

cross section variable → id

time variable → year

Let us create our own panel data





xtset unit year

- To estimate the fixed model, we write as follows:
- xtreg n w k ys, fe → command for fixed effect model

fixed effect

explanatory variables

Dependent variable

```
. xtreg n w k ys, fe

Fixed-effects (within) regression               Number of obs     =      1,031
Group variable: unit                            Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6143                                         min =          7
     Between = 0.8483                                         avg =        7.4
     Overall = 0.8348                                         max =          9

                                                F(3,888)          =     471.39
corr(u_i, Xb) = 0.5926                          Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |  -.3106426   .0499301    -6.22   0.000    -.4086374   -.2126479
           k |   .5489458   .0211507    25.95   0.000     .5074346     .590457
          ys |   .5370106   .0534193    10.05   0.000     .4321679    .6418533
       _cons |  -.2159125   .3108411    -0.69   0.487    -.8259814    .3941565
-------------+----------------------------------------------------------------
     sigma_u |  .66133383
     sigma_e |  .13015331
         rho |  .96271231   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 888) = 123.02                  Prob > F = 0.0000
```

- xtreg n w k ys, re → command for random effect model

```
. xtreg n w k ys, re

Random-effects GLS regression                   Number of obs     =      1,031
Group variable: unit                            Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6108                                         min =          7
     Between = 0.8479                                         avg =        7.4
     Overall = 0.8356                                         max =          9

                                                Wald chi2(3)      =    2018.16
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |  -.2900276   .0492318    -5.89   0.000    -.3865202   -.1935351
           k |    .639224   .0176213    36.28   0.000     .6046868    .6737611
          ys |   .4400793   .0529618     8.31   0.000     .3362761    .5438826
       _cons |   .2236536   .3125287     0.72   0.474    -.3888915    .8361986
-------------+----------------------------------------------------------------
     sigma_u |  .52415108
     sigma_e |  .13015331
         rho |   .9419219   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Between Effects Estimator

- xtreg n w k ys, re → command for between effects model

```
. xtreg n w k ys, be

Between regression (regression on group means)  Number of obs     =      1,031
Group variable: unit                            Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.5949                                         min =          7
     Between = 0.8488                                         avg =        7.4
     Overall = 0.8298                                         max =          9

                                                F(3,136)          =     254.58
sd(u_i + avg(e_i.)) = .5263562                  Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |  -.4553307   .1866796    -2.44   0.016    -.8245009   -.0861605
           k |   .8185982   .0296513    27.61   0.000     .7599609    .8772354
          ys |   1.586058   1.154752     1.37   0.172    -.6975352    3.869651
       _cons |  -4.496973    5.27889    -0.85   0.396     -14.9363    5.942353
------------------------------------------------------------------------------
```

- xtreg n w k ys, re → command for between effects model

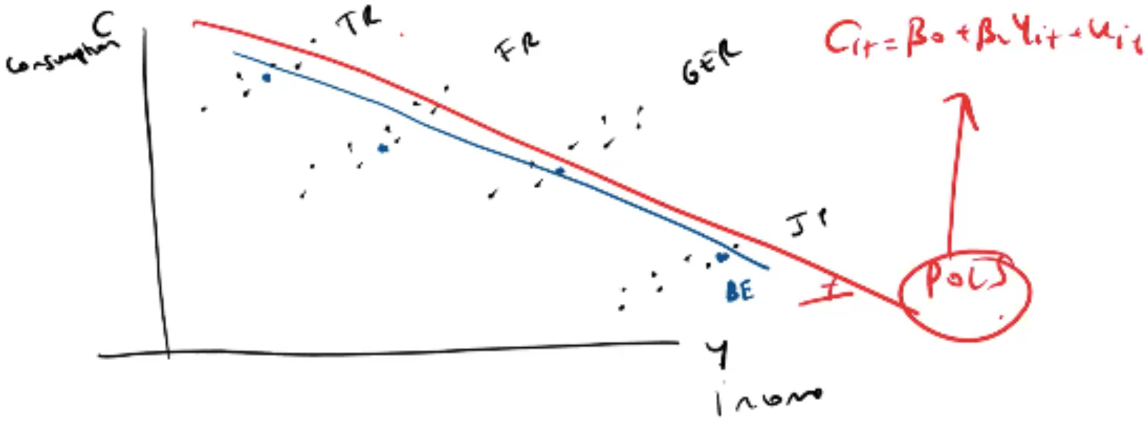
Pooled OLS (POLS estimator)

- regress n w k ys → command for pooled OLS model

```
. regress n w k ys

      Source |       SS           df       MS      Number of obs   =     1,031
-------------+----------------------------------   F(3, 1027)      =   1740.12
       Model |  1548.91136         3  516.303787   Prob > F        =    0.0000
    Residual |  304.717446     1,027  .296706374   R-squared       =    0.8356
-------------+----------------------------------   Adj R-squared   =    0.8351
       Total |  1853.62881     1,030  1.79963962   Root MSE        =    .54471

