UiO : University of Oslo

ECON 4160: Econometrics-Modelling and Systems Estimation

Lecture 8: Multiple equation models I

Ragnar Nymoen

Department of Economics University of Oslo

2 October 2018

The reference to this lecture is:

- Chapter 7.1-7.6 in the textbook: *Dynamic Econometrics for Empirical Macroeconomic Modelling*.

Notation for VAR-EX

- When we model real world data we often want to condition the *system* on one or more non-modelled random variables,
- and to include certain deterministic variables (constant terms, seasonal variables, dummies for structural breaks).
- Collect the m exogenous variables in the $\mathbf{x}_t$ vector, and the constant term etc in $\mathbf{D}_t$
- Notation for the Open-VAR, also called VAR-EX:

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{y}_{t-i} + \sum_{i=0}^{q} \mathbf{\Gamma}_i \mathbf{x}_{t-i} + \mathbf{Y}\mathbf{D}_t + \varepsilon_t, \quad (1)$$

$$\varepsilon_t \sim IIN(\mathbf{0}, \mathbf{\Sigma}). \quad (2)$$

Conditional plus marginal model

- In Lecture 6 (see Ch. 6.2.1) a main point was that the ADL(1,1) regression model can be interpreted as a valid conditional model of a VAR(1) where Y_t and X_t are endogenous variables.
- In particular, the OLS estimators of all the coefficients of the ADL(1,1) are *consistent* estimators.
- The ADL(2,2) that we looked at in Computer Class, is interpretable as a conditional model of Y_t given X_t, in a VAR(2) with Y_t and X_t as endogenous.

- Chapter 7.3 presents another example:
- VAR-EX with $n = 3, m = 1, p = 1, q = 0$
- If we choose Y_{1t} in $\mathbf{y}_t$ as the focus variable of the investigation, the multi-equation model of the VAR can be written as ((7.5)-(7.7) in the book):

$$
\begin{aligned}
Y_{1t} &= \phi_{0|2,3} + \phi_{11|2,3} Y_{1t-1} + \beta_{10|2,3} Y_{2t} + \beta_{11|2,3} Y_{2t-1} \\
&\quad + \beta_{20|2,3} Y_{3t} + \beta_{21|2,3} Y_{3t-1} + \beta_{40|2,3} X_{1t} + \epsilon_{1|2,3\,t} \qquad (3) \\
Y_{2t} &= \phi_{0|3} + \phi_{21|3} Y_{1t-1} + \phi_{22|3} Y_{2t-1} + \beta_{20|3} Y_{3t} \\
&\quad + \beta_{21|3} Y_{3t-1} + \beta_{40|3} X_{1t} + \epsilon_{2|3\,t} \qquad (4) \\
Y_{3t} &= \phi_{30} + \phi_{31} Y_{1t-1} + \phi_{32} Y_{2t-1} + \phi_{33} Y_{3t-1} + \gamma_{30} X_{1t} + \varepsilon_{3t} \qquad (5)
\end{aligned}
$$

  where the notation for the coefficients is chosen in order to highlight the conditioning.
- Estimation of (3)-(5), by OLS for each equation, gives exactly the same log likelihood as the estimation of the unrestricted VAR-EX.
- The same result that we showed in Lecture 7 for the cobweb model data.

"Conditional plus marginal" and dynamic multipliers

- A model like (3)-(5) is a structured way of estimating the dynamics effects on Y_{1t} of a change X_{1t}.
- Estimating only the ADL model equation (3) would make us miss the "indirect effects" through Y_{2t} and Y_{3t}.
- Easy to estimate (3)-(5), by OLS on each model equation.
- And more interpretability, given the choice to focus on Y_{1t} dynamic multipliers.
- In the next lectie we shall see that (3)-(5) is an example of a *recursive model*, aka SVAR (Structural VAR).

Right or wrong?

"When a right hand side variable in a regression model is endogenous, OLS estimators are inconsistent and must be replaced by IV estimators."

SEM

- Multiply the VAR-EX in (1) from the right ("pre-multiply") by a non-singular matrix $\mathbf{A}$:

$$\mathbf{A}\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{A}\mathbf{\Phi}_i \mathbf{y}_{t-i} + \sum_{i=0}^{q} \mathbf{A}\mathbf{\Gamma}_i \mathbf{x}_{t-i} + \mathbf{AYD}_t + \mathbf{A}\boldsymbol{\varepsilon}_t. \qquad (6)$$

- By normalization, the elements on the diagonal of $\mathbf{A}$ are 1.
- Clearly, if $\mathbf{A} = \mathbf{I}$, the pre-multiplication does not change the VAR-EX.
- But for $\mathbf{A} \neq \mathbf{I}$, the implication is that (6)has different parameters than the VAR-EX, including the contemporaneous covariance matrix of the error-terms:

$$\mathbf{\Omega} = Var(\boldsymbol{\epsilon}_t) = E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = \mathbf{A\Sigma A}' \neq \mathbf{\Sigma},$$

in general.

