E 4160: Econometrics–Modeling and Systems Estimation
Computer Class # 2

Ragnar Nymoen

Department of Economics, University of Oslo

12 September 2011

Data sets for today, posted on the web page:

- KonsDataSim2.zip
- ECON4160_Jorg121314.xls, see also Erik Biorn's note *FGLS for regression systems* which was posted 8 Sept
- ADLfromVAR_d.zip
- KonsData2Nor.zip

The simple bivariate case I

Assume that we have stochastic variables (y_i, x_i), $i = 1, 2, \ldots, n$ that are generated by the following system of linear equations:

$$y_i = \mu_y + \epsilon_{y,i} \tag{1}$$

$$x_i = \mu_x + \epsilon_{x,i} \tag{2}$$

where μ_y and μ_x are parameters and $\epsilon_{y,i}$ and $\epsilon_{x,i}$ are have a joint probability distribution. Without loss of generality (for the following derivations) we assume a normal distribution

$$\begin{pmatrix} \epsilon_{xi} \\ \epsilon_{yi} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_x^2 & \omega_{xy} \\ \omega_{xy} & \sigma_y^2 \end{pmatrix}\right). \tag{3}$$

The simple bivariate case II

$\epsilon_{x,i}$ and $\epsilon_{y,i}$ is bivariate normal with expectation zero and covariance matrix

$$\begin{pmatrix} \sigma_x^2 & \omega_{xy} \\ \omega_{xy} & \sigma_y^2 \end{pmatrix}.$$

The correlation coefficient between $\epsilon_{x,i}$ and $\epsilon_{y,i}$ is:

$$\rho_{xy} = \frac{\omega_{xy}}{\sigma_x \sigma_y}.$$

Since linear combination of normally distributed variables are also normally distributed, it follows that y_i and x_i are normally distributed (and correlated in general).

The simple bivariate case III

From the properties of the normal distribution we know that the distribution of y_i conditional on x_i is also normal, with expectation

$$\begin{aligned} E[y_i \quad | \quad x_i] &= \mu_y - \rho_{xy}\frac{\sigma_y}{\sigma_x}\mu_x + \rho_{xy}\frac{\sigma_y}{\sigma_x}x_i \\ &= \beta_1 + \beta_2 x_i \end{aligned} \tag{4}$$

If we define the stochastic variables ε_i, $i = 1, 2, \ldots, n$

$$\varepsilon_i = y_i - E[y_i \mid x_i] \tag{5}$$

we see that the simple regression model:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i. \tag{6}$$

gives y_i as the sum of the conditional expectations function (4) and the disturbance ε_t.

The simple bivariate case IV

Note that ε_i can also be written as

$$\begin{aligned} \varepsilon_i &= \mu_y + \epsilon_{yi} - \beta_1 - \beta_2(\mu_x + \epsilon_{xi}) \\ &= \epsilon_{yi} - \frac{\omega_{xy}}{\sigma_x^2}\epsilon_{xi} \end{aligned}$$

Which can be used to show:

$$\begin{aligned} E(\varepsilon_i) &= 0,\ E(\varepsilon_i \epsilon_{xi}) = 0 \\ Var(\varepsilon_i) &= \sigma_y^2 \left(1 - \frac{\omega_{xy}^2}{\sigma_y^2 \sigma_x^2}\right) \\ E(x_i \varepsilon_i) &= 0 \text{ for all } i \end{aligned}$$

The simple bivariate case V

The statistical system given by (1), (2) and (3) can be expressed in **model form** by:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \tag{7}$$

$$x_i = \mu_x + \epsilon_{x,i} \tag{8}$$

$$\begin{aligned} E(\varepsilon_i) &= 0, \text{ for all } i \\ Var(\varepsilon_i) &= \sigma_y^2 (1 - \frac{\omega_{xy}^2}{\sigma_y^2 \sigma_x^2}), \text{ for all } i \\ E(x_i \varepsilon_i) &= 0 \text{ for all } i \\ E(\epsilon_{x,ii} \varepsilon_i) &= 0 \text{ for all } i \end{aligned}$$

Regressions models I

1. When we estimate a linear regression we are estimating a conditional expectation that is derived from a system of variables.
2. When we estimate a linear regression model we can say that we are estimating a *partial* system!
3. The full econometric model of the system consists of the *conditional model* (7), the *marginal model* (8), and the disturbances ε_i and ϵ_{xi}.
4. With reference to $Cov(\varepsilon_i, x_i) = 0$, x_i is *exogenous* in the conditional model.
5. OLS estimation is efficient for Gaussian (i.e., normal) disturbances and gives the Maximum Likelihood estimators for β_1 and β_2.

