

W2 - 4.1 & 4.2

4.1 Overview of the single equation linear model

Structural model - Represents causal relationship (from theory)

Estimable model - Same as structural if it is estimable.

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u \quad \left\{ \begin{array}{l} \text{Omitted variables} \\ \text{Measurement error} \end{array} \right\} \quad \text{If } E[u] = 0 \text{ \& } \text{Cov}(u, x_i) = 0 \Rightarrow \lim_{n \rightarrow \infty} \hat{\beta} = \beta$$

$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$ + $\text{Cov}(u, x_i) = 0 \Rightarrow$ Defining linear projection of y onto $(1, x_1, \dots, x_k)$ as $(\beta_0, \beta_1 x_1, \dots, \beta_k x_k)$

Sufficient for $\text{Cov}(u, x_i) = 0$: $E[u | x_1, x_2, \dots, x_k] = E[u | \tilde{x}] = 0$

$$\Rightarrow E[y | x_1, \dots, x_k] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

x_i is endogenous in $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$ if $\text{Cov}(u, x_i) \neq 0$.

- 3 types:
- ① Omitted variables: For $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$ & x_2 is unobservable, we want $E[y | x_1, x_2]$ but have $E[y | x_1]$.
 $\hookrightarrow$ If $\text{Cov}(x_1, x_2) = 0 \Rightarrow E[y | x_1, x_2] = E[y | x_1]$
 $\hookrightarrow$ If $\text{Cov}(x_1, x_2) \neq 0 \Rightarrow E[y | x_1, x_2] \neq E[y | x_1]$
 - ② Measurement error: We want to measure x_k^* but we have an imperfect measure $x_k \Rightarrow$ Imperfection is included in u of estimable model.
 - ③ Simultaneity: At least one x is determined simultaneously w/ y . E.g: $x_k = f(y) \Rightarrow E[u | x] \neq 0$.

4.2 Asymptotic properties of OLS

Consider $y = x\beta + u$

- ① Consistency: $E[x'u] = 0 \rightarrow$ Population orthogonality condition ①
 $\hookrightarrow$ Sufficient: $E[u | x] = 0$

$\text{rank } E[x'x] = K \rightarrow x'x$ has full rank = $E[x'x]$ is positive definite
 $\hookrightarrow$ Fails $\Leftrightarrow x_i$ can be written as a linear f of x_j . ②

If $E[x'u] = 0$ & $\text{rank } E[x'x] = K \Rightarrow \beta$ is identified. $\Rightarrow$ Can be written in terms of population moments in observable variables.

Dem: Consider $y = x\beta + u$. Premultiply by $x' \Rightarrow x'y = x'x\beta + x'u$

$$\text{Take } E[x'y] = E[x'x]\beta + E[x'u] \Rightarrow \beta = (E[x'x])^{-1} E[x'y]$$

$x'y$ is observable $\Rightarrow \beta$ is identified.

Teo: Under assumptions ① & ②, the OLS $\hat{\beta}$ obtained from a random sample following $y = x\beta + u$ is consistent for β .

If u and x are independent $\Rightarrow E[u|x] = 0$.

② Asymptotic Inference Using OLS

Pregunta central: Efecto de x sobre y

- ↳ Educ - Salarios
- ↳ Cigarrillo - salud
- ↳ Programa - Beneficiarios

Econometría { Paramétrica: Impongo forma funcional
No paramétrica: Datos imponen forma función

EC. POBLACIONAL

$$Y = X\beta + u \rightarrow \text{Aleatorio, v. independientes, no observado, no medido}$$

↳ Parámetro poblacional de interés

↳ Variables independientes - Observo y puedo medir

V. Dependiente

$\hat{\beta}$ es una variable aleatoria

Pero no hay datos poblacionales, sino de muestras (e.g: GEIH)
↳ $\hat{\beta}$ estimador de β = Depende de la muestra = $f(X, Y)$ (lo que sí observo en mi muestra) = Es una v. aleatoria

Propiedades $\hat{\beta}$

$\hat{\beta} = f(X, Y)$ requiere ciertas propiedades:

(1) Función objetivo

① $f(\cdot)$ responde a una función objetivo que tiene sentido.

P.e: $Y = X\beta + u \rightarrow \hat{\beta} = (X'X)^{-1}X'Y$ es estimador de β

$\hat{Y} = X\hat{\beta} \rightarrow \hat{Y}$ es estimador de $E[Y|X] \Rightarrow \hat{u} = Y - \hat{Y}$

$\Rightarrow \text{MCO} \min \sum_{i=1}^n \hat{u}_i^2 \Rightarrow f(X, Y) = (X'X)^{-1}X'Y$

(2) Insensibilidad

② $E[\hat{\beta}] = \beta$ - Insensibilidad - no depende del tamaño muestral

(3) Consistencia

③ $\text{plim } \hat{\beta} = \beta$ - Consistencia - $\lim_{N \rightarrow \infty} \Pr(|\hat{\beta} - \beta| > \epsilon) = 0$ - si depende del tamaño de muestra

(4) Distribución

④ Distribución de $\hat{\beta}$ para llevar a cabo pruebas de hipótesis.
↳ Distribución asintótica

P.e: $Y = X\beta + u$. Asumimos $u \sim N(0, \sigma^2) \Rightarrow \hat{\beta} \sim N(\beta, \text{var}(\hat{\beta}))$

El punto es eliminar el supuesto $u \sim N(0, \sigma^2)$

MCO

$$Y = X\beta + u$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix}_{n \times (k+1)} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{n \times 1}$$

$$\hat{Y} = X\hat{\beta}, \hat{u} = Y - \hat{Y} \Rightarrow \hat{\beta}_{\text{MCO}} = \underset{\hat{\beta}}{\text{argmin}} \hat{u}'\hat{u}$$

$$\mathcal{L} = \min_{\hat{\beta}} (Y - X\hat{\beta})'(Y - X\hat{\beta}) = \min_{\hat{\beta}} (Y' - \hat{\beta}'X')(Y - X\hat{\beta})$$

$$\frac{\partial \mathcal{L}}{\partial \hat{\beta}} = \frac{d}{d\hat{\beta}} (Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta})$$

$$= -2X'Y + 2(X'X)\hat{\beta} = 0 \Rightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

Econometría: Conjunto de herramientas matemáticas y estadísticas para responder una pregunta económica

- Paramétrica: Impongo una forma funcional al modelo.
- No paramétrica: Los datos imponen la forma funcional del modelo.

Objetivo central: Obtener la magnitud de los efectos de variables independientes sobre dependientes = Estimar β .

↳ Sin embargo, estimamos sobre la base de una muestra, no una población
 $\Rightarrow \hat{\beta} = f(N)$

En consecuencia, debemos exigirle a $\hat{\beta}$ el cumplimiento de ciertas propiedades:

- ① Que se obtenga de una **función objetivo** con fundamentos teóricos, p.e.: $\min \sum u_i^2$.
- ② Que sea **insesgado**, es decir, que promediando los distintos $\hat{\beta}$ obtenidos de distintas muestras se obtenga el β poblacional.
- ③ Que sea **consistente**, es decir, que a medida que se aumente el tamaño de la misma muestra, se aproxime en probabilidad al β poblacional.
- ④ Que tenga una **distribución** para llevar a cabo pruebas de hipótesis.

W2. Chapter 3 - Basic Asymptotic Theory

I. Convergence of deterministic squares

Convergence

Def 1: ① A sequence of nonrandom numbers $\{a_n: n=1, 2, \dots\}$ converges to a if for all $\epsilon > 0$, $\exists N_\epsilon: n > N_\epsilon \Rightarrow |a_n - a| < \epsilon$. $a_n \rightarrow a$ when $n \rightarrow \infty$. *

Bounded

② A sequence of nonrandom numbers is bounded $\Leftrightarrow \exists b < \infty: |a_n| \leq b \forall n = 1, 2, \dots$ *

→ This applies to vectors & matrices element by element

 $O(N^\lambda)$

Si una secuencia converge $\Rightarrow$ tiene cota (no pasa al contrario).

 $O(N^\lambda)$

Def 2: ① A sequence $\{a_n\}$ is $O(N^\lambda)$ (at most of order N^λ) if $N^{-\lambda}a_n$ is bounded. When $\lambda = 0 \Rightarrow \{a_n\}$ is bounded & we write $a_n = O(1)$

② $\{a_n\}$ is $o(N^\lambda)$ if $N^{-\lambda}a_n \rightarrow 0$. When $\lambda = 0 \Rightarrow \{a_n\}$ converges to zero & we write $a_n = o(1)$.

E.g: $a_n = \log N \Rightarrow N^{-\lambda}a_n = N^{-\lambda} \log N \rightarrow 0$ if $N \rightarrow \infty \Rightarrow a_n = o(N^\lambda) \forall \lambda > 0$.

II. Convergence in probability & bounded in probability

Convergence in probability

Def 3: ① A sequence of random variables $\{x_n\}$ converges in probability to the constant a if $\forall \epsilon > 0$: $P[|x_n - a| \geq \epsilon] \rightarrow 0$ as $n \rightarrow \infty$.

 $Op(1)$

② If $a = 0 \Rightarrow \{x_n\}$ is $Op(1)$

Bounded in probability

③ A sequence of random variables $\{x_n\}$ is bounded in probability $\Leftrightarrow \forall \epsilon > 0 \exists b_\epsilon < \infty$ & N_ϵ integer: $P[|x_n| \geq b_\epsilon] < \epsilon \forall n \geq N_\epsilon$

→ If a sequence converges to any real number $\Rightarrow$ It is bounded in probability. *

$Op(1) \cdot O_p(1) = O_p(1)$
 $Op(1)$

Def 4: ① A random sequence $\{x_n: n=1, 2, \dots\}$ is $Op(a_n)$ where $\{a_n\}$ is a nonrandom, positive sequence, if $x_n/a_n = Op(1) \Rightarrow x_n = Op(a_n)$.

 $Op(1)$

② A random sequence $\{x_n: n=1, 2, \dots\}$ is $Op(a_n)$, where $\{a_n\}$ is a non-random, positive sequence if $x_n/a_n = Op(1)$

Slutsky's theorem

* Slutsky's theorem

Let $g: \mathbb{R}^K \rightarrow \mathbb{R}^J$ be a function continuous at some point $c \in \mathbb{R}^K$. Let x_n be a sequence of $K \times 1$ random vectors such that $x_n \xrightarrow{p} c \Rightarrow \text{plim } g(x_n) = g(\text{plim } x_n)$

With probability approaching 1 w.p.a.o.