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |  -.3669498   .0646708    -5.67   0.000    -.4938518   -.2400478
           k |   .8090177   .0112526    71.90   0.000      .786937    .8310984
          ys |   .4791146   .1810233     2.65   0.008     .1238969    .8343324
       _cons |   .3444243    .860552     0.40   0.689    -1.344217    2.033065
------------------------------------------------------------------------------
```

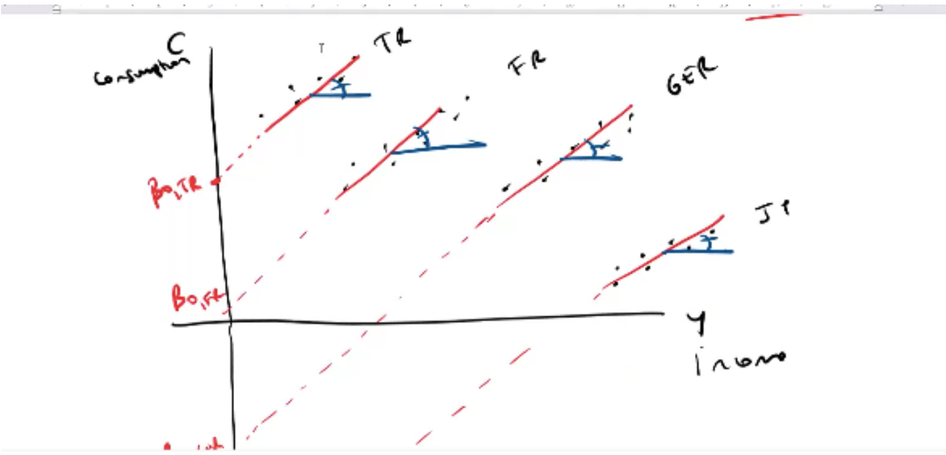
Heterogeneity Bias

Normalde income arttığında consumption artar ancak her ülkenin consumption ve income level'ları farklı ve panel data bu ülkelerin consumption ve income'larını bir olarak görüp bir bağlantı çıkardığı için sanki income arttıkça consumption azalıyor tarzı bir sonuç çıkartıyor. Buna heterogeneity bias deniliyor.

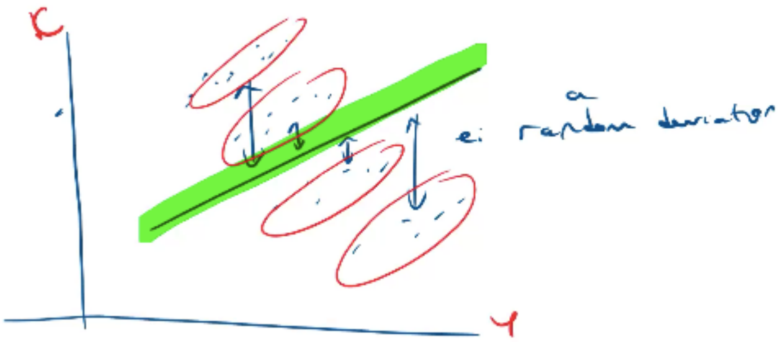


Fixed effect: Her ülkenin sabit teriminin farklılaşmasına izin veriyor.

$$C_{it} = \beta_{0,i} + \beta_1 Y_{it} + u_{it}$$



Random Effect:



$$C_{it} = \beta_{0,i} + \beta_1 Y_{it} + u_{it}$$ } Random effect

$$\beta_{0,i} = \beta_0 + \underbrace{e_i}_{\text{random term}}$$

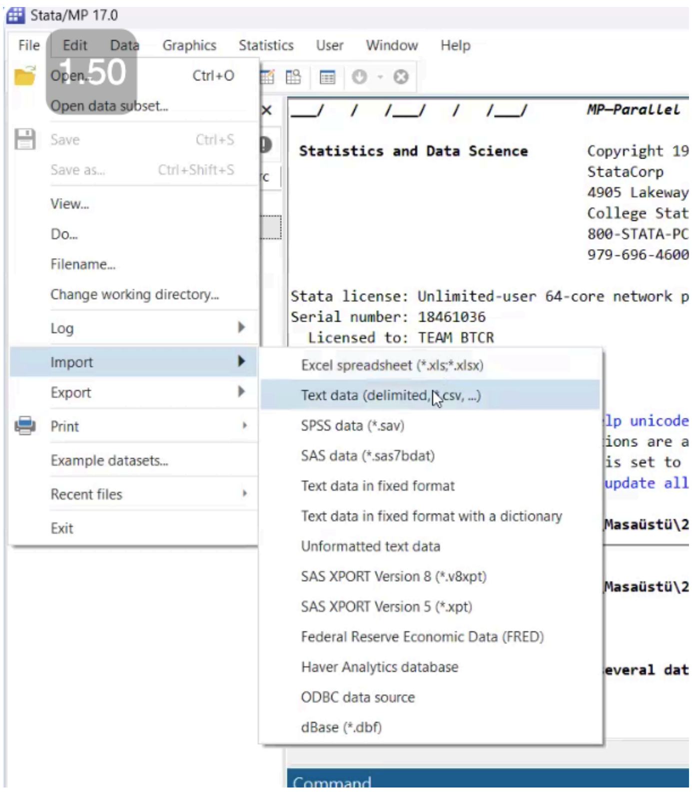
$$C_{it} = \beta_0 + \beta_1 Y_{it} + \underbrace{u_{it} + e_i}_{\epsilon_{it}}$$ } Random Effect

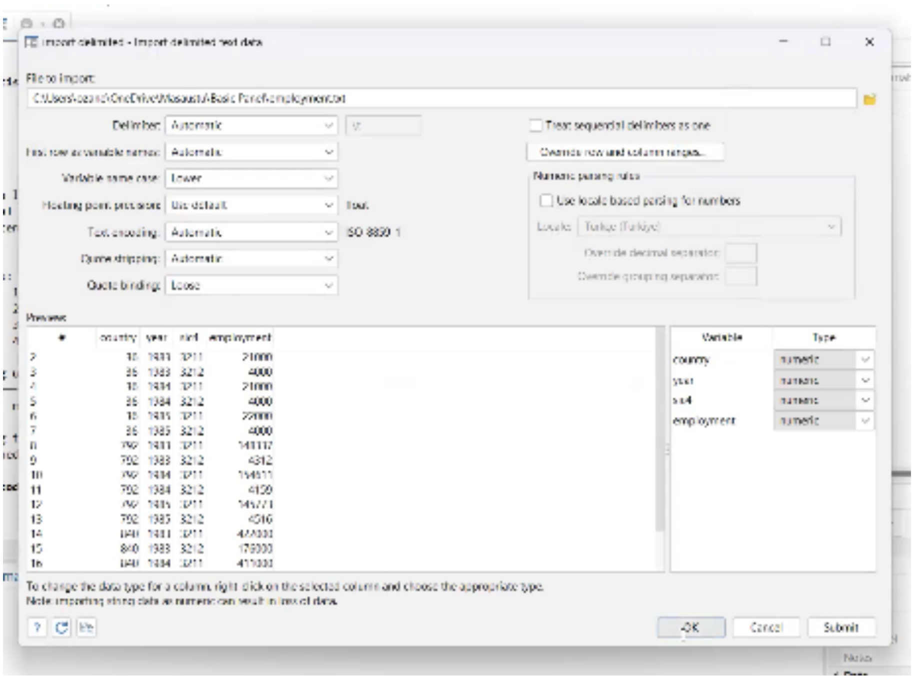
6^{th} Lecture
05.05.2023

country	year	sic4	employment
36	1983	3211	21000
36	1983	3212	4000
36	1984	3211	21000
36	1984	3212	4000
36	1985	3211	22000
36	1985	3212	4000
792	1983	3211	148337
792	1983	3212	4312
792	1984	3211	154611
792	1984	3212	4159
792	1985	3211	145773
792	1985	3212	4516
840	1983	3211	422000
840	1983	3212	176000
840	1984	3211	411000
840	1984	3212	178000
840	1985	3211	376000
840	1985	3212	175000

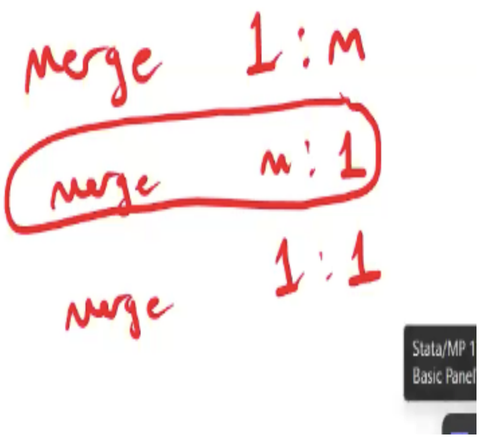
Bu panel data olmuyor çünkü bir de sektör değişkeni var ve her ülke için aynı değil

Today we will learn how to merge several data files and create a unique panel data file





sort country sic4




```
. merge 1:1 country year sic4 using output.dta

    Result                      Number of obs
    -----------------------------------------
    Not matched                             4
        from master                         1  (_merge==1)
        from using                          3  (_merge==2)

    Matched                                17  (_merge==3)
    -----------------------------------------
```

```
. use "C:\may5\all.dta"

. merge 1:1 country year sic4 using wages.dta

    Result                           Number of obs
    -----------------------------------------
    Not matched                                 1
        from master                             1  (_merge==1)
        from using                              0  (_merge==2)

    Matched                                    20  (_merge==3)
    -----------------------------------------
```

```
. merge m:1 country using name.dta

    Result                           Number of obs
    -----------------------------------------
    Not matched                                 0
    Matched                                    21  (_merge==3)
    -----------------------------------------

. drop _merge

. save "all.dta", replace
file all.dta saved

. merge 1:1 country year sic4 using valueadded.dta

    Result                           Number of obs
    -----------------------------------------
    Not matched                                 2
        from master                             2  (_merge==1)
        from using                              0  (_merge==2)

    Matched                                    19  (_merge==3)
    -----------------------------------------

. drop _merge

. save "C:\may5\all.dta", replace
file C:\may5\all.dta saved
```

```
. xtset country year
repeated time values within panel
r(451);

. * here since our current data format is in multipanel type (we have more than 1 cross sectional unit: country, sic4) we can
> not run the xtset command. Here we need to choose one of the corss sectioanl unit