Regressions models II

6. This means that there is no information in the marginal model that can help us improve on the estimates of β_1 and β_2 that we get from the conditional model.
7. We say that x_i is a *weakly exogenous* variable for *parameters of interest* β_1, β_2 and $Var[\varepsilon_i]$. (Greene defines weak exogeneity on page 357)

Generalizations I

- The results 1.-4. above do not depend on normality (it is just a simplification)
- In particular: normality of x_i is *not* required for $E[\varepsilon_i \mid x_i] = 0$ and $E[\varepsilon_i x_i] = 0$.
- The results $E[\varepsilon_i \mid x_i] = 0$ and $E[\varepsilon_i x_i] = 0$ do not depend on linearity. More generally we have

$$y_i = E[y_i \mid x_i] + \varepsilon_i$$

with $E[\varepsilon_i \mid x_i] = 0$ and $E[\varepsilon_i x_i] = 0$ for a *non-linear* conditional expectation function $E[y_i \mid x_i]$.

Generalizations II

- Generalization from one to k explanatory variables is straight-forward: we get

$$y_i = E[y_i \mid x_{1i}, x_{2i}, \ldots, x_{ki}] + \varepsilon_i$$

and linear multiple regression as a special case.

Omitted variables bias I

- Consider the three-variable system y, x_1 and x_2.
- The conditional expectations function for y is then a function of x_1 and x_2
- Analyze this in OxMetrics/PcGive with the use of the artificial data set in *KonsDataSim2.zip* on the web page.
- We will only use the variables named C (consumption), I (income) and F (households' wealth).
- DGP is

$$C = 45 + 0.8I + 0.05F$$

- Choose the same sample size as in the first computer class: 1960-2006.

The SUR model I

- Sometimes we are not interested in the parameters of the conditional model, but in the parameters of the system.
- We may be interested in μ_y and μ_x in (1) and (2) or more generally:

$$y_i = \mu_{y0} + \mu_{y1} z_{1i} + \epsilon_{yi} \tag{9}$$

$$x_i = \mu_{x0} + \mu_{x1} z_{2i} + \epsilon_{xi} \tag{10}$$

with $Cov(z_{1i}, \epsilon_{yi}) = 0$ $Cov(z_{2i}, \epsilon_{yi}) = 0$.

- Since $\omega_{yx} \neq 0$ in general, the two equations are Seemingly Unrelated Regressions, SUR.

The SUR model II

- From the lectures we know that the efficient estimator is Feasible Generalized Least Squares, FGLS, the SUR estimator,
- However, if $z_{1i} = z_{2i}$ or if $\omega_{yx} = 0$, the SUR estimator is identical to the OLS estimator for each equation
- Use the data set *ECON4160_JORG121314.xls* to estimate an example. This will also be commented in the Lecture later in the week.

The VAR I

- The above assumes implicitly that we can condition on *any* x when we model y conditionally.
  Hence we have in effect that

$$y_i = E[y_i \mid x_j] + \varepsilon_i \ \forall \ i, j$$

  with $E[\varepsilon_i \mid x_j] = 0, \ \forall \ i, j$ and $E[\varepsilon_i x_j] = 0, \ \forall \ i, j$.
- However, for time-series variables (x_t, ε_s) not all combinations are relevant: There is a natural ordering of time!

The VAR II

Let $\mu_{x,t-1}$, $\mu_{x,t-1}$ denote the expectations of x_t and y_t conditional on the pre-history:

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

$$\mu_{xt-1} = \mathrm{E}[x_t \mid x_{t-1}, y_{t-1}] = a_{21} y_{t-1} + a_{22} x_{t-1} \tag{12}$$

a_{ij} are the parameters of the conditional expectation of the economic system. 1) and (2) are replaced by:

$$y_t = \mu_{yt-1} + \epsilon_{yt} \tag{13}$$

$$x_t = \mu_{xt-1} + \epsilon_{xt} \tag{14}$$

The VAR III

where ϵ_{yt} and ϵ_{xt} have a joint probability distribution, for example

$$\begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_x^2 & \omega_{xy} \\ \omega_{xy} & \sigma_y^2 \end{pmatrix}\right). \tag{15}$$

