Econometric models

Paolo Zacchia

Econometric Theory

Lecture 9

Structure, identification and causality: roadmap

- This lecture develops some theoretical concepts at the core of econometric theory.

- It starts with a general definition of a **structural model**, while also providing some examples.

- It then develops a general definition of **identification** as it applies to structural models.

- To better substantiate these concepts, the lecture illustrates examples of linear **simultaneous equations models**.

- Finally, the lecture develops the concept of **causality** as it applies to structural models, and highlights the connection between identification and causality.

Structural models

A **“structural” econometric model** is a set of relationships that link together some socio-economic variables of interest. In such a model, one usually distinguishes between:

- some P **endogenous** variables: $\boldsymbol{y}_i = (Y_{1i}, Y_{2i}, \ldots, Y_{Pi})$;
- some Q **exogenous** variables: $\boldsymbol{z}_i = (Z_{1i}, Z_{2i}, \ldots, Z_{Qi})$;
- and some R **unobserved** variables $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \varepsilon_{2i}, \ldots, \varepsilon_{Ri})$, sometimes also called unobserved *factors*.

A structural model relates endogenous variables to themselves, to exogenous variables as well as to unobserved factors through P functional relationships.

$$\boldsymbol{y}_i = \boldsymbol{s}\left(\boldsymbol{y}_i, \boldsymbol{z}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}\right)$$

As usual, the subscript i denotes units of observation.

Structural model: discussion

- The P relationship are assumed *a priori* and motivated by economic theory or by other knowledge about the economy.

- These model is characterized by a vector of **parameters** $\boldsymbol{\theta}$ ($|\boldsymbol{\theta}| = K$) and whose parameter space is written as $\boldsymbol{\Theta}$. The "true" parameter vector is written as $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

- The objective is to conduct **statistical inference** about $\boldsymbol{\theta}$ using a **data sample** $\{(\mathbf{y}_i, \mathbf{z}_i)\}_{i=1}^N$ of size N.

- To this end, **distributional assumptions** are necessary.

- In a **fully parametric** approach, an econometrician makes assumptions about the entire joint distribution of $(\boldsymbol{y}_i, \boldsymbol{z}_i)$.

- In a **semi-parametric** approach, an econometrician makes more limited assumptions, e.g. about selected moments.

Linear simultaneous equations models (1/2)

A leading example of a structural model is the **linear** version of a **simultaneous equations model** (SEM). For $P = R$ this is:

$$\begin{aligned}
\gamma_{11}Y_{1i} + \ldots + \gamma_{1P}Y_{Pi} &= \phi_{11}Z_{1i} + \ldots + \phi_{1Q}Z_{Qi} + \varepsilon_{1i} \\
\gamma_{21}Y_{1i} + \ldots + \gamma_{2P}Y_{Pi} &= \phi_{21}Z_{1i} + \ldots + \phi_{2Q}Z_{Qi} + \varepsilon_{2i} \\
\ldots &= \ldots \\
\gamma_{P1}Y_{1i} + \ldots + \gamma_{PP}Y_{Pi} &= \phi_{P1}Z_{1i} + \ldots + \phi_{PQ}Z_{Qi} + \varepsilon_{Pi}
\end{aligned}$$

where the parameters θ are as follows:

- $\theta = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_P; \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_P)$;
- $\boldsymbol{\gamma}_p = (\gamma_{p1}, \ldots, \gamma_{pP})$;
- $\boldsymbol{\phi}_p = (\phi_{p1}, \ldots, \phi_{pQ})$.

For convenience, some parameters are typically normalized, e.g. $\gamma_{pp} = 1$ for $p = 1, \ldots, P$.

Linear simultaneous equations models (2/2)

A SEM can be written in **compact vectorial notation**:

$$\boldsymbol{\Gamma}\mathbf{y}_i = \boldsymbol{\Phi}\mathbf{z}_i + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are, respectively, matrices of dimension $P \times P$ and $P \times Q$, which collect the $2P$ $\boldsymbol{\gamma}_p$ and $\boldsymbol{\phi}_p$ parameter vectors along their rows; while:

$$\mathbf{y}_i = \begin{bmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{Pi} \end{bmatrix}, \ \mathbf{z}_i = \begin{bmatrix} z_{1i} \\ z_{2i} \\ \vdots \\ z_{Qi} \end{bmatrix}, \ \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{Pi} \end{bmatrix}$$

collect the observation-specific realizations of $\boldsymbol{y}_i$ and $\boldsymbol{z}_i$ as well as the values of $\boldsymbol{\varepsilon}_i$.

Example: the Mincer model as a SEM

The Mincer equation from Lecture 7:

$$\log W_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 S_i + \varepsilon_{1i}$$

together with the equation for education S_i from Lecture 7:

$$S_i = \gamma_0 + \gamma_1 Z_i + \phi_1 X_i + \phi_2 X_i^2 + \varepsilon_{2i}$$

are a (linear) SEM. Here:

- log-wages W_i and S_i are the endogenous variables;
- the random vector $\boldsymbol{z}_i = (X_i, X_i^2, Z_i)$ collects the exogenous variables;
- while $\varepsilon_{1i} = \alpha_i + \epsilon_i$ and $\varepsilon_{2i} = \psi_0 \alpha_i + \eta_i$ are two "combined" unobserved factors.

More examples of structural models follow.

Example: the Kline I model (1/2)

SEMs were introduced by the 'Cowles commission' in the '40s, at a time when econometrics was developed with the ambition of creating a large macroeconomic model of the economy.

A primitive but celebrated model is the Klein I (1950) model:

$$\begin{aligned} C_t &= \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 \left(W_t^p + W_t^g\right) + \varepsilon_{1t} \\ I_t &= \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + \varepsilon_{2t} \\ W_t^p &= \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 A_t + \varepsilon_{3t} \end{aligned}$$

which was paired with the following "identities" (restrictions).

$$\begin{aligned} X_t &= C_t + I_t + G_t \\ P_t &= X_t - T_t - W_t^p \\ K_t &= K_{t-1} + I_t \end{aligned}$$

The Keynesian consumption function from Lecture 7 is a very simplified version of this model.

