# ECON 4160, autumn term 2014. Lecture 6

## Exogeneity

Ragnar Nymoen

Department of Economics

18 Sep 2014

- References HN
  - Ch 10 which introduced exogeneity in the context of independent cross section data and the bivariate normal distribution
  - We will assume that material known
  - Ch 11.4: Hidden under the subchapter title "Congruency" is a very interesting discussion of *invariance* and *super exogeneity*, and how it relates to causaility (see page 168).
  - Ch 14
- References to DM,
  - Ch 8.1-8.3 about IV estimation, since tests of exogeneity make use of IV-estimation (but just need the rudimentary here), we shall return to IV estimation in a separate lecture later)
  - Ch 8.7 in particular, Durbin-Wu-Hausman test
  - Ch 15.3 on the relationship between tests of exogeneity and *encompassing* tests (for non-nested hypotheses).

- Lecture Note 4 and 5 may be of help as well.
- BN2014: Kap 9.

# The VAR and exogeneity I

- We have seen that a VAR (intercepts omitted for simplicity):

$$\underbrace{\begin{pmatrix} Y_t \\ X_t \end{pmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}}_{\mathbf{\Pi}} \underbrace{\begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix}}_{\mathbf{y}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \qquad (1)$$

is a stationary vector process and a stable dynamic system if and only if $\boldsymbol{\varepsilon}_t$ is stationary and the eigenvalues of $\mathbf{\Pi}$ are less than one in magnitude The eigenvalues are the roots of

$$|\mathbf{\Pi} - \lambda \mathbf{I}| = 0 \qquad (2)$$

- Examples of stationary $\boldsymbol{\varepsilon}_t$ processes are
  - $\varepsilon_{yt}$ and $\varepsilon_{xt}$ are stationary AR or ARMA processes

# The VAR and exogeneity II

- Both $\varepsilon_{yt}$ and $\varepsilon_{xt}$ are "white-noise" but possibly correlated in period $t$
- $\boldsymbol{\varepsilon}_t \sim IN(\mathbf{0}, \boldsymbol{\Sigma})$
- For simplicity we subsume the two last in the term Gaussian VAR

- Stationary VARs are "easy to estimate" and use for inference since OLS on each row gives (condtional) MLE
- In each "row regressions" the regressors $Y_{t-1}$ and $X_{t-1}$in a Gaussian VAR are **predetermined:** They are correlated with past disturbances, but not current and future disturbances
- There are no strictly exogenous regressors (meaning independent of prestent, future and past disturbances) in a VAR.

# Exogeneity in a conditional model of VAR I

- In Lecture 5, and Lecture Note 3, we saw that the ADL model

$$Y_t = \phi_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \tag{3}$$

together with the second row in the VAR:

$$X_t = \pi_{21} Y_{t-1} + \pi_{22} X_{t-1} + \varepsilon_{xt} \tag{4}$$

give a regression model representation of the VAR, in terms of a **conditional model** (3) and a **marginal model** (4).

# Exogeneity in a conditional model of VAR II

- HN §14.1 shows the VAR likelihood can be fatorized as

$$L_{VAR} = L_{(3)} \times L_{(4)}$$
$$\Longleftrightarrow$$
$$l_{VAR} = l_{(3)} + l_{(4)}$$

- This means that the full VAR likelihood can be maximized by maximizing the likelihoods of the conditional model seperately.
- This is becasue there are no cross equation restrictions between the parameters of (3) and (4)

# Exogeneity in a conditional model of VAR III

- *Disturbances*: Note that

$$E(\varepsilon_t \mid \varepsilon_{xt}) = 0 \tag{5}$$

by the construction of the model:

$$\varepsilon_t \equiv \varepsilon_{yt} - \frac{\sigma_{xy}}{\sigma_x^2}\varepsilon_{xt} \tag{6}$$

But then (for the Gaussian VAR)

$$Cov(\varepsilon_{t-i}, \varepsilon_{xt-j}) = 0 \quad \text{for all } i \text{ and } j \tag{7}$$

- Consequence: The regressors in (3), including $X_t$, are predetermined.
- If $\pi_{21} = 0$, $X_t$ is also uncorrelated with past $\varepsilon_{t-j}$ disturbances, and $X_t$ is a strictly exogenous regressor (for the gaussian VAR).

