ECON-C4200 - Econometrics II

Lecture 11: Time series IV

Otto Toivanen

Learning outcomes

- At the end of lecture 11, you know

1. what **cointegration** is
2. what an **error-correction** model is
3. how to estimate a cointegrated model and an error-correction model
4. what an (**G**eneralized) **A**uto**R**egressive **C**onditional **H**eteroskedastic ((G)ARCH) model is
5. and how to estimate one

Cointegration and (G)ARCH model cause for Nobel prize 2003

- Robert F. Engle shared the Nobel prize (2003) "for methods of analyzing economic time series with time-varying volatility" (ARCH)

with

- Clive W. J. Granger who received the prize "for methods of analyzing economic time series with common trends (cointegration)".

What is cointegration?

- We know the danger of spurious regression when regressing two non-stationary variables (Y_t, X_t).
- However, if (Y_t, X_t) are **cointegrated**, a spurious regression does not arise.
- **Order of integration**: time-series Y_t is **integrated of order** d (denoted $Y_t \sim I(d)$), if differencing Y_t d times yields a stationary process.
- Example: Differencing the EU unemployment series once yielded a stationary process. EU UE is therefore integrated of order 1, i.e., $I(1)$.

What is cointegration?

- Assume that (Y_t, X_t) are integrated of **order** d $Y_t, X_t \sim I(d)$.
- If there exists a β such that

$$Y_t - \beta X_t = u_t$$

  and such that u_t is integrated of order less than d (say, $d - b$), then Y_t and X_t are **cointegrated of order** d, b: d $Y_t, X_t \sim CI(d, b)$.
- Example: Assume $Y_t, X_t \sim I(1)$. Then taking first differences yields a stationary process for both.
- If there exists a β such that $Y_t - \beta X_t = u_t$ and $u_t \sim I(0)$ (=u_t is stationary), then Y_t, X_t are cointegrated of order $I(1, 1)$.

Cointegration

- If (Y_t, X_t) are integrated of **order** 1 $(Y_t, X_t \sim I(1))$ **and** are cointegrated, you do not have to difference the data, but can run the following OLS regression:

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

  β_1 is (**super**)consistent (=convergence at rate T), but the standard errors will be inconsistent.

- Note: If e.g. X_t is stationary $(X_t \sim I(0))$ and Y_t has a unit root $(Y_t \sim I(1))$, there cannot be cointegration.

Estimation in difference versus levels - what do we learn?

- In a levels regression, $\beta_1 = \partial Y_t / \partial X_t$.
- In a difference regression, $\beta_1 = \partial \Delta Y_t / \partial \Delta X_t$.
- One can think of the former as the long run, the latter as the short run effect.

Error Correction Model

- An **Error Correction Model** (ECM) allows us to estimate both the short and the long term effects in one go.
- You can estimate an ECM if the series are cointegrated.
- **Granger representation theorem**: For any set of $I(1)$ variables, error correction and cointegration are equivalent.

Deriving an ECM

- Let's assume the following model, with $Y_t, X_t \sim I(1)$:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 X_t + \beta_2 X_{t-1} + u_t$$

- Substract Y_{t-1} from both sides to get

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1) Y_{t-1} + \beta_1 X_t + \beta_2 X_{t-1} + u_t$$

- now add $\beta_1 X_{t-1} - \beta_1 X_{t-1}$ to get

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1) Y_{t-1} + \beta_1 \Delta X_t + (\beta_2 + \beta_1) X_{t-1} + u_t$$

$$\Delta Y_t = \alpha_0 + \rho Y_{t-1} + \beta_1 \Delta X_t + \theta_1 X_{t-1} + u_t$$

- Notice: ΔY_t and ΔX_t are stationary. If Y_t, X_t are cointegrated, then u_t is stationary, too.

The Engle-Granger test for cointegration

1. Pre-test your variables (Y_t, X_t) for the presence of unit roots using ADF and check whether they are integrated of the same order (e.g., both are stationary after first differencing, i.e., $I(1)$).
2. (If the series are integrated of the same order), regress the long-run equilibrium model (i.e., regress Y_t on X_t).
3. obtain the residuals $\hat{u}_t$ from the regression.
4. plot the residuals against time.
5. plot $\hat{u}_t$ against $\hat{u}_{t-1}$.
6. Run the ADF test on the residuals to check for a unit root. Note: you are using generated regressors instead of "original" data. Therefore need to use different critical values.