Example: the Kline I model (2/2)

The variables in the model are as follows:

- C_t is consumption;
- I_t is investment;
- G_t is the government's nonwage expenditure;
- X_t is the aggregate demand or GDP;
- T_t are the indirect business tax plus net exports;
- K_t is the aggregate capital stock and K_{t-1} is its *lagged* value;
- P_t is the aggregate level of profits realized in the private sector and P_{t-1} is its *lagged* value;
- W_t^p are wages paid in the private sector;
- W_t^g are wages paid in the business sector;
- A_t is a constant time trend.

Example: entry models (1/2)

The dominant "structural" subfield in Economics is **Industrial Organization** (IO). An archetypical class of structural models in IO are the **entry models**. Here is a stylized one.

Consider N markets indexed as $i = 1, \ldots, N$, each populated by an *endogenous* number F_i of identical firms. The average **profit** of a firm in market i, as a function of F_i is:

$$\pi_i\left(F_i\right) = \pi_{Vi}\left(F_i, \boldsymbol{z}_i, \nu_i; \boldsymbol{\theta}_M\right) - C_i$$

where:

- $\pi_{Vi}\left(\cdot\right)$ is a **variable profits function**;
- $\boldsymbol{z}_i$ are *exogenous* market variables with parameters $\boldsymbol{\theta}_M$;
- ν_i are some *unobserved factors*;
- and C_i are the market-specific **fixed costs**.

In standard economic models, $\pi_{Vi}\left(\cdot\right)$ is **decreasing** in F_i.

Example: entry models (2/2)

Economic theory predicts that, under complete information, **in equilibrium** firms will enter the market insofar they can make positive profits. This yields:

$$F_i \in \arg\min_{F \in \mathbb{N}} \pi_{Vi}\left(F, \boldsymbol{z}_i, \nu_i; \boldsymbol{\theta}_M\right) \quad \text{s.t. } \pi_{Vi}\left(F_i, \boldsymbol{z}_i, \nu_i; \boldsymbol{\theta}_M\right) - C_i \geq 0$$

F_i is a **non-linear** step function of the exogenous variables.

If the demand function has a constant elasticity ζ, it is directly proportional to a measure of "market size" $\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{\theta}_D + \nu_i$, where $\boldsymbol{z}_i$ are factors affecting demand; $\boldsymbol{\theta}_M = (\boldsymbol{\theta}_D, \zeta)$; firms have constant marginal costs and compete *à la* Cournot, then:

$$\pi_{Vi}\left(F_i, \boldsymbol{z}_i, \nu_i; \boldsymbol{\theta}_M\right) = \frac{\zeta\left(\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{\theta}_D + \nu_i\right)}{F_i^2}$$

à la Berry (1992), and convenient for the sake of estimation.

Before estimation

Structural econometric models can be different to one another, and they are estimated in different ways. Before proceeding to estimation, however, econometricians are advised to make sure their model is conceptually sound.

Specifically, they should address the following questions.

1. *Is it possible to use the results of my estimates for the sake of attributing unique values to each parameter belonging to the set* $\boldsymbol{\theta}$*?* [**Identification**]

2. *Is it possible then to use these estimates in order to answer questions about the "effect" that certain variables have upon the others?* [**Causality**]

Questions like these lie at the core of econometric analysis. The concepts of **identification** and **causality** are introduced in the remainder of this Lecture.

Identification: intuition

- There are competing informal definitions of identification.
- At the core, a certain parameter set $\theta \in \Theta$ is **identified** in a statistical model if no other set $\theta' \in \Theta$ is equally capable to generate the data, in a probabilistic sense.
- Intuitively: if education S_i and wages W_i are related in the data, is it due to some "effect" of education on wages (β_3) or to the indirect effect of ability α_i?
- The formal definition was developed by Rothenberg (1971) in the context of a **fully parametric** model.
- Quoting Rothenberg, *"the identification problem concerns the existence of a unique inverse association"* from the data to the parameters of interest.
- The idea is easily generalized to semi-parametric models.

Some basic definitions

Definition 1

A **data generation process** (DGP) is the joint probability distribution $F_{\boldsymbol{\theta}}(\boldsymbol{z}_i, \boldsymbol{\varepsilon}_i)$ parametrized by $\boldsymbol{\theta}$ or, given the P relationships $\boldsymbol{s}(\cdot;\cdot)$ implied by the structural model, $G_{\boldsymbol{\theta}}(\boldsymbol{z}_i, \boldsymbol{y}_i)$.

Definition 2

A **family** $\mathcal{P}$ of DGPs is some given set of similar DGPs.

Definition 3

A **structure** $\boldsymbol{\theta}'$, is a specific restriction on $\boldsymbol{\theta}$ that uniquely determines a particular DGP $\mathcal{P}_{\boldsymbol{\theta}'}(\boldsymbol{z}_i, \boldsymbol{\varepsilon}_i) \in \mathcal{P}$.

Definition 4

A **statistical model** $\mathcal{M}$ the set of valid structures, which needs not to be equivalent with the family of DGPs $\mathcal{P}$. A statistical model $\mathcal{M}$ is best understood as the set of structures $\mathcal{M} \subset \mathcal{P}$ that are compatible with the restrictions implied in the structural model.

Example: parametric bivariate regression

Consider a simple bivariate linear model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where the data are generated according to a well-known **family** $\mathcal{P}$ of **DGPs**, a bivariate normal distribution:

$$\begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_\varepsilon \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & \sigma_{x\varepsilon} \\ \sigma_{x\varepsilon} & \sigma_\varepsilon^2 \end{pmatrix}\right)$$

implying the following.

$$Y_i \sim \mathcal{N}\left(\beta_0 + \beta_1 \mu_x + \mu_\varepsilon, \beta_1^2 \sigma_x^2 + 2\beta_1 \sigma_{x\varepsilon} + \sigma_\varepsilon^2\right)$$

- The **restriction** $\mu_\varepsilon = \mathbb{E}[\varepsilon_i] = 0$ is usually imposed.
- The statistical **model** $\mathcal{M}$ is the set of **structures** that are allowed by the model: $\theta = (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_\varepsilon^2, \sigma_{x\varepsilon})$.
- An example is the restriction $\theta_0 = (5, 2, 0, 2, 2, 1)$.