. * suppose we only work with sector 3211 in sec4 variable
```

```
. keep if sic4==3211
(12 observations deleted)

. save "C:\may5\panel1.dta"
file C:\may5\panel1.dta saved
```

- xtset country year

```
. xtset country year

Panel variable: country (strongly balanced)
 Time variable: year, 1983 to 1985
         Delta: 1 unit
```

- xtreg output employment wages, fe

xtreg→ regresyon yazmak için

Eğer 0'a yakınsa fixed yerine random effect yapılması daha iyi

```
. xtreg output employment wages, fe

Fixed-effects (within) regression               Number of obs     =          8
Group variable: country                         Number of groups  =          3

R-squared:                                      Obs per group:
     Within  = 0.9999                                         min =          2
     Between = 1.0000                                         avg =        2.7
     Overall = 0.9938                                         max =          3

                                                F(2,3)            =   22857.74
corr(u_i, Xb) = -0.9526                         Prob > F          =     0.0000

------------------------------------------------------------------------------
      output | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  employment |  -54200.47   79775.31    -0.68   0.546    -308081.1    199680.2
       wages |   10.85633    .050878   213.38   0.000     10.69442    11.01825
       _cons |  -1.18e+11   1.73e+10    -6.82   0.006    -1.73e+11   -6.31e+10
-------------+----------------------------------------------------------------
     sigma_u |  1.702e+11
     sigma_e |  2.754e+09
         rho |  .99973826   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(2, 3) = 982.79                      Prob > F = 0.0001
```

<0.05 so reject the H_0

FE is good

H_0: All intercepts are the same. (No need for different intercepts for countries.)

H_a: different intercepts must exist

- xtreg output employment wages, re

```
. xtreg output employment wages, re

Random-effects GLS regression                   Number of obs     =          8
Group variable: country                         Number of groups  =          3

R-squared:                                      Obs per group:
     Within  = 0.9999                                         min =          2
     Between = 0.9999                                         avg =        2.7
     Overall = 0.9938                                         max =          3

                                                Wald chi2(2)      =     807.22
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      output | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  employment |  -24874.03   128402.9    -0.19   0.846      -276539      226791
       wages |   8.720793   .3190596    27.33   0.000     8.095447    9.346138
       _cons |  -1.76e+10   4.01e+10    -0.44   0.661    -9.61e+10    6.09e+10
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  2.754e+09
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

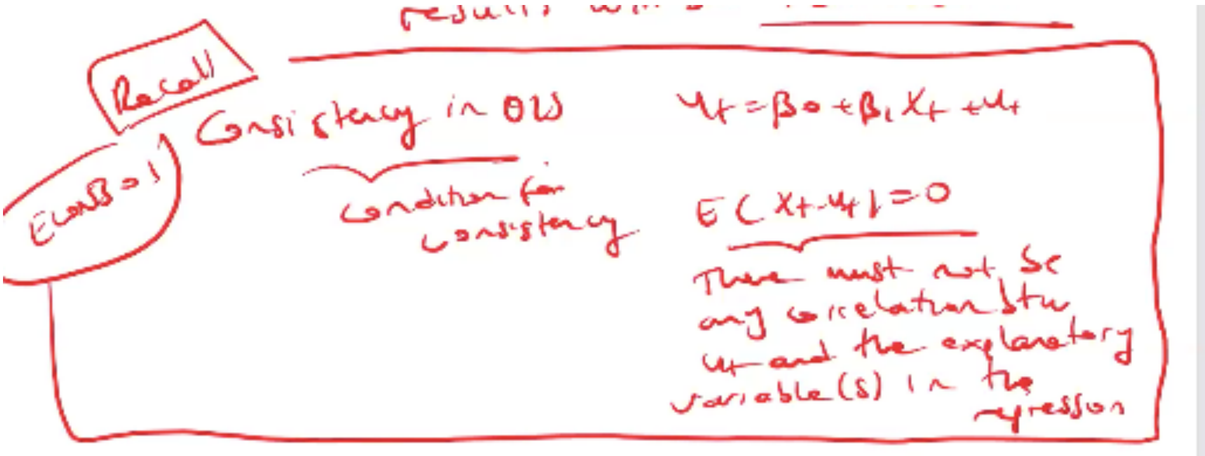
7th Lecture
26.05.2023

Selecting Appropriate Model Form

Random Effects or Fixed Effects

Info: FE can always be used. FE estimator is always consistent. But we cannot say the opposite. RE cannot always be used. If the appropriate estimation method is FE but you use RE estimator, in this case RE estimation results will be inconsistent.

Recall:



Types of Panel Data

T → Time Series Dimension
N → Cross Sectional Dimension

T > N → Long Panel (Macroeconomic studies)
If T>20 and T>N, then we can carry out Panel Cointegration

T>N → Short Panel (Microeconomic studies)

- webuse abdata
- xtreg n k w, fe

```
. xtreg n k w ys, fe

Fixed-effects (within) regression               Number of obs     =      1,031
Group variable: id                              Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6143                                         min =          7
     Between = 0.8483                                         avg =        7.4
     Overall = 0.8348                                         max =          9

                                                F(3,888)          =     471.39
corr(u_i, Xb) = 0.5926                          Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5489458   .0211507    25.95   0.000     .5074346     .590457
           w |  -.3106425   .0499301    -6.22   0.000    -.4086372   -.2126478
          ys |   .5370108   .0534193    10.05   0.000      .432168    .6418535
       _cons |  -.2159137   .3108411    -0.69   0.487    -.8259826    .3941552
-------------+----------------------------------------------------------------
     sigma_u |  .66133388
     sigma_e |  .13015331
         rho |  .96271232   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 888) = 123.02                  Prob > F = 0.0000

.
```

➔ To test FE vs POLS, we check if the reported F stat is significant. Here, reported F stat is F(139, 888) = 123.02 Prob=0.00 < 0.05 ➔ H_0: All firms has the same B_0 (POLS), H_a: all firms has different B_0 (FE)

- xtreg n k w ys, re

```
. xtreg n k w ys, re

Random-effects GLS regression                   Number of obs     =      1,031
Group variable: id                              Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6108                                         min =          7
     Between = 0.8479                                         avg =        7.4
     Overall = 0.8356                                         max =          9

                                                Wald chi2(3)      =    2018.16
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |    .639224   .0176213    36.28   0.000     .6046868    .6737611
           w |  -.2900276   .0492318    -5.89   0.000    -.3865201    -.193535
          ys |   .4400795   .0529618     8.31   0.000     .3362762    .5438828
       _cons |   .2236526   .3125287     0.72   0.474    -.3888925    .8361977
-------------+----------------------------------------------------------------
     sigma_u |  .52415108
     sigma_e |  .13015331
         rho |  .94192191   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

We can also estimate a two-way fixed effect model. In this case we also let the B_0 change for each year not only for each firm.

To carry out two-way fixed effect regression, we need to add a dummy variable for each years as follows:

- xtreg n k w ys i.year, fe

```
                                                F(11,880)         =     137.45
corr(u_i, Xb) = 0.5930                          Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5475598   .0217733    25.15   0.000     .5048261    .5902934
           w |  -.2968768   .0553473    -5.36   0.000     -.405505   -.1882485
          ys |   .2648254   .0819989     3.23   0.001     .1038893    .4257616
             |
        year |
        1977 |  -.0382327   .0187438    -2.04   0.042    -.0750204    -.001445
        1978 |  -.0638061   .0190658    -3.35   0.001    -.1012259   -.0263864
        1979 |  -.0746483   .0189938    -3.93   0.000    -.1119267   -.0373699
        1980 |  -.0763939   .0192734    -3.96   0.000    -.1142212   -.0385667
        1981 |  -.1071345   .0211646    -5.06   0.000    -.1486734   -.0655955
        1982 |  -.1233866   .0219401    -5.62   0.000    -.1664477   -.0803256
        1983 |  -.1274072   .0246385    -5.17   0.000    -.1757642   -.0790502
        1984 |   -.101978   .0290425    -3.51   0.000    -.1589787   -.0449774
             |
       _cons |   1.081303   .4040474     2.68   0.008     .2882943    1.874312
-------------+----------------------------------------------------------------
     sigma_u |  .66232129
     sigma_e |  .12768701
         rho |    .964165   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 880) = 127.28                  Prob > F = 0.0000
```

Bu her yıl için ayrı bir dummy koy demek

- testparm i.year

