9: Vector Auto-Regression (VAR) and Auto-Regressive Distributed Lag (ARDL) Analysis

GECO 6281 Advanced Econometrics 1 (Lab)

Patrick Mokre

Fall 2020

Recapitulation

- Dealing with Endogeneity: Instrumental Variables (IV)
- Static Panel Data
- Instrumental Variables in Dynamic Panel Data: Anderson-Hsiao, Arellano-Bond, Arellano-Bover
- Stationarity
- Panel Stationarity

Vector Autoregression

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{1}$$

$$\begin{aligned} y_t &= b_{11} y_{t-1} + b_{12} z_{t-1} + \epsilon_{1,t} \\ z_t &= b_{21} y_{t-1} + b_{22} z_{t-1} + \epsilon_{2,t} \end{aligned}$$

- Captures mutual dependencies between time series
- Intuitive Forecasting on all variables (ARDL/DL only allows forecasts for $y_{i,t}$)
- Takes account of the usual endogeneity between economic processes
- No simultaneity bias because the “other” variable appears in **lags** only.

VAR: Assumptions

- y_t and z_t are dynamically related, but not contemporaneously related.
- Error terms $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are contemporaneously uncorrelated
- y_t and z_t are stationary

$\Rightarrow$ Estimation by OLS is efficient, consistent, and SUR estimation does not improve efficiency. Standard Errors and Covariances can be estimated in standard form. AIC/BIC are appropriate for lag selection.

VAR with exogenous variables

Adopt **Lütkepohl** notation, because it underlies the corresponding STATA package (makes it easier to understand help files etc.)

$$y_t = AY_{t-1} + B_0 x_t + u_t \tag{2}$$

where y_t the vector of endogenous variables, A the matrix of AR coefficients, B_0 the matrix of exogenous coefficients, x_t the vector of exogenous covariates and u_t a vector of white noise disturbances ("innovations").

Intercepts are included in x_t.

VAR: Lütkepohl

Figure 1: Helmut Lütkepohl: *"I see that you have **not** adopted full matrix notation yet."*

VAR: Advanced Lütkepohl Notation

$$Y = BZ + U \tag{3}$$

$$
\begin{aligned}
Y &= (y_1, ..., y_t) & \text{Y is } K \times T \\
B &= (A, B_0) & \text{B is } K \times (Kp + M) \\
Z &= \begin{bmatrix} Y_0 & ... & Y_{T-1} \\ x_1 & ... & x_T \end{bmatrix} & \text{Z is } (Kp + M) \times T \\
U &= (u_1, ..., u_T) & \text{U is } K \times T
\end{aligned}
$$

In STATA, this model is estimated using iterative seemingly unrelated regressions (**SUR**).

Lag Selection for VAR

As for single equation auto-regressions, a variety of **goodness of fit** criteria can be calculated for VAR.

- Akaike's Final Prediction Error (FPE): the **determinant** of the average squared prediction error matrix, normalized by $((1+m/N)/(1-m/N))^K$ (m is the average number of coefficients between all models, K is the number of equations).
- Akaike Information Criterion (AIC): $-2\frac{LL}{T}+\frac{2t_P}{T}$ where LL is the log-likelihood of the model, and t_P is the total number of parameters in the evaluated model.
- Schwartz' Bayesian Information Criterion (BIC): $-2\frac{LL}{T}+\frac{ln(T)}{T}t_P$
- Hannan-Quin Information Criterion (HQIC): $-2\frac{LL}{T}+\frac{2ln[ln(T)]}{T}t_P$

The corresponding STATA command is `varsoc`, it can be used pre- and postestimation.

Multivariate Impulse Response Function

Impulse Response Functions (IRF) estimate how a time series reacts to a disturbance in the error terms.

Suppose that the estimated error structure ϵ is related to an underlying structural shock vector u_t.

$$\begin{aligned} \epsilon_t &= Au_t \\ E(u_t, u_t') &= I \end{aligned} \tag{4}$$

A is related to the error covariance matrix Σ:

$$\begin{aligned} \Sigma &= E[\epsilon_t \epsilon_t'] \\ &= E(Au_t u_t' A') \\ &= AE[u_t u_t']A' \\ &= AA' \end{aligned}$$

Because $\hat{\Sigma}$ can be estimated in the regression (e.g. via VAR), $\hat{A}$ can be retrieved.

In STATA, you can create post-estimation IRFs using `irf create` and `irf graph`.