Identification definitions

Definition 5

Observational Equivalence. Two structures θ' and θ'' are *observationally equivalent* if $\mathbb{P}(\boldsymbol{y}_i, \boldsymbol{z}_i | \theta') = \mathbb{P}(\boldsymbol{y}_i, \boldsymbol{z}_i | \theta'')$.

Definition 6

Global Identification. A Structure $\theta' \in \boldsymbol{\Theta}$ is *globally point identified* if there is no other structure $\theta \in \boldsymbol{\Theta}$ that is observationally equivalent to it.

Definition 7

Local Identification. A Structure $\theta' \in \boldsymbol{\Theta}$ is *locally point identified* if there is no other structure in an open neighborhood of θ' that is observationally equivalent to it.

Definition 8

Model Identification. An econometric model $\mathcal{M}$ is identified if all its structures $\theta \in \boldsymbol{\Theta}$ are identified.

Non-identification: an example (1/4)

Theorem 1

Identification of a fully parametric bivariate regression. *The statistical model $\mathcal{M}$ from the previous example is not identified, but the restricted model given by $\mathcal{M}' = \{\theta \in \mathcal{M} : \sigma_{x\varepsilon} = 0\}$ is identified, specifically, globally point identified.*

Proof.

(Sketched.) The proof proceeds as follows.

1. One establishes a *rule* to select parameter values from the data, akin to *estimating* θ. In this case MLE is applied, thanks to the fully parametric assumptions and the likelihood principle.
2. One then shows that $\mathcal{M}$ **does not** allow for a unique solution of the MLE problem.
3. One finally shows that $\mathcal{M}'$, instead, **has** a unique solution.

Here only point 2. is shown. This is broken in two parts. First, it is shown that $\widehat{\mu}_x$ and $\widehat{\sigma}_x^2$ are identified; then, that $\left(\beta_0, \beta_1, \sigma_\varepsilon^2, \sigma_{x\varepsilon}\right)$ are *not* identified. Point 3. is then easy to show. **(Continues. . .)**

Non-identification: an example (2/4)

Theorem 1

Proof.

(Continued.) Focus on $\left(\widehat{\mu}_x, \widehat{\sigma}_x^2\right)$: their identification is based on the observations of X_i only. So long as the realizations $\{x_1, \dots, x_N\}$ are not identical to each other, the log-likelihood function

$$\log \mathcal{L}\left(\mu_x, \sigma_x^2 \middle| x_1, \dots, x_N\right) = -\frac{N}{2} \log 2\pi\sigma_x^2 - \sum_{i=1}^{N} \frac{\left(x_i - \mu_x\right)^2}{2\sigma_x^2}$$

has a solution (see Lecture 5). The Hessian matrix, when evaluated at the solution, is:

$$\mathbf{H}\left(\widehat{\mu}_x, \widehat{\sigma}_x^2 \middle| x_1, \dots, x_N\right) = -N \begin{bmatrix} \widehat{\sigma}_x^{-2} & 0 \\ 0 & 2\widehat{\sigma}_x^{-4} \end{bmatrix}$$

and since $\widehat{\sigma}_x^2 \neq 0$ its determinant is nonzero. By the Implicit Function Theorem, the MLE solution is thus unique. **(Continues...)**

Non-identification: an example (3/4)

Theorem 1

Proof.

(Continued.) Now consider the entire sample $\{(y_i, x_i)\}_{i=1}^N$ and the log-likelihood function

$$\log \mathcal{L}\left(\vartheta \middle| y_1, \dots, y_N, x_1, \dots, x_N\right) = -\frac{N}{2} \log 2\pi \left(\beta_1^2 \sigma_x^2 + 2\beta_1 \sigma_{x\varepsilon} + \sigma_\varepsilon^2\right) - \sum_{i=1}^{N} \frac{\left(y_i - \beta_0 - \beta_1 x_i\right)^2}{2\left(\beta_1^2 \sigma_x^2 + 2\beta_1 \sigma_{x\varepsilon} + \sigma_\varepsilon^2\right)}$$

(one can abstract from μ_x and σ_x^2 as they are shown to be identified). Define

$$\widehat{\sigma}_y^2 \equiv \widehat{\beta}_1^2 \widehat{\sigma}_x^2 + 2\widehat{\beta}_1 \widehat{\sigma}_{x\varepsilon} + \widehat{\sigma}_\varepsilon^2$$

and

$$e_i \equiv y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$$

for $i = 1, \dots, N$. **(Continues. . .)**

Non-identification: an example (4/4)

Theorem 1

Proof.

(Continued.) The First Order Conditions are:

$$\frac{\partial \log \mathcal{L}\left(\widehat{\vartheta}\middle| \ldots\right)}{\partial \beta_0} = \sum_{i=1}^{N} \frac{e_i}{\widehat{\sigma}_y^2} = 0$$

$$\frac{\partial \log \mathcal{L}\left(\widehat{\vartheta}\middle| \ldots\right)}{\partial \beta_1} = -\frac{N\left(\widehat{\beta}_1 \widehat{\sigma}_x^2 + \widehat{\sigma}_{x\varepsilon}\right)}{\widehat{\sigma}_y^2} + \sum_{i=1}^{N} \frac{e_i x_i}{\widehat{\sigma}_y^2} + \sum_{i=1}^{N} \frac{e_i^2 \left(\widehat{\beta}_1 \widehat{\sigma}_x^2 + \widehat{\sigma}_{x\varepsilon}\right)}{\widehat{\sigma}_y^4} = 0$$

$$\frac{\partial \log \mathcal{L}\left(\widehat{\vartheta}\middle| \ldots\right)}{\partial \sigma_\varepsilon^2} = -\frac{N}{2\widehat{\sigma}_y^2} + \sum_{i=1}^{N} \frac{e_i^2}{2\widehat{\sigma}_y^4} = 0$$

$$\frac{\partial \log \mathcal{L}\left(\widehat{\vartheta}\middle| \ldots\right)}{\partial \sigma_{x\varepsilon}} = -2\widehat{\beta}_1 \left(\frac{N}{2\widehat{\sigma}_y^2} - \sum_{i=1}^{N} \frac{e_i^2}{2\widehat{\sigma}_y^4}\right) = 0$$

and the last two are linearly dependent, hence no unique solution can be attained. □

Non-identification: discussion of the example

- The identification condition $\sigma_{x\varepsilon} = 0$ in $\mathcal{M}'$ states that the covariance between X_i and ε_i must be zero.

