Introduction to Structural Econometrics

Class 1: What is Structural Econometrics?

Guillaume Wilemme

Aix-Marseille School of Economics - PhD course

Sept 2017

Table of contents

Plan

Schedule

- October 3rd, room 23, 10-12am
  What is Structural Econometrics?
- October 10th, room 23, 10-12am
  Dynamic Programming and Recursive Models
- October 24th, room 23, 10-12am
  Monte-Carlo Methods
- November 7th, room 23, 10-12am
  Simulation-based Estimation
- November 14th, room 15, 10-12am
  State-Space Representation and Latent Variables
- November 21st, room 16, 10-12am
  Numerical tools

Resources

This course will come with

- a pdf file (that I am writing)
- slides for each course
- two .R files per course
  - one as a sand box
  - one as a complete script

Evaluation

- Choose an "empirical issue" that
  1. leads to an estimation bias due to mispecification
  2. e.g. heterogeneous effects, omitted variables, non-linearities, self-selection, simultaneity
  3. I can help you find a subject related to your research
  4. It can be based upon a research article
- Build a pedagogical .R file with comments to illustrate the issue
- The objective is to apply something seen in class
- I will try to be more precise next time

Structural Econometrics (in short)

- Econometrics based on explicit *economic theory*
- useful to solve problems of *endogeneity*
  1. note: endogeneity is by definition related to the unobserved but guessed structure
  2. e.g. endogeneous omitted variable bias
- Linear regressions are sometimes structural econometrics

Plan

The Cowles Foundation

1. The Commission was crerated by Alfred Cowles in 1932 to promote the use of statistics in economics.
   - It was located at the University of Chicago from 1939 to 1955.
   - Koopmans led the Cowles Commission (Foundation) when it moves from Chicago to Yale in 1955.
   - 2 reasons behind this move: i) Tobin was at Yale, ii) conflict with the Department of Economics.
2. Koopmans argues that there is no measurement without theory. The motto of the Cowles Foundation is "Theory and Measurement" (before it was "Science is measurement").
3. The Foundation is related to the Econometric Society (that publishes *Econometrica*).

Simultaneous linear equations

- Consider typically a supply-demand system
- Koopmans introduces the term "structural equations"
- For instance:

$$p = \beta_1 q + \alpha_1' x + \epsilon_1, \qquad \text{(Supply)}$$
$$p = -\beta_2 q + \alpha_2' x + \epsilon_2. \qquad \text{(Demand)}$$

- Draw a graph!
- How to estimate the parameters?

Simultaneous linear equations (II)

- Koopmans introduces the term "reduced-form equations"
- Write the endogenous variable as the solutions of the system,

$$p = \left(\frac{\beta_2}{\beta_1 + \beta_2} \alpha_1 + \frac{\beta_1}{\beta_1 + \beta_2} \alpha_2 \right)' x + \left(\frac{\beta_2}{\beta_1 + \beta_2} \epsilon_1 + \frac{\beta_1}{\beta_1 + \beta_2} \epsilon_2 \right), \tag{1}$$

$$q = \left(\frac{1}{\beta_1 + \beta_2} \alpha_1 - \frac{1}{\beta_1 + \beta_2} \alpha_2 \right)' x + \left(\frac{1}{\beta_1 + \beta_2} \epsilon_1 - \frac{1}{\beta_1 + \beta_2} \epsilon_2 \right). \tag{2}$$

- q is correlated to the error terms, you cannot run the structural equations as a regression
- One needs more assumptions to estimate β_1 and β_2, identification problem

The Lucas critique

- In 1976, Lucas questioned the problem of driving policy implications from time series
- He claims that it is necessary to know the economic structure, defined as what remains stable over time
- For instance, econometrician have established a robust empirical relationship between inflation and unemployment, the Philipps curve
- If a government wants to decrease unemployment, then why shouldn't it increase long-run inflation?

The Lucas critique (II)

- Such a policy is likely to affect inflation expectations, and therefore would be inefficient
- There is a dynamic structure,

$$y_{t+1} = F(y_t, x_t, \theta, \epsilon_t), \tag{3}$$

where F is known and θ fixed.