```
. testparm i.year

 ( 1)  1977.year = 0
 ( 2)  1978.year = 0
 ( 3)  1979.year = 0
 ( 4)  1980.year = 0
 ( 5)  1981.year = 0
 ( 6)  1982.year = 0
 ( 7)  1983.year = 0
 ( 8)  1984.year = 0

       F(  8,   880) =    5.33
            Prob > F =    0.0000
```

Bu dummy'lerin toplu olarak anlamlı olduğu testini yapmak için

0.05'ten küçük, H_0'ı reddediyoruz. Yani year dummies are jointly significant
H_0: Hepsi sıfır
Yani Two way fixed effect (i.year diye yaptığımız) one way fixed effectten daha iyi

Let us carry out the Hausman Test

We need to estimate the model by both RE and FE estimators.

- xtreg n k w ys, fe

```
Fixed-effects (within) regression               Number of obs     =      1,031
Group variable: id                              Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6143                                         min =          7
     Between = 0.8483                                         avg =        7.4
     Overall = 0.8348                                         max =          9

                                                F(3,888)          =     471.39
corr(u_i, Xb) = 0.5926                          Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5489458   .0211507    25.95   0.000     .5074346     .590457
           w |  -.3106425   .0499301    -6.22   0.000    -.4086372   -.2126478
          ys |   .5370108   .0534193    10.05   0.000      .432168    .6418535
       _cons |  -.2159137   .3108411    -0.69   0.487    -.8259826    .3941552
-------------+----------------------------------------------------------------
     sigma_u |  .66133388
     sigma_e |  .13015331
         rho |  .96271232   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 888) = 123.02                  Prob > F = 0.0000

.
```

- estimates store sabit

Son bulduğum estimate sonuçlarını hafızanda tut anlamına geliyor

Sabit ismini biz verdik, başka isim de verilebilir

- xtreg n k w ys, re

```
. xtreg n k w ys, re

Random-effects GLS regression                   Number of obs     =      1,031
Group variable: id                              Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6108                                         min =          7
     Between = 0.8479                                         avg =        7.4
     Overall = 0.8356                                         max =          9

                                                Wald chi2(3)      =    2018.16
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |    .639224   .0176213    36.28   0.000     .6046868    .6737611
           w |  -.2900276   .0492318    -5.89   0.000    -.3865201    -.193535
          ys |   .4400795   .0529618     8.31   0.000     .3362762    .5438828
       _cons |   .2236526   .3125287     0.72   0.474    -.3888925    .8361977
-------------+----------------------------------------------------------------
     sigma_u |  .52415108
     sigma_e |  .13015331
         rho |  .94192191   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- estimates store rassal
- hausman sabit rassal

Önce fixed effect estimator'ın yazılması gerekiyor. Tersten yazılırsa test yanlış çıkar

```
. hausman sabit rassal

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     sabit        rassal        Difference       Std. err.
-------------+----------------------------------------------------------------
           k |   .5489458      .639224       -.0902782        .0116979
           w |  -.3106425     -.2900276       -.020615        .0083212
          ys |   .5370108      .4400795        .0969313        .0069758
------------------------------------------------------------------------------
                          b = Consistent under H0 and Ha; obtained from xtreg.
           B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

    chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =  62.76
Prob > chi2 = 0.0000
(V_b-V_B is not positive definite)
```

H_0: Random Effects (RE)
H_a: Fixed Effects (FE)

Prob: 0.00 <0.05 → Reject the null hypothesis → You reject Random Effect Estimator

You must not use Random Effects

- xtreg n k ys i.year, re

```
    Overall = 0.8355                                         max =          9

                                                Wald chi2(11)    =    2137.45
corr(u_i, X) = 0 (assumed)                      Prob > chi2      =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .6458372   .0178711    36.14   0.000     .6108105    .6808638
           w |   -.299485   .0544527    -5.50   0.000    -.4062103   -.1927598
          ys |   .1783213   .0833213     2.14   0.032     .0150145    .3416281
             |
        year |
        1977 |  -.0364859   .0192939    -1.89   0.059    -.0743011    .0013294
        1978 |  -.0666255   .0195998    -3.40   0.001    -.1050404   -.0282107
        1979 |  -.0803878   .0195241    -4.12   0.000    -.1186543   -.0421212
        1980 |  -.0845393   .0198067    -4.27   0.000    -.1233598   -.0457189
        1981 |  -.1125847   .0217924    -5.17   0.000    -.1552971   -.0698724
        1982 |  -.1180702   .0225988    -5.22   0.000     -.162363   -.0737773
        1983 |  -.1170761   .0253322    -4.62   0.000    -.1667264   -.0674259
        1984 |  -.0844213   .0298113    -2.83   0.005    -.1428504   -.0259923
             |
       _cons |   1.549606   .4113648     3.77   0.000     .7433454    2.355866
-------------+----------------------------------------------------------------
     sigma_u |   .5055356
     sigma_e |  .12768701
         rho |  .94003031   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Let us carry Hausman test for the model with time dummies

- xtreg n k w ys i.year, fe

```
                                                F(11,880)         =     137.45
corr(u_i, Xb) = 0.5930                          Prob > F          =     0.0000

           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5475598   .0217733    25.15   0.000     .5048261    .5902934
           w |  -.2968768   .0553473    -5.36   0.000     -.405505   -.1882485
          ys |   .2648254   .0819989     3.23   0.001     .1038893    .4257616
             |
        year |
       1977  |  -.0382327   .0187438    -2.04   0.042    -.0750204    -.001445
       1978  |  -.0638061   .0190658    -3.35   0.001    -.1012259   -.0263864
       1979  |  -.0746483   .0189938    -3.93   0.000    -.1119267   -.0373699
       1980  |  -.0763939   .0192734    -3.96   0.000    -.1142212   -.0385667
       1981  |  -.1071345   .0211646    -5.06   0.000    -.1486734   -.0655955
       1982  |  -.1233866   .0219401    -5.62   0.000    -.1664477   -.0803256
       1983  |  -.1274072   .0246385    -5.17   0.000    -.1757642   -.0790502
       1984  |   -.101978   .0290425    -3.51   0.000    -.1589787   -.0449774
             |
       _cons |   1.081303   .4040474     2.68   0.008     .2882943    1.874312
-------------+----------------------------------------------------------------
     sigma_u |  .66232129
     sigma_e |  .12768701
         rho |    .964165   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 880) = 127.28                  Prob > F = 0.0000
```

- estimates store sabit_dum
- xtreg n k ys i.year, re
- estimates store rassal_dum
- hausman sabit_dum rassal_dum