- This is akin to the "exogeneity" condition in linear models! Namely, that $\mathbb{E}\left[\varepsilon_i | X_i = x_i\right] = 0$ for all $x_i \in \mathbb{X}$

- This implies linearity of the CEF of Y_i given X_i, and:

$$\begin{aligned}\sigma_{x\varepsilon} = \mathbb{C}\text{ov}\left(X_i, \varepsilon_i\right) &= \mathbb{E}\left[X_i \varepsilon_i\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[\varepsilon_i\right] \\ &= \mathbb{E}_X\left[\mathbb{E}\left[X_i \varepsilon_i | X_i\right]\right] \\ &= 0\end{aligned}$$

  because $\mathbb{E}\left[\varepsilon_i\right] = 0$ and the Law of Iterated Expectations.

- Intuitively, $\sigma_{x\varepsilon} = 0$ makes sure that the co-movement of X_i and Y_i that is observed in the data is due to β_1 alone – and not to other factors.

Example: semi-parametric bivariate regression

Identification definitions extend to **semi-parametric models**. Consider again the bivariate linear model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, but devoid of fully parametric assumptions.

In this case one can re-define the concept of "model" $\mathcal{M}$ as a set of structures of the kind

$$\theta = \left(\beta_0, \beta_1, \mathcal{P}_x, \mathcal{P}_\varepsilon, \mathcal{P}_{\varepsilon|x}\right)$$

where $\mathcal{P}_x$, $\mathcal{P}_\varepsilon$ and $\mathcal{P}_{\varepsilon|x}$ are **families** of probability distributions – respectively of X_i, of ε_i, and of ε_i conditional on X_i – that are allowed by the model $\mathcal{M}$.

An obvious restriction here is that all elements of $\mathcal{P}_\varepsilon$ conform to $\mathbb{E}\left[\varepsilon_i\right] = 0$; clearly an unrestricted mean is indistinguishable from the constant parameter β_0.

Identification is achieved here by placing a restriction on $\mathcal{P}_{\varepsilon|x}$ – and not a surprising one: it must be $\mathbb{E}\left[\varepsilon_i \middle| X_i\right] = 0$.

Semi-parametric identification: an example

Theorem 2

Identification of a semi-parametric bivariate regression. *Consider the semi-parametric bivariate linear model $\mathcal{M}$ incorporating the restriction $\mathbb{E}[\varepsilon_i] = 0$; this model is not identified. Instead, the restricted model $\mathcal{M}' = \{\theta \in \mathcal{M} : \mathbb{E}[\varepsilon_i | X_i] = 0\}$ is identified.*

Proof.

(Outline.) A heuristic proof is based on the analysis of *cross moments* of the model (like covariances) that involve X_i and ε_i, and it leverages the analogy principle (see Lecture 5) to evaluate identification.

- $\mathcal{M}$ is not identified because it allows for $\mathbb{E}[X_i\varepsilon_i] = g(X_i)$, where $g(X_i)$ is some function of X_i. Hence, a zero moment condition of the form $\mathbb{E}[X_i\varepsilon_i - g(X_i)] \neq 0$ does not lead to a unique solution.

- Instead, $\mathcal{M}'$ is identified: the moment conditions implied by the model lead to a unique association from the data to β_0 and β_1.

In model $\mathcal{M}'$ the families $\mathcal{P}_x$, $\mathcal{P}_\varepsilon$, and $\mathcal{P}_{\varepsilon|x}$ are trivially identified: all non-degenerate joint distributions allowing for mean independence of the error term comply with the definition of identification. □

Identification: summary

- Identification must always be evaluated on a case-by-case basis. Identification failures can occur in various ways.

- The two previous examples describe failures of **statistical** identification: in order to obtain identification, restrictions on the DGP are typically necessary.

- Singularity of matrix $\mathbf{K}_0$ and its sample analogue in linear regression (occurring e.g. due to the dummy variable trap) is a **mechanical** or **algebraic** failure of identification. Issues of this sort also occur in other models.

- In elaborate structural models, algebraic non-identification can derive from the relationships specified in $\boldsymbol{s}(\cdot; \boldsymbol{\theta})$.

- Linear SEMs provide instructive examples of the latter: the classic example of *demand and supply* is developed next.

Reduced form and separable models

Some additional definitions are necessary.

Definition 9
The **reduced form** of a structural econometric model is its solution for $\boldsymbol{y}_i$.

$$\boldsymbol{y}_i = \boldsymbol{r}\left(\boldsymbol{z}_i, \boldsymbol{\varepsilon}_i; \theta\right)$$

Definition 10
A **separable structural model** is one that possesses a reduced form representation.

For example, SEMs are separable if $\boldsymbol{\Gamma}$ is invertible, where:

$$\begin{aligned}\mathbf{y}_i &= \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\Phi}\mathbf{z}_i + \boldsymbol{\varepsilon}_i\right) \\ &= \boldsymbol{\Pi}\mathbf{z}_i + \boldsymbol{\eta}_i\end{aligned}$$

$\boldsymbol{\Pi} \equiv \boldsymbol{\Gamma}^{-1}\boldsymbol{\Phi}$ is a $P \times Q$ **matrix of reduced form parameters** π_{pq} (p indexes rows, q columns) and $\boldsymbol{\eta}_i \equiv \boldsymbol{\Gamma}^{-1}\boldsymbol{\varepsilon}_i$.

Demand and supply in equilibrium

- An econometrician observes a sample of N **markets** which feature information about the **price** P_i and **quantity** Q_i of a given good or service; $i = 1, \ldots, N$ indexes markets.