# Simultaneous equations model and lack of exogeneity I

- An simultaneous equations model (SEM) of the 2-variable $\mathbf{y}_t$ process is

$$\begin{bmatrix} 1 & b_{12,0} \\ b_{21,0} & 1 \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} b_{11,1} & b_{12,1} \\ b_{21,1} & b_{22,1} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{y,t} \\ \epsilon_{x,t} \end{bmatrix} \tag{8}$$

  where $\epsilon_{y,t}$ and $\epsilon_{x,t}$ are contemporaneously **uncorrelated** Gaussian disturbances $\boldsymbol{\varepsilon}_t \sim IN(\mathbf{0}, \boldsymbol{\Omega})$ where the off-diagonal elements are zero: $\omega_{xy} = \omega_{yx} = 0$,
- If we write out the first row of this SEM we get:

$$Y_t = b_{11,1} Y_{t-1} + b_{12,0} X_t + b_{12,1} X_{t-1} + \epsilon_{y,t} \tag{9}$$

- From of (8) we see that $X_t$ **must be** correlated with $\epsilon_{y,t}$

# Simultaneous equations model and lack of exogeneity II

- In (9) $X_t$ **cannot be an exogenous or predetermined variable**, even if (9) looks like a dynamic regression model.
- The only exception is when $b_{21,0} = 0$ (we shall come back to this later under the heading recursive system)

# Exogeneity paradox I

- We find ourselves in the paradoxial situation that a variable $X_t$ can be "*exogenous*" in one econometric model, but "*not exogenous*" in another econometric model!
- In order to clarify this conundrum, at the conceptual level, modern econometrics distinguishes between different concepts of exogeneity:
- Weak exogeneity (WE)
- Strong exogeneity (StE)
- Super exogeneity (SuE)
- Strict exogeneity or pre-determinedness

# Weak exogeneity I

(1) can be re-parameterized as:

$$Y_t = \phi_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t. \quad (10)$$

$$X_t = \pi_{21} Y_{t-1} + \pi_{22} X_{t-1} + \varepsilon_{xt} \quad (11)$$

where $(\phi_1, \beta_0, \beta_1)$ depend on the parameters of the joint distribution of $Y_t$ and $X_t$ as shown, and $\varepsilon_t$ is derived from the VAR disturbances $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$. See *Lecture note 3* and *HN* § *14.2* for the details.

# Weak exogeneity II

- This model of the VAR corresponds to the factorization of the joint density:

$$f(X_t, Y_t; \theta) = f(Y \mid X_t; \theta_1) \cdot f_x(X_t; \theta_2) \tag{12}$$

where the explicit conditioning on $X_{t-1}$ and $Y_{t-1}$ is omitted to save notation

- Let $\theta$ denote the parameters of the joint density. $\theta_1$and $\theta_2$ are the parameters of conditional and marginal densities.

$$\theta = [\pi_{11}, \ldots, \pi_{22}, \sigma_y^2, \sigma_x^2, \sigma_{xy}]' \tag{13}$$

$$\theta_1 = [\ \phi_1, \beta_0, \beta_1, \ \sigma^2]' \tag{14}$$

$$\theta_2 = [\pi_{21}, \pi_{22}, \sigma_x^2]'. \tag{15}$$

# Weak exogeneity III

- *Weak exogeneity* (WE) is the case where statistically efficient estimation and inference can be achieved by only considering the conditional model and not taking the rest of the system into account.
  - With WE there is no loss of information by abstracting from the marginal model.
  - WE is defined relative to the **parameters of interest**.
  - The parameter of interest can be $\theta$ or a sub-set.