Def 5: Let $(\Omega, \mathcal{F}, P)$ be a probability space. A sequence of events $\{Q_n\} \subset \mathcal{F}$ is said to occur with probability approaching 1 $\Leftrightarrow P(Q_n) \rightarrow 1$ as $n \rightarrow \infty$

-3.12

Método
DeltaIf $Z_N \xrightarrow{p} A$
 $\Rightarrow Z_N^{-1} \xrightarrow{p} A^{-1}$

Corolario 3.1 Let $\{Z_N: N=1, 2, \dots\}$ be a sequence of random $K \times K$ matrices, and let A be a nonrandom, invertible $K \times K$ matrix. If $Z_N \xrightarrow{p} A \Rightarrow$ ① Z_N^{-1} exists w.p.a.1
② $Z_N^{-1} \xrightarrow{p} A^{-1}$ or $\text{plim } Z_N^{-1} = A^{-1}$

III. Convergence in distribution

Convergence in
distribution
(variables)

Def 6: A sequence of random variables $\{x_N\}$ converges in distribution to the continuous random variable $x \Leftrightarrow F_N(\xi) \rightarrow F(\xi)$ as $N \rightarrow \infty \forall \xi \in \mathbb{R}$, where F_N is the cumulative distribution function of x_N and F is the cumulative distribution function of x . We write $x_N \xrightarrow{d} x$.

Convergence in
distribution
(vectors)

Def 7 A sequence of $K \times 1$ random vectors $\{x_N\}$ converges in distribution to the continuous random vector $x \Leftrightarrow \forall K \times 1$ nonrandom vector $c: c'c = 1, c'x_N \xrightarrow{d} c'x$ & we write $x_N \xrightarrow{d} x$

? ¿Qué tiene la condición de c ?

↳ When $x \sim N(m, V)$, Def 7 requires $c'x_N \xrightarrow{d} N(c'm, c'Vc) \forall c \in \mathbb{R}^K: c'c = 1$.

? Si $x_N \xrightarrow{d} x \Rightarrow$
Converge en p?
Continuous
mapping
theorem

Lemma 5 If $x_N \xrightarrow{d} x \Rightarrow x_N = O_p(1)$

Continuous mapping theorem Let $\{x_N\}$ be a sequence of $K \times 1$ random vectors such that $x_N \xrightarrow{d} x$. If $g: \mathbb{R}^K \rightarrow \mathbb{R}^3$ is a continuous function, then $g(x_N) \xrightarrow{d} g(x)$

Corollary 3.2 If $\{z_N\}$ is a sequence of $K \times 1$ random vectors such that $z_N \xrightarrow{d} N(0, V) \Rightarrow$ ① $\forall K \times M$ nonrandom $A: A'z_N \xrightarrow{d} N(0, A'VA)$
② $z_N'V^{-1}z_N \xrightarrow{d} \chi_K^2$

Asymptotic
equivalence
lemma

Asymptotic equivalence lemma Let $\{x_N\}$ and $\{z_N\}$ be sequences of $K \times 1$ random vectors. If $z_N \xrightarrow{d} z$ and $x_N - z_N \xrightarrow{p} 0 \Rightarrow x_N \xrightarrow{d} z$

IV. Limit theorems for random samples

Two classic limit theorems for i.i.d sequences of random vectors. These apply when sampling is done randomly from a population.

Weak Law of
Large Numbers

Theo 1 Let $\{w_i\}$ be a sequence of i.i.d $G \times 1$ random vectors such that $E(|w_i|) < \infty, g = 1, 2, \dots, G \Rightarrow$ The sequence satisfies the weak law of large numbers (WLLN):
 $N^{-1} \sum_{i=1}^N w_i \xrightarrow{p} \mu_w$, where $\mu_w = E[w_i]$.

Lindeberg Levy
Theorem
(Central Limit
Theorem)

Theo 2 (Lindeberg Levy) Let $\{w_i\}$ be a sequence of independent, identically distributed $G \times 1$ random vectors such that $E(w_i^2) < \infty, g = 1, \dots, G$ and $E(w_i) = 0$. Then $\{w_i\}$ satisfies the central limit theorem (CLT):

$$N^{-\frac{1}{2}} \sum_{i=1}^N w_i \xrightarrow{d} \text{Normal}(0, B)$$

Where $B = \text{Var}(w_i) = E(w_i w_i')$ is necessarily positive semidefinite.

V. Limiting behaviour of Estimators and Test Statistics

Application of previous concepts to sequences of estimators.

(i) Asymptotic properties of estimators

Consistent estimator

Def 8 Let $\{\hat{\theta}_N\}$ be a sequence of estimators of the $P \times 1$ vector $\theta \in \Theta$ where N is the sample size. If $\hat{\theta}_N \xrightarrow{p} \theta$ for any value of $\theta \Rightarrow \hat{\theta}_N$ is a consistent estimator of θ .

$\sqrt{N}$ -asymptotically normally distributed

Def 9 Let $\{\hat{\theta}_N\}$ be a sequence of estimators of θ . Suppose that $\sqrt{N}'(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$ where V is a positive semidefinite $P \times P$ matrix. Then we say that $\hat{\theta}_N$ is $\sqrt{N}$ -asymptotically normally distributed and V is the asymptotic variance of $\sqrt{N}'(\hat{\theta}_N - \theta)$, denoted $AVar \sqrt{N}'(\hat{\theta}_N - \theta) = V$.

Asymptotic standard error

Def 10 If $\sqrt{N}'(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$ where V is positive definite with j th diagonal v_{jj} , and $\hat{V}_N \xrightarrow{p} V \Rightarrow$ The asymptotic standard error of $\hat{\theta}_N$; $se(\hat{\theta}_N)$ is $(\hat{v}_{Nj}/N)^{1/2}$.

$\sqrt{N}$ -consistent estimator

$\hookrightarrow$ If $\sqrt{N}'(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$ (*) lemma 5 $\Rightarrow \sqrt{N}(\hat{\theta}_N - \theta) = o_p(1)$ or $\hat{\theta}_N - \theta = o_p(N^{-1/2}) \Rightarrow \hat{\theta}_N$ is a $\sqrt{N}$ -consistent estimator of θ .
 $\hookrightarrow$ Implies $\text{plim } \hat{\theta}_N \rightarrow \theta$

Def 11 Let $\hat{\theta}_N$ and $\tilde{\theta}_N$ be estimators of θ satisfying (*) with asymptotic variances $V = AVar \sqrt{N}'(\hat{\theta}_N - \theta)$ and $D = AVar \sqrt{N}'(\tilde{\theta}_N - \theta)$

Asymptotically efficient relative to...

(1) $\hat{\theta}_N$ is asymptotically efficient relative to $\tilde{\theta}_N$ if $D - V$ is positive semidefinite $\forall \theta$.

$\sqrt{N}$ equivalent

(2) $\hat{\theta}_N$ and $\tilde{\theta}_N$ are $\sqrt{N}$ -equivalent if $\sqrt{N}(\hat{\theta}_N - \tilde{\theta}_N) = o_p(1)$.

(ii) Asymptotic properties of test statistics

Asymptotic size

Def 13 (1) The asymptotic size of a testing procedure is the limiting probability of rejecting H_0 when it is true: $\lim_{N \rightarrow \infty} P_N(\text{reject } H_0 | H_0)$.

consistent

(2) A test is said to be consistent against the alternative H_1 if H_0 is rejected with probability approaching one when H_1 is true: $\lim_{N \rightarrow \infty} P_N(\text{reject } H_0 | H_1) = 1$

Lemma 8 Suppose that $\sqrt{N}'(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$ holds, where V is positive definite. Then for any nonstochastic $Q \times P$ matrix R ($Q \leq P$), with $\text{rank}(R) = Q$,

$$\Rightarrow \sqrt{N}' R(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, RVR')$$

$$\& [\sqrt{N}' R(\hat{\theta}_N - \theta)]^{-1} (RVR')^{-1} (\sqrt{N}' R(\hat{\theta}_N - \theta)) \xrightarrow{d} \chi^2_Q$$

In addition, if: $\text{plim } \hat{V}_N = V \Rightarrow [\sqrt{N}' R(\hat{\theta}_N - \theta)]^{-1} (R\hat{V}_N R')^{-1} (\sqrt{N}' R(\hat{\theta}_N - \theta)) = [\sqrt{N}' R(\hat{\theta}_N - \theta)]^{-1} R' [\hat{V}_N / N]^{-1} R (\hat{\theta}_N - \theta) \xrightarrow{d} \chi^2_Q$

Wald
statistic

(For testing H_0 against $H_1: R\theta \neq r$)
$$W_N = (R\hat{\theta}_N - r)' [R(\hat{V}_N/N)R']^{-1} (R\hat{\theta}_N - r)$$

↳ Under H_0 , $W_N \xrightarrow{d} \chi^2_Q$

↳ If we treat $\hat{\theta}_N \sim N(\theta, \hat{V}_N/N)$, we get W_N exactly

Lemma 9

Suppose $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$. Let $c: \Theta \rightarrow \mathbb{R}^Q$ be a continuously differentiable function on the parameter space $\Theta \subset \mathbb{R}^P$ where $Q \leq P$, and assume that θ is in the interior of the parameter space. Define $C(\theta) = \nabla_{\theta} c(\theta)$ as the $Q \times P$ Jacobian of c . Then

$$\begin{aligned} \sqrt{N}[c(\hat{\theta}_N) - c(\theta)] &\xrightarrow{d} N(0, C(\theta)VC(\theta)') \\ &\Leftrightarrow \{\sqrt{N}[c(\hat{\theta}_N) - c(\theta)]\}' [C(\theta)VC(\theta)']^{-1} \{\sqrt{N}[c(\hat{\theta}_N) - c(\theta)]\} \\ &\sim \chi^2_Q \end{aligned}$$

Define $\hat{C}_N = C(\hat{\theta}_N)$. Then $\text{plim } \hat{C}_N = C(\theta)$. If $\text{plim } \hat{V}_N = V$
 $\Rightarrow \{\sqrt{N}[c(\hat{\theta}_N) - c(\theta)]\}' [\hat{C}_N \hat{V}_N \hat{C}_N']^{-1} \{\sqrt{N}[c(\hat{\theta}_N) - c(\theta)]\} \sim \chi^2_Q$

Delta method

Econometría = Economía + Estadística → Traducción de hipótesis sociales a lenguaje económico.

MICRO

- Datos panel
- Corte transversal

MACRO

- Series de tiempo

#delimit

Para decirle a Stata que el comando acaba con ; :

#delimit ;

```
import excel
using "Casos Dengue",
clear ;
```

#delimit cr → Para que Stata sepa que quiero que vuelva a leer normalmente los comandos

Tipos de variables

STATA

String: Letras, se ponen entre comillas

Números:

Considere una base de datos

	x	y
1	x ₁	y ₁
2	x ₂	y ₂
3	x ₃	y ₃
4	x ₄	y ₄
5	x ₅	y ₅
6	x ₆	y ₆
7	x ₇	y ₇
8	x ₈	y ₈
9	x ₉	y ₉
10	x ₁₀	y ₁₀
11	x ₁₁	y ₁₁
12	x ₁₂	y ₁₂
13	x ₁₃	y ₁₃

→ list x y in 1/3

→ list x y if x == "x₇"

Uso de 'list'

Organizar obs

Organizar observaciones: sort [var] → Ordena en ascendente según el valor.

Google

Findit: Google de Stata

Archivos log

Crear archivos log: log using xx.log, replace

Ley de grandes números

Ley de Grandes Números (WLLN)

Sea $\{w_i: i=1, 2, \dots\}$ una secuencia de vectores aleatorios de $G \times 1$ independientes e idénticamente distribuidos (i.i.d, con el mismo proceso generador / distribución), donde $E(|w_i|) < \infty$, $g=1, \dots, G$. Entonces, la secuencia satisface la ley de grandes números:

$$\frac{1}{N} \sum_{i=1}^N w_i \xrightarrow{P} E(w_i) = \mu$$

E.g: Lanzamiento de un dado. X : Resultado del lanzamiento.