Granger Causality

Granger Causality is not causality. Rather it measures which event happens first: z_t is said to "Granger cause" y_t if $(z_{t-1}, ..., z_{t-p})$ contains information that helps predict y_t better than only $(y_{t-1}, ..., y_{t-p})$ does.

A simple way of testing Granger Causality is to compare the tests for joint insignificance with and without $(z_{t-1}, ..., z_{t-p})$ in predicting y_t. The corresponding STATA command is `vargranger`.

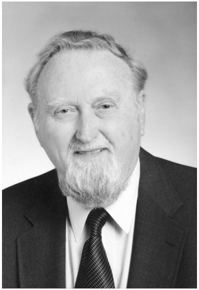


Figure 2: Clive Granger: *However, several writers stated that "of course, this is not real causality, it is only Granger causality."*

Vector Error Correction

Let x_t and y_t be two first difference stationary processes, i.e. Δy and Δx are **covariance stationary**.

According to Granger and Newbold (1974), OLS regression of y on x provides **spurious results**, i.e. t-tests suggest significance of the coefficients where there is none in the data generating process. Phillips (1986) shows this is due to the asymptotic OLS properties not holding for first difference stationary processes.

If y_t and x_t cointegrate, a regression of Δy_t on Δx_t is also misspecified.

Remember **cointegration**:

- x_t, y_t are first difference stationary
- $e_t = y_t - \alpha - \beta x_t$ is covariance stationary

Engle-Granger VECM: Intuition

Re-Define the relationship between y_t and x_t as:

$$y_t + \beta x_t = \epsilon_t \quad \epsilon_t = \epsilon_{t-1} + \xi_t \tag{5}$$

$$y_t + \alpha x_t = v_t \quad v_t = \rho v_{t-1} + \zeta_t \quad |\rho| < 1 \tag{6}$$

Here ξ_t and ζ_t are i.i.d. but mutually correlated processes responsible for the co-integration. ϵ_t is I(1), so consequently, so must be y_t and x_t.

Define $\delta = (1-\rho)/(\alpha - \beta)$ and $z_t = y_t + \alpha x_t$.

$$\Delta y_t = \beta \delta z_{t-1} + \eta_{1,t} \tag{7}$$

$$\Delta x_t = -\delta z_{t-1} + \eta_{2,t} \tag{8}$$

In $z_t = 0$, y_t and x_t are in **equilibrium**, and coefficients on z_{t-1} show how y_t and x_t react to deviations from equilibrium.

Engle-Granger: Nobel Prize Winners 2002

Figure 3: Engle and Granger: Winning the Nobel Medal in 2003 for being really careful about which relationships they call "causal" or spurious.

Engle-Granger VECM: General Case

Any VAR can be written and estimated as a VECM.

$$y_t = v_t + A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + \epsilon_t \quad (9)$$

$$\Delta y_t = v_t + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \epsilon_t \quad (10)$$

One last re-writing for **Johansen** maximum likelihood estimation:

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + v + \delta t + \epsilon_t \quad (11)$$

The important STATA commands are `varsoc` for lag selection, `vecrank` for the number of cointegrating equations and `vec` for the estimation.

Auto-Regressive Distributed Lag Models

- Model the relationship between variables in a single-equation setup
- Error Correction Representation (EC) is equivalent to co-integration of non-stationary variables
- EC representation is used to test for a long-run cointegrating relationship
- This allows for testing without knowing if the co-integrating variables are I(0) or I(1) themselves
- Examples: Wages and Labor Productivity, Foreign Direct Investment and Capital Intensity

Engle-Granger (1987) Test for long-run relationships

Assume $(y_t, x_t)'$ is a vector of I(1) variables

First Step: Run levels OLS $y_t = \alpha_1 + x_t'\beta + v_t$

Test if v_t is stationary (e.g. Adjusted Dickey Fuller or KPSS test)

Second Step: Estimate an error correction model and include lagged residuals $\hat{v}_{t-1}$ (if they are stationary):

$$\Delta y_t = \alpha_2 + \gamma \hat{v}_{t-1} + \sum_{i}^{p-1} \phi_{yi} \Delta y_{t-i} + \sum_{j}^{p-1} \phi_{xj} \Delta x_{t-j} + u_t$$

Test whether $-1 \leq \gamma < 0$.

Engle-Granger (1987): Downsides

variables must be I(1) and tested beforehand.

In short panels, first-step OLS estimates may be biased because of omitted short-run dynamics (no x_t as covariate), which influences the second step.

