ECON 4160: Econometrics–Modelling and Systems Estimation:

Computer Class

André Anundsen

Department of Economics, University of Oslo

September 9, 2013

Practical information

Who am I? $\rightarrow$ André K. Anundsen (PhD-student)
Email: *a.k.anundsen@econ.uio.no*
Office: 1143
Responsible for the rest of the CPU classes + the first half of the seminar series (1–3)

Outline

Data sets for today, posted on the web page:

- KonsDataSim2.zip
- ADLfromVAR_d.zip

A conditional model of the VAR

As economists we will typically be interested in building econometric *models* of the VAR *system*

Today, we will consider the conditional model of the VAR!

From VAR to ARDL I

Consider a bi-variate VAR model of first order, i.e.:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{1}$$

A more compact notation gives:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{2}$$

Let us now assume that $\boldsymbol{\varepsilon}_t \sim MVN(\mathbf{0}_{2\times 1}, \boldsymbol{\Sigma})$, i.e.:

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim MVN\left(\mathbf{0}_{2\times 1}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right) \tag{3}$$

From VAR to ARDL II

It follows from (1) and (3) that $\mathbf{y}_t|\mathbf{y}_{t-1} \sim MVN\left(E(\mathbf{y}_t|\mathbf{y}_{t-1}), \mathbf{\Sigma}\right)$, where it is easy to show that:

$$\begin{aligned} E(\mathbf{y}_t|\mathbf{y}_{t-1}) &= E\left(\left(\begin{array}{c} y_{1,t} \\ y_{2,t} \end{array}\right) \Big| y_{1,t-1}, y_{2,t-1}\right) \\ &= \left(\begin{array}{c} \mu_1 + a_{11}y_{1,t-1} + a_{12}y_{2,t-1} \\ \mu_2 + a_{21}y_{1,t-1} + a_{22}y_{2,t-1} \end{array}\right) \end{aligned} \quad (4)$$

From VAR to ARDL III

Since $\mathbf{y}_t | \mathbf{y}_{t-1} \sim MVN\left(E(\mathbf{y}_t|\mathbf{y}_{t-1}), \boldsymbol{\Sigma}\right)$, we know from the properties of the MVN distribution that:

$$y_{1,t}|y_{2,t} \sim N\left(\underbrace{E(y_{1,t}|y_{1,t-1}, y_{2,t-1}) - \rho_{12}\frac{\sigma_1}{\sigma_2}\left(E(y_{2,t}|y_{1,t-1}, y_{2,t-1}) - y_{2,t}\right)}_{=E(y_{1,t}|y_{2,t},y_{1,t-1},y_{2,t-1})}, \underbrace{(1-\rho_{12}^2)\sigma_1^2}_{=Var(y_{1,t}|y_{2,t},y_{1,t-1},y_{2,t-1})}\right) \quad (5)$$

with $\rho_{12} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$

From VAR to ARDL IV

Using the expressions for $E(\mathbf{y}_t|\mathbf{y}_{t-1})$ as derived in (4) in combination with the expression for $E(y_{1,t}|y_{2,t}, y_{1,t-1}, y_{2,t-1})$ in (5), we get that:

$$E(y_{1,t}|y_{2,t}, y_{1,t-1}, y_{2,t-1}) = \underbrace{\mu_1 + a_{11}y_{1,t-1} + a_{12}y_{2,t-1}}_{E(y_{1,t}|y_{1,t-1},y_{2,t-1})} - \rho_{12}\frac{\sigma_1}{\sigma_2}\left(\underbrace{\mu_2 + a_{21}y_{1,t-1} + a_{22}y_{2,t-1}}_{E(y_{2,t}|y_{1,t-1},y_{2,t-1})} - y_{2,t}\right) \tag{6}$$

just collecting some terms, we get:

$$E(y_{1,t}|y_{2,t}, \mathbf{y}_{t-1}) = \underbrace{\phi_0}_{=\mu_1 - \frac{\sigma_{12}}{\sigma_2^2}\mu_2} + \underbrace{\beta_0}_{=\frac{\sigma_{12}}{\sigma_2^2}} y_{2,t} + \underbrace{\phi_1}_{=a_{11} - \frac{\sigma_{12}}{\sigma_2^2}a_{21}} y_{1,t-1} + \underbrace{\beta_1}_{=a_{12} - \frac{\sigma_{12}}{\sigma_2^2}a_{22}} y_{2,t-1} \tag{7}$$