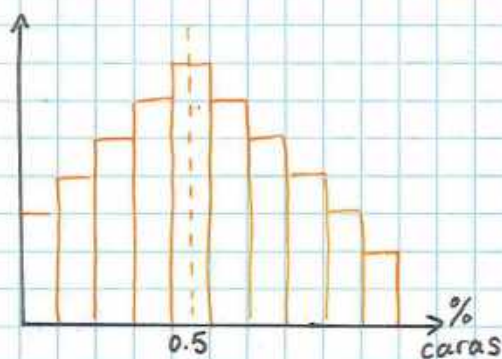
$$E(X) = \sum_x x \Pr(X=x) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \dots + 6\left(\frac{1}{6}\right) = 3.5$$

Ley de grandes números $\frac{1}{N} \sum_i w_i \xrightarrow{P} E[w_i] = \mu = 3.5$

Teorema del límite central

Teorema del Límite Central (CLT)

Frecuencias



Experimento aleatorio: Lanzar moneda 10 veces.

↳ Variable aleatoria: % de caras

Explica cuándo se distribuye una secuencia como normal.

Sea $\{w_i\}$ una secuencia de vectores aleatorios de $G \times 1$ que cumple:

① Son i.i.d

② $E(w_i^2) < \infty$, $g=1, 2, \dots, G$ y $E(w_i) = 0$

↳ $E(w_i) = 0$ no es un supuesto tan fuerte, se puede transformar la secuencia.

↳ Implica $\text{Var}(w_i) < \infty$ porque $\text{Var}(w_i) = E(w_i^2) - E(w_i)^2$

Entonces $\{w_i\}$ satisface el teorema del límite central.

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N w_i \xrightarrow{d} N(0, B) \quad \text{donde } B = \text{var}(w_i) = E(w_i w_i')$$

¿De dónde sale el $\frac{1}{\sqrt{N}}$?

Dem: $x_1, \dots, x_n$ es una variable aleatoria iid tal que $E[x] = \mu$, $\text{Var}(x) = \sigma^2$.

$$\sum_i x = z \quad E[z] = E\left[\sum_i x_i\right] = N\mu$$

$$\text{Var}[z] = \text{Var}\left[\sum_i x_i\right] = \sum_i \text{Var}(x_i) = N\sigma^2$$

Estandarizar z : $y = \frac{z - N\mu}{\sqrt{N}\sigma} \xrightarrow{\text{CLT}} N(0, 1)$

CLT $Y \xrightarrow{d} N(0,1)$

$\frac{Z - N\mu}{\sqrt{N}\sigma} \xrightarrow{d} N(0,1)$. Sin pérdida de generalidad, $\mu = 0$

$\frac{Z}{\sqrt{N}\sigma} \xrightarrow{d} N(0,1) \Rightarrow \frac{Z}{\sqrt{N}} \xrightarrow{d} N(0, \sigma^2)$

Consistencia

Def 8 - Consistencia

Sea $\{\hat{\theta}_N : N = 1, 2, \dots\}$ es una secuencia de estimadores de $\theta_{p \times 1}$ y N es el tamaño muestral.

Si $\hat{\theta}_N \xrightarrow{p} \theta$ ($\lim_{N \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$) $\Rightarrow \theta$ es un estimador consistente de θ .

Distribución normal asintótica a tasa $\sqrt{N}$

Def 9 - Distribución normal asintótica a tasa $\sqrt{N}$

Sea $\{\hat{\theta}_N : N = 1, 2, \dots\}$ un estimador de $\theta_{p \times 1}$. Si $\sqrt{N}'(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ donde V es la varianza asintótica de $\sqrt{N}'(\hat{\theta} - \theta)$ y es positiva semi definida (garantiza varianzas positivas).

$\Rightarrow \hat{\theta}_N$ se distribuye asintóticamente normal con $\text{Avar}[\sqrt{N}'(\hat{\theta} - \theta)] = V$

La media es cero si $\hat{\theta}$ es consistente, y la normalidad se mantiene.

$$\begin{aligned} \text{Avar}(\sqrt{N}'[\hat{\theta} - \theta]) &= N \text{Avar}(\hat{\theta} - \theta) = N(\text{Avar}[\hat{\theta}] + \text{Avar}[\theta] - 2\text{Cov}[\hat{\theta}, \theta]) \\ \text{Avar}(\sqrt{N}'[\hat{\theta} - \theta]) &= N \text{Avar}(\hat{\theta}) \\ V &= N \text{Avar}(\hat{\theta}) \\ \text{Avar}(\hat{\theta}) &= \frac{V}{N} \end{aligned}$$

Error estándar asintótico

Def 10 - Error estándar asintótico

Si $\sqrt{N}'(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$, V positiva semi definida y v_{jj} es el elemento diagonal de V , y $\hat{V}_N \xrightarrow{p} V \Rightarrow$ El error estándar asintótico de $\hat{\theta}_N$ es

$$\text{se}(\hat{\theta}_{Nj}) = \sqrt{\frac{v_{jj}}{N}}$$

$\sqrt{N}$ -consistencia

Lema: Si $\sqrt{N}'(\hat{\theta} - \theta) \xrightarrow{d} N(0, V) \Rightarrow \sqrt{N}'(\hat{\theta} - \theta) = O_p(1)$
 $\Rightarrow \hat{\theta}_N$ es un estimador $\sqrt{N}$ -consistente (mejor que consistente)

Eficiencia asintótica

Def 11 - Eficiencia asintótica

Sean $\hat{\theta}_N$ y $\tilde{\theta}_N$ estimadores de $\theta_{p \times 1}$ tal que ambos son consistentes y tienen varianza asintótica $V = \text{Avar}(\sqrt{N}'[\hat{\theta} - \theta])$ y $D = \text{Avar}(\sqrt{N}'[\tilde{\theta} - \theta])$

$\Rightarrow$ ① $\hat{\theta}_N$ es asintóticamente eficiente si $D - V$ es positiva semi-definida.

② $\hat{\theta}_N$ y $\tilde{\theta}_N$ son $\sqrt{N}$ equivalentes si $\sqrt{N}'(\hat{\theta} - \tilde{\theta}) = o_p(1)$.

Consistencia de MCO

Consistencia de MCO

$$Y = X\beta + u. \text{ Independencia Condicional implica } E[x'u] = 0 \\ E[u|x] = 0.$$

- ① β está identificado (lo puedo escribir como función de las variables observadas).

$$X'Y = X'X\beta + X'u \rightarrow E[X'Y] = E[X'X\beta] + E[X'u]$$

$$E[X'Y] = E[X'X]\beta \Rightarrow \beta = E[X'X]^{-1} E[X'Y]$$

- ② Método de momentos: Estimar β sustituyendo el promedio muestral en los momentos.

$$\hat{\beta} = (X'X)^{-1} (X'Y) = \left[\frac{1}{N} \sum_i x_i' x_i \right]^{-1} \left[\frac{1}{N} \sum_i x_i' y_i \right]$$

Donde x_i : i-ésima fila de X , y_i : i-ésimo elemento de Y

$$\hat{\beta} = (X'X)^{-1} X'Y = \beta + (X'X)^{-1} X'u$$

$$\hat{\beta} = \beta + \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_i x_i' u_i \right), u_i: i\text{-ésimo elemento de } u.$$

Por WLLN y el teorema de Slutsky:

$$\left(\frac{1}{N} \sum_i x_i' x_i \right) \xrightarrow{p} E(X'X) \Rightarrow \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \xrightarrow{p} E(X'X)^{-1}$$

$$\text{Por WLLN: } \left(\frac{1}{N} \sum_i x_i' u_i \right) \xrightarrow{p} E(X'u) = 0$$

$$\text{plim } \hat{\beta} = \text{plim} \left[\beta + \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_i x_i' u_i \right) \right]$$

Por teorema de Slutsky:

$$\begin{aligned} \text{plim } \hat{\beta} &= \text{plim } \beta + \text{plim} \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} * \text{plim} \left(\frac{1}{N} \sum_i x_i' u_i \right) \\ \text{plim } \hat{\beta} &= \beta + E(X'X)^{-1} * E(X'u) \\ \text{plim } \hat{\beta} &= \beta \end{aligned}$$

- ③ Supuestos necesarios: Independencia condicional e invertibilidad de $X'X$. (X debe ser de rango completo)

Distribución asintótica

Distribución asintótica

$$\hat{\beta} = \beta + \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_i x_i' u_i \right) \rightarrow \hat{\beta} - \beta = \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_i x_i' u_i \right)$$

$$\sqrt{N}(\hat{\beta} - \beta) = \sqrt{N} \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_i x_i' u_i \right)$$

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_i x_i' u_i \right)$$

Supuestos necesarios: IC, X de rango completo, $E(uu') = \sigma^2 I_{n \times n}$ (homoscedasticidad y no autocorrelación).

- ① Por WLLN y corolario 3.1: $\left(\frac{1}{N} \sum_i x_i' x_i \right)^{-1} \xrightarrow{p} E(X'X)^{-1} = A^{-1}$

- ② $\{x_i' u_i\}: i = 1, 2, \dots\}$ es i.i.d con media cero y asume varianza finita.

$$\Rightarrow \frac{1}{\sqrt{N}} \sum x_i' u_i \xrightarrow{d} N(0, B), \quad B = \text{var}(x_i' u_i) = E[u^2 x' x]$$

$$B = E[u^2] E[x' x]$$

$$B = \sigma^2 A$$

RESUMEN

Clases 2-3

Convergencia

Una secuencia determinística $\{a_n\}$ converge a a si $\forall \varepsilon > 0, \exists N_\varepsilon$ tal que si $N > N_\varepsilon \Rightarrow |a_n - a| < \varepsilon$. Además $a_n \rightarrow a$ cuando $N \rightarrow \infty$.

Acotada

Una secuencia determinística $\{a_n\}$ está acotada $\Leftrightarrow \exists b < \infty$ tal que $|a_n| \leq b \quad \forall N = 1, 2, \dots$

 $O(N^\alpha)$

Una secuencia determinística $\{a_n\}$ es $O(N^\alpha)$ si $N^{-\alpha} a_n$ está acotada.

 $o(N^\alpha)$

Una secuencia determinística $\{a_n\}$ es $o(N^\alpha)$ si $N^{-\alpha} a_n \rightarrow 0$

Convergencia en probabilidad

Una secuencia de variables aleatorias $\{x_n\}$ converge en probabilidad a la constante a si $\forall \varepsilon > 0, P[|x_n - a| > \varepsilon] \rightarrow 0$ si $N \rightarrow \infty$.

 $o_p(1)$

Si $P[|x_n| > \varepsilon] \rightarrow 0$ cuando $N \rightarrow \infty \Rightarrow x_n$ es $o_p(1)$.