Let $\psi$ denote the vector with parameters of interest. $X_t$ in the conditional model is weakly exogenous if

1. $\psi = g(\theta_1)$, $\psi$ depends functionally on $\theta_1$ and not on $\theta_2$.
2. $\theta_1$ and $\theta_2$ are *variation free* (free to take any values)

# Weak exogeneity IV

Heuristically we will think of 1. as the condition that secures that there is no *direct dependence* of $\psi$ on $\theta_2$ and 2. as a condition that secures that there are no *indirect* (e.g. cross-restrictions) dependence between $\theta_2$ and $\psi$.

## Example

Set $\psi = \beta_0$. $X_t$ is WE because both 1 and 2 is fulfilled.

(In fact, $x_t$ in (10) is WE with respect to the whole vector $\psi = \theta_1 = [\ \phi_1, \beta_0, \beta_1, \sigma^2]'$.)

## Example

If $\psi = (\lambda_1, \lambda_2)$, the eigenvalues of the companion matrix, then $X_t$ is not WE, since $\psi$ is a function of $\pi_{12}$ and $\pi_{22}$ which belongs to $\theta_2$.

# Weak exogeneity V

REMARK:

1. Refering back to HN Chapter 10, we now see that Weak Exogeneity is the time series counterpart to what was called Strong exogeneity for independent cross section data in Ch 10.
2. However, the concept of strong exogenity is already defined in time series econometrics, and we turn to that concept next.

# Strong exogeneity and Granger non-causality I

The purpose of an econometric study is often to find the dynamic effects on one economic variable ($Y_t$) of a change in a variable ($X_t$) "elsewhere in the economy".
As we have seen, these effects can be found as

$$\frac{\partial Y_{t+s}}{\partial X_t}$$

from the solution of (10) for period $t+s$, conditional on period $t$:

$$\begin{aligned} Y_{t+s} = {} & \beta_0 X_{t+s} + (\beta_1 + \phi_1 \beta_0) X_{t+s-1} + \phi_1 (\beta_1 + \phi_1 \beta_0) X_{t+s-2} \\ & + \phi_1^2 (\beta_1 + \phi_1 \beta_0) X_{t+s-3} + \ldots + \phi_1^s Y_t \end{aligned} \quad (16)$$

# Strong exogeneity and Granger non-causality II

$$s = 0,\ \frac{\partial Y_t}{\partial X_t} = \beta_0$$

$$s = 1,\ \frac{\partial Y_{t+1}}{\partial X_t} = (\beta_1 + \phi_1\beta_0)$$

$$s = 2,\ \frac{\partial Y_{t+2}}{\partial X_t} = \phi_1(\beta_1 + \phi_1\beta_0)$$

$$s = j,\ \frac{\partial Y_{t+j}}{\partial X_t} = \phi_1^{j-1}(\beta_1 + \phi_1\beta_0)$$

as long as $Y_t$ is not **Granger-causing** $X_t$, meaning

$$Y_{t-1} \nrightarrow x_t \Longleftrightarrow \pi_{21} = 0 \text{ in } (1)$$

the multipliers give the correct effect on $Y_{t+s}$ of an independent change in $X_t$.

# Strong exogeneity and Granger non-causality III

### Definition (Strong exogeneity)

$X_t$ is strongly exogenous, (StE) if $X_t$ is WE in (10) and $Y_t$ is not Granger-causing $X_t$.

# Super exogeneity (autonomy and invariance) I

- If a change in $\theta_2$ does not affect $\theta_1$ we say that $\theta_1$ is *invariant* or *autonomous* with respect to che change in $\theta_2$.
- Autonomy implies that the parameter $\theta_1$ of the conditional model remains a constant also when the parameter of the marginal model is a non-constant function of time.
- For example $\theta_2$ can be constant over one time period, corresponding to one "regime", and then change to a new level, temporarily, or more permanently. The change can be fast or slow. In such cases we speak of **structural breaks** in the marginal model. The term **intervention** is also common.

