E 4160: Econometrics–Modelling and Systems Estimation
Computer Class # 6

Ragnar Nymoen

Department of Economics, University of Oslo

14 November 2011

- *KeynesID2.fl*
  This is a batch file that uses the same data set as in *KeynesID1.zip* from computer class 5
- *Wage2_excel.xls*
  A data set with micro data from USA.
- *ADLfromVAR_d.in7/bn7* from computer class 2 (*ADLfromVAR_d.zip*)

We will use *KeynesID2.fl* to "run trough":

- the Specification test for the validity of the over-identifying instruments in a single equation, estimated by IV, and
- the LR test for the validity of the over-identifying restriction implies by a multi-equation structural model on the reduced form system
- See the last slides of CC5

- In our example, the situation is that we are interested in a theoretical relationship

$$C_i = \theta Y_i \tag{1}$$

where θ is the true derivative coefficient.

- The motivation of the Hausman-test is that we test

$$H_0\colon \text{plim}(\hat{\theta}_{OLS}) = \text{plim}(\hat{\theta}_{IV})$$

recognizing that $\text{plim}(\hat{\theta}_{IV}) = \theta$.

The test uses the information implied by the statistical model of C and Y under the alternative hypothesis, to test the equivalent H_0

$$\delta = 0$$

in the regression model:

$$C_i = \beta Y_i + \delta \hat{v}_i + \varepsilon_i \tag{2}$$

where $\hat{v}_i$ is the residual from the marginal (reduced form) model of Y_i, for example

$$Y_i = \gamma Z_i + e_i$$

which can only be formulated by specifying the statistical system for C, Y and Z.

- The Hausman test can be interpreted as a test of *Weak Exogeneity*.
- The definition of WE is that it allows us to do efficient inference about the *parameter of interest* (θ) by only considering the conditional relationship.
- If the test rejects, the parameter of interest is not "in" the regression model—we need a different statistical model, where the marginal model for Y is allowed to play a role in the estimation of θ.

Models of the VAR I

Referring back to CC2 we had:

$$y_t = \mu_{yt-1} + \epsilon_{yt} \tag{3}$$

$$x_t = \mu_{xt-1} + \epsilon_{xt} \tag{4}$$

$$\mu_{yt-1} = E[y_t \mid x_{t-1}, y_{t-1}] = \mu_{y0} + a_{11} y_{t-1} + a_{12} x_{t-1} \tag{5}$$

$$\mu_{xt-1} = E[x_t \mid x_{t-1}, y_{t-1}] = \mu_{x0} + a_{21} y_{t-1} + a_{22} x_{t-1} \tag{6}$$

$$\begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_x^2 & \omega_{xy} \\ \omega_{xy} & \sigma_y^2 \end{pmatrix}\right). \tag{7}$$

which is a bivariate VAR with normally distributed errors.

Models of the VAR II

- If μ_{yt-1} and μ_{yt-1} depend on (a vector of) exogenous variables we have an open VAR, also called VAR-X
- We have seen that we can formulate different econometric models of this type of VAR
  - A simultaneous equations model
  - A recursive model
  - A model with a conditional model for y_t and a marginal model for x_t.

Models of the VAR III

- To review some important concepts and properties, we now choose the conditional model representation, meaning that we have

$$\begin{aligned} y_t &= \beta_0 + \phi_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t. &(8)\\ x_t &= \mu_{x0} + a_{21} y_{t-1} + a_{22} x_{t-1} + \epsilon_{xt} &(9)\end{aligned}$$

  see CC2 for derivations, in particular:

$$E(\varepsilon_t \epsilon_{xt}) = 0 \Rightarrow E(\varepsilon_t x_t) = 0$$

(8) is often referred to as the Autoregressive Distributed Lag Model (ARDL) for y_t given x_t.

- We have that

Models of the VAR IV

- x_t is *strictly exogenous*
- x_t is *weakly exogenous* since *OLS* estimation of (8) efficient.
- x_t is *strongly exogenous* if $a_{21} = 0$ is a valid restriction on the system (one-way Granger Causality)
- x_t is *super-exogenous* if the parameters of (8) are invariant with respect to structural changes in the marginal model.

Single equation typology: special cases of ARDL

$\phi_1 = 0$	Distributed lag (DL $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t.$
$\phi_1 = 1$ $\beta_1 + \beta_2 = 0$	Growth rate (GR) $\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \varepsilon_t.$
$\phi_1 = 1$ $\beta_1 = \beta_2 = 0$	Random Walk (RW) $y_t = \beta_0 + y_{t-1} + \varepsilon_t.$
$\phi_1 = 0$ $\beta_2 = 0$	Static model $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t.$
$\lvert\phi_1\rvert < 1$	ECM $\Delta y_t = \beta_0 + \beta_1 \Delta x_t + (\phi_1 - 1) y_{t-1}$ $+ (\beta_1 + \beta_2) x_{t-1} \varepsilon_t.$

And many more, including polynomial lag distributions and geometric lag distribution, see EB note DL which can be used to represent longer lags by a small number of parameters.

- Note in particular that ECM does not impose any restrictions on ARDL, except $|\phi_1| < 1$ (dynamic stability).
- The partial derivatives
$$\frac{\partial y_t}{\partial x_{t-j}}$$
also called *dynamic multipliers,* or l*ag-weights*, are easy to obtain in PcGive after estimation, as are the *long-run multipliers*, as we have seen (for example CC3).
- As all parameters of an econometric model, also dynamic and long-run multipliers can become badly biased if a mis-specified model is estimated.

Estimation of correctly specified AR and ARDL models I

Yet another special case: If $\beta_1 = \beta_2 = 0$ is true, then the AR model

$$y_t = \beta_0 + \phi_1 y_{t-1} + \varepsilon_t$$

is correctly specified.

- The only caveat is that y_{t-1} is predetermined rater than exogenous.
- OLS estimator of ϕ_1 is consistent but biased in small samples.
- We can check this directly using Monte Carlo Simulation.
- Choose Model category **Monte Carlo** in the **Model Formulation menu**.