```
. hausman sabit_dum rassal_dum

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |   sabit_dum    rassal_dum     Difference         Std. err.
-------------+----------------------------------------------------------------
           k |    .5475598     .6458372       -.0982774        .0124379
           w |   -.2968768     -.299485        .0026083        .0099114
          ys |    .2648254     .1783213        .0865042               .
        year |
        1977 |   -.0382327    -.0364859       -.0017468               .
        1978 |   -.0638061    -.0666255        .0028194               .
        1979 |   -.0746483    -.0803878        .0057395               .
        1980 |   -.0763939    -.0845393        .0081454               .
        1981 |   -.1071345    -.1125847        .0054503               .
        1982 |   -.1233866    -.1180702       -.0053165               .
        1983 |   -.1274072    -.1170761       -.0103311               .
        1984 |    -.101978    -.0844213       -.0175567               .
------------------------------------------------------------------------------
                          b = Consistent under H0 and Ha; obtained from xtreg.
           B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

    chi2(11) = (b-B)'[(V_b-V_B)^(-1)](b-B)
             =  63.39
Prob > chi2 = 0.0000
(V_b-V_B is not positive definite)
```

Again, FE is pointed out by Hausman test. RE must not be used. The best choice here seems to use a two-way fixed effect model estimation

Cross Sectional Dependence: Yatay kesitler arası otokorelasyon

Korelasyon çıkınca da RE ve FE çöp oluyor. Cross Sectional Dependence çıkıyorsa, buna dikkat edilerek tahmin yapılması gerekiyor.

The usual RE and FE estimators that we did up to now are implicitly assuming that there is no cross-sectional dependence.

If any CD is detected, the RE and FE estimators must be run with Driscoll Kraay standard error (They are robust to CD problem.)

- ssc install xtcsd

Let us test if there is CD problem in usual FE estimation.

- xtreg n k w ys, fe

- xtcsd, pesaran abs

```
R-squared:                                      Obs per group:
     Within  = 0.6143                                         min =          7
     Between = 0.8483                                         avg =        7.4
     Overall = 0.8348                                         max =          9

                                                F(3,888)          =     471.39
corr(u_i, Xb) = 0.5926                          Prob > F          =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5489458   .0211507    25.95   0.000     .5074346     .590457
           w |  -.3106425   .0499301    -6.22   0.000    -.4086372   -.2126478
          ys |   .5370108   .0534193    10.05   0.000      .432168    .6418535
       _cons |  -.2159137   .3108411    -0.69   0.487    -.8259826    .3941552
-------------+----------------------------------------------------------------
     sigma_u |  .66133388
     sigma_e |  .13015331
         rho |  .96271232   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(139, 888) = 123.02                  Prob > F = 0.0000

. xtcsd, pesaran abs

Pesaran's test of cross sectional independence =     5.387, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.512

.
```

0.00 <0.05
Reject the H_0
There is CD

H_0: No CD

H_A: There is CD

We must use Driscoll Krayy standard error to solve the CD problem in FE estimation (This option is only available in Stata)

- ssc install xtscc
- xtscc n k w ys, fe

```
. xtscc n k w ys, fe

Regression with Driscoll-Kraay standard errors   Number of obs     =      1031
Method: Fixed-effects regression                 Number of groups  =       140
Group variable (i): id                           F(  3,       8)   =    204.01
maximum lag: 2                                   Prob > F          =    0.0000
                                                 within R-squared  =    0.6143

------------------------------------------------------------------------------
             |             Drisc/Kraay
           n | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |   .5489458   .0390942    14.04   0.000     .4587943    .6390972
           w |  -.3106425   .1417172    -2.19   0.060     -.637443    .0161579
          ys |   .5370108   .0706657     7.60   0.000     .3740553    .6999662
       _cons |  -.2159137   .7299604    -0.30   0.775    -1.899205    1.467378
------------------------------------------------------------------------------
```

These values can be used. Let us also test if there is CD problem in RE estimation.

- xtreg n k w ys, re

```
. xtreg n k w ys, re

Random-effects GLS regression                   Number of obs     =      1,031
Group variable: id                              Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6108                                         min =          7
     Between = 0.8479                                         avg =        7.4
     Overall = 0.8356                                         max =          9

                                                Wald chi2(3)      =    2018.16
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |    .639224   .0176213    36.28   0.000     .6046868    .6737611
           w |  -.2900276   .0492318    -5.89   0.000    -.3865201    -.193535
          ys |   .4400795   .0529618     8.31   0.000     .3362762    .5438828
       _cons |   .2236526   .3125287     0.72   0.474    -.3888925    .8361977
-------------+----------------------------------------------------------------
     sigma_u |  .52415108
     sigma_e |  .13015331
         rho |  .94192191   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- xtcsd, pesaran abs

```
Pesaran's test of cross sectional independence =     5.238, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.500
```

0.00 <0.05
Reject the H_0
There is CD

- xtscc n k w ys, re

```
. xtscc n k w ys, re

Regression with Driscoll-Kraay standard errors   Number of obs     =      1031
Method: Random-effects GLS regression            Number of groups  =       140
Group variable (i): id                           Wald chi2(3)      =   4127.15
maximum lag: 2                                   Prob > chi2       =    0.0000
corr(u_i, Xb) = 0 (assumed)                      overall R-squared =    0.8356

------------------------------------------------------------------------------
             |             Drisc/Kraay
           n | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           k |    .639224   .0151356    42.23   0.000     .6043212    .6741267
           w |  -.2900276   .1367317    -2.12   0.067    -.6053314    .0252762
          ys |   .4400795   .0917566     4.80   0.001     .2284885    .6516705
       _cons |   .2236526   .7391751     0.30   0.770    -1.480888    1.928194
-------------+----------------------------------------------------------------
     sigma_u |  .52415108
     sigma_e |  .13015331
         rho |  .94192191   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- estimates store rassal_dk
- xtscc n k w ys, fe
- estimates store sabit_dk
- hausman sabit_dk rassal_dk

```
. hausman sabit_dk rassal_dk

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    sabit_dk    rassal_dk      Difference       Std. err.
-------------+----------------------------------------------------------------
           k |    .5489458     .639224       -.0902782        .0360454
           w |   -.3106425    -.2900276       -.020615        .0372587
          ys |    .5370108     .4400795        .0969313              .
------------------------------------------------------------------------------
                          b = Consistent under H0 and Ha; obtained from xtscc.
           B = Inconsistent under Ha, efficient under H0; obtained from xtscc.

Test of H0: Difference in coefficients not systematic

    chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =  11.25
Prob > chi2 = 0.0104
(V_b-V_B is not positive definite)
```

Prob = 0.0104 < 0.05

We reject the null hypothesis
H_0: RE so FE must be used.

If we do not use Driscoll Kraay standard errors, we can also use another alternative. But firstly, I advise you to use Driscoll Kraay since this option allows us still using the RE estimation if it is necessary.

The other two option to solve the CD, AC, HC problems are

1. Panel GLS estimator
2. PCSE panel estimation

Both are using POLS if we add manually cross-sectional dummies, we can get their FE similar versions. But we cannot estimate them with RE forms.

Warning: Driscoll Kraay standard error is not only solving the CD problem but also it solves the AC and HC problems.

7th Lecture
09.06.2023

Random effects ve pooled OLS arasında seçim yapmak gerekiyorsa:

ssc install xttest0

- webuse abddata
- xtreg n w k ys, re

```
Group variable: id                              Number of groups  =        140

R-squared:                                      Obs per group:
     Within  = 0.6108                                         min =          7
     Between = 0.8479                                         avg =        7.4
     Overall = 0.8356                                         max =          9

                                                Wald chi2(3)      =    2018.16
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           n | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           w |  -.2900276   .0492318    -5.89   0.000    -.3865201    -.193535
           k |    .639224   .0176213    36.28   0.000     .6046868    .6737611
          ys |   .4400795   .0529618     8.31   0.000     .3362762    .5438828
       _cons |   .2236526   .3125287     0.72   0.474    -.3888925    .8361977
-------------+----------------------------------------------------------------
     sigma_u |  .52415108
     sigma_e |  .13015331
         rho |  .94192191   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- xttest0

```
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        n[id,t] = Xb + u[id] + e[id,t]

        Estimated results:
                         |       Var     SD = sqrt(Var)
                ---------+-----------------------------
                       n |    1.79964       1.341506
                       e |   .0169399       .1301533
                       u |   .2747344       .5241511

        Test: Var(u) = 0
                             chibar2(01) =  3044.54
                          Prob > chibar2 =   0.0000
```