- If x_t can be considered as exogenous, one can estimate θ to derive policy implications. However, x_t may be endogenous ("optimal decisions of agents"). The true model might actually be

$$y_{t+1} = F(y_t, x_t, \theta(\lambda), \epsilon_t) \tag{4}$$

$$x_t = G(y_t, \lambda, \eta_t). \tag{5}$$

McFadden

- At the time McFadden wrote his first papers (1965-1975), economic theory was tested empirically only through representative-agent models.
- He made popular the use of microeconometric data to study individual discrete choices, using a revealed preferences argument (in particular the multi-nomial logit model).
- Assume there are multiple discrete choices j from 1 to J, with characteristics x_j.
  - The choice of several individuals i is observed, $\mathcal{J}(i)$.
  - Possibly, workers may have different characteristics s_i.
- How can you predict the choice probabilities of an individual (with or without theory)?
- We look for a mathematical relationship to relate x_j and s_i to the realized choice $\mathcal{J}(i)$

The multinomial logit model

- The multinomial logit is a good candidate,

$$\mathbb{P}_i(\mathcal{J} = j) = \frac{e^{F(x_j, s_i, \theta)}}{\sum_k e^{F(x_k, s_i, \theta)}}, \tag{6}$$

- McFadden's contribution is to bridge the gap between this ad-hoc econometric structure and an economic model (the random-utility model)
  - He proposed a full axiomatization of this modelling,
  - meaning the necessary and sufficient conditions (assumptions on the individual's behavior) for a RUM to have a MNL form
- Axiomatization enables the hypothetico-deductive reasoning
  - If the model does not fit the data, then one can point one of the assumptions

Heckman according to Heckman, the Nobel lecture

In the 1970's, Heckman "developed index models of potential outcomes to unite Cowles econometrics and discrete choice theory as well as to unify the disjointed and scattered literature on sample selection, truncation and limited dependent variables that characterizes the literature of the day" (see his Nobel lecture).

Heckman 1974, self-selection in participation

- Consider the model

$$\ln(W) = \alpha h + \beta' X + \epsilon, \qquad \text{(LaborSupply)}$$
$$\ln(W) = \gamma' Z + \mu. \qquad \text{(MarketWage)}$$

- The first equation is a labor supply equation relating h to a wage W.
- The second equation determines the wage (Z contains education and experience for instance).
- ϵ and μ are not necessarily correlated.

Heckman 1974, self-selection in participation (II)

- We can write the reduced-form equations

$$h = \frac{\gamma' Z - \beta' X + \mu - \epsilon}{\alpha}, \tag{7}$$

$$\ln(W) = \gamma' Z + \mu, \tag{8}$$

  if α is not zero.

- we only observes (h_i, W_i) for those who decided to work:

$$\beta' X_i + \epsilon_i < \gamma' Z_i + \mu_i$$

  .

- *In the sample observed*, there is an endogeneity issue like $\mathbb{E}(Z\mu) \neq 0$ which prevents us using an OLS method.
- draw a graph

Heterogeneity, Deaton's example

- We are interested in the impact of building railways on poverty in Chinese cities. Consider the econometric model:

$$\text{Poverty}_c = \gamma + \theta \text{Railways}_c + v_c, \tag{9}$$

$$\text{Railways}_c = \phi + \beta Z_c + \eta_c. \tag{10}$$

- Z_c is a proper instrument, like being randomly selected by the Chinese government.
- Imagine that, in fact, θ varies other cities
  - This is why the estimate interprets as the *local average treatment effect* (under conditions), which differs from the *average treatment effect*
  - Is it useful? External validity

More on Deaton's example

- Suppose *Railways* and Z are dummy variables
- Denote $\bar{\theta} = \mathbb{E}(\theta_c | \mathit{Railways}_c = 1)$, the LATE
- The econometric model rewrites

$$\text{Poverty}_c = \gamma + \bar{\theta}\text{Railways}_c + \left[(\theta_c - \bar{\theta})\text{Railways}_c + v_c\right], \quad (11)$$

$$\text{Railways}_c = \phi + \beta Z_c + \eta_c. \quad (12)$$

- You can show that being designated randomly is exogenous only if one of the two conditions is satisfied:
  - no city not selected build railways; the event $R_c = 1$ and $Z_c = 0$ never occurs
  - $\mathbb{E}(\theta_c | R_c = 1, Z_c = 1) = \mathbb{E}(\theta_c | Rs_c = 1, Z_c = 0)$
- famous criticism against Angrist (1990)

Plan

An attempt of a definition

- No consensus, personal definition.
- A structural model is a mapping from a set of parameters to a probability distribution of observed variables. Parameters are deterministic elements like constants and probabilistic distributions.
  - Denote the space of parameters Θ and the space of observed variables S.
  - A model $\mathcal{M}$ associates any element $\theta \in \Theta$ to a probability distribution $\mathcal{M}(\theta)$ on the space S.
- A structural model can be decomposed in three parts.
  - Sampling, defines exogenous variables. Variables like shocks are drawn from distributions defined by the parameters.
  - Economic modeling, endogenous variables derive from the exogenous variables through economically-meaningful mathematical operations.
  - Measurement, distinguishes the latent variables from the observed variables.