The (Adjusted) Engle-Granger test for cointegration

- You could save the residuals from the first stage regression of Y_t on X_t and add the lagged residual to a first-differenced equation. The coefficient on $\hat{u}_{t-1}$ would be an estimate of the adjustment speed.
- Stata provides a command (`egranger` ; type `ssc install egranger`) to run the Engle-Granger test and to estimate ECM.
- The adjusted version of the test allows for serial correlation in the error term.
- The Engle-Granger test is suited for a situation where you have two time series.
- If you have more than two variables (think VAR), then the **Johansen** test needs to be used. We do not cover it.

Variable volatility - (G)ARCH

- Some time series – stock prices (indices) – being a prime example, exhibit time-varying volatility.
- Volatility is linked to risk, and therefore of direct interest to investors.
- Some financial instruments' value based on volatility –e.g. options.

Variable volatility - (G)ARCH

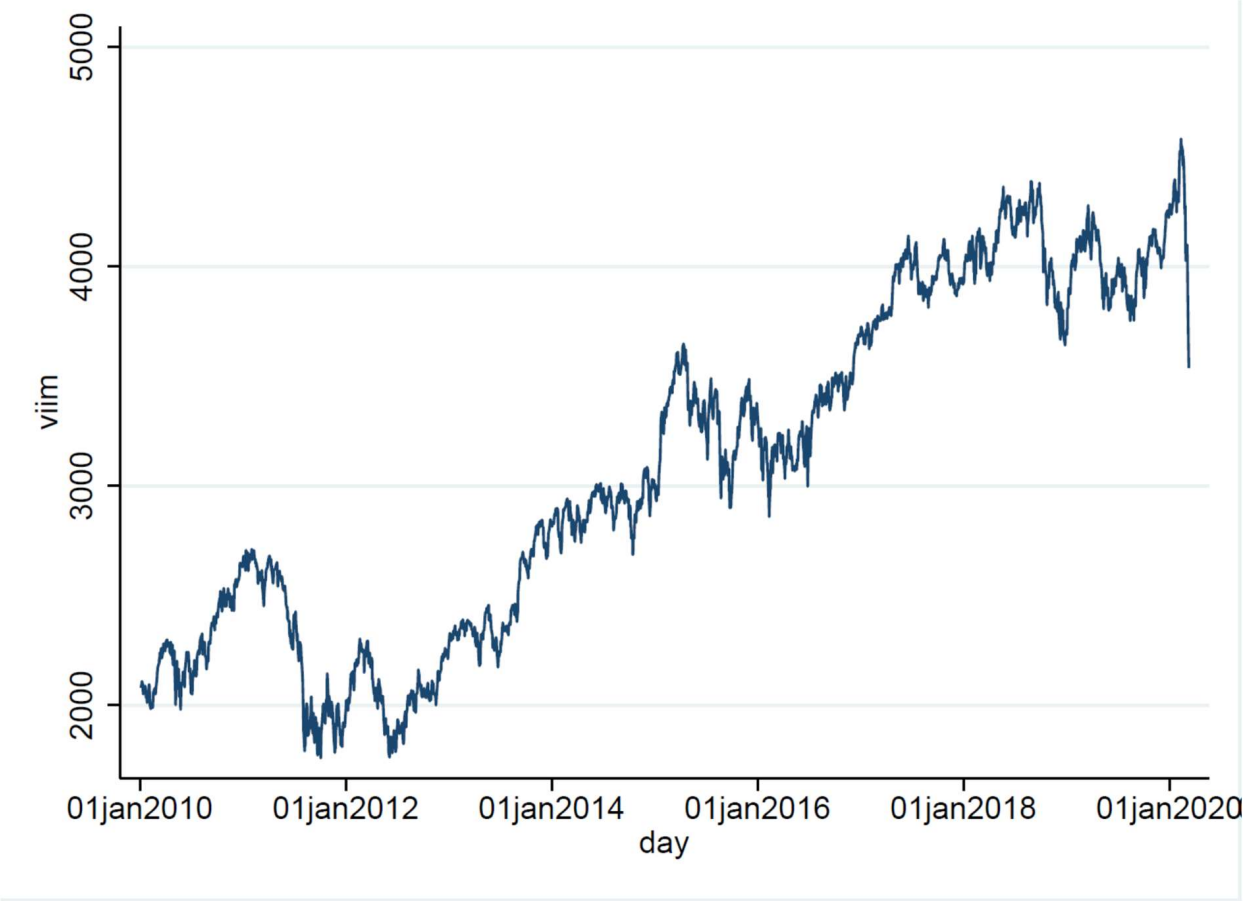
- Generalized Autoregressive Conditional Heteroscedasticity models.
- Goal is to model volatility.
- Understanding volatility is one of the objects of the modeling exercise.
- Modeling volatility may yield better forecast intervals.
- Let's study the OMX25 index.