H_0 : POLS
H_A : Random Effect estimator

Prob < 0.05 --> Reject null hypothesis
POLS is rejected
Random effect is better

There is also xttest1 version which handles unbalanced panels as long as there are no gaps in the series.

ssc install xttest1

- xttest1

```
Tests for the error component model:

        n[id,t] = Xb + u[id] + v[id,t]
           v[id,t] = lambda v[id,(t-1)] + e[id,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                       n |    1.79964       1.341506
                       e |   .0169399       .13015331
                       u |   .2747344       .52415108

        Tests:
           Random Effects, Two Sided:
           ALM(Var(u)=0)          = 1940.78 Pr>chi2(1) =  0.0000   < 0.05

           Random Effects, One Sided:
           ALM(Var(u)=0)          =   44.05 Pr>N(0,1)  =  0.0000   < 0.05

           Serial Correlation:
           ALM(lambda=0)          =   36.13 Pr>chi2(1) =  0.0000

           Joint Test:
           LM(Var(u)=0,lambda=0) = 3080.67 Pr>chi2(2) =  0.0000
```

H_0: POLS
H_a: Random Effect estimator

Prob < 0.05 --> Reject null hypothesis
POLS is rejected
Random effect is better

Wooldridge AC Test (It is mainly designed for POLS, do not use it after RE and FE)

findit xtserial

```
Search of official help files, FAQs, Examples, and Stata Journals

FAQ     . . . . Testing for panel-level heteroskedasticity and autocorrelation
        . . . . . . . . . . . . . . . . . . . . . . . . .  V. Wiggins and B. Poi
        6/13    How do I test for panel-level heteroskedasticity
                and autocorrelation?
                http://www.stata.com/support/faqs/statistics/panel-level-
                heteroskedasticity-and-autocorrelation/

SJ-3-2  st0039  . . Testing for serial correlation in linear panel-data models
        (help xtserial if installed)  . . . . . . . . . . . . . .  D. M. Drukker
        Q2/03   SJ 3(2):168--177
        test for serial correlation in random- or fixed-effects
        one-way models that can be applied under general conditions

Search of web resources from Stata and other users

(contacting http://www.stata.com)

5 packages found (Stata Journal listed first)
---------------------------------------------
```

```
+

package st0039 from http://www.stata-journal.com/software/sj3-2

TITLE
      SJ3-2 st0039.  Testing for serial correlation in linear ...

DESCRIPTION/AUTHOR(S)
      Testing for serial correlation in linear panel-data models
      by David M. Drukker, Stata Corporation
      Support:  ddrukker@stata.com
      After installation, type help xtserial

INSTALLATION FILES                          (click here to install)
      st0039/xtserial.ado
      st0039/xtserial.hlp

ANCILLARY FILES                             (click here to get)
      st0039/xtserial.do

(click here to return to the previous screen)
```

- webuse abdata
- xtserial n w k ys (sonuna ,output yazarsak regresyonu da veriyor)

```
. xtserial n w k ys

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,      139) =     142.715
           Prob > F =      0.0000
```

H_0: no autocorrelation

Prob < 0.05 --> Reject null hypothesis

There is autocorrelation

AC Test for Random Effect

- xtreg n w k ys, re
- xttest1

Bu sefer üstteki ilk ikiye değil alttaki ikiye bakıyoruz

```
Tests:
   Random Effects, Two Sided:
   ALM(Var(u)=0)             = 1940.78 Pr>chi2(1) =  0.0000

   Random Effects, One Sided:
   ALM(Var(u)=0)             =   44.05 Pr>N(0,1)  =  0.0000

   Serial Correlation:
   ALM(lambda=0)             =   36.13 Pr>chi2(1) =  0.0000

   Joint Test:
   LM(Var(u)=0,lambda=0)     = 3080.67 Pr>chi2(2) =  0.0000
```

H_0: no autocorrelation
Prob < 0.05 --> Reject null hypothesis

AC Test for Fixed Effect

- findit st0514

```
Search of official help files, FAQs, Examples, and Stata Journals

Search of web resources from Stata and other users

(contacting http://www.stata.com)

1 package found (Stata Journal listed first)
-----------------------------------------------

st0514 from http://www.stata-journal.com/software/sj18-1
    SJ18-1 st0514. Testing for serial correlation in... / Testing for serial
    correlation in fixed-effects / models / by Jesse Wursten, Faculty of
    Economics and / Business, KU Leuven, Leuven, Belgium / Support:
    jesse.wursten@kuleuven.be / After installation, type help {cmd:xtqptest},

(click here to return to the previous screen)

(end of search)
```

```
TITLE
      SJ18-1 st0514. Testing for serial correlation in...

DESCRIPTION/AUTHOR(S)
      Testing for serial correlation in fixed-effects
        models
      by Jesse Wursten, Faculty of Economics and
           Business, KU Leuven, Leuven, Belgium
      Support:  jesse.wursten@kuleuven.be
      After installation, type help xtqptest,
        xthrtest, and xtistest

INSTALLATION FILES                               (click here to install)
      st0514/xthrtest.ado
      st0514/xtistest.ado
      st0514/xtqptest.ado
      st0514/xthrtest.sthlp
      st0514/xtistest.sthlp
      st0514/xtqptest.sthlp

ANCILLARY FILES                                  (click here to get)
      st0514/xtqptest.do
      st0514/serialcorrelationpaper_preprint_example.do
      st0514/serialcorrelationpaper_preprint_example_sctestgraphs.do
      st0514/serialcorrelationpaper_preprint_example_timeseriesgraphs.do
      st0514/serialcorrelationpaper_preprint_montecarlo.do
```

- xthrtest:
  - heteroscedasticity robust HR-test for first order panel serial correlation introduced in Born & Breitung (Econometric Reviews, 2016). The test is suited only for fixed effects regressions without gaps

Heteroscedasticity varsa bunu yap

- xtistest:
  - Portmanteau IS-test for panel serial correlation introduced in Inoue & Solon (ET, 2006). The test is suited only for fixed effects regressions and can handle any sort of unbalancedness (e.g. gaps are allowed).

- xtqptest:
  - calculates the bias-corrected Q(P) statistic for serial correlation described in Born & Breitung (Econometric Reviews, 2016)

- xtreg n w k ys, fe
- predict artik_fe if e(sample), residual
- xtqptest artik_fe, lags(2)

```
. xtqptest artik_fe, lags(2)

Bias-corrected Born and Breitung (2016) Q(p)-test on variables artik_fe
Panelvar: id
Timevar: year
p (lags): 2
```

H0: No AC

Variable	Q(p)-stat	p-value	N	maxT	balance?
artik_fe	73.03	0.000	140	9	unbalanced

```
Notes: Under H0, Q(p) ~ chi2(p)
    H0: No serial correlation up to order p.
    Ha: Some serial correlation up to order p.
```

<0.05 → Reject H0

There is AR(2) type autocorrelation

$$u_{it} = \rho_1 u_{it-1} + \rho_2 u_{it-2} + \varepsilon_{it}$$

Heteroscedasticity robust HR-test

- xthrtest artik_fe

```
. xthrtest artik_fe

Heteroskedasticity-robust Born and Breitung (2016) HR-test on artik_fe
Panelvar: id
Timevar: year
```

