ECON-C4200 - Econometrics II: Capstone

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.

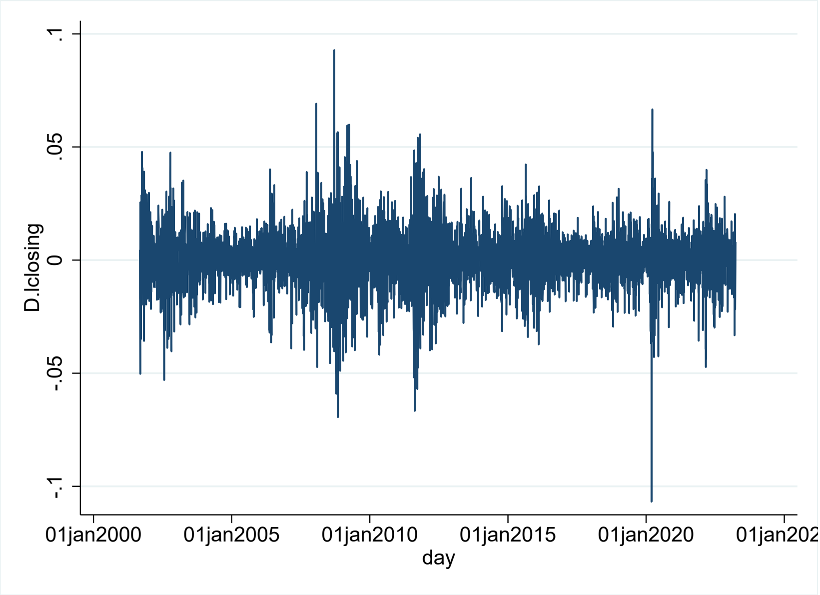
Plotting OMX25

Stata code

```
use "OMX_2023.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 3/9/2001 – 29/3/2023

$d \ln OMX25$ Helsinki 3/9/2001 – 29/3/2023



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
global end_point 5300
dfuller lclosing
dfuller d.lclosing
forvalues i = 1/8 {
  regress d.lclosing l(1/`i')d.lclosing if time < $end_point & time > 10, robust
  estimates store ar1_rob`i'
}
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 lclosing

Dickey-Fuller test for unit root        Number of obs  = 5,415
Variable: lclosing                      Number of lags =     0

H0: Random walk without drift, d = 0

                                       Dickey-Fuller
                   Test        -------- critical value ---------
              statistic           1%           5%          10%
-----------------------------------------------------------------
 Z(t)            -1.384       -3.430       -2.860       -2.570
-----------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5900.

. dfuller d.lclosing

Dickey-Fuller test for unit root        Number of obs  = 5,414
Variable: D.lclosing                    Number of lags =     0

H0: Random walk without drift, d = 0

                                       Dickey-Fuller
                   Test        -------- critical value ---------
              statistic           1%           5%          10%
-----------------------------------------------------------------
 Z(t)           -71.434       -3.430       -2.860       -2.570
-----------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000.
```

Effects of lag length

```
. estimates table ar1_rob* ar1_hac, 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	ar1_hac
lclosing									
LD.	0.0303	0.0308	0.0305	0.0306	0.0308	0.0297	0.0303	0.0299	0.0308*
L2D.		-0.0157	-0.0153	-0.0152	-0.0157	-0.0155	-0.0146	-0.0142	-0.0157
L3D.			-0.0137	-0.0138	-0.0145	-0.0148	-0.0149	-0.0142	-0.0145
L4D.				0.0042	0.0055	0.0051	0.0053	0.0052	0.0055
L5D.					-0.0409*	-0.0400*	-0.0397*	-0.0395*	-0.0409*
L6D.						-0.0271	-0.0277	-0.0275	
L7D.							0.0217	0.0211	
L8D.								0.0176	
_cons	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
N	5289	5289	5289	5289	5289	5289	5289	5289	5289
r2	0.0009	0.0012	0.0014	0.0014	0.0030	0.0038	0.0042	0.0046	
r2_a	0.0007	0.0008	0.0008	0.0006	0.0021	0.0026	0.0029	0.0030	
aic	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	
bic	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	-3.0e+04	

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

Determining lag length