Acotada en probabilidad

Una secuencia $\{x_n\}$ está acotada en probabilidad $\Leftrightarrow \forall \varepsilon > 0, \exists b_\varepsilon < \infty$ y un N_ε íntegro tal que $P[|x_n| \geq b_\varepsilon] < \varepsilon \quad \forall N \geq N_\varepsilon$.

 $o_p(a_n)$

Una secuencia $\{x_n\}$ es $o_p(a_n)$ si $\frac{x_n}{a_n}$ es $o_p(1)$.

 $O_p(1)$

Si $x_n \xrightarrow{p} a \Rightarrow x_n = O_p(1)$ (Lema 3.1)

 $O_p(a_n)$

Una secuencia $\{x_n\}$ es $O_p(a_n)$ si $\frac{x_n}{a_n}$ es $O_p(1)$.

Nota Una secuencia determinística a_n es $a_n = O_p(1) \Leftrightarrow a_n = O(1)$. ($\lambda = 0$)
 $a_n = o_p(1) \Leftrightarrow a_n = o(1)$. ($\lambda = 0$)

! Si una secuencia converge $\Rightarrow$ Está acotada. No al revés.

Teorema de Slutsky

Sea g una función continua en un punto c . Sea $\{x_n\}$ una secuencia de vectores $K \times 1$ tal que $x_n \xrightarrow{p} c$

$\Rightarrow g(\text{plim } x_n) = \text{plim } g(x_n)$ aunque g sea no lineal.

Corolario 3.1

Sea $\{z_n\}$ una secuencia aleatoria de matrices $K \times K$ y sea A una matriz determinística invertible de $K \times K$.

Si $z_n \xrightarrow{p} A \Rightarrow$ (1) z_n^{-1} existe con probabilidad cercana a 1
 (2) $z_n^{-1} \xrightarrow{p} A^{-1}$ o $\text{plim } z_n^{-1} = A^{-1}$

Convergencia en distribución (variables)

Una secuencia aleatoria $\{x_n\}$ converge en distribución a la variable aleatoria continua $x \Leftrightarrow F_n(\frac{x}{N}) \rightarrow F(\frac{x}{N})$ cuando $N \rightarrow \infty \quad \forall \frac{x}{N} \in \mathbb{R}$.
 F_n es la función de distribución acumulada de x_n y F es la función de distribución acumulada de x . Se escribe $x_n \xrightarrow{d} x$.

Convergencia en distribución (vectores)

Una secuencia $\{x_N\}$ de vectores $K \times 1$ (aleatoria) converge en distribución al vector aleatorio y continuo $x \Leftrightarrow \forall c$ tal que $c'c=1$ (c no aleatorio) se cumple $c'x_N \xrightarrow{d} c'x$. Escribimos $x_N \xrightarrow{d} x$

Lema 5

Si $x_N \xrightarrow{d} x \Rightarrow x_N = O_p(1)$

Teorema del mapeo continuo

Sea $\{x_N\}$ una secuencia aleatoria tal que $x_N \xrightarrow{d} x$. Si $g: \mathbb{R}^K \rightarrow \mathbb{R}^1$ es una función continua

$$\Rightarrow g(x_N) \xrightarrow{d} g(x) \text{ aunque } g \text{ sea no lineal}$$

Corolario 3.2

Si $\{z_N\}$ es una secuencia de vectores aleatorios de tamaño $K \times 1$ tal que $z_N \xrightarrow{d} N(0, V)$

$$\Rightarrow (1) \forall \text{ matriz no aleatoria } A \text{ de } K \times M: A'z_N \xrightarrow{d} N(0, A'VA) \\ (2) z_N'Vz_N \xrightarrow{d} \chi^2_K$$

Lema de equivalencia asintótica

Sean $\{z_N\}$ y $\{x_N\}$ dos secuencias aleatorias de vectores $K \times 1$. Si $z_N \xrightarrow{d} z$ y $x_N - z_N \xrightarrow{p} 0 \Rightarrow x_N \xrightarrow{d} z$.

Ley de los grandes números

Sea $\{w_i\}$ una secuencia de vectores aleatorios i.i.d de tamaño $G \times 1$ donde $E[|w_{ig}|] < \infty$, $g = 1, 2, \dots, G$. Entonces, la secuencia satisface la ley de los grandes números:

$$\frac{1}{N} \sum_{i=1}^N w_i \xrightarrow{p} E[w_i] = \mu$$

Teorema del límite central

Sea $\{w_i\}$ una secuencia de vectores aleatorios i.i.d de tamaño $G \times 1$ donde $E[w_{ig}^2] < \infty$, $g = 1, \dots, G$. Entonces, la secuencia satisface el teorema del límite central:

¿De dónde sale? $\frac{1}{\sqrt{N}} \sum_{i=1}^N w_i \xrightarrow{d} N(0, B)$, $B = \text{Var}(w_i) = E[w_i w_i']$

Considere $z = \sum_{i=1}^N x_i$. $E[z] = E[\sum_{i=1}^N x_i] = \sum_{i=1}^N E[x_i] = \sum_{i=1}^N \mu = N\mu$
 $\text{Var}[z] = \text{Var}[\sum_{i=1}^N x_i] = \sum_{i=1}^N \text{Var}(x_i) = \sum_{i=1}^N \sigma^2 = N\sigma^2$

Estandarizamos z : $y = \frac{z - N\mu}{\sqrt{N}\sigma} \xrightarrow{d} N(0, 1)$. Sin pérdida de generalidad, $\mu = 0$.

$$\frac{z}{\sqrt{N}\sigma} \xrightarrow{d} N(0, 1)$$

$$\frac{z}{\sqrt{N}} \xrightarrow{d} N(0, \sigma)$$

$$N^{-1/2} \sum_{i=1}^N x_i \xrightarrow{d} N(0, \sigma)$$

Consistencia

Sea $\{\hat{\theta}_N\}$ una secuencia de estimadores de $\theta_{p \times 1}$. Si $\hat{\theta}_N \xrightarrow{p} \theta$, es decir $P[|\hat{\theta}_N - \theta| > \epsilon] \rightarrow 0$ cuando $N \rightarrow \infty$, $\hat{\theta}_N$ es un estimador consistente de θ .

Distribución normal asintótica a tasa $\sqrt{N}$

Sea $\{\hat{\theta}_N\}$ una secuencia de estimadores de $\theta_{p \times 1}$. Si $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$, donde $V = A \text{Var}(\sqrt{N}(\hat{\theta}_N - \theta))$ y es positiva semi-definida (las varianzas son positivas)

$$\Rightarrow \hat{\theta}_N \text{ se distribuye asintóticamente } N(0, V) \rightarrow \hat{\theta}_N \overset{?}{\sim} N(0, V) \text{ a una tasa } \sqrt{N}.$$

Error estándar
asintótico.

Si $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$, V es positiva semi-definida y v_{jj} es el elemento diagonal de V , y $\hat{V}_N \xrightarrow{p} V \Rightarrow$ El error estándar asintótico de $\hat{\theta}_N$ es: $ee(\hat{\theta}_N) = \sqrt{\frac{v_{jj}}{N}}$

 $\sqrt{N}$ -Consistencia

Si $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V) \Rightarrow \sqrt{N}(\hat{\theta}_N - \theta) = O_p(1) \Rightarrow \hat{\theta}_N - \theta = O_p(N^{-1/2})$
 y $\hat{\theta}_N$ es un estimador $\sqrt{N}$ -consistente ($\text{plim } \hat{\theta}_N \rightarrow \theta$)

Eficiencia
asintótica

Sean $\hat{\theta}_N$ y $\tilde{\theta}_N$ estimadores de $\theta_{p \times 1}$ que cumplen $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, V)$ y $\sqrt{N}(\tilde{\theta}_N - \theta) \xrightarrow{d} N(0, D)$. Tienen varianza asintótica $V = \text{Avar}(\sqrt{N}(\hat{\theta}_N - \theta))$ y $D = \text{Avar}(\sqrt{N}(\tilde{\theta}_N - \theta))$

(1) $\hat{\theta}_N$ es asintóticamente eficiente en relación a $\tilde{\theta}_N$ si $D - V$ es positiva semidefinida $\forall \theta$.

(2) $\hat{\theta}_N$ y $\tilde{\theta}_N$ son $\sqrt{N}$ -equivalentes si $\sqrt{N}(\hat{\theta}_N - \tilde{\theta}_N) = o_p(1)$

Consistencia de MCO

Considere el modelo $y = X\beta + u$. Asumimos que: (1) Hay independencia condicional $\Rightarrow E[X'u] = 0$ y (2) X es de rango completo (no hay multicolinealidad perfecta).

$y = X\beta + u$ Multiplicando por X' a ambos lados:

$X'y = X'X\beta + X'u$ Tomando valor esperado:

$E[X'y] = E[X'X\beta] + E[X'u]$ Considerando que $E[X'u] = 0$ y β no aleatoria:

$E[X'y] = E[X'X]\beta$ Reacomodando términos:

$$\beta = (E[X'X])^{-1} E[X'y]$$

$\Rightarrow \beta$ está identificado: Se puede escribir como función de las variables observadas.

Usando el método de momentos (reemplazando los momentos muestrales en los poblacionales:

Reemplazando y :

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N x_i' y_i \right) = (X'X)^{-1} X'y \\ \hat{\beta} &= (X'X)^{-1} X'(X\beta + u) = (X'X)^{-1} X'X\beta + (X'X)^{-1} X'u \\ \hat{\beta} &= \beta + (X'X)^{-1} X'u = \beta + \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N x_i' u_i \right)\end{aligned}$$

Por ley de los grandes números: $\frac{1}{N} \sum x_i' x_i \xrightarrow{p} E[X'X]$
 Por corolario 3.1: $\left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \xrightarrow{p} (E[X'X])^{-1}$

Por ley de los grandes números: $\frac{1}{N} \sum x_i' u_i \xrightarrow{p} E[X'u] = 0$ por sup. (1)

Por tma. Slutsky $\Rightarrow \text{plim } \hat{\beta} = \text{plim } \left\{ \beta + \left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum x_i' u_i \right) \right\}$
 $\Rightarrow \text{plim } \hat{\beta} = \text{plim } \beta + \text{plim} \left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \text{plim} \left(\frac{1}{N} \sum x_i' u_i \right)$
 $\text{plim } \hat{\beta} = \beta + \text{plim} (E[X'X])^{-1} (E[X'u])$
 $\Rightarrow \text{plim } \hat{\beta} = \beta$

Distribución Asintótica de $\hat{\beta}$

Considere

$$\begin{aligned}\hat{\beta} &= \beta + \left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum x_i' u_i \right) \\ \hat{\beta} - \beta &= \left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum x_i' u_i \right) \\ \sqrt{N}(\hat{\beta} - \beta) &= \left[\left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum x_i' u_i \right) \right] \sqrt{N} \\ \Rightarrow \sqrt{N}(\hat{\beta} - \beta) &= \left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum x_i' u_i \right)\end{aligned}$$

Suponemos: (1) Independencia condicional, (2) X de rango completo, (3) Homoscedasticidad

y (4) no autocorrelación $\Rightarrow E[u'u] = \sigma^2$, $E[uu'] = \sigma^2 I_n$

