ECON 4160, 2015. Lecture 3

Exogeneity; Empirical models; and Introducing Time Series

Ragnar Nymoen

University of Oslo

3 September 2015

References to Lecture 3

- HN Ch 10,11,12
- DM 13.1-13.2
- (BN2011: kap 10, HN2014: kap 6.1-6.2;7.1-7.2)

Exogeneity I

- So far, our regression models (for cross section data) have been specified by a set of assumptions about the conditional distribution function of Y given k regressors.
- As noted, there is nothing "invalid" or "inferior" about a conditional distribution function! It is just as valid as a simultaneous distribution function or a marginal pdf.
- Importantly, in practical modelling, it is often easier to be formulate (realistic) assumptions about the conditional distribution of Y given $\mathbf{X}$,than to specify the simultaneous probability density function (pdf) of all the $(k+1)$ random variables in our project.

Exogeneity II

- If the parameters of the condition model (the expectation function $E(Y \mid \mathbf{X})$) and the conditional variance function $Var(Y \mid \mathbf{X})$) are relevant for our research purpose (formally the *parameters of interest*), then we can work happily with regression models and the maximum likelihood estimation (MLE)n and inference methods (following the first 9 chapters in HN)
- At least up to a point.
- That point is when we ask the following question: "Do we loose information about the parameters of interest by only estimating the conditional for Y given $\mathbf{X}$, and not also the marginal model for $\mathbf{X}$?"

Exogeneity III

- If we answer "no, we do not loose useful information" we have the case of exogeneity of the explanatory variables. If the answer is "yes" we concede that there is a loss of information, and that $\mathbf{X}$ is not exogenous.
- In Chapter 10 in HN there is a precise definition of exogeneity with reference to models of cross section data. We will use that as reference when we get to dynamic models.

Exogeneity defined I

- Heuristically, a variable X_i is strongly exogenous with respect to the parameters of interest if they can be efficiently estimated without taking into account the marginal distribution (marginal model) of X_i. (I continue to use X for regressor even though HN change to Z at this point).
- Without loss of generality, we can make the idea precise by considering variable pairs (Y_i, X_i) $i = 1, 2, \ldots, n$ that satisfy the IID assumption.
- Let ψ ("psi") denote the parameters of the conditional pdf $f_\psi(y \mid x)$ and let λ denote the parameters of the marginal pdf $f_\lambda(x)$.

Exogeneity defined II

- By definition:

$$f_{\psi,\lambda}(y,x) = f_{\psi}(y \mid x) f_{\lambda}(x) \tag{1}$$

- But it **not always** true that the MLE that we obtain for ψ by maximizing the conditional likelihood function, is identical to the MLE of ψ that we get from maximization of the joint likelihood function based on $f_{\psi,\lambda}(y,x)$.
- When the conditional MLE of ψ is equal to the unconditional MLE of ψ, X is exogenous.
- This is the same as observing equality between the maximum of the simultaneous likelihood function, and the product of two separately maximized likelihoods: The conditional for Y given X and the marginal for X. See HN page 141 for notation.

Exogeneity defined III

- A sufficient condition for exogeneity, is that both ψ and λ can take all values within their respective parameter spaces.
  - There are no cross restrictions that link ψ to λ, (like ψ to λ).
  - Both ψ to λ are "free-to-vary".

Inference in the presence of exogeneity

- For the cross section models that we have reviewed, the correctness of the model specification and exogeneity goes hand in hand.
- Otherwise the BLUE theorem for regression models would not have been true!
- When we get to dynamic regression model, we will see that this aspect of exogeneity does not necessarily apply (although it can)

- Use for example Ch 10.2 in HN to solve Question A in Seminar exercise 2 to formulate the normal (also called gaussian) regression model:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \tag{2}$$
$$X_i = \mu_x + \varepsilon_{x_i} \tag{3}$$

with

$$\boldsymbol{\psi} = (\beta_1, \beta_2, \sigma^2), \;\; \boldsymbol{\lambda} = (\mu_x, \sigma^2_{xx})$$

where $\sigma^2 = Var(Y_i \mid X_i)$, $\mu_x = E(X), \sigma^2_{xx} = Var(X_i)$ and where ε_i and ε_{x_i} and bivariate normal and independent (and uncorrelated).

- In the model (2)-(3) X_i is strongly exogenous since the marginal and conditional parameters in $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ are allowed so vary freely.
- See HN § 10.3.1. for a more detailed argument.