From VAR to ARDL V

Define the stochastic variable ϵ_t in the following way:

$$\epsilon_t = y_{1,t} - E(y_{1,t} | y_{2,t}, \mathbf{y}_{t-1}) \tag{8}$$

Combining (7) and (8) yields

$$y_{1,t} = \phi_0 + \beta_0 y_{2,t} + \phi_1 y_{1,t-1} + \beta_1 y_{2,t-1} + \epsilon_t \tag{9}$$

which is nothing but a conditional model for $y_{1,t}$, and now you see how this may be derived from the VAR

The ARDL disturbance I

Consider the expression for the ARDL disturbance:

$$\epsilon_t = y_{1,t} - E(y_{1,t} | y_{2,t}, \mathbf{y}_{t-1})$$

Substitute in for $E(y_{1,t} | y_{2,t}, \mathbf{y}_{t-1})$ from (6) to get (after some simple re-arrangements):

$$\begin{aligned}\epsilon_t = & \left(y_{1,t} - \mu_1 - a_{11} y_{1,t-1} - a_{12} y_{2,t-1}\right) \\ & - \rho_{12} \frac{\sigma_1}{\sigma_2} \left(y_{2,t} - \mu_2 - a_{21} y_{1,t-1} - a_{22} y_{2,t-1}\right)\end{aligned}$$

If you have a quick look at where we started out (equation (1)), you will recognize that the terms in the parentheses are nothing but the two VAR disturbances! Hence:

$$\epsilon_t = \varepsilon_{1,t} - \rho_{12} \frac{\sigma_1}{\sigma_2} \varepsilon_{2,t}$$

The ARDL disturbance II

The expression in (11) can be used to show:

$$
\begin{aligned}
E(\epsilon_t) &= 0,\ E(\epsilon_t \varepsilon_{2,t}) = 0 \\
Var(\epsilon_t) &= Var(\varepsilon_{1,t}) + Var(\rho_{12}\frac{\sigma_1}{\sigma_2}\varepsilon_{2,t}) - 2Cov\left(\varepsilon_{1,t}, \rho_{12}\frac{\sigma_1}{\sigma_2}\varepsilon_{2,t}\right) \\
&= \sigma_1^2 + \rho_{12}^2\sigma_1^2 - 2\rho_{12}\frac{\sigma_1}{\sigma_2}\sigma_{12} \\
&= \sigma_1^2 + \rho_{12}^2\sigma_1^2 - 2\rho_{12}\frac{\sigma_1}{\sigma_2}\underbrace{\rho_{12}\sigma_1\sigma_2}_{=\sigma_{12}} = \sigma_1^2(1-\rho_{12}) \\
E(y_{2,t}\epsilon_t) &= 0 \ \forall \text{ (for all) } t
\end{aligned}
$$

The statistical system described by (1) and (3) can now be expressed in **model form** by:

$$y_{1,t} = \phi_0 + \beta_0 y_{2,t} + \beta_1 y_{2,t-1} + \phi_1 y_{1,t-1} + \epsilon_t \qquad (10)$$
$$y_{2,t} = \mu_2 + a_{21} y_{1,t-1} + a_{22} y_{2,t-1} + \varepsilon_{2,t} \qquad (11)$$

$$\begin{aligned} E(\epsilon_t) &= 0 \; \forall \; t \\ Var(\epsilon_t) &= \sigma_1^2 (1 - \rho_{12}) \; \forall \; t \\ E(y_{2,t}\epsilon_t) &= 0 \; \forall \; t \\ E(\varepsilon_{2,t}\epsilon_t) &= 0 \; \forall \; t \end{aligned}$$