$\Rightarrow$ Por ley de los grandes números y corolario 3.1: $(\frac{1}{N} \sum x_i' x_i)^{-1} \xrightarrow{p} (E[x'x])^{-1} = A^{-1}$

Considere el término $(\frac{1}{\sqrt{N}} \sum x_i' u_i)$. $\{x_i' u_i\}$ es una secuencia i.i.d por el supuesto (4). Por el supuesto (1), tiene valor esperado cero y varianza finita

$\Rightarrow$ Por teorema del límite central: $(\frac{1}{\sqrt{N}} \sum x_i' u_i) \xrightarrow{d} N(0, B)$, $B = \text{Var}(x_i' u_i)$

$$B = \text{Var}(x_i' u_i) = E[(x_i' u_i)(x_i' u_i)'] = E[x_i' u_i u_i' x_i] = E[x_i' \sigma^2 I x_i] = \sigma^2 E[x_i' x_i]$$

$$B = \sigma^2 E[x'x] = \sigma^2 A$$

③

Dado que $(\frac{1}{N} \sum x_i' x_i)^{-1} \xrightarrow{p} A^{-1} \Rightarrow (\frac{1}{N} \sum x_i' x_i)^{-1} - A^{-1} \xrightarrow{p} 0 \Rightarrow (\frac{1}{N} \sum x_i' x_i)^{-1} - A^{-1} = o_p(1)$

Como $(\frac{1}{\sqrt{N}} \sum x_i' u_i) \xrightarrow{d} N(0, B) \Rightarrow$ Por lema 5 $(\frac{1}{\sqrt{N}} \sum x_i' u_i) = O_p(1)$

$$o_p(1) * O_p(1) = o_p(1) = \left[\left(\frac{1}{N} \sum x_i' x_i \right)^{-1} - A^{-1} \right] \left[\frac{1}{\sqrt{N}} \sum x_i' u_i \right]$$

$$o_p(1) = \underbrace{\left(\frac{1}{N} \sum x_i' x_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum x_i' u_i \right)}_{\sqrt{N}(\hat{\beta} - \beta)} - A^{-1} \frac{1}{\sqrt{N}} \sum x_i' u_i$$

$$o_p(1) = \sqrt{N}(\hat{\beta} - \beta) - A^{-1} \frac{1}{\sqrt{N}} \sum x_i' u_i \xrightarrow{p} 0 \Rightarrow \text{Equivalencia asintótica}$$

Si $\frac{1}{\sqrt{N}} \sum x_i' u_i \xrightarrow{d} N(0, B)$, por teorema del mapa continuo: $A^{-1} \frac{1}{\sqrt{N}} \sum x_i' u_i \xrightarrow{d} N(0, A^{-1} B A^{-1})$

$$\Rightarrow \sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1}) = N(0, E[x'x] \sigma^2 E[x'x]^{-1}) \\ = N(0, \sigma^2 E[x'x]^{-1})$$

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 E[x'x]^{-1})$$

$$\hat{\beta} - \beta \xrightarrow{d} N(0, \frac{\sigma^2 E[x'x]^{-1}}{N})$$

$$\hat{\beta} \xrightarrow{d} N(\beta, \frac{\sigma^2 E[x'x]^{-1}}{N})$$

$$\hat{\beta} \xrightarrow{a} N(\beta, \frac{\sigma^2 E[x'x]^{-1}}{N})$$

Usando el método de momentos:

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2 (\frac{1}{N} \sum x_i' x_i)^{-1}}{N} = \sigma^2 (x'x)^{-1}$$

Continuación
Distribución
Asintótica
Estimador
MCO

Distribución asintótica del estimador de MCO

$$\sqrt{N}'(\hat{\beta} - \beta) = \left(\frac{1}{N} \sum_i x_i' x_i\right)^{-1} (N^{-1/2} \sum_i x_i' u_i)$$

Supuestos: (1) IC, (2) X rango completo, (3) $E[uu'] = \sigma^2 I$.

① Por WLLN y corolario 3.1: $\left(\frac{1}{N} \sum_i x_i' x_i\right)^{-1} \xrightarrow{p} E(x'x)^{-1} = A^{-1}$

② $\{(x_i' u_i)\}$ es una secuencia i.i.d porque el modelo se asume como no autocorrelacionado (supuesto 3). Tiene valor esperado cero (IC) y varianza finita.

$$\frac{1}{\sqrt{N}} \sum_i (x_i' u_i) \xrightarrow{d} N(0, B), \quad B = \text{Var}(x_i' u_i) = E[u_i^2 x_i' x_i]$$

$$B = \sigma^2 E(x'x) = \sigma^2 A$$

De ①: $\left(\frac{1}{N} \sum_i x_i' x_i\right)^{-1} - A^{-1} \xrightarrow{p} 0 \Rightarrow \left(\frac{1}{N} \sum_i x_i' x_i\right)^{-1} - A^{-1} = o_p(1)$

De ②: $\frac{1}{\sqrt{N}} \sum_i (x_i' u_i) = o_p(1)$

Teniendo en cuenta que $o_p(1) + O_p(1) = O_p(1)$:

$$\left[\left(\frac{1}{N} \sum_i x_i' x_i\right)^{-1} - A^{-1}\right] \left[\frac{1}{\sqrt{N}} \sum_i (x_i' u_i)\right] = o_p(1)$$

$$\underbrace{\left[\left(\frac{1}{N} \sum_i x_i' x_i\right)^{-1} + \frac{1}{\sqrt{N}} \sum_i (x_i' u_i)\right]}_{\sqrt{N}'(\hat{\beta} - \beta)} - A^{-1} \frac{1}{\sqrt{N}} \sum_i (x_i' u_i) = o_p(1)$$

$$\Rightarrow \sqrt{N}'(\hat{\beta} - \beta) - A^{-1} N^{-1/2} \sum_i (x_i' u_i) \xrightarrow{p} 0 \Rightarrow \text{Equivalentes asintóticamente}$$

Sé que $A^{-1} N^{-1/2} \sum_i (x_i' u_i) \xrightarrow{d} N(0, A^{-1} B A^{-1})$ por teorema del mapa continuo.

↳ Como $\sqrt{N}'(\hat{\beta} - \beta)$ es asintóticamente equivalente $\Rightarrow$ Deben tener la misma distribución asintótica.

$$\sqrt{N}'(\hat{\beta} - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1})$$

$$\sqrt{N}'(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 E[x'x]^{-1})$$

$$\hat{\beta} \approx N\left(\beta, \frac{\sigma^2 E[x'x]^{-1}}{N}\right)$$

Usando método de momentos: $\text{Var}(\hat{\beta}) = \frac{\sigma^2 \left(\frac{1}{N} \sum_i x_i' x_i\right)^{-1}}{N} = \sigma^2 (X'X)^{-1}$

Derivación de
 σ^2

Derivación de $\hat{\beta}$ v. Econometría 1

$$\hat{\beta} = \beta + (X'X)^{-1} X' u \Rightarrow \hat{\beta} - \beta = (X'X)^{-1} X' u$$

$$\text{Var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1} X' u u' X (X'X)^{-1}]$$

$$(X'X)^{-1} X' X (X'X)^{-1} E[uu'] = \sigma^2 I_{N \times N} (X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

¿Cuánto es σ^2 ? $\sigma^2 = \text{Var}(u) = E[u'u]$. Por el método de momentos:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_i u_i^2 \quad \text{Sin embargo, los } u \text{ no se observan. Hay que sustituir por los residuales.}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_i \hat{u}_i^2 = \frac{\hat{u}'\hat{u}}{N} \quad \text{Queremos ver si este estimador es insesgado. } E[\hat{\sigma}^2] = \sigma^2?$$

$$\hat{\sigma}^2 = \frac{(Y - \hat{Y})'(Y - \hat{Y})}{N} \Rightarrow \begin{aligned} \hat{u} &= Y - \hat{Y} = X\beta + u - X\hat{\beta} \\ \hat{u} &= X\beta + u - X(X'X)^{-1}X'(X\beta + u) \\ \hat{u} &= X\beta + u - X(X'X)^{-1}X'X\beta - X(X'X)^{-1}X'u \end{aligned}$$

$$\begin{aligned} \hat{u} &= X\beta + u - X\beta - X(X'X)^{-1}X'u = u - X(X'X)^{-1}X'u \\ \hat{u} &= (I_N - X(X'X)^{-1}X')u \end{aligned}$$

M es idempotente $\Rightarrow M'M = M$

$$E[\hat{u}'\hat{u}] = E[u'MMu] = E[u'Mu] \rightarrow \text{Escalar}$$

Traza: Suma de los elementos de la diagonal de una matriz.
Traza de un escalar = escalar

$$E[\hat{u}'\hat{u}] = \text{tr}(E[u'Mu]) = E[\text{tr}(u'Mu)]. \quad \text{Por propiedades de la traza:}$$

$$\begin{aligned} E[\hat{u}'\hat{u}] &= E[\text{tr}(Mu u')] = \text{tr}(ME[uu']) = \text{tr}(M\sigma^2 I_N) \\ &= \text{tr}(\sigma^2 I_N - \sigma^2 X(X'X)^{-1}X') \\ &= \sigma^2 \text{tr}(I_N) - \sigma^2 \text{tr}(X'(X'X)^{-1}X) = \sigma^2 \text{tr}(I_N) - \sigma^2 \text{tr}(X'X(X'X)^{-1}) \\ E[\hat{u}'\hat{u}] &= \sigma^2 [\text{tr} I_N - \text{tr} I_{K+1}] = \sigma^2 (N - K - 1) \end{aligned}$$

$$\Rightarrow E[\hat{\sigma}^2] = E\left[\frac{\hat{u}'\hat{u}}{N}\right] = \frac{\sigma^2(N-K-1)}{N} \neq \sigma^2 \quad \text{sesgado.}$$

Para convertirlo en insesgado: $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N-K-1}$

Teorema de Gauss-Markov

Gauss - Markov (MELI)

En la clase de estimadores lineales e insesgados: $\hat{\beta} = A'Y$ (cualquier estimador lineal).

$$E[\hat{\beta}] = E[A'Y] = \beta \Rightarrow \hat{\beta}_{MCO} \text{ es el mejor estimador en términos de eficiencia.}$$

$$\begin{aligned} \text{Dem: } E(\hat{\beta}) &= E(A'Y) = E(A'[X\beta + u]) = E(A'X\beta + A'u) \\ &= E(A'X\beta) + E(A'u) = E(A'X\beta) = \beta \end{aligned}$$

$O \text{ por IC} \quad A'X\beta = \beta$
 $A'X = I_{K+1}$

D = Matriz de diferencias entre un estimador lineal cualquiera y MCO

$$\Rightarrow \text{Sea } D' = A' - (X'X)^{-1}X' \Rightarrow D'X = A'X - (X'X)^{-1}X'X = 0$$

$$\begin{aligned} \beta^* &= (D' + [X'X]^{-1}X')Y = (D' + [X'X]^{-1}X')(X\beta + u) \\ \beta^* &= D'X\beta + D'u + (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \end{aligned}$$

$$\beta^* = D'u + \beta + (X'X)^{-1}X'u \Rightarrow \beta^* - \beta = D'u + (X'X)^{-1}X'u$$

Demstración
eficiencia
estimador HCO

$$\text{Var}(\beta^*) = E[(\beta^* - \beta)(\beta^* - \beta)']$$

$$\text{Var}(\beta^*) = E[(D'u + (X'X)^{-1}X'u)(u'D + u'X(X'X)^{-1})]$$

$$\text{Var}(\beta^*) = E[D'u u'D] + E[D'u u'X(X'X)^{-1}] + E[(X'X)^{-1}X'u u'D] + E[(X'X)^{-1}X'u u'X(X'X)^{-1}]$$

$$\text{Var}(\beta^*) = E[D'u \sigma^2 I_n D] + E[D'u \sigma^2 I_n X(X'X)^{-1}] + E[(X'X)^{-1}X'u \sigma^2 I_n D] + E[(X'X)^{-1}X'u \sigma^2 I_n X(X'X)^{-1}]$$

$$\text{Var}(\beta^*) = \sigma^2 D'D + D'X(X'X)^{-1}0 + \sigma^2(X'X)^{-1}X'D + \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$\text{Var}(\beta^*) = \sigma^2 D'D + \sigma^2(X'X)^{-1} = \sigma^2(D'D + [X'X]^{-1})$$

$$\text{Var}(\beta^*) = \sigma^2 f(A) + \text{Var}(\hat{\beta})$$

Como $f(A) = D'D$ es una matriz al cuadrado, no puede asumir valores negativos.