Empirical models I

- Ch 11 in HN: "Empirical models and modelling"
- The chapter discusses the implications of "the fact of life" that the Data Generating Process (DGP) is in practice unknown to us as econometricians.
- First of all: Estimated econometric models are **empirical models** which have properties that depend on the observed data.
- For example: If the variance of the Y is clearly heteroskedastic, the model residuals will become manifestly heteroskedastic unless we account for that variation in variance in the model. If we fail in our empirical modelling, estimation and testing based on the assumptions of the IID regression model may not be reliable.

Empirical models II

- Ch 11.3 is an interesting discussion of four different interpretations of *linear single equation models*:

  1. Regression, meaning conditional expectations.
     Although the conditional expectation function exists under weak assumptions about the joint pdf, it can only be derived as a linear function in a few special cases (the normal, as given above, is the best known, as).
     Hence the specification of the functional form is an important modelling task.
  2. Linear least squares approximation
  3. Contingent plan
     A plan which is implemented by an economic agent after observing the outcome of the conditioning variable.
     Contingent plan models can therefore be estimated at least consistently by OLS

Empirical models III

4. Behavioural model (could have used "expectations model")
   In this case, the agents' plan depends on expectations, X_i^e, about the variable X_i. OLS estimation can give **biased estimates** to the parameters of this type of model. It depends on how expectations are formed.
   For the past 25 years the most popular of this type is the Rational Expectations model, and we therefore return to it under the *Lucas-critique* of OLS estimated time series models later.

Note that our Lecture note 1 introduced the point about linear econometric equation being subject to interpretation. It can be regression equation. But not always.

Congruence and Encompassing (ch 11.4 and 11.5) I

- From the premise that the true DGP is complex and unknown to us, at least two implications follow
  - It is non-trivial to match the theoretical framework to the observations. An empirical model that achieves that aim is called a congruent model. Mis-specification testing is necessary to test for congruency.
  - it follows that there can exist two or more **competing models** of the same variable.
- *Encompassing* has been developed as a concept and research ideal to tackle that kind of situation.
- Encompassing means literary: “putting a fence around”.

Congruence and Encompassing (ch 11.4 and 11.5) II

- In econometrics, this entails that if there is an existing (incumbent) empirical model **A** of the variable Y and you build a new model **B**, then your model M_B should explain the results and properties of M_A:$M_B \ \mathcal{E} \ M_A$.
- Parsimonious encompassing (explaining more by less) gives the most value-added to you model.
- There is a handful of formal tests of encompassing. The simplest is to form the union model of M_A and M_B and then use the LR test (in Chi-square or F-test form) to test the two sets of restrictions.

Congruence and Encompassing (ch 11.4 and 11.5) III

- Simple regression example:

$$M_A : Y_i = \beta_1 + \beta_2 X_i + \varepsilon_{Ai}$$
$$M_B : Y_i = \gamma_1 + \gamma_2 Z_i + \varepsilon_{Bi}$$

then the *nesting-model* (M_0) is simply

$$M_0 : Y_i = \lambda_1 + \lambda_2 X_i + \lambda_3 Z_i + \varepsilon_{0i}$$

Time series I

- We define a time series Y_t as the realization of a *stochastic process* $\{Y_t; t \epsilon\ T\}$. In any period t the variable Y_t can take a number of values consistent with the the sample space. (Norwegian: "utfallsrom").
- A stochastic process has therefore a random distribution for each Y_t. It is consistent with this definition that T can be $\{0, \pm 1, \pm 2, \ldots\}$, $\{1, 2, 3, \ldots\}$, $[0, \infty\}$ or $(-\infty, \infty)$.
- When there is no room for misunderstanding, we follow convention and use the term *time series* both for a data series, and for the process of which it is a realization.
- We will begin by examining the autoregressvie model of order one: AR(1), the simplest dynamic process with properties that carry over to more general/complicated models.