Regressions models I

1. When estimating a linear regression model, we are estimating a conditional expectation that is derived from a system of equations, e.g. a simple bivariate VAR
   - We are therefore estimating a *partial* system!
2. The full econometric model of the system consists of the *conditional model* (10), the *marginal model* (11), and the disturbances ϵ_t and $\varepsilon_{2,t}$
3. When $Cov(\epsilon_t, y_{2,t}) = 0$, $y_{2,t}$ is *exogenous* in the conditional model
4. OLS estimation is efficient for Gaussian (i.e., normal) disturbances and gives the Maximum Likelihood estimators for ϕ_0, β_0, ϕ_1 and β_1

Regressions models II

5. This means that there is no information in the marginal model that can help us improve on the estimates of ϕ_0, β_0, ϕ_1 and β_1 that we get from the conditional model
6. We say that $y_{2,t}$ is a *weakly exogenous* variable for the *parameters of interest*. In our case: ϕ_0, β_0, ϕ_1 and β_1 and $Var[\epsilon_t]$

Let's see if we can confirm the results we just derived! I

Load in the in7 file contained in the **KonsDataSim2.zip** (on the web). Now, let us do the following:

1. Back rows: Estimate a bi-variate VAR of first order in consumption (C) and income (I), where $y_{1,t} = C_t$ and $y_{2,t} = I_t$
2. First rows: Estimate an $ARDL(1,1)$ model for a conditional consumption equation, i.e. $C_t | I_t, C_{t-1}, I_{t-1}$

Both groups use the full sample (1959–2007)!

Let's see if we can confirm the results we just derived! II

I have cheated, and calculated the empirical variance-covariance matrix of the VAR residuals to save some time. It is given by:

$$\begin{pmatrix} 10290.0736 & 4655.303596 \\ 4655.303596 & 4310.579025 \end{pmatrix}$$

Can we by combining the VAR estimates obtained by the guys at the back rows and the cov-matrix above guess what results the guys at the front row got?

Let's see if we can confirm the results we just derived! III

Again, I have cheated and calculated in advance, but I guess the front rows got the following:

$$\hat{\phi}_0 = \hat{\mu}_1 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_2^2}\hat{\mu}_2 = 72.0987 - \frac{4655.303596}{4310.579025}22.0776 = 48.25551557$$

$$\hat{\beta}_0 = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_2^2} = \frac{4655.303596}{4310.579025} = 1.079971755$$

$$\hat{\phi}_1 = \hat{a}_{11} - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_2^2}\hat{a}_{21} = 0.702311 - \frac{4655.303596}{4310.579025}0.120492 = 0.572183043$$

$$\hat{\beta}_1 = \hat{a}_{21} - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_2^2}\hat{a}_{22} = 0.351689 - \frac{4655.303596}{4310.579025}0.784072 = -0.49508661$$

Generalizations

- The results 1-3 above do – in general – not depend on normality (it is just a simplification)
- In particular: Normality of $y_{2,t}$ is *not* required for $E[\epsilon_t] = 0$ and $E[\epsilon_t y_{2,t}] = 0$
- The results $E[\epsilon_t] = 0$ and $E[\epsilon_t y_{2,t}] = 0$ do not depend on linearity. More generally, we have

$$y_{1,t} = E[y_{1,t} \mid y_{2,t}, y_{1,t-1}, y_{2,t-1}] + \epsilon_t$$

  with $E[\epsilon_t] = 0$ and $E[\epsilon_t y_{2,t}] = 0$ for a *non-linear* conditional expectation function $E[y_{1,t} \mid y_{2,t}, y_{1,t-1}, y_{2,t-1}]$
- As Lecture note # 4 demonstrates, generalizations from one to k explanatory variables and p lags is straight-forward. We get:

$$y_{1,t} = E[y_{1,t} \mid y_{1,t-1}, \ldots, y_{1,t-p}, y_{2,t}, \ldots, y_{2,t-p}, \ldots, y_{k+1,t}, \ldots y_{k+1,t-p}]$$

and the linear multiple regression is a special case

Another ARDL example

We specify a DGP in accordance with equation (1) :