- The demand Q_i^D and the supply Q_i^S are specified as **linear** functions of the price; $(\upsilon_i^D, \upsilon_i^S)$ are their error terms.

$$Q_i^D = \alpha_0 + \alpha_1 P_i + \upsilon_i^D$$
$$Q_i^S = \beta_0 + \beta_1 P_i + \upsilon_i^S$$

- We learn from economic theory that, in a market, demand and supply meet in equilibrium.

$$Q_i^D = Q_i^S = Q_i$$

- Prices P_i and quantities Q_i are determined simultaneously and interdependently: they are both **endogenous**.

Demand and supply in equilibrium

- The parameters of this model are **not identified**.

- To see why, consider the **reduced form** of the model.

$$Q_i = \frac{\beta_1\alpha_0 - \alpha_1\beta_0}{\beta_1 - \alpha_1} + \frac{\beta_1 \upsilon_i^D - \alpha_1 \upsilon_i^S}{\beta_1 - \alpha_1}$$
$$P_i = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\upsilon_i^D - \upsilon_i^S}{\beta_1 - \alpha_1}$$

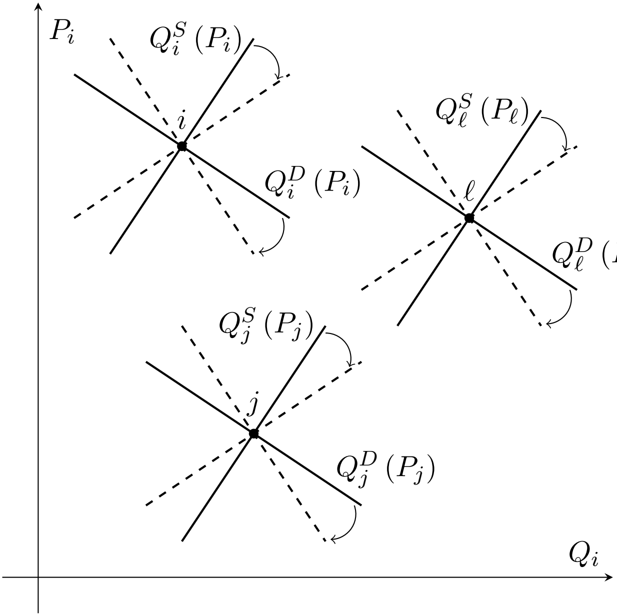
- *By construction*, $\mathbb{E}\left[\upsilon_i^D, \upsilon_i^S \middle| P_i\right] \neq 0$ and $\mathbb{E}\left[\upsilon_i^D, \upsilon_i^S \middle| Q_i\right] \neq 0$.

- Neither fully nor semi-parametric approaches would work.

- Only the *constants of the reduced form* are identified.

- Intuitively, **infinitely many pairs** of supply and demand curves meet in the equilibrium points.

Infinitely many demand and supply curves



Exact, partial and over-identification

By expanding the model one can obtain better identification results. Again, some definitions are necessary.

Definition 11

Exact, or just identification. An econometric model is *exactly* or *just* identified (the two expressions are interchangeable) if there exists a *unique* association from the data to the parameter set θ.

Definition 12

Partial identification. An econometric model is *partially* identified if there exists a *unique* association from the data to a *subset* of the parameter set θ ($\theta^* \subset \theta$), but not so for the other parameters.

Definition 13

Overidentification. In an econometric model a *subset* θ^{**} of the parameter set θ ($\theta^{**} \subset \theta$) is *overidentified* if there exist *multiple* associations from the data to the parameter subset in question.

Partial identification and overidentification can coexist in a model.

Partial identification with a demand shifter (1/2)

- Now add a new variable: the **exogenous** consumers' income; denote it as M_i.

- According to theory, this affects demand but not supply.

- Hence the model becomes:

$$\begin{aligned} Q_i &= \alpha_0 + \alpha_1 P_i + \alpha_2 M_i + \upsilon_i^D \\ Q_i &= \beta_0 + \beta_1 P_i + \upsilon_i^S \end{aligned}$$

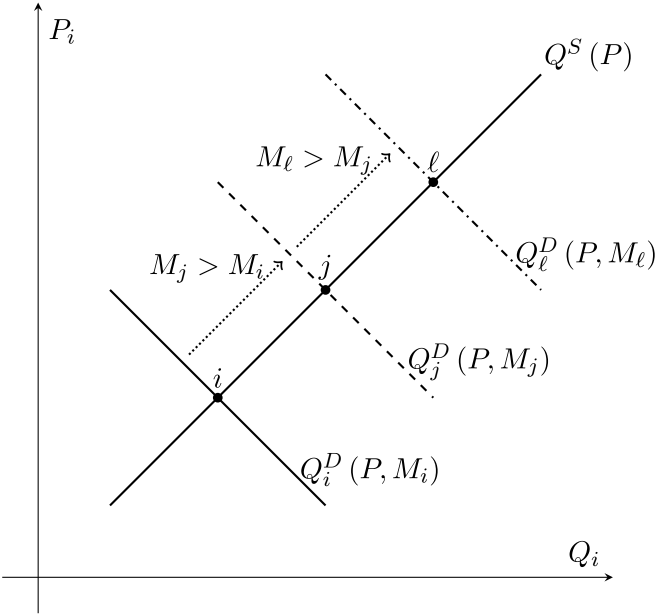
- ... and the reduced form is now as follows.

$$\begin{aligned} Q_i &= \frac{\beta_1 \alpha_0 - \alpha_1 \beta_0}{\beta_1 - \alpha_1} + \frac{\beta_1 \alpha_2}{\beta_1 - \alpha_1} M_i + \frac{\beta_1 \upsilon_i^D - \alpha_1 \upsilon_i^S}{\beta_1 - \alpha_1} \\ P_i &= \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} M_i + \frac{\upsilon_i^D - \upsilon_i^S}{\beta_1 - \alpha_1} \end{aligned}$$

Partial identification with a demand shifter (2/2)

- If $\mathbb{E}[\upsilon_D, \upsilon_S | M_i] = 0$, the two reduced form equations are identified!
- Hence, β_1 (the supply curve's slope) is **identified** as the ratio of the two coefficients for M_i!
- The model is **partially identified**: other parameters are not identified.
- The intuition is that M_i is a **demand shifter**: it allows us to account for changes in demand that do not depend upon other factors (a *ceteris paribus* thought experiment).
- This, in turn, allows to "trace out" the shape of the supply curve, which stays constant.
- The intuition is best appreciated graphically.