Time series II

- Time series variables are often heavily correlated, so the independency assumption of cross section data needs to be replaced by something else.
- What we use s the assumption of *conditional independence*, where the conditioning is on the history of the time series.
- We make this idea clearer by explaining the concept of Markov process

Markov proceses I

- The population correlation of Y_t and Y_{t-2} given Y_{t-1} is zero if
$$f(y_t, y_{t-2} \mid y_{t-1}) = f(y_t \mid y_{t-1}) \cdot f(y_{t-2} \mid y_{t-1}) \tag{4}$$
which is called conditional independence (cf. §7.3 in HN)
- An important implication is that the conditional distribution of Y_t given (Y_{t-1}, Y_{t-2}) does not depend on Y_{t-2}:
Start with the general decomposition

$$\underbrace{f(y_t, y_{t-2} \mid y_{t-1})}_{\text{joint pdf}} = \underbrace{f(y_t \mid y_{t-2}, y_{t-1})}_{\text{cond pdf}} \cdot \underbrace{f(y_{t-2} \mid y_{t-1})}_{\text{marg pdf}} \tag{5}$$

Markov proceses II

and equate the right hand sides of (4) and (5):

$$f(y_t \mid y_{t-2}, y_{t-1}) f(y_{t-2} \mid y_{t-1}) = f(y_t \mid y_{t-1}) \cdot f(y_{t-2} \mid y_{t-1})$$

$$\Downarrow$$

$$f(y_t \mid y_{t-2}, y_{t-1}) = f(y_t \mid y_{t-1}) \tag{6}$$

- If the more general property holds that the conditional density of Y_t given the entire past of Y_{t-1}, Y_{t-2},..., depends only on Y_{t-1} the time series is a said to be a first order Markov process.

The autoregressive model of order 1, AR(1) I

The statistical model is:

1. *Conditional independence*: $(Y_t \mid Y_0, Y_1, \ldots, Y_{t-1}) \overset{D}{=} (Y_t \mid Y_{t-1})$;
2. *Conditional distribution*: $Y_t \overset{D}{=} N(\phi_0 + \phi_1 Y_{t-1}, \sigma^2)$;
3. *Parameter space*: γ_0, γ_1,$\sigma^2 \in \mathbb{R}^2 \times \mathbb{R}_+$

The *model equation* is

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t, \text{ for } t = 1, \ldots, T \tag{7}$$

The autoregressive model of order 1, AR(1) II

- ε_t are called innovations: Conditionally on past vales of Y_t, the time series variables ε_1,$\varepsilon_2, \ldots, \varepsilon_T$ are independent, and the IID conditional distributions are:

$$\varepsilon_t \overset{D}{=} N(0, \sigma_\varepsilon^2) \ \forall \ t \tag{8}$$

- Another name is a gaussian white-noise process. And a slightly weaker assumption is that ε_t is white-noise ($E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2 \ \forall \ t$, $Cov(\varepsilon_t, \varepsilon_{t-j}) = 0$, $j = 1, 2, \ldots$).
- Y_0 is called the initial condition. For simplicity we will treat it as a fixed parameter (treating it as a random variable only complicates the formalities, does not matter much with even a moderate sample size).

The autoregressive model of order 1, AR(1) III

- For reasons that will become clear we will restrict the parameter ϕ_1 in the following way:

$$-1 < \phi_1 < 1 \tag{9}$$

- The case of $\phi_1 = 1$ we will later get to know as the Random Walk Model (with drift if $\phi_1 \neq 0$).

,

The Autoregressive Likelihood I

Step 1: Factorization of the joint pdf (general, no assumptions used):

$$
\begin{aligned}
f(y_T, \ldots, y_1 \mid y_0) &= f(y_T \mid y_{T-1}, \ldots, y_1, y_0) f(y_{T-1}, \ldots, y_1 \mid y_0) \\
&= f(y_T \mid y_{T-1}, \ldots, y_1, y_0) f(y_{T-1} \mid y_{T-2}, \ldots, y_1, y_0) f(y_{T-2}, \ldots \\
&\vdots \\
&= \textstyle\prod_{t=1}^{T} f(y_t \mid y_{t-1}, \ldots, y_1, y_0) \qquad (10)
\end{aligned}
$$

Step 2: Use the Markov-property assumption of the model:

$$
f(y_T, \ldots, y_1 \mid y_0) = \textstyle\prod_{t=1}^{T} f(y_t \mid y_{t-1}) \qquad (11)
$$

The Autoregressive Likelihood II

Step 3: Apply the normality assumption of the model to $f(y_t \mid y_{t-1})$ to obtain the log likelihood function

$$l(\phi_0, \phi_1, \sigma^2) = -\frac{T}{2}(\ln(2\pi/\sigma_\varepsilon^2)) - \sum_{t=1}^{T} \frac{[Y_t - \phi_0 - \phi_1 Y_{t-1}]^2}{2\sigma_\varepsilon^2}. \tag{12}$$

which is the conditional log likelihood function given Y_0.