Variable	HR-stat	p-value	N	maxT	balance?
artik_fe	0.95	0.344	140	9	unbalanced

```
Notes: Under H0, HR ~ N(0,1)
    H0: No first-order serial correlation.
    Ha: Some first order serial correlation.
```

>0.05 → Do not reject H0

There is no autocorrelation

It is a sign that there are harmful effects of heteroscedasticity problems (çünkü sonucu çok değiştirdi, normalde AC yokmuş, HC varmış. HC olduğu için test sonuçları zarar görmüş)

- xtistest artik_fe

```
. xtistest artik_fe, lags(2)

Inoue and Solon (2006) LM-test on variables artik_fe
Panelvar: id
Timevar: year
p (lags): 2
```

H0: No AC

Variable	IS-stat	p-value	N	maxT	balance?
artik_fe	85.36	0.000	140	9	unbalanced

```
Notes: Under H0, LM ~ chi2(p*T-p(p+1)/2)
    H0: No auto-correlation of any order.
    Ha: Auto-correlation up to order 2.
```

<0.05 → Reject H0 → AC there is

Poi-Wiggins HC Test

- webuse abdata
- xtgls n w k ys, panels(heterosk) igls
- estimates store hetero
- xtgls n w k ys, igls
- estimates store homosk
- local df = e(N_g) - 1
- lrtest hetero homosk , df(`df')

```
. lrtest hetero homosk , df(`df')

Likelihood-ratio test
Assumption: homosk nested within hetero

LR chi2(139) = 1147.20
 Prob > chi2 =  0.0000
```

H_0: no heteroscedasticity
Prob < 0.05 --> Reject null hypothesis

There is heteroscedasticity.

Leven-Brown-Forsythe

- webuse grunfeld
- xtreg mvalue invest kstock, re
- predict artik_re, e
- robvar artik_re, by(company)

```
. robvar artik_fe, by(company)

             |     Summary of e[company,t]
     Company |        Mean   Std. dev.       Freq.
-------------+------------------------------------
           1 |  -1.526e-06   598.66356          20
           2 |  -1.144e-06   309.24462          20
           3 |  -7.629e-07    392.4164          20
           4 |  -1.907e-07   123.96299          20
           5 |   2.384e-07   133.83164          20
           6 |  -1.246e-06   154.28704          20
           7 |  -4.768e-08   40.765497          20
           8 |   2.861e-07   202.20971          20
           9 |  -2.444e-07   82.558421          20
          10 |   1.341e-08   10.212858          20
-------------+------------------------------------
       Total |  -4.624e-07   261.20041         200

W0  =  21.701188   df(9, 190)     Pr > F = 0.00000000

W50 =  20.249392   df(9, 190)     Pr > F = 0.00000000

W10 =  21.667477   df(9, 190)     Pr > F = 0.00000000
```

If any of them is less than 0.05 --> reject null hypothesis

Hence, there is HC.

Groupwise HC Test for FE (xttest3)

- ssc install xttest3
- xttest3

```
. xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (10)  =     7.4e+06
Prob>chi2 =      0.0000
```

Reject null hypothesis

There is HC.

Cross Sectional Dependence Test (CD)

- ssc install xtcsd
- xtreg invest mvalue kstock, fe
- xtcsd, pesaran abs

(Run only after FE)

```
. xtcsd, pesaran abs


Pesaran's test of cross sectional independence =     7.542, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.419
```

H_0 : No Cross-Sectional Dependence

Prob <0.05

Reject null hypothesis
There is cross sectional dependence

For any variable (including residuals) to test CD,we can use xtcd

- ssc install xtcd
- xtcd artik_fe

```
Average correlation coefficients & Pesaran (2004) CD test

Variables series tested: artik_fe
                        Group variable: company
                      Number of groups: 10
             Average # of observations: 22.22
                              Panel is: unbalanced
```

Variable	CD-test	p-value	corr	abs(corr)
artik_fe	7.54	0.000	0.251	0.419

Notes: Under the null hypothesis of cross-section independence CD ~ N(0,1)

< 0.05 → Reject Ho
⇒ there is CD problem in FE estimation

Bu testi random effects için de kullanabiliriz

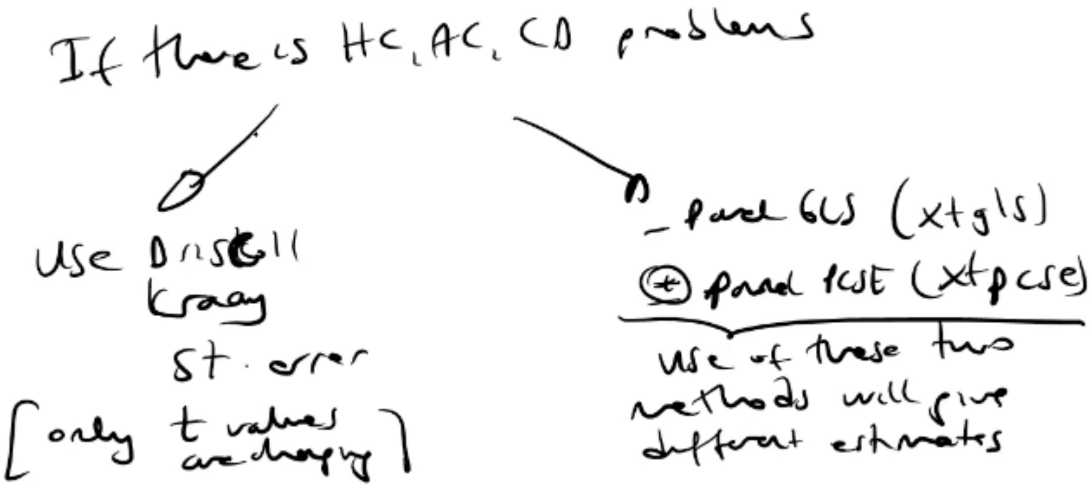
- xtreg mvalue invest kstock, re
- predict, artik_re, ue
- xtcd artik_re

```
Average correlation coefficients & Pesaran (2004) CD test

Variables series tested: artik_re
                        Group variable: company
                      Number of groups: 10
             Average # of observations: 22.22
                              Panel is: unbalanced
```

Variable	CD-test	p-value	corr	abs(corr)
artik_re	7.10	0.000	0.237	0.421

Notes: Under the null hypothesis of cross-section independence CD ~ N(0,1)

< 0.05 → Reject Ho
there is CD problem in RE estimation



Panel GLS (xtgls)

T N’den ne kadar fazla ise o kadar iyi çalışır.

- xtgls mvalue invest kstock, panels(iid) corr(psar1) → no HC, no CD, AC var

AC varsa bunu kullanıyoruz

```
. xtgls mvalue invest kstock, panels(iid) corr(psar1)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        homoskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =         1          Number of obs     =        200
Estimated autocorrelations =        10          Number of groups  =         10
Estimated coefficients     =         3          Time periods      =         20
                                                Wald chi2(2)      =     198.53
                                                Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      mvalue | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      invest |   4.851986   .3587305    13.53   0.000     4.148887    5.555084
      kstock |   -1.16384   .2743802    -4.24   0.000    -1.701616   -.6260649
       _cons |   615.8822    142.853     4.31   0.000     335.8955    895.8689
------------------------------------------------------------------------------
```

Here, xtgls and xtpcse are estimating POLS models. We can convert this equation into a FE sense by adding dummy variables for cross sections.