Standard significance testing in the first step is not available because asymptotic distribution of $\hat{\beta}$ is non-normal.

Engle-Granger (1987): Application

```
use usa.dta, clear
gen date = tq(1984q1) + _n-1
tsset date

dfuller f
dfuller D.f     //  f is integrated of order 1

dfuller b
dfuller D.b     //  b is integrated with order 1

reg b f
predict e, resid
dfuller e, noconstant
tsline e
```

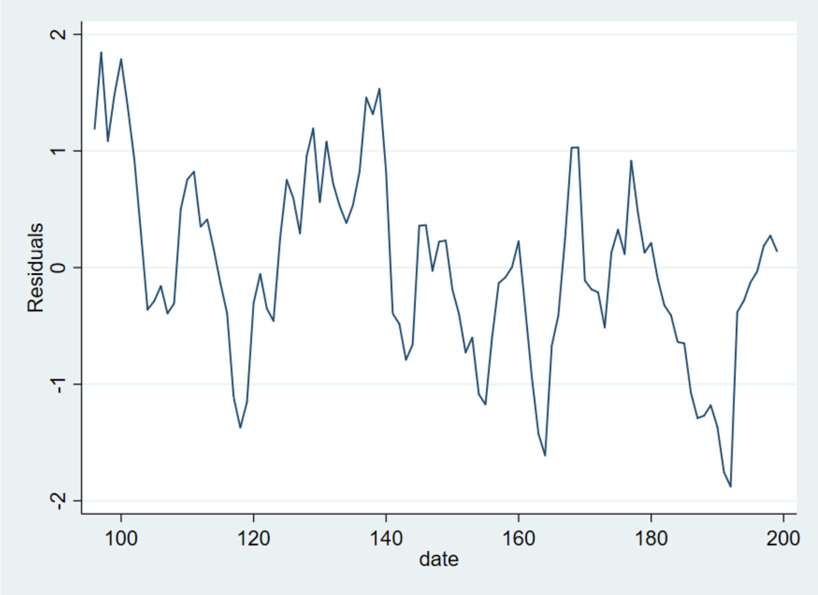
Engle-Granger (1987): Results

```
. dfuller e, noconstant

Dickey-Fuller test for unit root                   Number of obs   =       103

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.188            -2.600            -1.950            -1.610
```

Engle-Granger (1987): Results 2



Possible Models for long-run relationships

(https://davegiles.blogspot.com/2013/06/ardl-models-part-ii-bounds-tests.html)

If one wants to understand the dynamic relationship between two variables, there is a number of possible cases:

- Both are I(0), i.e. stationary. Then an OLS on the variable levels will be unbiased and efficient.
- The variables are integrated of the same order (eg. I(1)) but *not cointegrated*. Appropriate differentiation (i.e. first difference for first order integration) allows for OLS estimation.
- The variables are integrated of the same order *and* co-integrated. Then a level OLS provides the long-run relationship, whereas an Error Correction Model (ECM) (which can be estimated using OLS) represents the short-run dynamics.
- Data might be of different orders and/or co-integrated ("things are not as clear cut"). ARDL analyzes *both* short-run dynamics and long-run relationships.

ARDL: Pre-Requisites and Procedure

- none of the variables must be I(2)
- The model is written as an unrestricted ECM
  $\Delta y_t = \alpha + \sum_i^{p-1} \beta_1 \Delta y_{t-i} + \sum_j^{p-1} \beta_2 \Delta x_{t-j} + \gamma_1 y_{t-1} + \gamma_2 x_{t-1} + \epsilon_t$
- an appropriate lag structure is determined, e.g. using information criteria
- test for *serially independent errors*
- test for *dynamic stability*
- Pesaran-Shin-Smith Bounds test for long-run relationship (later in semester)
- estimate long-run "levels" model and short-run ECM

ARDL in STATA

```
. ardl eur us, aic //Use Akaike Information Criterion to decide on optimal mode

ARDL(4,0) regression

Sample:       424 -       614                     Number of obs     =        191
                                                  F(   5,    185)   =    3004.38
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.9878
                                                  Adj R-squared     =     0.9875
Log likelihood = -75.256023                       Root MSE          =     0.3646

------------------------------------------------------------------------------
         eur |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         eur |
         L1. |   1.055477   .0698777    15.10   0.000     .9176176    1.193337
         L2. |   .0733688   .1042555     0.70   0.482    -.1323138    .2790513
         L3. |   .1138271   .1057457     1.08   0.283    -.0947953    .3224496
         L4. |  -.2835677   .0690439    -4.11   0.000    -.4197823    -.147353
             |
          us |   .0526194   .0142578     3.69   0.000     .0244906    .0807482
       _cons |  -.0077393   .0609519    -0.13   0.899    -.1279894    .1125109
------------------------------------------------------------------------------
```