Plotting OMX25

Stata code

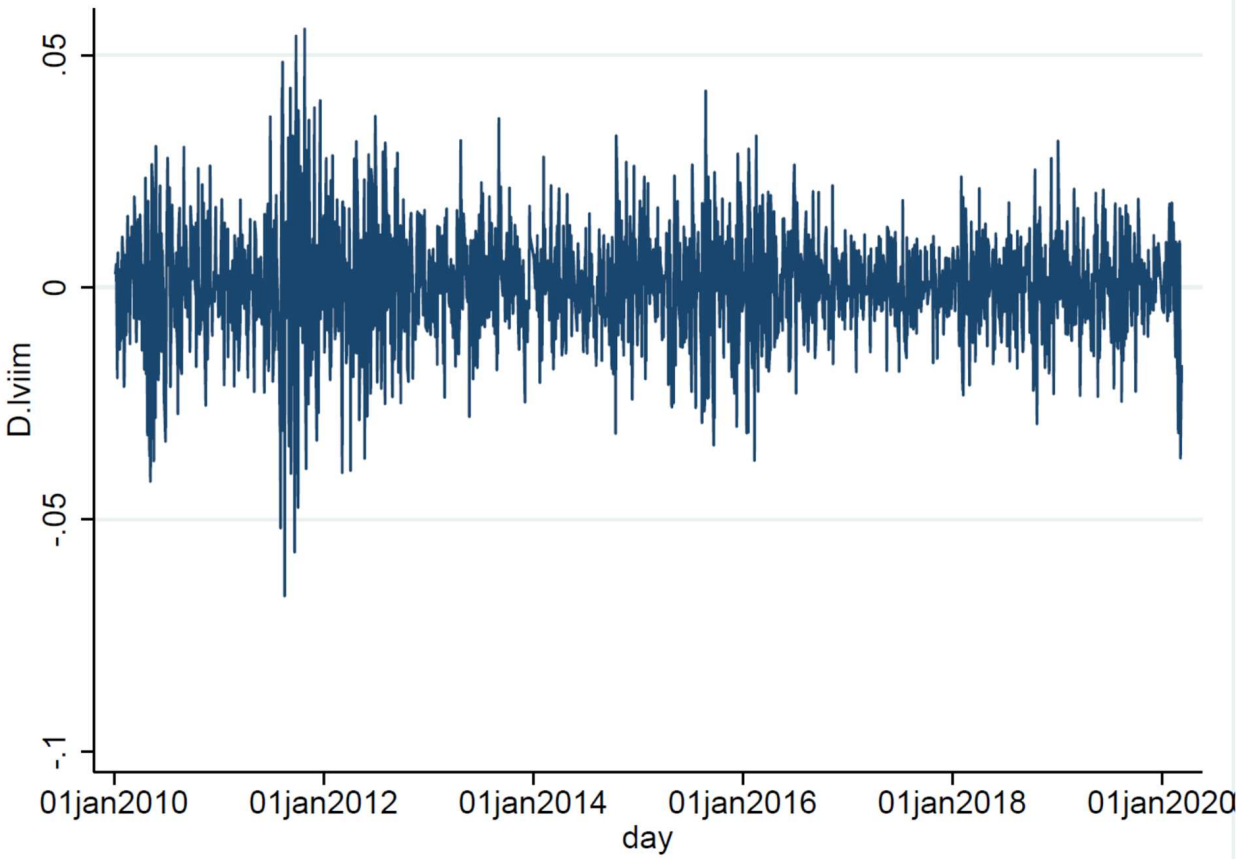
```
use  "OMX_25HKI.dta", clear
gen day        = date(date, "DMY")
format day %td
sort day
gen time       = _n
tsset day
tsline viim , ///
    graphregion(fcolor(white))
graph export  "OMX25_graph.pdf", replace
gen lviim      = ln(viim)
tsline d.lviim , ///
    graphregion(fcolor(white))

graph export  "dlnOMX25_graph.pdf", replace
```