A figure

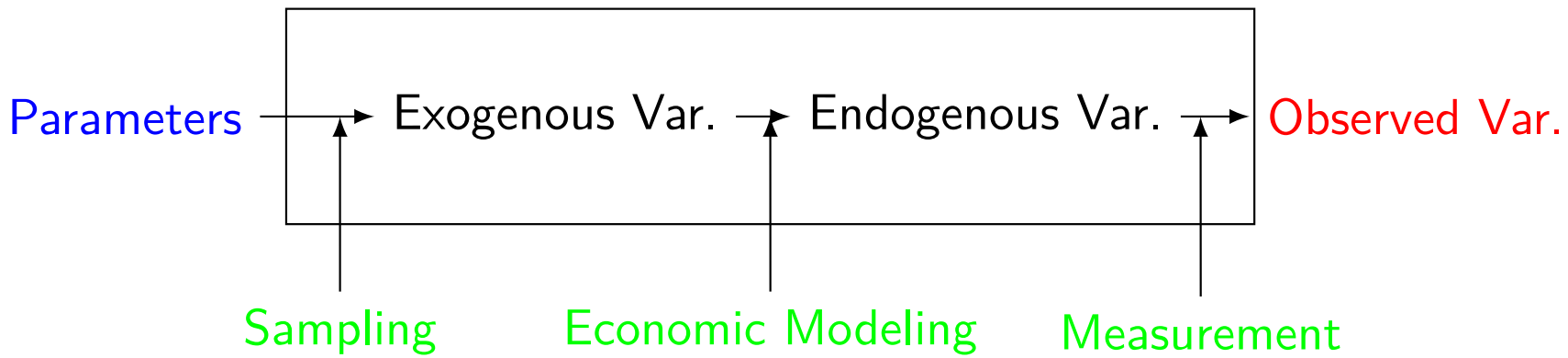

Figure: Structural model

Example 1

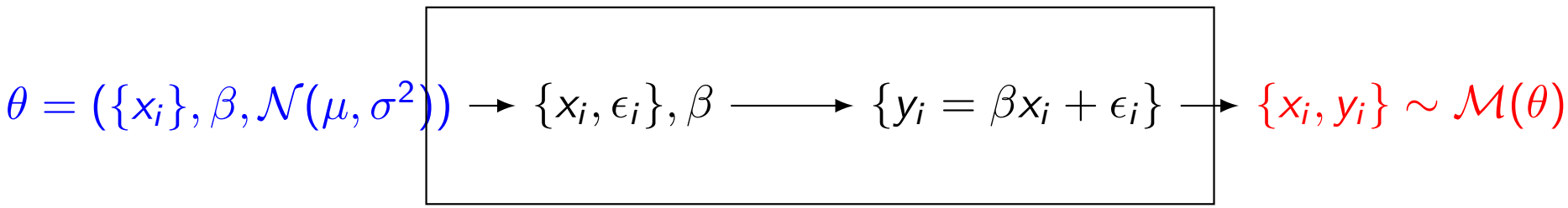

Figure: The univariate linear model with gaussian errors

Example 2

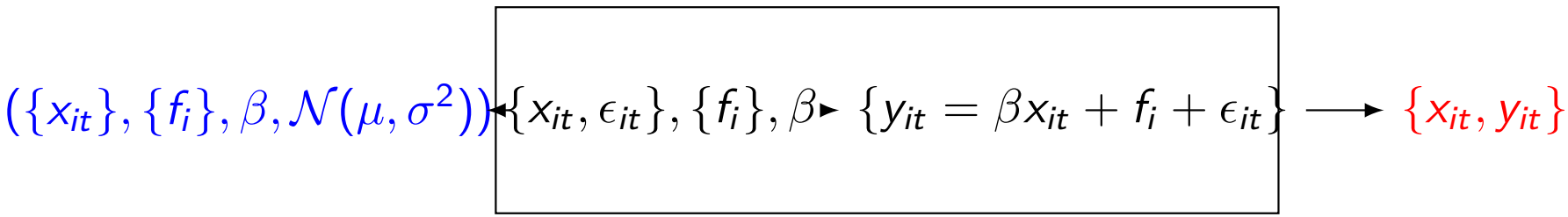

Figure: Fixed effects model

Example 3

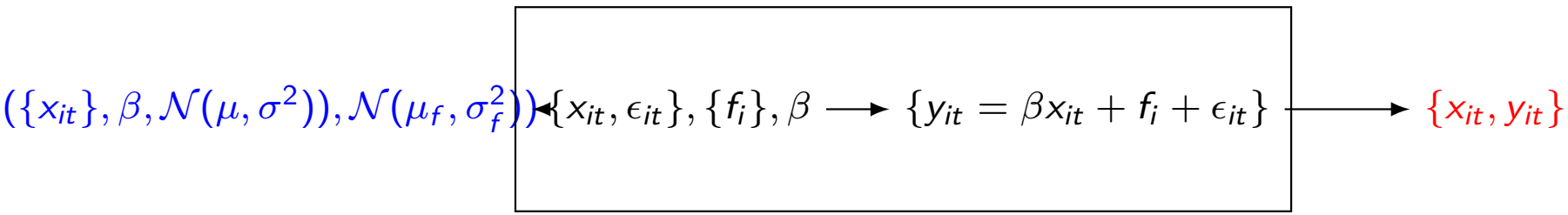

Figure: Random effects model

The data-generating process (DGP)

- The structural model is assumed to be the *data-generating process* (DGP) of the dataset in hands.
  - The first assumption in structural econometrics is to believe that the structural model is the "true" model.
  - This is a source of criticism. What if it is not the case and the model is misspecified? Why if there is no such thing as the "true" model?
  - Rust gives us a way to reply: econometrics is not the quest for truth.
  - *does your model fit the data better and provide more accurate out of sample forecasts than any other competing model?", Rust (2010)*

An estimation strategy

- The econometrician observes a dataset $\mathcal{D}$
- Assume there is a model $\mathcal{M}$ and a set of parameters $\theta \in \Theta$ such that $\mathcal{D}$ is a draw from $\mathcal{M}(\theta)$
