Lecture 1

Identification

Michael Curran

Trinity College Dublin

JS Econometrics

Lecture 1 Outline

Introduction

Overview of HT Modules

1. First Half – Michael Curran (Further Topics in Econometrics)
2. Second Half – Agustín Bénétrix (Time Series Econometrics)

Topics to be Covered

Lecturer: Michael Curran

Lec 1: Identification (slides)

- i) Incomplete Data
- ii) Treatment Response

Lec 2–6: Limited Independent & Dependent Variables (Wooldridge, 7 & 17)

- Lec 2: Binary (Dummy) Explanatory Variables
- Lec 3: Binary Response I: Dummy Dependent Variables (LPM)
- Lec 3: Application: Policy Analysis
- Lec 4: Binary Response II: Logit & Probit Models
- Lec 5: Corner Solutions / Threshold Models: Tobit Model
- Lec 5: Count Models: Poisson Model
- Lec 6: Censored & Truncated Models
- Lec 6: Sample Selection Corrections

Lec 7–10: Endogeneity (Wooldridge, 15 & 16)

- Lec 7–8: Instrumental Variable Estimation & Two Stage Least Squares
- Lec 9–10: Simultaneous Equation Models – early studies on identification

Lecture 1 Outline

Identification

- Combining models and data, we draw conclusions.
- The credibility of our conclusions typically diminishes with the strength of the assumptions of our models.
- **Identification** problems concern conclusions we could draw from models where data is at the **population** level ($N = \infty$), while **inference** problems concern conclusions we draw using models with **sample** data.
- Examples of identification problems: reflection problem, death penalty, missing data – not disappear by increasing the size of the sample.
- Extrapolation, counterfactuals and external validity.

Identification

$$y = x'\beta + \epsilon \quad E(\epsilon|x) = 0 \tag{1}$$

Parameter $b \in \mathbb{R}^k$ is **identified relative to** β if

$$P_X\{x : x'b \neq x'\beta\} > 0$$

In model (1), β is **point identified** if $\forall b \neq \beta$, b is identified relative to β.

See example in class.

Conditional Prediction

Goal: predict $P(y|x)$.
Example: death penalty.
The **best predictor** p of the random variable Y given other random variables X minimises a **loss function** $\mathcal{L}(\cdot)$, say

$$\min_p E[\mathcal{L}(y-p)|x]$$

Let $u = y - p$. Then

$$p = \begin{cases} \mu \text{ (mean)} & \text{if } \mathcal{L}(u) = u^2 \\ m \text{ (median)} & \text{if } \mathcal{L}(u) = |u| \end{cases}$$

$$m = \min_\theta \left\{ \theta : P(y \le \theta) \ge \frac{1}{2} \right\}$$

Conditional Prediction

$$P_N[(y,x) \in A] = \frac{1}{N}\sum_{i=1}^{N} 1[(y_i, x_i) \in A] \xrightarrow{as} P[(y,x) \in A]$$

t is in the **support of** P if

$$P(t - \delta \leq y \leq t + \delta) > 0 \; \forall \delta > 0$$

$$P_N(y \in B | x = x_0) = \frac{\frac{1}{N}\sum_{i=1}^{N} 1[y_i \in B, x_i = x_0]}{\frac{1}{N}\sum_{i=1}^{N} 1[x_i = x_0]} \xrightarrow{as} P(y \in B | x = x_0)$$

$$E_N(y | x = x_0) = \frac{\frac{1}{N}\sum_{i=1}^{N} y_i \cdot 1[x_i = x_0]}{\frac{1}{N}\sum_{i=1}^{N} 1[x_i = x_0]} \xrightarrow{as} E(y | x = x_0)$$

Conditional Prediction

Bandwidth: d_N.

Local average / uniform kernel estimate:

$$\theta_N(x_0, d_N) = E_N(y|x = x_0) = \frac{\frac{1}{N}\sum_{i=1}^{N} y_i \cdot 1[\rho(x_i, x_0) < d_N]}{\frac{1}{N}\sum_{i=1}^{N} 1[\rho(x_i, x_0) < d_N]}$$

Local weighted average / kernel estimate:

$$E_N(y|x = x_0) = \frac{\frac{1}{N}\sum_{i=1}^{N} y_i K\left[\frac{\rho(x_i, x_0)}{d_N}\right]}{\frac{1}{N}\sum_{i=1}^{N} K\left[\frac{\rho(x_i, x_0)}{d_N}\right]}$$

Example: predicting high school graduation – see in class.

Incomplete Data

For a sample space Ω (set of all outcomes of an experiment), events $B_1, \dots, B_N$ where $B_i \in \Omega \; \forall i$ **partition** Ω if:

1. $B_i \cap B_j = \emptyset \; \forall i \neq j$ (pairwise disjoint).
2. $\cup_i B_i = \Omega$ (cover).

Let $P(B_i) > 0$ for all events in the partition. Then for any event A, we have the **law of total probability**:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Let y be the outcome to be predicted, x be covariates and define $z = 1$ if y is observed and $z = 0$ otherwise.
Express the missing data problem via law of total probability:
$P(y|x) =$

Incomplete Data

$P(y|x) =$
Let $P(y|x, z = 0) = \gamma \in \Gamma_Y$
$\Gamma_Y =$ set of all probability distributions on the set Y.
Identification region for $P(y|x)$:

$$H[P(y|x)] = [P(y|x, z = 1)P(z = 1|x) + \gamma P(z = 0|x); \gamma \in \Gamma_Y]$$

$$P(z = 0|x) < 1 \Longrightarrow H[\cdot] \subsetneq \Gamma_Y$$

$P(y|x)$ **partially identified** when $0 < P(z = 0|x) < 1$ and **point identified** when $P(z = 0|x) = 0$.