Simultaneity and recursiveness

- We call the multiple equation model (6) a *Simultaneous Equations Model*, SEM.
- The term is used broadly (in the same way as in PcGive) since it covers the case where **A** is a triangular matrix, in which case there is *no contemporaneous feed-back* in the model.
- Whether **A** is a triangular a matrix or not is however important for the discussion of identification of SEMs, and for the choice of estimation method for identified equations.
- We therefore need to be clear about when we assume an unrestricted **A** (apart from normalization (1s along the diagonal)), and when we chose to use a diagonal **A** (for example through Conditional plus marginal modelling).
- We start with the case of unrestricted **A**, Ch 7.5 in the book.

Identification

- We shall understand the identification issue (or "problem") as the question about the logical possibility of consistent estimation of a model equation's parameters.
- We know from before that the following model types are identified:
  - dynamic conditional regression models,
  - marginal models (AR(p) for example),
  - VARs,
  - VAR-EXs,
  - multiple equations models of the "conditional plus marginal type" (cf. above).
- However, care must be taken before concluding about the identification of equations in a SEM, which we now turn to. Ch. 7.5.

SEM identification example

- Consider (1)-(2) in the simplest possible case where $n = 2$, $m = 0$, and $p = q = 0$.
- Hence, we abstract from dynamics, in order concentrate on the new phenomenon of contemporaneous feed-back or simultaneity.
- No loss of generality: because the identification issue that we now focus, on only arises because of that simultaneity.
- To aid intuition, we can think of a partial equilibrium model of the market for a product, and define Y_1 as quantity, Q, and Y_2 as price P.

SEM unidentification example I

- In matrix notation this model can be written as:

$$\underbrace{\begin{pmatrix} 1 & a_{12} \\ a_{21} & 1 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} Q_t \\ P_t \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix}}_{\mathbf{A}\boldsymbol{\Upsilon}\mathbf{D}_t} + \underbrace{\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}}_{\boldsymbol{\epsilon}_t}.$$

- Written in model equation form, this SEM is:

$$Q_t + a_{12}P_t = a_{10} + \epsilon_{1t} \text{ (demand, } a_{12} > 0) \tag{7}$$

$$a_{21}Q_t + P_t = a_{20} + \epsilon_{2t} \text{ (supply, } a_{21} < 0). \tag{8}$$

- To make a mapping between the 1st and 2nd order moments of the observable variables and SEM parameters, we need the reduced form.

SEM unidentification example II

- Reduced Form:

$$Q_t = \underbrace{\frac{a_{10} - a_{12}a_{20}}{1 - a_{12}a_{21}}}_{\phi_{10}} + \underbrace{\frac{\epsilon_{1t} - a_{12}\epsilon_{2t}}{1 - a_{12}a_{21}}}_{\varepsilon_{1t}}, \tag{9}$$

$$P_t = \underbrace{\frac{a_{20} - a_{21}a_{10}}{1 - a_{12}a_{21}}}_{\phi_{20}} + \underbrace{\frac{\epsilon_{2t} - a_{21}\epsilon_{1t}}{1 - a_{12}a_{21}}}_{\varepsilon_{2t}}, \tag{10}$$

- Following our definition of identification, we assume that we have perfect knowledge of the 5 RF parameters: ϕ_{10}, ϕ_{10} and the three coefficients in $\mathbf{\Sigma}$.
- However, the SEM has $4 + 3 = 7$ coefficients
- The 7 parameters of the SEM cannot be determined from knowledge of the 5 RF coefficients.

SEM undentification example III

- Mathematically, we have a under-determined equation system, between known 5 RF coefficients and 7 unknown SEM coefficients.
- The information we have available can be summarized in two equation:

$$\phi_{10} = \frac{a_{10} - a_{12}a_{20}}{1 - a_{12}a_{21}}, \quad (11)$$

$$\phi_{20} = \frac{a_{20} - a_{21}a_{10}}{1 - a_{12}a_{21}}, \quad (12)$$

  with the known reduced form parameters on the left hand sides of the equations.
- (11)-(12) are two equations in four unknowns.
- Note that knowing the elements of $\boldsymbol{\Sigma}$ does not "help identification", as we also need to determine the elements of $\boldsymbol{\Sigma}$, as long as the matrix is unrestricted.