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.4 \\ 0.2 & 0.7 \end{pmatrix}$$

and the following distribution for the disturbances

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N\left(\mathbf{0}_{2\times 1}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

Then we expect:

$$\begin{aligned}
\phi_1 &: a_{11} - \frac{\sigma_{12}}{\sigma_2^2} a_{21} \Rightarrow 0.5 - 0.5 * 0.2 = 0.4 \\
\beta_0 &: \frac{\sigma_{12}}{\sigma_2^2} \Rightarrow 0.5 \\
\beta_1 &: a_{12} - \frac{\sigma_{12}}{\sigma_2^2} a_{22} \Rightarrow 0.4 - 0.5 * 0.7 = 0.05
\end{aligned}$$

What do we find?

- Data from this DGP is found in the file *ADLfromVAR_d.in7/bn7*.
- In that file *YA* corresponds to $y_{1,t}$ above and *YB* corresponds to $y_{2,t}$ above.
- Use *PcGive* to estimate the conditional model and see what you get!

Weak exogeneity of $y_{2,t}$ in the conditional model

OLS gives ML estimates of the parameters of the ARDL model

$$y_{1,t} = \phi_0 + \beta_0 y_{2,t} + \phi_1 y_{1,t-1} \beta_1 y_{2,t-1} + \epsilon_t \tag{12}$$

$y_{2,t}$ is therefore weakly exogenous in (12) despite the fact that $y_{2,t}$ is an endogenous variable in the VAR:

$$\begin{aligned} y_{1,t} &= a_{11} y_{1,t-1} + a_{12} y_{2,t-1} + \varepsilon_{1,t} \qquad (13) \\ y_{2,t} &= a_{21} y_{1,t-1} + a_{22} y_{2,t-1} + \varepsilon_{2,t} \qquad (14) \end{aligned}$$

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim MVN \left(\mathbf{0}_{2\times 1}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right) \tag{15}$$

- There is a difference between a variable being endogenous in a statistical system like (13)-(15) and being endogenous in a *model* of the statistical system, such as (12)
- $y_{2,t}$ is *weakly exogenous* for the parameters in (12) because we do not gain anything in terms of efficiency by estimating (12) jointly with the marginal equation (14).

Again, this is a consequence of the conditioning, which also gives

$$E(\epsilon_t \varepsilon_{2,t}) = 0 \Rightarrow E(\epsilon_t y_{2,t}) = 0$$

so $y_{2,t}$ is exogenous in the econometric sense that is used in most textbooks (sometimes referred to as the condition of *strict exogeneity*.)

Parameters of interest and weak exogeneity

- How helpful and relevant is the weak exogeneity of a variable in a conditional (regression) model?
- It *is* relevant if the parameters that we want to estimate, *the parameters of interest*, are the parameters of the conditional model!
- If the parameters of interest are not the conditional model, then the weak exogeneity of $y_{2,t}$ is not very helpful
- The solution is to change to a *different econometric model* of the system
- The other model is estimated by *other methods* than OLS

Predeterminedness

Consider again the two-variable ARDL model that we have derived from the bivariate VAR(1) model

$$y_{1,t} = \phi_0 + \beta_0 y_{2,t} + \phi_1 y_{1,t-1} + \beta_1 y_{2,t-1} + \epsilon_t \tag{16}$$

we have that

$$E(y_{1,t-1}\epsilon_{t+j}) = 0, \text{and } E(y_{2,t-1}\epsilon_{t+j}) = 0 \ \forall\, j > 0$$

by conditioning on history of the system, and

$$E(y_{2,t}\epsilon_{t+j}) = 0 \ \forall\, j > 0$$

by conditioning on $y_{2,t}$

Heuristically, we cannot claim strict exogeneity

$$E(y_{1,t-1}\epsilon_{t\pm j}) = 0 \; \forall \; j \qquad (17)$$

Intuitively, this is because $y_{1,t-1}$ must be correlated with ϵ_{t-1},ϵ_{t-2} and older disturbances through the solution of the equation for $y_{1,t}$