Incomplete Data

Let $\theta(\cdot)$ map probability distributions on Y into $\mathbb{R}$ and consider parameter $\theta[P(y|x)]$.
Identification region: $H\{\theta[P(y|x)]\} = \{\theta(\eta), \eta \in H[P(y|x)]\}$.
Event probabilities:

$$P(y \in B|x) \overset{\text{LTP}}{=} P(y \in B|x, z=1)P(z=1|x) + \underbrace{P(y \in B|x, z=0)}_{\in[0,1]} P(z=0|x)$$

$$H[P(y \in B|x)] = [P(y \in B|x, z=1)P(z=1|x), P(y \in B|x, z=1)P(z=1|x) + P(z=0|x)]$$

Interval width: $P(z=0|x)$; hence, data is **informative** unless y is always missing.
See example in class.

Incomplete Data

Given the existence of expectations, the **law of iterated expectations** states $E_X(E(Y|X)) = E(Y)$. Equivalently:

$$E(Y) = E_X(E(Y|X)) = \sum_{x \in Supp(X)} E(Y|X=x)P(X=x)$$

Corollary of LIE: $E(E(h(Y,X)|X)) = E(h(Y,X))$, so

$$E[g(y)|x] \overset{\text{LIE}}{=} E[g(y)|x, z=1]P(z=1|x) + E[g(y)|x, z=0]P(z=0|x)$$

Incomplete Data

Data is assumed to be **missing at random** (MAR) or to be **conditionally statistically independent** if:

$$P(y|x, z = 1) = P(y|x, z = 0) = P(y|x)$$

Note that MAR $\implies E[y|x, z = 1] = E[y|x, z = 0] = E[y|x]$.
MAR is a **nonrefutable** assumption – restricts the distribution $P(y|x, z = 0)$ of missing data.

$$H_0[P(y|x)] \overset{\text{MAR}}{\equiv} P(y|x, z = 1)$$
$$H_0[E(y|x)] \overset{\text{MAR}}{\equiv} E(y|x, z = 1)$$

Incomplete Data

Data alone imply that $P(y|x) \subset H[P(y|x)]$. Combining the data with the assumption $P(y|x) \subset \Gamma_{1Y}$:

$$H_1[P(y|x)] \equiv H[P(y|x)] \cap \Gamma_{1Y}$$

$H_1 = 0 \implies$ assumption is refutable; $H_1 \neq 0 \implies$ assumption is nonrefutable, but we are not saying that the assumption is true!
$H_1[P(y|x)] \subsetneq H[P(y|x)] \implies$ assumption has **identifying power**.

Treatment Response

- What would the outcomes be if we were to apply some (possibly the same) treatment to a population?
- Examples: life-span if treatment is drugs or surgery, retraining versus job assistance.

Treatment Response

Notation:

- T: set of all feasible treatments $t \in T$ mutually exclusive and exhaustive.
- Each member j possesses covariates $x_j \in X$.
- Outcomes $y_j(t) \in Y$.
- Response function $y_j(\cdot) : T \longrightarrow Y$.
- $z_j \in T$ is j's **received** treatment so $y_j = y_j(z_j)$ are **realised** outcomes and $[y_j(t), t \neq z_j]$ are **counterfactual** outcomes.

Goal: infer $P[y(t)|x]$ given $P(y,z|x)$ – **selection problem** – problem of identification of $P[y(t)|x]$ given $P(y,z|x)$.

$P[y(t)|x] \overset{\text{LTP}}{=}$

$$H[P(y(t)|x)] = [P(y|x, z = t)P(z = t|x) + \gamma P(z \neq t|x); \gamma \in \Gamma_Y]$$

$$H\{P[y(t)|x], t \in T\} = \times_{t \in T} H\{P[y(t)|x]\}$$

Treatment Response

For two treatments t and t', the **average treatment effect** (ATE) is:

$$E[y(t)|x] - E[y(t')|x]$$

Hypothesis: $ATE = 0$ is nonrefutable.
Randomisation of treatment:

$$P(y(t)|x) = P(y|x, z = t) = P(y(t)|x, z \neq t)$$

gives point identification and so hypothesis that $ATE = 0$ becomes refutable.
Exercise: assuming $y(t) \in [y_0, y_1]\ \forall t$, bound ATE. Hint: use LIE and observability of realised outcomes. What is the identification region for ATE? Does it always contain zero and what is its width?

Lecture 1 Outline

Summary

- Problem of identification: population level.
- Conditional prediction:
    - Mean (median) is best predictor for square (absolute) loss function.
    - Empirical nonparametric estimation uses analogy principle via kernels and choice of bandwidth.
- Incomplete data:
    - Use law of total probability and law of iterated expectations to bound probability.
    - Identification is not binary and data is informative unless outcomes always missing.
    - Missingness at random is a common nonrefutable assumption.
    - Combining data and assumptions: strong assumptions tend to have identifying power.
- Treatment response:
    - ATE is a useful statistic for the selection problem where randomisation of treatment is the analog to missing at randomness.

References

- Identification: these slides and material from week 1 laboratory session.