Cross sectional'ın boyutunu görmek için:

- xtset

cross-sectional series

```
. xtset

Panel variable: company (strongly balanced)
 Time variable: year, 1935 to 1954
         Delta: 1 year
```

- xtgls mvalue invest kstock i.company, panels(corr) corr(psar1)

Dummy variable yaptık

Bu sayede fixed effect haline getirdik

```
Estimated covariances      =         1          Number of obs     =       200
Estimated autocorrelations =        10          Number of groups  =        10
Estimated coefficients     =        12          Time periods      =        20
                                                Wald chi2(11)     =   4074.96
                                                Prob > chi2       =    0.0000

------------------------------------------------------------------------------
      mvalue | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      invest |   3.463826   .2924483    11.84   0.000     2.890638    4.037014
      kstock |  -.8476416   .1437889    -5.90   0.000    -1.129463   -.5658205
             |
     company |
          2  |  -1967.045   142.1807   -13.83   0.000    -2245.714   -1688.376
          3  |   -847.279   161.0433    -5.26   0.000    -1162.918   -531.6399
          4  |  -2280.228   119.0881   -19.15   0.000    -2513.636   -2046.819
          5  |  -2317.808   891.4373    -2.60   0.009    -4064.993   -570.6233
          6  |  -3522.934   2007.409    -1.75   0.079    -7457.383    411.5149
          7  |  -2529.783   167.8609   -15.07   0.000    -2858.784   -2200.781
          8  |  -2174.153   216.0108   -10.07   0.000    -2597.526   -1750.779
          9  |  -2315.309   267.2869    -8.66   0.000    -2839.181   -1791.436
         10  |  -2714.548   148.6267   -18.26   0.000    -3005.851   -2423.245
             |
       _cons |   2779.428    122.054    22.77   0.000     2540.206    3018.649
------------------------------------------------------------------------------
```

Panel GLS estimation corrected for AR(1) type AC problem in a FE model framework

We cannot manually run RE within panel GLS command.

```
xtgls mvalue invest kstock,
panels(het) corr(psar1)
```

there HC, AC but no CD

```
. xtgls mvalue invest kstock, panels(het) corr(psar1)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =        10          Number of obs     =        200
Estimated autocorrelations =        10          Number of groups  =         10
Estimated coefficients     =         3          Time periods      =         20
                                                Wald chi2(2)      =      62.54
                                                Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      mvalue | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      invest |   3.160785   .4021134     7.86   0.000     2.372657    3.948913
      kstock |  -.1690298   .2056214    -0.82   0.411    -.5720403    .2339807
       _cons |   344.9042   74.51265     4.63   0.000     198.8621    490.9463
------------------------------------------------------------------------------
```

POLS corrected for HC and AC problems.

To run a FE type regression with these corrections.

- xtgls mvalue invest kstock i.company, panels(het) corr(psar1)

```
Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =        10          Number of obs     =        200
Estimated autocorrelations =        10          Number of groups  =         10
Estimated coefficients     =        12          Time periods      =         20
                                                Wald chi2(11)     =    2443.48
                                                Prob > chi2       =     0.0000
```

mvalue	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
invest	2.109918	.3153107	6.69	0.000	1.491921	2.727916
kstock	-.1846967	.1194599	-1.55	0.122	-.4188337	.0494404
company						
2	-2020.926	194.6343	-10.38	0.000	-2402.402	-1639.45
3	-1371.225	233.7375	-5.87	0.000	-1829.342	-913.1074
4	-2636.634	181.5907	-14.52	0.000	-2992.545	-2280.723
5	-2954.93	306.8592	-9.63	0.000	-3556.363	-2353.497
6	-3915.563	547.4167	-7.15	0.000	-4988.48	-2842.646
7	-3065.465	196.5442	-15.60	0.000	-3450.685	-2680.246
8	-2569.88	218.1457	-11.78	0.000	-2997.438	-2142.323
9	-2854.105	207.0592	-13.78	0.000	-3259.933	-2448.276
10	-3107.432	195.8858	-15.86	0.000	-3491.361	-2723.502
_cons	3172.465	196.4227	16.15	0.000	2787.483	3557.446

Hepsinin correct edilmiş hali için:

xtgls mvalue invest kstock, panels(corr) corr(psar1)

```
. xtgls mvalue invest kstock, panels(corr) corr(psar1)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   panel-specific AR(1)

Estimated covariances      =        55          Number of obs     =       200
Estimated autocorrelations =        10          Number of groups  =        10
Estimated coefficients     =         3          Time periods      =        20
                                                Wald chi2(2)      =    345.75
                                                Prob > chi2       =    0.0000

------------------------------------------------------------------------------
      mvalue | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      invest |   4.397232   .2372556    18.53   0.000      3.93222    4.862245
      kstock |  -.6319713   .1241686    -5.09   0.000    -.8753373   -.3886053
       _cons |   430.5154   41.06988    10.48   0.000     350.0199    511.0109
------------------------------------------------------------------------------
```

Fixed effect yapmak istiyorsak:

xtgls mvalue invest kstock i.company, panels(corr) corr(psar1)

xtpcse mvalue invest kstock, correlation(psar1) nmk

```
. xtpcse mvalue invest kstock, correlation(psar1) nmk
note: estimates of rho outside [-1,1] bounded to be in the range [-1,1].

Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable:   company                       Number of obs     =       200
Time variable:    year                          Number of groups  =        10
Panels:           correlated (balanced)         Obs per group:
Autocorrelation:  panel-specific AR(1)                        min =        20
                                                              avg =        20
                                                              max =        20
Estimated covariances      =        55          R-squared         =    0.5232
Estimated autocorrelations =        10          Wald chi2(2)      =     51.21
Estimated coefficients     =         3          Prob > chi2       =    0.0000

------------------------------------------------------------------------------
             |           Panel-corrected
      mvalue | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      invest |   4.843699   .6827452     7.09   0.000     3.505543    6.181855
      kstock |   -1.15131   .5700103    -2.02   0.043    -2.268509   -.0341099
       _cons |   616.9006   172.3364     3.58   0.000     279.1275    954.6737
-------------+----------------------------------------------------------------
        rhos =   .8563595   .8597952        1  .7225431  .9076928 ...  .9310709
------------------------------------------------------------------------------
```

Fixed effect yapmak için:
xtpcse mvalue invest kstock i.company, correlation(psar1) nmk

```
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable:   company                     Number of obs     =        200
Time variable:    year                        Number of groups  =         10
Panels:           correlated (balanced)       Obs per group:
Autocorrelation:  panel-specific AR(1)                      min =         20
                                                            avg =         20
                                                            max =         20
Estimated covariances      =        55        R-squared         =     0.9696
Estimated autocorrelations =        10        Wald chi2(10)     =    1955.24
Estimated coefficients     =        11        Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |           Panel-corrected
      mvalue | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      invest |    3.46527    .604193     5.74   0.000     2.281073    4.649467
      kstock |  -.8474568   .3258642    -2.60   0.009    -1.486139   -.2087747
             |
     company |
          2  |  -1966.682   185.0997   -10.62   0.000    -2329.471   -1603.893
          3  |  -846.5072   273.4686    -3.10   0.002    -1382.496   -310.5185
          4  |  -2279.378   236.2689    -9.65   0.000    -2742.456   -1816.299
          5  |  -2316.991    402.606    -5.75   0.000    -3106.085   -1527.898
          6  |          0  (omitted)
          7  |  -2528.915   288.2782    -8.77   0.000     -3093.93     -1963.9
          8  |  -2173.235   261.0553    -8.32   0.000    -2684.894   -1661.576
          9  |  -2314.428   292.2358    -7.92   0.000    -2887.199   -1741.656
         10  |  -2713.557   269.3682   -10.07   0.000    -3241.509   -2185.606
             |
       _cons |   2778.432   271.1482    10.25   0.000     2246.991    3309.872
-------------+----------------------------------------------------------------
        rhos = -.0334459   .6263819   .4755092   -.137362   .9720083 ...        1
------------------------------------------------------------------------------
```

HC ve CD var kabul ediyor.