# Super exogeneity (autonomy and invariance) II

## Definition

$X_t$ is super exogenous (SuE) in (10) if $X_t$ is WE and the parameters $(\phi_1, \beta_0, \beta_1, \sigma^2)$ are invariant with respect to structural breaks in the marginal model (11).

- For the bivariate normal case ($\phi_1 = \beta_1 = 0$ in ADL) we have that SuE of $X_t$ requires

$$\sigma_{xy} = \beta_0 \sigma_x^2, \tag{17}$$

  since only then can $\beta_0$ be unaffected by changes in $\sigma_x^2$, for example an intervention in the marginal model. Similar conditions is true for ADL and other multivariabe models.
- Note that super-exogeneity does not require strong exogeneity.

# Super exogeneity (autonomy and invariance) III

Further remarks:

- While there is nothing hindering that a condition like (17) *may* hold, there also nothing that "makes it hold".
  - Invariance is a relative concept: A conditional model can have parameters that are super exogeneous with respect to certain interventions structural breaks, but not all.
    - All models break down sooner or later!
  - It it not obvious that all structural breaks (in the marginal model) affect $\beta_0$ or other "derivative coefficients". Might be a strong incidence of structural breaks that mainly affect conditional mean, i.e., the constant term (which we have abstracted from for simplicity her)—Return to that when we discuss forecasting.

# Super exogeneity (autonomy and invariance) IV

- The *Lucas-critique* states that (17) never holds: Policy analysis should never be based on a conditional model—it gives the wrong answer to the question "what happens to $Y_t$ when $X_t$ is changed?"
  - See Lecture Note 5 about the Lucas-critique
- If the conditional model does not have super exogenous variables, it may well be that another parameterization, i.e., another econometric model of the VAR has parameters that are invariant. This is the constructive part of the Lucas' critique:
  - Estimate models where the parameters of interest are coefficients of variables that are subject to rational expectations

# Super exogeneity (autonomy and invariance) V

  - These coefficients will (according to this theory) be "*deep structural parameters*" and will have a high degree of invariance.
- We understand that invariance is a more general property than SuE, which only to conditional models,
- Invariance of the parameters of a structural equation with respect to structural breaks elsewhere in the economic system is a desireable property of any econometric model of parts of the system.

# Strict exogeneity and pre-determinedness I

- WE, StE and SuE are different from the older concepts of exogeneity in that they are defined **relative to the purpose of the econometric model** and also relative to parameters of interest.
- For reference, those older concepts that we are now fell familiar with are:
- *Strict exogeneity* (disturbances uncorrelated with *any* randomness in the DGP that generated $X_t$)
- and the *pre-determinedness* secured by sequential conditioning (the work-horse of time series econometrics)
- One reconciliation of views may be that in several situations it pays off to be clear about *parameters of interest*—as the Lucas critique shows:

# Strict exogeneity and pre-determinedness II

- If the parameters of interest is given by the rational expectations model then $X_t$ cannot be weakly exogenous
- Even if $X_t$ is predetermined in the condition model

# Testing exogeneity—overview I

- Weak exogeneity.
  - In the case of stationary time series one could say that WE is implied by the model specification:
  - If the parameters of interest are "in" the conditional model (ADL for example), then the variables of the model are WE
  - That said, the a well known exogeneity test like the *Durbin-Wu-Hausman* test (DM Ch 8.7) can be interpreted as a test of WE (see below)
  - In the case of non-stationary time series modelling WE is testable more generally (hope to have time to return to this under cointegration).
- *Strong exogeneity*
  - Granger non-causality is testable in a stationary VAR