ϵ_{xt} and ϵ_{yt} are bivariate normal with expectation zero and covariance matrix

$$\begin{pmatrix} \sigma_x^2 & \omega_{xy} \\ \omega_{xy} & \sigma_y^2 \end{pmatrix}.$$

The correlation coefficient between ϵ_{xt} and ϵ_{yt} is:

$$\rho_{xy} = \frac{\omega_{xy}}{\sigma_x \sigma_y}. \tag{16}$$

The VAR IV

- The specification of the system that begins with (11) and ends with (16) is a Vector AutoRegressive model, VAR.
- The name reflects that we here have a so called autoregressive model for a vector variable:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{yt} \\ \epsilon_{xt} \end{pmatrix} \qquad (17)$$

- VAR models are popular in macroeconomics.
- Note that the VAR above is an example of a *system of regression equations* (Biorn ch 4) or *SUR* model (Greene Ch 10) when we consider y_{t-1} and x_{t-1} as exogenous variables
- In fact, we will need the more precise term predetermined variable which we define below.

A conditional model of the VAR I

As economists we will typically be interested in building econometric *models* of the VAR *system*.
In this course we will consider mainly two models of the VAR

- A conditional model of the VAR
- A simultaneous equations model of the VAR

Today we will consider the conditional model of the VAR
With the new notation for time series, the derivation used at the start of the slide set goes can be used directly to give

$$y_t = \phi_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t. \tag{18}$$

A conditional model of the VAR II

$$E[\varepsilon_t \mid x_t, x_{t-1}, y_{t-1}] = 0$$

$$Var[\varepsilon_t \mid x_t, x_{t-1}, y_{t-1}] \equiv \sigma^2 = \sigma_y^2(1-\rho_{xy}^2),$$

$$\phi_1 = a_{11} - \frac{\omega_{xy}}{\sigma_x^2} a_{21},$$

$$\beta_1 = \frac{\omega_{xy}}{\sigma_x^2},$$

$$\beta_2 = a_{12} - \frac{\omega_{xy}}{\sigma_x^2} a_{22}.$$

$$x_t = a_{21} y_{t-1} + a_{22} x_{t-1} + \epsilon_{xt} \quad (19)$$

$$Cov(\varepsilon_t, \epsilon_{xt}) = 0.$$

A conditional model of the VAR III

- (18)
$$y_t = \phi_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t$$
can be estimated by OLS, which is also conditional FIML for the case of normal disturbances.
- As in the simplest case, (18) is a conditional model and (2) is a marginal model.

Granger causality

- We see that in one sense (18) and (19) define a recursive model, see Biorn page 6.5, since given the history (represented by y_{t-1} and x_{t-1}), then x_t is determined first, and given this, y_t is determined. Moreover, $Cov(\varepsilon_t, \epsilon_{xt}) = 0$.
- However, the two variables are clearly jointly determined over time, since in general $x_{t-1} \longrightarrow y_t$ and $y_{t-1} \longrightarrow x_t$. In econometrics we call this joint Granger causality. Only if $a_{21} = 0$ can we say that we have a recursive causal chain.
- With $a_{21} = 0$ imposed we say that y_{t-1} is not Granger-causing x while x_{t-1} is Granger-causing y_t.

The ADL model

Dynamic models of the form (18):

$$y_t = \phi_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t.$$

are referred to as Autoregressive Distributed Lag models in the literature, and in OxMetrics/PcGive in particular.

- ADL models can be readily generalized to k explanatory variables and to “any” lag lengths (how far back we condition in the first step)
- Such large ADL equations then implies k marginal models (although they are often not estimated, since we want to focus on the conditional model)

An example ADL model I

We specify a DGP in accordance with (11) (13) :

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

and

$$\begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

Then we expect:

An example ADL model II

$$\begin{aligned}
\phi_1 &: \quad a_{11} - \frac{\omega_{xy}}{\sigma_x^2} a_{21} \Rightarrow 0.5 - 0.5 * 0.2 = 0.4 \\
\beta_1 &: \quad \frac{\omega_{xy}}{\sigma_x^2} \Rightarrow 0.5 \\
\beta_2 &: \quad a_{12} - \frac{\omega_{xy}}{\sigma_x^2} a_{22} \Rightarrow 0.4 - 0.5 * 0.7 = 0.05
\end{aligned}$$

What do we find?