```
. estimates stats ar1_rob*
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
ar1_rob1	5,289	15198.67	15201.1	2	-30398.21	-30385.06
ar1_rob2	5,289	15198.67	15201.76	3	-30397.51	-30377.79
ar1_rob3	5,289	15198.67	15202.25	4	-30396.5	-30370.2
ar1_rob4	5,289	15198.67	15202.3	5	-30394.59	-30361.73
ar1_rob5	5,289	15198.67	15206.72	6	-30401.44	-30362
ar1_rob6	5,289	15198.67	15208.67	7	-30403.33	-30357.32
ar1_rob7	5,289	15198.67	15209.91	8	-30403.82	-30351.23
ar1_rob8	5,289	15198.67	15210.73	9	-30403.46	-30344.3

Note: BIC uses N = number of observations. See **[R] BIC note**.

ADL OMX25

Stata code

```
newey d.lclosing dl(1/5).lclosing   if time < $end_point & time > 10,  lag(2)
estimates store ar1_hac
estimates table ar1_rob* ar1_hac, b(%7.4f) star(0.1 0.05 0.001) stat(N r2 r2_a aic bic)
estimates stats ar1_rob* ar1_hac
```

ADL

```
. newey d.lclosing dl(1/5).lclosing       if time < $end_point & time > 10,  lag(2)

Regression with Newey-West standard errors          Number of obs     =      5,289
Maximum lag = 2                                     F(  5,      5283) =       1.59
                                                    Prob > F          =     0.1589
```

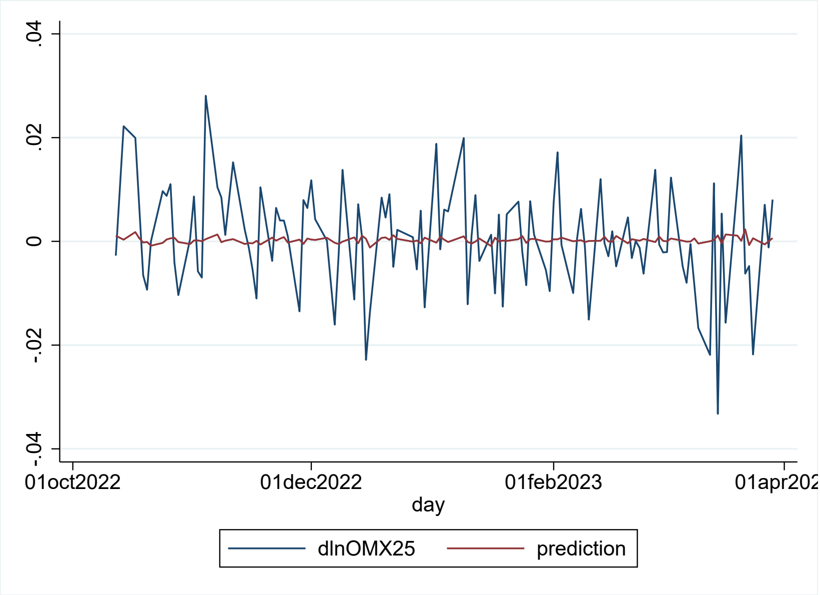
D.lclosing	Coefficient	Newey-West std. err.	t	P>|t|	[95% conf.	interval]
lclosing						
LD.	.0307946	.0180656	1.70	0.088	-.0046214	.0662106
L2D.	-.0156739	.0222014	-0.71	0.480	-.0591978	.0278501
L3D.	-.0144625	.0217677	-0.66	0.506	-.0571362	.0282111
L4D.	.0054905	.0199633	0.28	0.783	-.0336458	.0446269
L5D.	-.0408527	.0222346	-1.84	0.066	-.0844418	.0027364
_cons	.0002436	.0001915	1.27	0.203	-.0001319	.0006191