El mínimo valor posible es $f(A) = 0$.

En consecuencia, no es posible que $\text{Var}(\beta^*)$ sea igual o menor a $\text{Var}(\hat{\beta})$, a menos de que $f(A) = D'D = 0$.

$$\text{En ese caso: } D' = 0 = A' - (X'X)^{-1}X' \\ A' = (X'X)^{-1}X'$$

Q.E.D de $\hat{\beta}_{HCO}$ es un estimador eficiente.

Método Delta

Método Delta

si $\sqrt{N}'(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ y $c(\cdot)$ que es continua,

$$\sqrt{N}'[c(\hat{\theta}) - c(\theta)] \xrightarrow{d} N(0, J(\theta)N J(\theta)')$$

Donde $J(\theta)$ = Jacobiano de $c(\theta)$.

E.g: $\hat{\theta}$ es $\sqrt{N}$ -asintóticamente normal con $\text{Avar}[\sqrt{N}'(\hat{\theta} - \theta)] = V$ y θ es un escalar. Quiero aplicar P.H. sobre $\log \hat{\theta} = \hat{\gamma} = c(\hat{\theta})$

$$\Rightarrow \sqrt{N}'(\hat{\gamma} - \gamma) \sim N(0, \text{Avar}(\hat{\gamma}))$$

$$J(\theta) = \frac{d}{d\theta} \gamma = \frac{1}{\theta} \Rightarrow \text{Avar}(\hat{\gamma}) = \frac{1}{\theta^2} \text{Avar}(\hat{\theta})$$

E.g: $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$ $\sqrt{N}$ -asintóticamente normal de $\theta = (\theta_1, \theta_2)'$ con $\theta_2 \neq 0$.

$$\text{Sea } \gamma = \theta_1/\theta_2 \Rightarrow \text{Avar}(\hat{\gamma}) = f(\text{Avar}[\hat{\theta}]) = f(\Sigma) = f \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

$$J(\gamma) = \begin{pmatrix} d\gamma/d\theta_1 & d\gamma/d\theta_2 \end{pmatrix} = \begin{pmatrix} 1/\theta_2 \\ -\theta_1/\theta_2^2 \end{pmatrix}$$

$$\Rightarrow \text{Avar}(\hat{\gamma}) = \begin{pmatrix} 1/\theta_2 & -\theta_1/\theta_2^2 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} 1/\theta_2 \\ -\theta_1/\theta_2^2 \end{pmatrix}$$

Para encontrar la varianza de transformaciones no lineales de parámetros que ya estimamos.

Derivación de $\text{Var}(\hat{\beta})$ - Versión Econometría 1

Considere $\hat{\beta} = \beta + (X'X)^{-1}X'u \Rightarrow \hat{\beta} - \beta = (X'X)^{-1}X'u$

$$\text{Var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1}X'u u'X(X'X)^{-1}] = (X'X)^{-1}X'X(X'X)^{-1}E[uu']$$

$$\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}I_N = \sigma^2(X'X)^{-1}$$

¿Cuánto es σ^2 ?

$\sigma^2 = \text{Var}(u) = E[u'u]$. Por el método de momentos, tenemos que: $\frac{1}{N} \sum_i u_i^2 = \hat{\sigma}^2$. No obstante, u_i no se observa, y hay que aproximarlos usando $\hat{u}_i = y_i - \hat{y}_i$

$$\frac{1}{N} \sum_i \hat{u}_i^2 = \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N} = \frac{(Y - \hat{Y})'(Y - \hat{Y})}{N} = \frac{(X\beta + u - X\hat{\beta})'(X\beta + u - X\hat{\beta})}{N}$$

$$\hat{\sigma}^2 = \frac{(X\beta + u - X(X'X)^{-1}X'(X\beta + u))'(X\beta + u - X(X'X)^{-1}X'(X\beta + u))}{N}$$

$$\hat{\sigma}^2 = \frac{(X\beta + u - X\beta - X(X'X)^{-1}X'u)'(X\beta + u - X\beta - X(X'X)^{-1}X'u)}{N}$$

$$\hat{\sigma}^2 = \frac{(u - X(X'X)^{-1}X'u)'(u - X(X'X)^{-1}X'u)}{N} = \frac{[I_N - X(X'X)^{-1}X'u]'[I_N - X(X'X)^{-1}X'u]}{N}$$

Sea $M = I_N - X(X'X)^{-1}X'$. Note que $M' = (I_N - X(X'X)^{-1}X')' = (I_N)' - (X[X'X]X')' = I_N - X(X'X)^{-1}X' = M$

$$M'M = (I_N - X[X'X]^{-1}X')(I_N - X[X'X]^{-1}X') = I_N I_N - I_N X(X'X)^{-1}X' - X(X'X)^{-1}X' I_N + X(X'X)^{-1}X'X(X'X)^{-1}X'$$

$M'M = I_N + X(X'X)^{-1}X'$?! M no es idempotente

$$\begin{aligned} M'M &= I'I - I'P_X - I'P_X + P_X \\ &= I'I - P_X - P_X + P_X \\ M'M &= I'I - P_X \neq M \end{aligned}$$

$$\Rightarrow \hat{\sigma}^2 = \frac{(Mu)'(Mu)}{N} = \frac{u'M'Mu}{N} = E[u'M'Mu] = E[\text{escalar}]$$

Matriz $1 \times n$ Matriz $n \times n$ Matriz $n \times 1$

$$\begin{aligned} E[\hat{u}'\hat{u}] &= \text{tr}(E[u'M'Mu]) \stackrel{\text{Por propiedades de la traza}}{=} E[\text{tr}(u'M'Mu)] = E[\text{tr}(M'u u')] = \text{tr}(E[M'u u']) \\ &= \text{tr}(M'M E[uu']) = \text{tr}(M'M \sigma^2 I_N) \end{aligned}$$

En adelante, asumo que M sí es idempotente: $M'M = M$

$$\begin{aligned} E[\hat{u}'\hat{u}] &= \text{tr}(M'M \sigma^2 I_N) = \text{tr}(M \sigma^2 I_N) = \text{tr}(\sigma^2 [I_N - X(X'X)^{-1}X']) = \text{tr}(\sigma^2 I_N - \sigma^2 X(X'X)^{-1}X') \\ &= \text{tr}(\sigma^2 I_N) - \text{tr}(\sigma^2 X[X'X]^{-1}X') = \sigma^2 \text{tr}(I_N) - \sigma^2 \text{tr}(X[X'X]^{-1}X') \\ &= \sigma^2 (\text{tr}[I_N] - \text{tr}[X'X(X'X)^{-1}]) = \sigma^2 (\text{tr}[I_N] - \text{tr}[I_{K+1}]) \end{aligned}$$

$$E[\hat{u}'\hat{u}] = \sigma^2(N - [K + 1]) = E[\hat{\sigma}^2] \Rightarrow \hat{\sigma}^2 \text{ es un estimador sesgado de } \sigma^2$$

Para que sea insesgado: $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N - K - 1}$

Demostración eficiencia HELI (Gauss Markov)

Considere cualquier estimador lineal insesgado $\hat{\beta} = A'y$ (donde A es una función de X).

Definimos $D' = A' - (X'X)^{-1}X'$ como la diferencia entre un estimador lineal cualquiera

y el $\hat{\beta}_{MCO}$. (Note que $(X'X)^{-1}X' = A'_{MCO}$).

Si multiplicamos por X : $D'X = A'X - \underbrace{(X'X)^{-1}X'X}_{I_{K+1}}$

Tomando el supuesto de inesgamiento: $E[\tilde{\beta}] = \beta$

$$E[A'Y] = \beta$$

$$E[A'(X\beta + u)] = \beta$$

$$E[A'X\beta] + E[A'u] = \beta$$

o porque A es $f(X)$

$$A'X\beta = \beta$$

$$A'X = I_{K+1}$$

Por tanto: $D'X = I_{K+1} - I_{K+1} = 0$.

Suponga que $\tilde{\beta}$ es un estimador distinto al de MCO, definido por $\tilde{\beta} = A'Y$

$$\Rightarrow \tilde{\beta} = (D' + [X'X]^{-1}X')Y = (D' + [X'X]^{-1}X')(X\beta + u) = D'X\beta + D'u + (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$$

$$\tilde{\beta} = D'u + \beta + (X'X)^{-1}X'u \Rightarrow \tilde{\beta} - \beta = D'u + (X'X)^{-1}X'u$$

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] = E[(D'u + [X'X]^{-1}X'u)(u'D + u'X[X'X]^{-1})] \\ &= E[D'u u'D + (X'X)^{-1}X'u u'D + D'u u'X(X'X)^{-1} + (X'X)^{-1}X'u u'X(X'X)^{-1}] \\ &= D'D E[u u'] + (X'X)^{-1}X'D E[u u'] + D'X(X'X)^{-1}E[u u'] + (X'X)^{-1}X'X(X'X)^{-1}E[u u'] \end{aligned}$$

$$\text{Var}(\tilde{\beta}) = \sigma^2 D'D I_N + \sigma^2 (X'X)^{-1} I_N = \sigma^2 D'D + \underbrace{\sigma^2 (X'X)^{-1}}_{\text{Var}(\hat{\beta}_{MCO})}$$

Si $\tilde{\beta}$ es más eficiente que $\hat{\beta}_{MCO} \Rightarrow \text{Var}(\tilde{\beta}) < \text{Var}(\hat{\beta}_{MCO})$. Si esto se cumple, entonces $\sigma^2 D'D$ debe ser negativa semidefinida. Sin embargo, esto no es posible porque $D'D$ es una matriz "al cuadrado" y necesariamente es positiva semidefinida.