OMX25 Helsinki 4/1/2010 – 10/3/2020



1

$d \ln OMX25$ Helsinki 4/1/2010 – 10/3/2020



2

Volatility - ARCH

- Think of volatility as clustering of variance over time: High- and low-variance periods follow each other.
- To illustrate, let's study an ADL(1,1) model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t$$

- But now let's model the variance of u_t as follows:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 ... + \alpha_p u_{t-p}^2$$

- Volatility now is a weighted average of the squared past residuals, with the weights (the αs) estimated from the data.

Difference between ARCH and GARCH

- GARCH is also based on the idea of modeling variance of the error term using a weighted average of past residuals.
- Instead of making σ_t^2 a function of only past squared residuals, it also makes it a function of lags of the variance itself.
- A GARCH(p,q) model of variance would be

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 ... + \alpha_p u_{t-p}^2$$
$$+ \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 ... + \beta_q \sigma_{t-q}^2$$

Volatility of OMX25

- Let's
  1. first study an ADL model
  2. (We have already plotted the levels and differenced data.)
  3. run first DF tests, then test for the appropriate lag length using AIC / BIC
  4. then compare the ADL and ARCH model results

ADL of OMX25

Stata code

```
tsset time
dfuller lviim
dfuller d.lviim
regress d.lviim dl(1/1).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob1
regress d.lviim dl(1/2).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob2
regress d.lviim dl(1/3).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob3
regress d.lviim dl(1/4).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob4
regress d.lviim dl(1/5).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob5
regress d.lviim dl(1/6).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob6
regress d.lviim dl(1/7).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob7
regress d.lviim dl(1/8).lviim if time < 2511 & time > 10, robust
estimates store ar1_rob8
estimates table ar1_rob*, b(%7.4f) star(0.1 0.05 0.001) stat(N r2 r2_a aic bic)
```

Dickey-Fuller tests

```
. dfuller lviim

Dickey-Fuller test for unit root                   Number of obs   =      2557

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.521            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5230

. dfuller d.lviim

Dickey-Fuller test for unit root                   Number of obs   =      2556

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -48.250            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

3

Determining lag length

```
estimates table ar1_rob*, b(%7.4f) star(0.1 0.05 0.001) stat(N r2 r2_a aic bic)
```

Variable	ar1_rob1	ar1_rob2	ar1_rob3	ar1_rob4	ar1_rob5	ar1_rob6	ar1_rob7	ar1_rob8
lviim								
LD.	0.0332	0.0334	0.0332	0.0310	0.0262	0.0259	0.0259	0.0258
L2D.		-0.0087	-0.0077	-0.0082	-0.0102	-0.0105	-0.0112	-0.0112
L3D.			-0.0315	-0.0292	-0.0298	-0.0300	-0.0306	-0.0313
L4D.				-0.0682**	-0.0661**	-0.0661**	-0.0664**	-0.0671**
L5D.					-0.0709**	-0.0708**	-0.0709**	-0.0712**
L6D.						-0.0041	-0.0039	-0.0040
L7D.							-0.0091	-0.0088
L8D.								-0.0103
_cons	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
N	2500	2500	2500	2500	2500	2500	2500	2500
r2	0.0011	0.0012	0.0022	0.0068	0.0118	0.0118	0.0119	0.0120
r2_a	0.0007	0.0004	0.0010	0.0052	0.0098	0.0094	0.0091	0.0088
aic	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04
bic	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04	-1.5e+04

legend: * p<.1; ** p<.05; *** p<.001

4

Determining lag length

```
. estimates stats ar1_rob* ar1_hac

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
ar1_rob1	2,500	7430.089	7431.464	2	-14858.93	-14847.28
ar1_rob2	2,500	7430.089	7431.56	3	-14857.12	-14839.65
ar1_rob3	2,500	7430.089	7432.799	4	-14857.6	-14834.3
ar1_rob4	2,500	7430.089	7438.636	5	-14867.27	-14838.15
ar1_rob5	2,500	7430.089	7444.939	6	-14877.88	-14842.93
ar1_rob6	2,500	7430.089	7444.96	7	-14875.92	-14835.15
ar1_rob7	2,500	7430.089	7445.062	8	-14874.12	-14827.53
ar1_rob8	2,500	7430.089	7445.194	9	-14872.39	-14819.97
ar1_hac	2,500	.	.	6	.	.