Prediction based on ADL

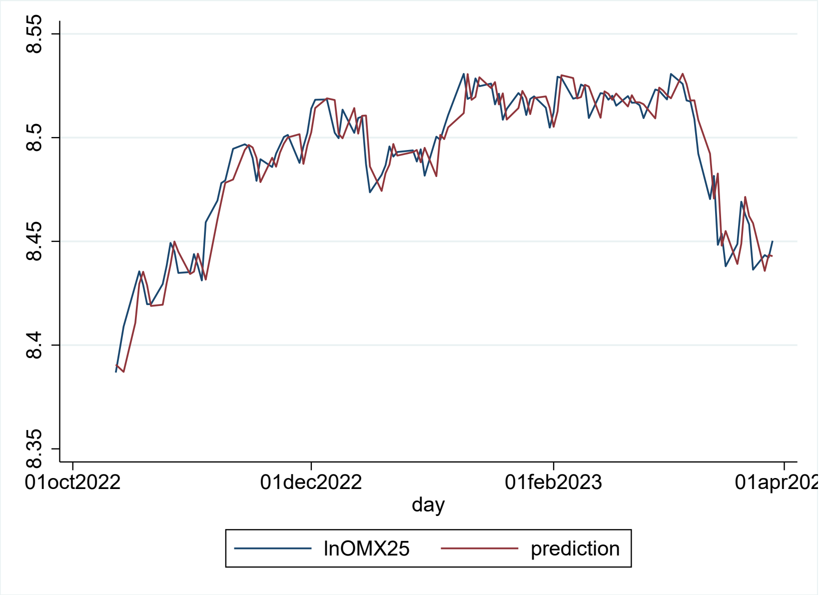
Stata code

```
gen lclosing_pred   = lclosing
replace lclosing_pred = . if time != $end_point - 1
replace dlclosing_ownhat = _b[_cons] + _b[LD.lclosing] * LD.lclosing
+ _b[L2D.lclosing] * L2D.lclosing + _b[L3D.lclosing] * L3D.lclosing
+ _b[L4D.lclosing] * L4D.lclosing + _b[L5D.lclosing] * L5D.lclosing
if time == $end_point -1

scalar bigT       = r(max)
local bigT        = bigT
forvalues i = $end_point/`bigT' {
  replace dlclosing_ownhat = _b[_cons] + _b[LD.lclosing] * LD.lclosing
    + _b[L2D.lclosing] * L2D.lclosing + _b[L3D.lclosing] * L3D.lclosing
    + _b[L4D.lclosing] * L4D.lclosing + _b[L5D.lclosing] * L5D.lclosing
    if time == `i'

    replace lclosing_pred = lclosing[_n-1] + dlclosing_ownhat
  }
```

Prediction - differences



Prediction - (log) levels



ARCH

```
. arch d.lclosing dl(1/5).lclosing if time <  $end_point & time > 10, arch(1/5)

(setting optimization to BHHH)
Iteration 0:   log likelihood =  15859.791
Iteration 1:   log likelihood =  15900.585
Iteration 2:   log likelihood =  15908.662
Iteration 3:   log likelihood =  15909.977
Iteration 4:   log likelihood =  15910.192
(switching optimization to BFGS)
Iteration 5:   log likelihood =  15910.238
Iteration 6:   log likelihood =  15910.255
Iteration 7:   log likelihood =  15910.256
Iteration 8:   log likelihood =  15910.256
```

ADL and ARCH

```
. 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
_		
lclosing		
LD.	0.0308*	
L2D.	-0.0157	
L3D.	-0.0145	
L4D.	0.0055	
L5D.	-0.0409*	
_cons	0.0002	
lclosing		
lclosing		
LD.		0.0339**
L2D.		-0.0228
L3D.		-0.0112
L4D.		-0.0208
L5D.		-0.0616***
_cons		0.0008***
ARCH		
arch		
L1.		0.1147***
L2.		0.1520***
L3.		0.1668***
L4.		0.1398***
L5.		0.1699***
_cons		0.0001***
Statistics		
N	5289	5289
r2		
r2_a		
aic	.	-3.2e+04
bic	.	-3.2e+04