ARDL in STATA 2: Interpretation

(http://repec.org/usug2018/uk18_Kripfganz.pdf)

$$y_t = \alpha_0 + \alpha_1 t + \sum_{i}^{p} \phi_i y_{t-i} + \sum_{j}^{q} \beta_j' x_{t-j} + u_t$$

- Coefficients represent the long-term relationship between variable levels
- Include auto-regressive terms
- Include a time trend (trend stationarity)

ARDL in STATA 3 (ECM)

```
. ardl eur us, bic ec

ARDL(4,0) regression

Sample:        424 -       614                 Number of obs     =        191
                                               R-squared         =     0.2943
                                               Adj R-squared     =     0.2753
Log likelihood = -75.256023                    Root MSE          =     0.3646

------------------------------------------------------------------------------
       D.eur |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ADJ          |
         eur |
         L1. |  -.0408945   .0103098    -3.97   0.000    -.0612345   -.0205546
-------------+----------------------------------------------------------------
LR           |
          us |   1.286711   .3021187     4.26   0.000     .6906697    1.882751
-------------+----------------------------------------------------------------
SR           |
         eur |
         LD. |   .0963718   .0681707     1.41   0.159    -.0381202    .2308637
        L2D. |   .1697405   .0678472     2.50   0.013     .0358869    .3035941
        L3D. |   .2835677   .0690439     4.11   0.000      .147353    .4197823
             |
       _cons |  -.0077393   .0609519    -0.13   0.899    -.1279894    .1125109
```

ARDL in STATA 4: Interpretation of the Conditional EC Formulation

$$\Delta y_t = \alpha_0 + \alpha_1 t + \alpha_2 (y_{t-1} - \theta x_t) + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \sum_{i=0}^{q-1} \psi'_{xi} \Delta x_{t-i} + u_t$$

- α_2 is the *speed-of-adjustment* parameter, measuring how fast the system returns to equilibrium. It is denoted as a negative ("ADJ").
- $\alpha_2 = 1 - \sum_{j=1}^{p} \phi_j$ (from the level-ARDL regression).
- $\theta = \frac{\sum_{j=0}^{q} \beta_j}{\alpha_2}$ denotes the long run coefficients from the same *first step*. ("LR")
- ψ just denote the short-run coefficients from the second, error-correcting step ("SR")

ARDL in STATA 5: Alternative Error Correction Representation

```
. ardl eur us, bic ec1

ARDL(4,0) regression

Sample:       424 -       614                     Number of obs     =       191
                                                  R-squared         =    0.2943
                                                  Adj R-squared     =    0.2753
Log likelihood = -75.256023                       Root MSE          =    0.3646

------------------------------------------------------------------------------
       D.eur |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ADJ          |
         eur |
         L1. |  -.0408945   .0103098    -3.97   0.000    -.0612345   -.0205546
-------------+----------------------------------------------------------------
LR           |
          us |
         L1. |   1.286711   .3021187     4.26   0.000     .6906697    1.882751
```

ARDL in STATA 6: Alternative Error Correction Representation

```
------------+----------------------------------------------------------------
SR          |
        eur |
        LD. |   .0963718   .0681707     1.41   0.159    -.0381202    .2308637
       L2D. |   .1697405   .0678472     2.50   0.013     .0358869    .3035941
       L3D. |   .2835677   .0690439     4.11   0.000      .147353    .4197823
            |
         us |
        D1. |   .0526194   .0142578     3.69   0.000     .0244906    .0807482
            |
      _cons |  -.0077393   .0609519    -0.13   0.899    -.1279894    .1125109
-----------------------------------------------------------------------------
```

ARDL in STATA 7: Alternative Error Correction Representation

$$\Delta y_t = \alpha_0 + \alpha_1 t + \alpha_2(y_{t-1} - \theta x_{t-1}) + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \omega' \Delta x_{t-1} + \sum_{i=1}^{q-1} \psi'_{xi} \Delta x_{t-i} + u_t$$

- Δx_{t-1} is isolated with coefficient ω ("SR": "D1")
- Thus, the long-run dynamics only include lag levels ("LR": "L1").