El mínimo valor posible que puede asumir $\sigma^2 D'D$ es cero. Suponiendo que esto se cumple:

$$\sigma^2 D'D = 0 = D'D \Rightarrow D' = 0$$

$$D' = A' - (X'X)^{-1}X'$$

$$A' = (X'X)^{-1}X' \Rightarrow \hat{\beta}_{MCO} \text{ es MELI.}$$

Método Delta

Se usa para encontrar la varianza de transformaciones no lineales de parámetros que ya estimamos.

Suponga $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ y tenemos una función continua $c(\theta)$.

$$\Rightarrow \sqrt{N}(c(\hat{\theta}) - c(\theta)) \rightarrow N(0, J(\theta)' V J(\theta)) \text{ donde } J(\theta) = \text{Jacobiano de } c(\theta)$$

Ejemplo 1

Suponga un $\hat{\theta}$ asintóticamente normal (a tasa $\sqrt{N}$), estimador de $\theta = (\theta_1, \theta_2)'$ con $\theta_2 \neq 0$. $\hat{\theta}$ tiene $\text{Avar}(\sqrt{N}(\hat{\theta} - \theta)) = V$

$$\text{Sea } \hat{\gamma} = c(\theta) = \frac{\theta_1}{\theta_2} \Rightarrow \text{Avar}(\hat{\gamma}) = f(\text{Avar}(\hat{\theta})) = f(J_1) = f\left(\begin{matrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{matrix}\right)$$

$$J(\hat{\theta}) = \begin{pmatrix} d\hat{\theta}/d\theta_1 \\ d\hat{\theta}/d\theta_2 \end{pmatrix} = \begin{pmatrix} 1/\theta_2 \\ -\theta_1/\theta_2^2 \end{pmatrix} \Rightarrow \text{Avar}(\hat{\theta}) = \begin{pmatrix} 1/\theta_2 \\ -\theta_1/\theta_2^2 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} 1/\theta_2 & -\theta_1/\theta_2^2 \end{pmatrix}$$

Ejemplo 2

Considere un $\hat{\theta}$ estimador de $\theta \in \mathbb{R}$, que se distribuye $\sqrt{N}$ -asintóticamente normal con $\text{Avar}(\sqrt{N}(\hat{\theta} - \theta)) = V$. $\text{cl}(\theta) = \log \theta = \hat{\gamma}$

$$\text{Avar}(\hat{\gamma}) = J(\theta) \text{Avar}(\hat{\theta}) J(\theta)'$$

$$\text{Avar}(\hat{\gamma}) = \left(\frac{1}{\theta}\right)^2 \text{Avar}(\hat{\theta}) = \frac{1}{\theta^2} \text{Avar} \hat{\theta} \Rightarrow \sqrt{N}(\hat{\gamma} - \gamma) \hat{\sim} N(0, \frac{1}{\theta^2} \text{Avar}(\hat{\theta}))$$

Supuestos de Gauss-Markov

1. Linealidad en los parámetros: $Y = X\beta + u$
2. La muestra es i.i.d
3. X tiene rango completo - no multicolinealidad
4. $E[u|x] = 0 \Rightarrow E[u] = 0$
5. Homoscedasticidad $\text{Var}(u) = \sigma^2$
6. $u \sim N(0, \sigma^2)$

Simulaciones

`runiform` Genera una variable uniforme en el intervalo $[0, 1)$
`gen uniforme1 = runiform(1)`

↳ `gen uniforme2 = uniforme1 * (4 - 2) + 2` Crea una variable
 uniforme $[2, 4)$.

TALLER 1

1. Sea $\{y_i\}$ secuencia i.i.d con $E[y_i^2] < \infty$, $\mu = E[y_i]$ y $\sigma^2 = \text{Var}(y_i)$

(a) Hallar $\text{Var}[\sqrt{N}(\bar{y}_N - \mu)]$ Considere que $\bar{y}_N = \frac{1}{N} \sum_i y_i$

$$\begin{aligned}
 \text{Var}(\sqrt{N}[\bar{y}_N - \mu]) &= N \text{Var}(\bar{y}_N - \mu) = N[\text{Var}(\bar{y}_N) + \text{Var}(\mu) - 2\text{Cov}(\bar{y}_N, \mu)] \\
 &= N \text{Var}(\bar{y}_N) = N \text{Var}\left(\frac{1}{N} \sum_i y_i\right) \\
 &= \frac{1}{N} \text{Var}\left(\sum_i y_i\right) = \frac{1}{N} \sum_i \text{Var}(y_i) = \frac{1}{N} \sum_i \sigma^2 = \frac{N}{N} \sigma^2 = \sigma^2
 \end{aligned}$$

Porque $\{y_i\}$ es i.i.d $\Rightarrow \text{Cov}(y_i, y_j) = 0 \forall i \neq j$.

(b) Hallar $\text{Avar}(\sqrt{N}[\bar{y}_N - \mu])$ Necesito que $\sqrt{N}(\bar{y}_N - \mu) \xrightarrow{d} N(0, v)$ donde $v = \text{Var}(y_i - \mu) = \sigma^2$

Quieren que aplique TLC: $\frac{1}{\sqrt{N}} \sum_{i=1}^N (y_i - \mu) \xrightarrow{d} N(0, v)$ $v = \text{Var}(y_i - \mu) = \sigma^2$

Se puede porque $\{y_i\}$ es i.i.d $\Rightarrow \{y_i - \mu\}$ es i.i.d
 $E[(y_i - \mu)^2] = \text{Var}(y_i) = \sigma^2 < \infty$

$$\begin{aligned}
 \bar{y}_N - \mu &= \frac{1}{N} \sum_i y_i - \mu \\
 \sqrt{N}(\bar{y}_N - \mu) &= \frac{1}{\sqrt{N}} \sum_i y_i - \sqrt{N} \mu = \sqrt{N} \left(\frac{1}{N} \sum_i y_i - \mu \right) = \frac{1}{\sqrt{N}} \sum_i (y_i - \mu) \\
 \sqrt{N}(\bar{y}_N - \mu) &= \frac{1}{\sqrt{N}} \sum_i (y_i - \mu) \xrightarrow{d} N(0, v) \\
 \Rightarrow \sqrt{N}(\bar{y}_N - \mu) &= \frac{1}{\sqrt{N}} \sum_i (y_i - \mu) = O_p(1)
 \end{aligned}$$

$$\sqrt{N}(\bar{y}_N - \mu) \xrightarrow{d} N(0, v)$$

$$\sqrt{N}(\bar{y}_N - \mu) \overset{a}{\sim} N(0, \sigma^2) \Rightarrow \text{Avar}(\sqrt{N}[\bar{y}_N - \mu]) = \sigma^2$$

(c) ¿Cuáles $\text{Avar}(\bar{y}_N)$?

$$\bar{y}_N - \mu \xrightarrow{d} N(0, \frac{v}{N})$$

$$\bar{y}_N \xrightarrow{d} N(\mu, \frac{\sigma^2}{N})$$

$$\begin{aligned}
 \text{Var}(\bar{y}_N) &= \frac{1}{N^2} \text{Var}\left(\sum_i y_i\right) \\
 &\stackrel{\text{i.i.d}}{\Rightarrow} \frac{1}{N^2} \sum_i \text{Var}(y_i) \\
 &= \frac{N \sigma^2}{N^2} \\
 &= \frac{\sigma^2}{N}
 \end{aligned}$$

1. (d) ¿cuál es la desviación estándar de

3. Usted estimó $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e$

(a) Demuestre que si el modelo poblacional es $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e$, $\hat{\beta}_1$ es insesgado. ¿Qué pasa con $\text{Var}(\hat{\beta}_1)$ si x_2 no está correlacionado con y_i ? ¿Qué pasa si sí?

$$\hat{\beta}_1 = \frac{\text{Cov}(x_1, y) \text{Var}(x_2) - \text{Cov}(x_2, y) \text{Cov}(x_1, x_2)}{\text{Var}(x_1) \text{Var}(x_2) - \text{Cov}^2(x_1, x_2)} \quad \left. \begin{array}{l} \text{Lo que estimo si estimo el} \\ \text{modelo incorrecto.} \end{array} \right\}$$

Reemplazo $y = \beta_0 + \beta_1 x_1 + e$ (modelo correcto)

$$\hat{\beta}_1 = \frac{\text{Cov}(x_1, \beta_0 + \beta_1 x_1 + e) \text{Var}(x_2) - \text{Cov}(x_2, \beta_0 + \beta_1 x_1 + e) \text{Cov}(x_1, x_2)}{\text{Var}(x_1) \text{Var}(x_2) - \text{Cov}^2(x_1, x_2)}$$

$$\hat{\beta}_1 = \frac{[\beta_1 \text{Var}(x_1) + \text{Cov}(x_1, e)] \text{Var}(x_2) - [\beta_1 \text{Cov}(x_1, x_2) + \text{Cov}(x_2, e)] \text{Cov}(x_1, x_2)}{\text{Var}(x_1) \text{Var}(x_2) - \text{Cov}^2(x_1, x_2)}$$

$$\hat{\beta}_1 = \beta_1 \left[\frac{\text{Var}(x_1) \text{Var}(x_2) - \text{Cov}^2(x_1, x_2)}{\text{Var}(x_1) \text{Var}(x_2) - \text{Cov}^2(x_1, x_2)} \right] = \beta_1$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

$$y - \hat{y} = \hat{e} \text{ (MODELO INCORRECTO)} = \beta_0 + \beta_1 x_1 + e - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2$$

$$\hat{e} = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) x_1 - \hat{\beta}_2 x_2 + e$$

Todos los $\hat{\beta}$ son consistentes y insesgados

$$X_c = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{bmatrix} \Rightarrow (X'X) = \begin{bmatrix} 1 & 1 & 1 \\ x_{11} & x_{21} & x_{31} \end{bmatrix} \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{bmatrix} = \begin{bmatrix} 3 & x_{11} + x_{21} + x_{31} \\ x_{11} + x_{21} + x_{31} & x_{11}^2 + x_{21}^2 + x_{31}^2 \end{bmatrix}$$

$$X_1 = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \end{bmatrix} \Rightarrow$$

$$E[y] = \beta_0 + \beta_1 E[x_1] + \beta_2 E[x_2] = \beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2$$

$$\beta_1 = \frac{\bar{y} - \beta_0 - \beta_2 \bar{x}_2}{\bar{x}_1}$$

$$\hat{\beta}_1 = \frac{\sigma^2}{\sum (x_i - \bar{x}_1)^2} \text{ si no hay redundante vs. } \hat{\beta}_1 = \frac{\sigma^2}{\sum (x_i - \bar{x}_1)^2 (1 - \rho_{x_1, x_2}^2)} \text{ si hay } x_2$$

Demostración
suficiente

¿cómo llegar a

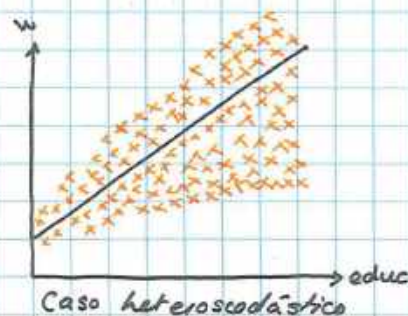
$$v(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x}_1)^2 (1 - \rho_{x_1, x_2}^2)}$$