- the goal of the econometrician is to estimate a subset of parameters $\theta^I \subseteq \theta$
- an estimation strategy is a mapping, $\hat{theta}^I(\mathcal{D})$
  - for any dataset, there is a an estimate of θ^I

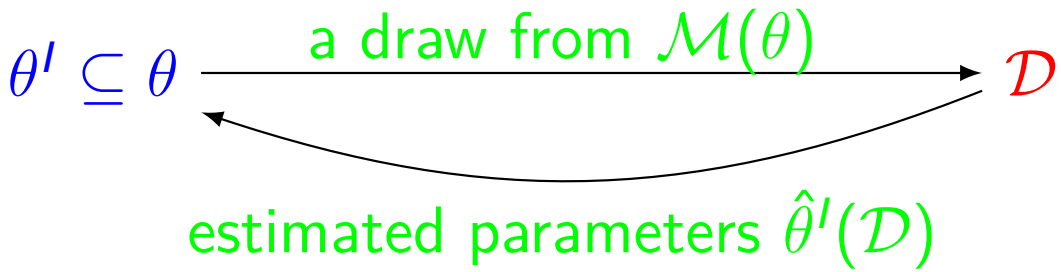

Figure: Structural model

Identification

- To have a chance of measuring a $\hat{\theta}^I(\mathcal{D})$ close to a θ^I, we need *identifiability* of parameters θ^I.
- θ^I is identified if: for any $\theta_1, \theta_2 \in \Theta$

$$\mathcal{M}(\theta_1) = \mathcal{M}(\theta_2) \Rightarrow \theta_1^I = \theta_2^I$$

Identification

- A *test* can be thought of as a comparison between two competing models.
- Most of the time, the two models are nested with one imposing a least one additional *over-identifying* restrictions.
- For instance, consider the linear regression ($y = \beta_0 + \beta_1 x + \epsilon$).
  - The test for $\beta_0 = 0$ consists in comparing the model $\{y = \beta_1 x + \epsilon\}$ to the model $\{y = \beta_0 + \beta_1 x + \epsilon\}$
  - In the first model, β_0 is not a parameter.
  - In the Wald test, it is not obvious. It is, however, in the likelihood-ratio test by definition!
- John Sargan also proposed a test for over-identifying restrictions.