# Testing exogeneity—overview II

- *Super exogeneity*
  - Lack of invariance with respect to structural breaks (interventions) that have occurred in the sample is a testable hypothesis. We will see specific examples later.
  - When the model "under test" is a conditional model, these invariance tests are tests of super exogeneity.
  - But invariance tests are also relevant for the parameters in an equations in a simultaneous equation model, and other deep structural parameters (Euler equations for consumption , NPC for inflation).
  - Given so called overidentification—testing is possible and the statistics have power but this is for coming lectures and Computer classes

# Durbin-Wu-Hausman test I

- The DWH test is presented in section 8.7 in DM, not covered by HN(?)
- In line with the original motivation of the test, the exposition is in terms of the difference between two MM estimators of $\boldsymbol{\beta}$ in

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{18}$$

- one is the OLS estimator $\hat{\boldsymbol{\beta}}_{OLS}$that we know well and the other is the Instrumental Variables, IV estimator $\hat{\boldsymbol{\beta}}_{IV}$ that many of you will have seen examples of in your introductory course in econometrics. (But since we have not covered it yet, this part is mainly for reference)

# Durbin-Wu-Hausman test II

- The test situation here is

$$H_0\text{: } plim(\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{OLS}) = 0 \text{ against } H_1 : plim(\hat{\boldsymbol{\beta}}_{IV} - \hat{\boldsymbol{\beta}}_{OLS}) \neq 0$$

- Without knowing too much about IV-estimation yet, we can ask: Where should any significant difference between $\hat{\boldsymbol{\beta}}_{OLS}$ and another MM estimator $\hat{\boldsymbol{\beta}}_{IV}$ come from?
- The answer must be: from the rest of the system! Or in other words, form the marginal model of the variables in $\mathbf{X}$ in (18).
- Intuitively we can therefore test $H_0$ by estimating the marginal model for $\mathbf{X}$ by OLS, calculate fitted value of $\mathbf{X}$ from that marginal model and then testing if those predicted values are significant when added to the original model (18) as additional regressors.

# Durbin-Wu-Hausman test III

- This is the interpretation of the algebra on page 340-341 of DM where $\mathbf{P_W}$ is a **prediction-maker** (orthogonal projection) matrix like in Lecture note 1 (DM Ch 1 and 2) but from the marginal model not (18). That's why $\mathbf{P_W}$ is in terms of a $\mathbf{W}$ matrix with instruments (not $\mathbf{X}$).

- Since the matrix $\mathbf{W}$ is a **residual-maker** which is orthogonal to $\mathbf{P_W}$, another way of implementing the test is to add the residuals from the marginal models to the regression and test of they are significant.
- In either version, the interpretation of a significant test outcome is that the marginal model contains information about $\beta$, meaning that **Weak-exogeneity** is rejected.

# Durbin-Wu-Hausman test IV

- In practice, the test is an OLS based **F-test** where the first degree of freedom is the number of "suspected" endogenous explanatory variables in (18).
- If we want we can interpret it as a LM-test since we only estimate the model under the null hypothesis of WE.
- Example in class.

# Automatizised Gets I

- Data based model development is often met with negative attitudes.
  - "Data-mining", "garbage-in, garbage-out"
- An undisputed basis for scepticism is the cumulation of Type-1 error probabilities, as a result of repeated testing that we discussed earlier, in §7.6 and § 9.5 in HN.
- Nevertheless, as noted in the important and novel chapter 19 in HN, there are now computer programs that performs General-to-Specific (Gets) modelling in an automatic way.
- Oxmetrics includes one such program which is called *Autometrics*.

# Automatizised Gets II

- The aim of Autometrics is to search for relevant variables with a high probability of success, at low *cost of search* (low probability of including many irrelevant variables).
- After reading Chapter 19 in HN you are able to start testing out Autometrics and see if it is useful for you.
- See the PcGive Vol I book for tutorial and documentation.
- We will also check it out in computer classes.