- Data from this DGP is found in the file *ADLfromVAR_d.in7/bn7*.
- In that file *YA* corresponds to y_t above and *YB* corresponds to x_t above.
- We use *PcGive* to check the theoretical results.

Weak exogeneity of X in the conditional model I

OLS gives ML estimates of the parameters of the ADL model

$$y_t = \phi_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t. \tag{20}$$

x_t is therfore weakly exogenous in (20) despite the fact that x_t is an endogenous variable in the VAR:

$$y_t = a_{11} y_{t-1} + a_{12} x_{t-1} + \epsilon_{yt} \tag{21}$$

$$x_t = a_{21} y_{t-1} + a_{22} x_{t-1} + \epsilon_{xt} \tag{22}$$

$$\begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_x^2 & \omega_{xy} \\ \omega_{xy} & \sigma_y^2 \end{pmatrix}\right) \tag{23}$$

Weak exogeneity of X in the conditional model II

- There is a difference between a variable being endogenous in a statistical system like (21)-(23) and being endogenous in a *model* of the statistical system.
- x_t is *weakly exogenous* for the parameters in (20) because we do not gain anything in terms of efficiency by estimating (20) jointly with the marginal equation (22).

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

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

so x_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 I

- 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 x_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 I

In the two variable ADL model

$$y_t = \phi_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t \tag{24}$$

we have that

$$E(y_{t-1}\varepsilon_{t+j}) = 0, \text{and } E(x_{t-1}\varepsilon_{t+j}) = 0 \text{ for } j > 0$$

by conditioning on history of the system, and

$$E(x_t \varepsilon_{t+j}) = 0 \text{ for } j > 0$$

by conditioning on x_t.

Predeterminedness II

Heuristically, we cannot claim strict exogeneity

$$E(y_{t-1}\varepsilon_{t\pm j}) = 0 \text{ for all } j \tag{25}$$

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

- (25) defines y_{t-1} as a pre-determined variable.
- x_t and x_{t-1} are either exogenous or predetermined (depending on Granger causality).
- With pre-determinedness OLS estimators are biased in small samples, but they are remain consistent estimators in stationary systems.
- The size of the bias is seldom very large, and it declines with ϕ_1.

Predeterminedness and misspecification

X_i	Disturbances ε_i are: heteroscedastic		autocorrelated	
X_i	$\hat{\beta}_1$	$\widehat{Var}(\hat{\beta}_1)$	$\hat{\beta}_1$	$\widehat{Var}(\hat{\beta}_1)$
exogenous	unbiased consistent	wrong	unbiased consistent	wrong
predetermined	unbiased consistent	wrong	biased inconsistent	wrong

Strong exogeneity

- In the ADL model

$$y_t = \phi_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t$$

- x_t is a strongly exogenous variable if there is no feed-back from y_{t-1} on x_t.
- x_t is Granger causing y_t, but y_t is **not** Granger causing x_t.
- In the ADL model, x_t is strongly exogenous if $a_{21} = 0$ in the marginal equation for x_t.
- We see that strong exogeneity is easy to test.
- But note that the test involves the system, we learn nothing about strong exogeneity from the conditional equation alone.

Invariance, super exogeneity and autonomy I

- x_t is super exogenous if the parameter of x_t in the conditional model for y_t is invariant to structural breaks in the marginal equation for x_t.
- Super-exogeneity is the property that we have constant parameters in the conditional model even in periods where there is a structural break in the marginal equation for x_t.
- We discussed this concept during the first seminar.
- Super exogeneity is defined for conditional models, but the concept is related to the more general idea of autonomy.
- Econometric models with parameters that are invariant in the face of wide range of structural breaks have a high degree of autonomy.

Testing invariance and super exogeneity and autonomy I

To test a hypothesis of lack of invariance we need to investigate two issues:

1. Test the null hypothesis of no-structural breaks in the marginal model
2. Test the null hypothesis of stability in the conditional model.

We have several tools available:

- Recursive estimation and recursive graphs
- Formal tests of structural breaks,
- Test the significance of dummy variables for structural breaks.