ML estimation I

- By direct inspection of (12) we see the MLEs for ϕ_1 and ϕ_0 are the least squares estimators:

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} Y_t (Y_{t-1} - \bar{Y}_{(-)})}{\sum_{s=1}^{T} (Y_{s-1} - \bar{Y}_{(-)})^2} \tag{13}$$

where $\bar{Y}_{(-)} = T^{-1} \sum_{t=1}^{T} Y_{t-1}$ and

$$\hat{\phi}_0 = \bar{Y} - \hat{\phi}_1 \bar{Y}_{(-)} \tag{14}$$

with $\bar{Y} = T^{-1} \sum_{t=1}^{T} Y_t$.

ML estimation II

- Furthermore

$$\hat{\sigma}_{\varepsilon}^2 = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 = T^{-1}\sum_{t=1}^{T}[\underbrace{Y_t - \hat{\phi}_0 - \hat{\phi}_1 Y_{t-1}}_{\hat{\varepsilon}_t}]^2 \qquad (15)$$

- We focus on the MLE of ϕ_1.
  DIY exercise: 3.1 Write $\hat{\phi}_1$as "true parameter plus bias" in the usual way:

$$\hat{\phi}_1 = \phi_1 + \frac{\sum_{t=1}^{T}\varepsilon_t(Y_{t-1} - \bar{Y}_{(-)})}{\sum_{s=1}^{T}(Y_{s-1} - \bar{Y}_{(-)})^2} \qquad (16)$$

- But we cannot prove unbiasedness ($E(\hat{\phi}_1 - \phi_1) = 0$) by using the law of iterated expectations any longer.

ML estimation III

- The conditioning variable in the model, Y_{t-1}, is not "independent enough".
- In fact, it seems plausible that there is going to be a finite sample bias: $\hat{\phi}_1 - \phi_1 > 0$ for a given T.
- But how large is the bias? And, can we claim consistency? How will t-ratios be affected?
- In order to answer these questions we use Monte Carlo in Computer Class.
- And, we will go a little bit into the solution to the difference equation (7) and some generalizations.

Solution of AR(1) I

DIY exercise: 3.2. Use repeated substitution (backward) in

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t, \tag{17}$$

to obtain

$$Y_t = \phi_0 \sum_{i=0}^{t-1} \phi_1^i + \phi_1^t Y_0 + \sum_{i=0}^{t-1} \phi_2^i \varepsilon_{t-i} \tag{18}$$

DIY exercise: 3.3. Prove that (18) is a solution by inserting the solution expressions for Y_t and Y_{t-1} in (17) and show that you get an identity.

- The solution is a function of t, the whole sequence ε_t,ε_{t-1},... ε_1 and the *initial condition* Y_0.

Solution of AR(1) II

- The mathematical *solution* does not require $\varepsilon_t \overset{D}{=} N0, \sigma_\varepsilon^2)$ or any other specific distributional assumption for ε_t. But the statistical properties of Y_t given by the solution will depend on distributional assumptions..

Note also that the *solution* (18) does not depend on the assumption $|\phi_1| < 1$. But ϕ_1 is nevertheless essential both for the nature of the solution, as stable, unstable or explosive, and for the statistical properties of the solution. Here be briefly look at the importance of ϕ_1 for the stability of $Y_t \sim$AR(1).

Expectation I

- From (18) and (8)

$$E(Y_t \mid Y_0) = \phi_0 \sum_{i=0}^{t-1} \phi_1^i + \phi_1^t Y_0 \tag{19}$$

- To find the unconditional expectation $E(Y_t)$ we apply $t \to \infty$.
- Now see why $-1 < \phi_1 < 1$ is essential, because only then does $t \to \infty$ give

$$E(Y_t \mid Y_0) \underset{t\to\infty}{\to} E(Y_t) = \frac{\phi_0}{1-\phi_1} =: \mu_Y \tag{20}$$

since the geometric progression $\sum_{i=0}^{t-1} \phi_1^i \underset{t\to\infty}{=} \frac{1}{1-\phi_1}$ and $\phi_1^t \underset{t\to\infty}{=} 0$.