SEM identification example

- Not all SEMs are unidentified. For example, a theory specify non-modelled variables that shift demand, X_{1t} say, and supply (X_{2t}):

$$Q_t + a_{12}P_t = a_{10} + \gamma_{11}X_{1t} + \epsilon_{1t}, a_{12} > 0, \gamma_{11} \neq 0 \quad (13)$$

$$a_{21}Q_t + P_t = a_{20} + \gamma_{22}X_{2t} + \epsilon_{2t},\ a_{21} < 0, \gamma_{22} \neq 0 \quad (14)$$

- In this SEM, there are 6 unknown parameters: a_{10},a_{12}, γ_{11},a_{20},a_{21} and γ_{22}.
- The RF of (13)-(14) contains six parameters that we can consider as known: Two constant terms and the four reduced form parameters of X_{1t} and X_{2t}.
- We can therefore formulate a determined equation system for the six unknowns $(a_{10}, a_{12}, \gamma_{11}, a_{20}, a_{21}, \gamma_{22})$ and the six known parameters of the reduced form.
- Both equations are identified.

Identified ?

$$Q_t + a_{12}P_t = a_{10} + a_{11,1}Q_{t-1} + \gamma_{11}X_{1t} + \epsilon_{1t}, a_{12} > 0, \gamma_{11} \neq 0$$
$$a_{21}Q_t + P_t = a_{20} + a_{221,1}P_{t-1} + \gamma_{22}X_{2t} + \epsilon_{2t},\ a_{21} < 0, \gamma_{22} \neq 0$$

Order condition

In a SEM model of n linear equations, and with no restrictions on the covariance matrix of the disturbances, to be identified an equation must exclude at least $n-1$ of the variables appearing in the equations of the model.

Rank condition

In a SEM model of n linear equations, and with no restrictions on the covariance matrix of the disturbances, an equation is identified if and only if at least one non-zero $(n-1) \times (n-1)$ determinant is contained in the array of coefficients with which those variables excluded from the equation in question appear in the other equations

- If the rank condition is satisfied, the order condition is automatically satisfied.
- Intuitively the rank condition is satisfied if all the excluded variables really have non-zero coefficients in those equations where they enter.
- The name “Rank” comes from how large the largest invertible matrix with coefficients of excluded variables is.

Alternative form of the conditions I

Write the first equation of a SEM as(almost like in Chapter 2):

$$\mathbf{y}_1 = \mathbf{X}_{1X}\boldsymbol{\beta}_{11} + \mathbf{X}_{1Y}\boldsymbol{\beta}_{21} + \boldsymbol{\epsilon}_1, \tag{15}$$

where:

- $\mathbf{y}_1$ is the $T \times 1$ vector with observations of the Y-variable that the first equation in the SEM is normalized on.
- $\mathbf{X}_{1X}$ is the $T \times k_{1X}$ matrix with observations of the k_{1X} included pre-determined or exogenous variables (incl constant term).
- $\mathbf{X}_{1Y}$ is the $T \times k_{1Y}$ matrix that holds observations of the k_{1Y} included endogenous explanatory variables, minus one.
- Order conditions says:

$$\underbrace{(k_X - (k_{1X} + 1))}_{\text{excluded exogenous}} + \underbrace{(n - k_{1Y})}_{\text{excluded endogenous}} \geq n - 1. \tag{16}$$

Equivalent order condition I

Re-arrange (16) to give:

$$k_X - k_{1X} \geq k_{1Y}. \tag{17}$$

Said with other words:

> *To be an identified equation, the number of excluded exogenous and pre-determined variables from the equation must be not less than the number of included right-hand side endogenous variables in the equation.*

Equivalent order condition II

- So far have expressed everything in terms of "zero restrictions" on the model coefficients
- a zero-restriction is a special case of *linear homogenous restriction*
- The first formulation of the order condition can be extended as follows:

  > *In a n equation linear model, a necessary condition for the identification of a single equation is that there is at least* $n-1$ *independent linear homogenous equations, on the parameters of the equation.*

What about model identities?

- An identity (*eg* general budget in a macro model) is always identified, but should be counted as one of the n equations of the SEM when the identification of the econometric model equations are discussed.
- Another method is to substitute in the identities, then the formulation of the order condition with linear homogenous restrictions must often be used.