- (17) defines $y_{1,t-1}$ as a pre-determined variable.
- $y_{2,t}$ and $y_{2,t-1}$ are either exogenous or predetermined (depending on Granger causality, which we will discuss in more detail the next time)
- With pre-determinedness OLS estimators are biased in small samples, but they remain consistent estimators in stationary systems
- The size of the bias is seldom very large, and it declines with ϕ_1

Consider two alternative simplifications of the ARDL model:

1. Mod. 1: $\phi_1 = \beta_1 = 0$, which is just a simple static regression model with an exogenous regressor
2. Mod. 2: $\beta_0 = \beta_1 = 0$, which is called an $AR(1)$ model

Predeterminedness and mis-specification

	Disturbances ϵ_t are:			
Mod.	heteroscedastic		autocorrelated	
Mod. 1: Static	$\hat{\beta}_0$ unbiased consistent	$\widehat{Var}(\hat{\beta}_0)$ wrong	$\hat{\beta}_0$ unbiased consistent	$\widehat{Var}(\hat{\beta}_0)$ wrong
Mod. 2 AR(1)	$\hat{\phi}_1$ biased consistent	$\widehat{Var}(\hat{\phi}_1)$ wrong	$\hat{\phi}_1$ biased inconsistent	$\widehat{Var}(\hat{\phi}_1)$ wrong

What is Monte Carlo simulation? I

Say that the process that has generated the date (the data generating process, the DGP) takes the following form:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (18)$$

where ε_t is normally distributed. Now, assume that we had some data $t = 1, \ldots, T$ on y_t and x_t, and that we want to pin down β_1 (our parameter of interest)

As long as $Cov(x_t, \varepsilon_t) = 0$, you know that the OLS estimator is BLUE! Can we confirm this by simulation?

What is Monte Carlo simulation? II

So, what do we do?

1. Fix β_0 and β_1 in (18) at some values, e.g. $\beta_0 = 2$ and $\beta_1 = 1.5$
2. Generate some numbers for the time series x_t on a sample $t = 1 \ldots, T$
3. Say that $\varepsilon_t \sim N(0, 1)$, and draw T numbers from the standard normal distribution
4. Then, y_t will follow by definition from the DGP!
5. Estimate an equation of the form (18) by OLS and collect your β_1 estimate; call it $\hat{\beta}_1^1$
6. Now, repeat the steps 1–5 M times, and calculate $\beta_1^{MC} = \frac{\sum_{m=1}^{M} \hat{\beta}_1^m}{M}$! This is just the mean estimator, which by the law of large numbers converges to $E(\hat{\beta}_1)$ as $M \to \infty$.

What is Monte Carlo simulation? III

But then, we know that the estimator is unbiased if $\beta_1^{MC} - \beta_1 = 0$. Let us vary the sample size from $T = 20$ to $T = 500$ in increments of 20 and do this experiment with $M = 1000$ to check the unbiasedness of the OLS estimator!

Our experiments

1. Show that OLS estimator in Mod. 1 is unbiased (DGP for Mod. 1: $y_{1,t} = 2 + 1.2y_{2,t} + \epsilon_t, \epsilon_t \sim N(0,1)$)
2. Show small sample bias of AR coefficient in Mod. 2 (DGP for Mod. 2: $y_{1,t} = 2 + 0.6y_{1,t-1} + \epsilon_{1,t}, \epsilon_{1,t} \sim N(0,1)$)
3. Show small sample bias of coefficients in ARDL(1,1) (DGP for ARDL(1,1):
$y_{1,t} = 2 + 0.6y_{1,t-1} + 0.2y_{2,t} - 0.3y_{2,t-1} + \epsilon_{1,t}, \epsilon_t \sim N(0,1)$)