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

ADL and ARCH of OMX25

Stata code

```
newey d.lviim dl(1/5).lviim   if time < 2511 & time > 10, lag(2)
estimates store newey_res
arch d.lviim dl(1/5).lviim if time < 2511 & time > 10, arch(1/5)
estimates store arch
estimates table newey_res arch, b(%7.4f) star(0.1 0.05 0.001) stat(N r2 r2_a aic bic)
```

ADL

```
. newey d.lviim dl(1/5).lviim     if time < 2511 & time > 10, lag(2)

Regression with Newey-West standard errors      Number of obs     =      2,500
maximum lag: 2                                  F(  5,      2494) =       3.30
                                                Prob > F          =     0.0056
```

D.lviim	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
lviim						
LD.	.0261769	.028069	0.93	0.351	-.0288641	.0812178
L2D.	-.0102372	.030303	-0.34	0.736	-.0696589	.0491845
L3D.	-.0298407	.0294114	-1.01	0.310	-.0875139	.0278325
L4D.	-.0660948	.026047	-2.54	0.011	-.1171707	-.0150188
L5D.	-.0708739	.0291785	-2.43	0.015	-.1280905	-.0136572
_cons	.0003258	.0002517	1.29	0.196	-.0001677	.0008193

3

ARCH

```
. arch d.lviim dl(1/5).lviim if time < 2511 & time > 10, arch(1/5)

(setting optimization to BHHH)
Iteration 0:   log likelihood =  7672.4235
Iteration 1:   log likelihood =  7682.8446
Iteration 2:   log likelihood =  7684.0823
Iteration 3:   log likelihood =  7684.4587
Iteration 4:   log likelihood =  7684.6886
(switching optimization to BFGS)
Iteration 5:   log likelihood =  7684.8563
Iteration 6:   log likelihood =  7685.0802
Iteration 7:   log likelihood =  7685.0915
Iteration 8:   log likelihood =  7685.0915

ARCH family regression
```

4

ARCH

```
Sample: 11 - 2510                                   Number of obs   =      2,500
Distribution: Gaussian                              Wald chi2(5)    =      26.45
Log likelihood =  7685.091                          Prob > chi2     =     0.0001
```

D.lviim	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
lviim						
lviim						
LD.	.0402096	.0210391	1.91	0.056	-.0010262	.0814454
L2D.	-.0229097	.0207769	-1.10	0.270	-.0636317	.0178123
L3D.	-.0017616	.0211728	-0.08	0.934	-.0432595	.0397363
L4D.	-.0482207	.0206618	-2.33	0.020	-.0887171	-.0077242
L5D.	-.0746251	.0198848	-3.75	0.000	-.1135986	-.0356516
_cons	.0006657	.0001966	3.39	0.001	.0002805	.001051
ARCH						
arch						
L1.	.1121811	.0230449	4.87	0.000	.067014	.1573483
L2.	.1201808	.0245835	4.89	0.000	.071998	.1683637
L3.	.1513574	.0241028	6.28	0.000	.1041168	.198598
L4.	.120337	.0116079	10.37	0.000	.0975859	.1430881
L5.	.1600169	.0257325	6.22	0.000	.1095821	.2104517
_cons	.0000524	3.12e-06	16.79	0.000	.0000463	.0000585

5

Comparison

```
. estimates table newey_res arch, b(%7.4f) star(0.1 0.05 0.001) stat(N r2 r2_a aic bic)
```

Variable	newey_res	arch
_		
lviim		
LD.	0.0262	
L2D.	-0.0102	
L3D.	-0.0298	
L4D.	-0.0661**	
L5D.	-0.0709**	
_cons	0.0003	
lviim		
lviim		
LD.		0.0402*
L2D.		-0.0229
L3D.		-0.0018
L4D.		-0.0482**
L5D.		-0.0746***
_cons		0.0007***
ARCH		
arch		
L1.		0.1122***
L2.		0.1202***
L3.		0.1514***
L4.		0.1203***
L5.		0.1600***
_cons		0.0001***
Statistics		
N	2500	2500
r2		
r2_a		
aic	.	-1.5e+04
bic	.	-1.5e+04

legend: * p<.1; ** p<.05; *** p<.001

6