Shifting demand and tracing supply



Exact identification of demand and supply

- Add another variable: an **exogenous** index of production costs, denoted as C_i.

- According to theory this affects supply but not demand, so:

$$Q_i = \alpha_0 + \alpha_1 P_i + \alpha_2 M_i + \upsilon_i^D$$
$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 C_i + \upsilon_i^S$$

- …and the reduced form is as follows.

$$Q_i = \frac{\beta_1\alpha_0 - \alpha_1\beta_0}{\beta_1 - \alpha_1} + \frac{\beta_1\alpha_2}{\beta_1 - \alpha_1}M_i - \frac{\alpha_1\beta_2}{\beta_1 - \alpha_1}C_i + \frac{\beta_1\upsilon_i^D - \alpha_1\upsilon_i^S}{\beta_1 - \alpha_1}$$
$$P_i = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1}M_i - \frac{\beta_2}{\beta_1 - \alpha_1}C_i + \frac{\upsilon_i^D - \upsilon_i^S}{\beta_1 - \alpha_1}$$

- If $\mathbb{E}\left[\upsilon_D, \upsilon_S \middle| M_i\right] = \mathbb{E}\left[\upsilon_D, \upsilon_S \middle| C_i\right] = 0$ **exact identification** is attained! The six structural parameters can be recovered from the six reduced form parameters.

Overidentification of the supply curve

- Add another variable: the **exogenous** price of a competing product, denoted as P_i^*. Drop C_i from the model.

- According to theory, P_i^* affects demand and not supply, so:

$$Q_i = \alpha_0 + \alpha_1 P_i + \alpha_2 M_i + \alpha_3 P_i^* + \upsilon_i^D$$
$$Q_i = \beta_0 + \beta_1 P_i + \upsilon_i^S$$

- ...and the reduced form is as follows.

$$Q_i = \frac{\beta_1\alpha_0 - \alpha_1\beta_0}{\beta_1 - \alpha_1} + \frac{\beta_1\alpha_2}{\beta_1 - \alpha_1} M_i + \frac{\beta_1\alpha_3}{\beta_1 - \alpha_1} P_i^* + \frac{\beta_1\upsilon_i^D - \alpha_1\upsilon_i^S}{\beta_1 - \alpha_1}$$
$$P_i = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} M_i + \frac{\alpha_3}{\beta_1 - \alpha_1} P_i^* + \frac{\upsilon_i^D - \upsilon_i^S}{\beta_1 - \alpha_1}$$

- Note that β_1 is **overidentified**: it can be recovered in two ways (intuitively, there are two demand shifters). Note that other parameters are instead **not identified**.

Identifying a linear SEM: general approach

- If $\mathbb{E}\left[\left.\boldsymbol{\varepsilon}_i\right|\boldsymbol{z}_i\right] = \mathbf{0}$, a reduced form equation of a linear SEM:

$$Y_{pi} = \pi_{p1}Z_{1i} + \pi_{p2}Z_{2i} + \cdots + \pi_{pQ}Z_{Qi} + \eta_{pi}$$

  is identified. This gives $P \times Q$ identified parameters in $\mathbf{\Pi}$.

- However, there are $P\left(P+Q\right)$ **structural parameters** in $(\mathbf{\Gamma}, \mathbf{\Phi})$. To achieve identification, **restrictions** leading to a one-to-one mapping $\mathbf{\Pi} \mapsto (\mathbf{\Gamma}, \mathbf{\Phi})$ are necessary.

- So-called **exclusion restrictions** (e.g. M_i, C_i, P_i^* showing up either in the demand or in the supply equation, but not in both) are most common in econometrics.

- Under the diagonal normalization $\gamma_{pp} = 1$ for $p = 1, \ldots, P$, at least $P\left(P-1\right)$ restrictions are necessary.

Rank and order condition: introduction

- The *position* of the exclusion restrictions matters, however: even with the right *number* of restrictions, some parameters may be overidentified, other not identified.

- Example: the model with M_i and P_i^* but without C_i. This is a problem of determinacy of the solution for $(\mathbf{\Gamma}, \mathbf{\Phi})$.

- To guide the analysis of linear SEMs, the **rank** and **order conditions** are borrowed from linear algebra.

- Both express **necessary** conditions for identification; they complement one another.

- They are based on the analysis of the $P \times (P + Q)$ matrix which "horizontally binds" $\mathbf{\Gamma}$ and $\mathbf{\Phi}$.

$$\mathbf{F} \equiv \begin{bmatrix} \mathbf{\Gamma} & \mathbf{\Phi} \end{bmatrix}$$

Order and rank conditions

Order condition for SEM identification

Define ϱ_p as be the number of restrictions applied to the p-th equation of the structural form.

- if $\varrho_p < P - 1$, the p-th equation is *not identified*;
- if $\varrho_p = P - 1$, the p-th equation is *exactly identified*, as long as the rank condition holds;
- if $\varrho_p > P - 1$, the p-th equation is *overidentified*, as long as the rank condition holds.

Rank condition for SEM identification

If the order condition holds, the p-th Equation is identified if at least one nonzero determinant of order $(P-1)(P-1)$ can be constructed out of the coefficients of the variables excluded from that equation but included in other equations in the model.

Order and rank conditions: discussion

To check the rank condition for the p-th equation, one should:

1. delete from $\mathbf{F}$ the columns corresponding to the variables *included* in the p-th equation; and
2. delete row p too, which results in some submatrix $\mathbf{F}_p$.

Submatrix $\mathbf{F}_p$ should have full row rank for the rank condition to hold in the p-th equation.

Note that while the two conditions are formulated with respect to the number of endogenous variables Q; they can alternatively be expressed in terms of the number of exogenous variables Q.

Taken together, the two conditions allow to express a **sufficient** condition for exact identification of a linear SEM, which can be linked to the analysis of **instrumental variables** (Lecture 10).