Expectation II

- Memo: (19) can be written as

$$E(Y_t \mid Y_0) = \frac{\phi_0(1-\phi_1^t)}{(1-\phi_1)} + \phi_1^t Y_0$$

by using the formulae for the first t elements a geometric progression.

Variance, autocovariance and autocorrelation (acf) I

- The variance of Y_t

$$Var(Y_t \mid Y_0) = \sum_{i=0}^{t-1} \phi_1^{2i} \sigma_\varepsilon^2 \tag{21}$$

$$Var(Y_t) = \frac{\sigma_\varepsilon^2}{1-\phi_1^2} \tag{22}$$

which also requires: $-1 < \phi_1 < 1$.

Variance, autocovariance and autocorrelation (acf) II

- The first **autocovariance** is defined as

$$Cov(Y_t, Y_{t-1}) =: E\left[(Y_t - \mu)(Y_{t-1} - \mu)\right] \tag{23}$$

Show (as **DIY Exercise 3.4**) that

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t,$$

can be re-parameterized as

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \varepsilon_t \text{ iff } -1 < \phi_1 < 1 \tag{24}$$

But then

$$(Y_t - \mu)(Y_{t-1} - \mu) = \phi_1(Y_{t-1} - \mu)^2 + \varepsilon_t(Y_{t-1} - \mu)$$

Variance, autocovariance and autocorrelation (acf) III

and taking expectations on both sides gives:

$$Cov(Y_t, Y_{t-1}) = E\left[(\phi_1(Y_{t-1} - \mu)^2\right] = \phi_1 Var(Y_t)$$

Hence, when $-1 < \phi_1 < 1$ holds, the covariance between Y_t and Y_{t-1} does not depend on time itself.

- For Y_t and Y_{t-2} we obtain

$$Cov(Y_t, Y_{t-2}) = \phi_1^2 Var(Y_t)$$

also independent of t. It is the **time difference** between Y_t and Y_{t-2} that matters.

- And generally (still assuming $-1 < \phi_1 < 1$)

$$Cov(Y_t, Y_{t-j}) = \phi_1^j Var(Y_t)\ ,\ j = 1, 2, \ldots$$

Variance, autocovariance and autocorrelation (acf) IV

- The theoretical **autocorrelation function** (ACF) is defined as :

$$\zeta_{j,t} =: \frac{Cov(Y_t, Y_{t-j})}{Var(Y_t)}, \tag{25}$$

- For the case of AR(1) with $-1 < \phi_1 < 1$, the ACF is only a function of j:

$$\zeta_j = \phi_1^j \text{ for } j = 1, 2, ... \tag{26}$$

Second order dynamics AR(2) I

- By a generalization of the above theory,autoregressive models with higher order dynamics can be defined
- For example the AR(2) process with gaussion (normal disturbances)

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{27}$$

$$\varepsilon_t \overset{D}{=} N(0, \sigma_\varepsilon^2) \ \forall \ t \tag{28}$$

- We can guess that the magnitude of the parameters ϕ_1 and ϕ_2 are important for the solution for Y_t in this more general model
- That turns out to be true, and we will present the theory in the next lecture.

Second order dynamics AR(2) II

- But now, we can use the computer to study the solution of the homogenous equation corresponding to (27)

$$Y_t^h = \phi_1 Y_{t-1}^h + \phi_2 Y_{t-2}^h \tag{29}$$

as well as the full solution of the non-homogenous difference equation (27).

- From mathematics we have that the solution of (29) depends on the roots of the associated characteristic polynomial

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0 \tag{30}$$

where λ denotes a root.

Second order dynamics AR(2) III

- Memo: For AR(1) the characteristic polynomial is ("real root")

$$\lambda - \phi_1 = 0$$

  which gives a single root which is a real number
- For AR(2) there can be to two real roots, or two roots that are complex numbers.
- The complex roots have the same real part (the norm) and is therefore called a complex pair.

AR(2) example I

$\phi_0 = 0$, $\phi_1 = 1.6$, $\phi_2 = -0.9$:

$$Y_t = 1.6Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t, \tag{31}$$

- In this case the two roots are a complex numbers. The norm (modulus) of the two roots is 0.94868. This number is comparable to the absolute value (magnitude) of a real root.
- Show graph of homogenous solution,
  and the particular (full) solution when $\varepsilon_t \sim IID(0, 1)$ in class