Sufficient condition for SEM identification

Theorem 3
Sufficient Condition for Exact Identification. *A SEM is at least exactly identified if every equation of the structural model features an exogenous variable that does not show up in any other equation.*

Proof.
(Exercise!) This proof is a straightforward and instructive application of the order and rank conditions, and it is best left as an exercise. □

This delivers a strategy for *estimating* a just -identified SEM:

1. verify that the model is identified;
2. estimate the reduced form parameters $\mathbf{\Pi}$ via OLS;
3. solve for the structural parameters $(\mathbf{\Gamma}, \mathbf{\Phi})$.

This approach is known as **Indirect Least Squares** (ILS).

The Two- and Three-stages Least Squares elaborated in Lecture 10 are however more general, and allow for overidentification.

Introduction to causality

- An important feature of econometric models is their ability to express the **causal effect** that some economic or social variables have on the others.

- There are several competing definitions of **causality**. The dominant approach is the "potential outcomes" framework in statistical causal inference.

- The ensuing discussion elaborates the traditional definition of causality in **separable, structural** econometric models.

- In addition, it develops its connections with **identification** and with the statistical causal inference framework.

- What follows assumes that the exogenous variables Z_{qi} for $q = 1, \ldots, Q$ can vary on their support $\mathbb{X}_{\boldsymbol{z}q}$, independently of other exogenous variables as well as of unobservables.

Individual causal effects

Definition 14

Individual Causal Effect. Consider the unit of observation i; the realizations of its observable and unobservable factors are written as $(\mathbf{z}_i, \boldsymbol{\varepsilon}_i)$. Let z_{qi} be the q-th element of $\mathbf{z}_i$ and $\mathbf{z}_{-i}$ be the collection of all other $Q-1$ elements in that vector. The *individual causal effect* of the exogenous variable Z_{qi} on the endogenous variable Y_{pi} *for unit* i is:

$$\mathcal{C}_{qpi}\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) = r_p\left(z'_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) - r_p\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right)$$

if $\mathbb{X}_{\boldsymbol{z}q}$ is a countable discrete set with $\left(z_{qi}, z'_{qi}\right) \in \mathbb{X}^2_{\boldsymbol{z}q}$ being two consecutive values; and:

$$\mathcal{C}_{qpi}\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) = \frac{\partial}{\partial z_{qi}} r_p\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right)$$

if $\mathbb{X}_{\boldsymbol{z}q}$ is a continuous set, and where $r_p(\cdot)$ is the p-th equation of the reduced form, the one that predicts Y_{pi}.

- The individual causal effect can be interpreted as the "effect" of a *ceteris paribus* marginal variation of Z_{qi} on Y_{pi}.

Example: causal effects in the Mincer equation

In the Mincer equation:

- the causal effect of *education* S_i is:

$$\mathcal{C}_{SWi}\left(s_i, x_i, \varepsilon_{1i}\right) = \beta_3$$

- ... whereas the causal effect of *experience* X_i is:

$$\mathcal{C}_{XWi}\left(s_i, x_i, \varepsilon_{1i}\right) = \beta_1 + 2\beta_2 x_i$$

  which is a function of the current experience x_i of the i-th observation.

Generally, in linear models the causal effect of variables without higher order terms and/or "interactions" equals their associated structural parameter.

Causality: discussion

- Causality and identification are distinct concepts. In a (say linear) model which is **not** identified, causal effects may be still well defined.

- However, **estimation** of causal effects typically requires an identified model.

- If the variable of interest is a binary **treatment** $S_i \in \{0, 1\}$ causal effects are usually expressed through the **potential outcomes** framework by Rubin (1974):

$$Y_i = \begin{cases} Y_i(1) & \text{if } S_i = 1 \\ Y_i(0) & \text{if } S_i = 0 \end{cases}$$

  so the causal effect in question is $\mathcal{C}_{SYi} = Y_i(1) - Y_i(0)$.

- Theoretical connections between the structural econometric framework and Rubin's framework exist.

Average causal effects

Individual causal effects are not too interesting, and hard to evaluate. The *population averages* of causal effects are more interesting/useful.

Definition 15

Average Causal Effect. In the population, the average causal effect associated to variable Z_q is the expected value of the individual causal effects *conditional* on the other exogenous variables $\boldsymbol{z}_{-q}$.

$$\text{ACE}_{qp}\left(z_{qi}, \mathbf{z}_{-qi}\right) \equiv \mathbb{E}_{\boldsymbol{\varepsilon}}\left[\mathcal{C}_{qpi}\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) \middle| \, z_{qi}, \mathbf{z}_{-qi}\right] =$$

$$= \int_{\mathbb{X}_{\boldsymbol{\varepsilon}}} \mathcal{C}_{qpi}\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) f_{\boldsymbol{\varepsilon}|\boldsymbol{z}}\left(\boldsymbol{\varepsilon}_i \middle| \, z_{qi}, \mathbf{z}_{-qi}\right) d\varepsilon_{1i} \ldots d\varepsilon_{Pi}$$

- For binary treatments, this is called **Average Treatment Effect**.

$$\text{ATE}_Y \equiv \mathbb{E}\left[Y_i(1) - Y_i(0) \middle| \, z_{1i}, \ldots, z_{(Q-1)i}\right]$$

- A related quantity is the **Average Treatment on the Treated**.

$$\text{ATT}_Y \equiv \mathbb{E}\left[Y_i(1) - Y_i(0) \middle| \, S_i = 1; z_{1i}, \ldots, z_{(Q-1)i}\right]$$

Causal effects and the CEF

Econometricians often estimate the **derivative of the CEF**:

$$\mu^{q}_{Y_p|\boldsymbol{z}}\left(z_{qi}, \boldsymbol{z}_{-qi}\right) \equiv \frac{\partial}{\partial z_q} \mathbb{E}\left[Y_{pi} \middle| Z_{1i}, Z_{2i}, \dots, Z_{Qi} \right] \Bigg|_{Z_{qi}=z_{qi}}$$

with $q = 1, \dots, Q$. How does this compare with the ACE?

The two quantities are *generally* **not** equal.

$$\begin{aligned}
\mu^{q}_{Y_p|\boldsymbol{z}}\left(z_{qi}, \boldsymbol{z}_{-qi}\right) = & \\
&= \frac{\partial}{\partial z_{qi}} \int_{\mathbb{X}_{\boldsymbol{\varepsilon}}} r_p\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) f_{\boldsymbol{\varepsilon}|\boldsymbol{z}}\left(\boldsymbol{\varepsilon}_i \middle| z_{qi}, \mathbf{z}_{-qi}\right) d\varepsilon_{1i} \dots d\varepsilon_{Pi} \\
&= \int_{\mathbb{X}_{\boldsymbol{\varepsilon}}} r_p\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) \left[\frac{\partial}{\partial z_{qi}} f_{\boldsymbol{\varepsilon}|\boldsymbol{z}}\left(\boldsymbol{\varepsilon}_i \middle| z_{qi}, \mathbf{z}_{-qi}\right) \right] d\varepsilon_{1i} \dots d\varepsilon_{Pi} + \\
&\quad + \underbrace{\int_{\mathbb{X}_{\boldsymbol{\varepsilon}}} \left[\frac{\partial}{\partial z_{qi}} r_p\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right) \right] f_{\boldsymbol{\varepsilon}|\boldsymbol{z}}\left(\boldsymbol{\varepsilon}_i \middle| z_{qi}, \mathbf{z}_{-qi}\right) d\varepsilon_{1i} \dots d\varepsilon_{Pi}}_{=\mathrm{ACE}_{qp}\left(z_{qi}, \mathbf{z}_{-qi}\right)}
\end{aligned}$$

Conditional Independence Assumption

Hence, the ACE and the CEF derivative coincide only if:

$$\int_{\mathbb{X}_{\boldsymbol{\varepsilon}}} r_p\left(z_{qi}, \mathbf{z}_{-qi}, \boldsymbol{\varepsilon}_i\right)\left[\frac{\partial}{\partial z_{qi}} f_{\boldsymbol{\varepsilon}|\boldsymbol{z}}\left(\boldsymbol{\varepsilon}_i \middle| z_{qi}, \mathbf{z}_{-qi}\right)\right] d\varepsilon_{1i} \dots d\varepsilon_{Pi} = 0$$

which is ensured under the following condition.

Definition 16

Conditional Independence Assumption (CIA). The CIA is the hypothesis that the the unobservables $\boldsymbol{\varepsilon}_i$ and a specific exogenous variable Z_{qi} are statistically independent, conditional upon all the other exogenous variables $\boldsymbol{z}_{-qi}$.

$$Z_{qi} \perp \boldsymbol{\varepsilon}_i | \, \boldsymbol{z}_{-qi}$$

- For binary treatments, the CIA can be written as follows.

$$Y_i\left(1\right), Y_i\left(0\right) \perp S_i | \, Z_{1i}, \dots, Z_{(Q-1)i}$$

Causality in linear models (1/2)

- If the CEF of interest in linear:

  $$\mathbb{E}\left[Y_{pi} \middle| Z_{1i}, Z_{2i}, \ldots, Z_{Qi}\right] = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \ldots \beta_Q Z_{Qi}$$

  and the CIA holds for Z_{qi}, then $\beta_q = \text{ACE}_{qp}\left(z_{qi}, \mathbf{z}_{-qi}\right)$ for any $q = 1, \ldots, Q$ of interest.

- The CIA is weaker than unconditional independence, but it is closely linked to the condition for statistical identification (mean independence of the error term, *à la* $\mathbb{E}\left[\varepsilon_i \middle| X_i\right] = 0$).

- Thus, "identification" and "causality" are often confused.

- An important special case is the one where the regressor of interest S_i is binary and satisfies $\mathbb{E}\left[S_i \middle| \boldsymbol{x}_i\right] = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\pi}_0$ for some exogenous variables $\boldsymbol{x}_i$ (like in fully saturated models).

  $$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \delta_0 s_i + \varepsilon_i$$

Causality in linear models (2/2)

- In this case, the CEF derivative *is* the ATE.

$$\mu'_{Y|S,\boldsymbol{x}} = \mathbb{E}\left[\,Y_i|\,S_i = 1, \boldsymbol{x}_i\right] - \mathbb{E}\left[\,Y_i|\,S_i = 0, \boldsymbol{x}_i\right] \equiv \Delta\left(\boldsymbol{x}_i\right)$$

- Thus, by the Yitzakhi-Angrist-Krueger decomposition:

$$\delta^* = \frac{\mathbb{E}_{\boldsymbol{x}}\left[\Delta\left(\boldsymbol{x}_i\right)\mathbb{P}\left(\,S_i = 1|\,\boldsymbol{x}_i\right)\left[1 - \mathbb{P}\left(\,S_i = 1|\,\boldsymbol{x}_i\right)\right]\right]}{\mathbb{E}_{\boldsymbol{x}}\left[\mathbb{P}\left(\,S_i = 1|\,\boldsymbol{x}_i\right)\left[1 - \mathbb{P}\left(\,S_i = 1|\,\boldsymbol{x}_i\right)\right]\right]}$$

  where $\phi\left(\boldsymbol{x}_i\right) = \mathbb{P}\left(\,S_i = 1|\,\boldsymbol{x}_i\right)\left[1 - \mathbb{P}\left(\,S_i = 1|\,\boldsymbol{x}_i\right)\right]$.

- The linear projection δ^* identifies the ACE of S_i on Y_i if:
  1. the causal effect $\Delta\left(\boldsymbol{x}_i\right)$ is constant over $\boldsymbol{x}_i$, *or*;
  2. the probability to "take up the treatment" ($S_i = 1$) is constant for $\boldsymbol{x}_i$.

  Else, δ^* carries the usual approximation interpretation.

Triangular models

In some cases, causal effects can also be defined for *endogenous* variables. This happens in so-called **triangular models**.

Definition 17

Triangular Models. A *triangular* structural model is one where its P equations and its P endogenous variables can be ordered in such a way that, for any natural number $P' < P$, the first P' endogenous variables never enter the last $P - P'$ equations, or vice versa.

The simplest triangular model is perhaps the following.

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i \\ X_i &= \pi_0 + \pi_1 Z_i + \eta_i \end{aligned}$$

Generally a linear SEM is triangular if matrix $\mathbf{\Gamma}$ is either upper- or lower-triangular. The Mincer model is triangular.

Causal effects like that of X_i on Y_i above, or education on wages in the Mincer case, are well-defined in models of this sort.