HeteroscedasticidadDefinición
Heteroscedasticidad

$$y_i = x_i \beta + u_i, \quad \text{Var}(u_i) = \sigma^2 \quad \forall i: \text{ caso homoscedástico}$$

$$\text{Var}(u_i) = \sigma_i^2 \quad \text{ caso heteroscedástico}$$

$\text{Var}(y_i | x_i) = \text{Var}(x_i \beta + u_i | x_i)$. Si se cumple IC:
 $= \text{Var}(u_i | x_i) + \text{Var}(x_i \beta | x_i)$ Pero $x_i \beta$ es conocido $\Rightarrow \text{Var}(x_i \beta | x_i) = 0$
 $\Rightarrow \text{Var}(y_i | x_i) = \text{Var}(u_i | x_i)$

Ejemplo

$$\text{Considere } \hat{\beta} = \beta + (X'X)^{-1}X'u.$$

Efectos

- ① Si hay heteroscedasticidad, ¿ $\hat{\beta}_{MCO}$ es insesgado?
 Sí. Aún se cumplen $\rightarrow$ Independencia condicional
 $\rightarrow$ X de rango completo
- ② Si hay heteroscedasticidad, ¿se cumple que $\text{var}(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}$?

$$\text{var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1}X'u u'X(X'X)^{-1}]$$

$$= (X'X)^{-1}X'E[uu']X(X'X)^{-1}$$

Homoscedasticidad
 $\Rightarrow \text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$

Heteroscedasticidad

$$E[uu'] = \text{Var-Cov}(u) = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix}$$

$$E[uu'] = \Sigma_{N \times N}$$

$$\Rightarrow \text{var}(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1} \quad *$$

SolucionesSolución

Deshacerse de la heteroscedasticidad (Mínimos cuadrados Generalizados Ponderados)
 Computar errores estándar robustos con base en *

MCGPMínimos cuadrados generalizados ponderados

$$y = x\beta + u$$

$$E[uu'] = \Sigma = \sigma^2 \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & w_N \end{bmatrix}$$

donde $w = f(x)$. Asumimos que conocemos la forma de f .

$$E[uu'] = \Sigma = \sigma^2 \Omega. \quad \text{Definimos:}$$

$$\text{Note que } P'P = \Omega^{-1}$$

$$P = \begin{bmatrix} \frac{1}{\sqrt{w_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{w_N}} \end{bmatrix}$$

Considere $Y = X\beta + u$. Premultiplicamos por P para volver el modelo homoscedástico.

$$PY = PX\beta + Pu \Rightarrow y^* = x^*\beta + u^*$$

$$\text{var}(u^*) = E[u^*u^{*'}] = E[PUu'P'] = PE[uu']P' = P\sigma^2\Omega P'$$

Note que P y Ω son matrices triangulares superior e inferior. En este caso, la multiplicación de matrices es conmutativa

$$\text{Var}(u^*) = \sigma^2 P'P\Omega = \sigma^2 \Omega^{-1}\Omega = \sigma^2 \Rightarrow \text{Homoscedástico.}$$

Corro el modelo $y^* = x^*\beta + u^*$ por MCO. El $\hat{\beta}$ es insesgado porque u^* sigue cumpliendo IC y x^* sigue siendo de rango completo.

Errores estándar robustos

Errores estándar robustos.

$$\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}$$

$$= \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{K1} & x_{K2} & \dots & x_{KN} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_N^2 \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \dots & x_{K1} \\ 1 & x_{12} & \dots & x_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & \dots & x_{KN} \end{bmatrix}$$

$$X' \Sigma X = \sum_i \sigma_i^2 x_i x_i' \Rightarrow \text{Estimador de White: } \hat{\Sigma} X = S_0 = \sum_i \hat{u}_i^2 x_i x_i'$$

$\Rightarrow \text{Var}(\hat{\beta}) = (X'X)^{-1} S_0 (X'X)^{-1}$ pero casi siempre es al menos tan grande como $\text{Var}(\hat{\beta})_{\text{Homo}} = \sigma^2 (X'X)^{-1}$. $\Rightarrow$ Ineficiencia $\Rightarrow$ Intervalos de confianza más pequeños.

Detección

Detección

Objetivo: Saber si la varianza del residual tiene un patrón con respecto a las características de los individuos.

Test de White

Test de White: Regresión de $\hat{u}_i^2$ sobre un polinomio flexible de las X .

$\hat{u}_i^2$ sobre $x_1, x_2, x_1^2, x_2^2, x_1 x_2$ (***)

H_0 : Homoscedasticidad

Estadístico $\chi^2_K = nR^2$

K : # de β menos el intercepto.

Si $\chi^2_K > \text{valor crítico} \Rightarrow R H_0 \Rightarrow$ Hay heteroscedasticidad.

La significancia individual de los β da luces sobre la forma de la heteroscedasticidad.

En el ejemplo de la educación, la heteroscedasticidad es lineal. $\Rightarrow w_i = x_i$.

Autocorrelación

AUTOCORRELACIÓN

$$\text{cov}(u_i, u_j) \neq 0 \quad \forall i \neq j$$

$$\Rightarrow E[uu'] = \Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{12} & \sigma^2 & \dots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \dots & \sigma^2 \end{bmatrix}$$

Efectos

- ① ¿ $\hat{\beta}_{MCO}$ es insesgado? Sí, porque se cumple IC y X es de rango completo.
- ② ¿Los errores estándar son correctos?

$$\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1} \neq \sigma^2 (X'X)^{-1}. \text{ No son correctos.}$$

Soluciones

Soluciones

- Deshacerse de la autocorrelación (MCO factibles)
- Estimar errores estándar robustos

Problema común en series de tiempo, p.e: $\text{Corr}(u_t, u_{t-1}) \neq 0$.

Modelo AR(1), autocorrelación conocida

$$y_t = X_t \beta + u_t, \text{ donde } u_t = \rho u_{t-1} + \varepsilon_t \text{ donde } \varepsilon_t \rightarrow E[\varepsilon_t] = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, \text{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0 \quad \forall j \neq 0.$$

$$\text{Cov}(u_t, u_{t-1}) = E[u_t u_{t-1}] - E[u_t]E[u_{t-1}] = \rho \sigma^2$$

$$u_{t-1} u_t = \rho u_{t-1}^2 + \varepsilon_t u_{t-1}$$

$$E[u_{t-1} u_t] = \rho E[u_{t-1}^2] + E[\varepsilon_t u_{t-1}] = \rho \sigma^2$$

$$\text{Cov}(u_t, u_{t-2}) = E[u_{t-2} u_t] - E[u_t]E[u_{t-2}]$$

$$u_{t-2} u_t = \rho u_{t-1} u_{t-2} + \varepsilon_t u_{t-2}$$

$$E[u_{t-2} u_t] = \rho E[u_{t-1} u_{t-2}] + E[u_{t-2} \varepsilon_t] = \rho + \rho \sigma^2 = \rho^2 \sigma^2$$

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho \sigma^2 & \rho^2 \sigma^2 & \rho^3 \sigma^2 \\ \rho \sigma^2 & \sigma^2 & \rho \sigma^2 & \rho^2 \sigma^2 \\ \rho^2 \sigma^2 & \rho \sigma^2 & \sigma^2 & \rho \sigma^2 \\ \rho^3 \sigma^2 & \rho^2 \sigma^2 & \rho \sigma^2 & \sigma^2 \end{bmatrix}$$

RESUMEN - CLASE 5

I. HETEROSCEDASTICIDAD

Efectos: $\hat{\beta}$ insesgado, pero $\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1} \neq \sigma^2 (X'X)^{-1}$.

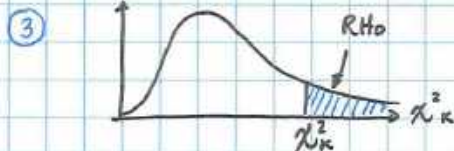
Detección: Objetivo: Saber si $\text{Var}(\hat{u}) = f(x)$, donde x son las características de los individuos.

TEST DE WHITE

- ① Considere $\hat{u}_i^2 = \alpha_0 + \alpha_1 x_i^2 + \alpha_2 x_i^3 + \alpha_3 x_i x_i^2 + \varepsilon$

H_0 : Homoscedasticidad $\alpha_0 = \alpha_1 = \alpha_2 = \dots = 0$

- ② Estadístico de prueba: $\chi^2_K = N R_{\hat{u}_i^2}^2$, $K = \#$ de α 's (sin contar intercepto)



Soluciones:

MÍNIMOS CUADRADOS GENERALIZADOS PONDERADOS

Para el modelo $Y = X\beta + u \rightarrow E[uu'] = \sigma^2 \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_N \end{bmatrix} = \sigma^2 \Omega$

donde $w = f(x)$. Asumimos que conocemos la forma de f .

Definimos $P = \begin{bmatrix} \frac{1}{\sqrt{w_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{w_N}} \end{bmatrix} \Rightarrow P'P = \Omega^{-1}$

Para convertir $Y = X\beta + u$: $PY = PX\beta + Pu \Rightarrow Y^* = X^*\beta + U^*$
 $\begin{matrix} PY \\ \text{"} \end{matrix}$ $\begin{matrix} PX \\ \text{"} \end{matrix}$ $\begin{matrix} Pu \\ \text{"} \end{matrix}$

$Var(U^*) = E[U^*U^{*'}] = E[PUU'P'] = P E[UU'] P' = P \sigma^2 \Omega P'$
 $\uparrow$
no aleatoriedad de P

Como P y Ω son matrices triangulares superior e inferior, es posible conmutar el orden de la multiplicación.

$Var(U^*) = \sigma^2 \Omega P'P = \sigma^2 \Omega \Omega^{-1} = \sigma^2$. Q.E.D que el modelo Y^* es homoscedástico.

ERRORES ESTÁNDAR ROBUSTOS

Considere $Var(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}$. El término $X' \Sigma X$ se puede escribir como $X' \Sigma X = \Sigma_i \sigma_i^2 x_i x_i'$.

Estimador de White: $\widehat{X' \Sigma X} = S_0 = \sum_i \hat{u}_i^2 x_i x_i'$

Por ende: $Var(\hat{\beta}) = (X'X)^{-1} S_0 (X'X)^{-1}$. Sin embargo, en general se cumplirá que $Var(\hat{\beta})_{hetero} = (X'X)^{-1} S_0 (X'X)^{-1} > Var(\hat{\beta})_{homo} = \sigma^2 (X'X)^{-1}$.

$\Rightarrow$ Intervalos de confianza serán más pequeños. Aumenta la probabilidad de que los parámetros del modelo no sean significativos.

II. AUTOCORRELACIÓN $cov(u_i, u_j) \neq 0 \forall i \neq j$.

Efectos: $\hat{\beta}$ insesgado, pero $Var(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1} \Rightarrow$ Ineficiente.

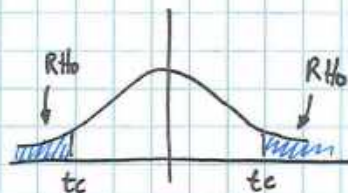
Detección: COCHRANNE - ORCUTT

Considere el modelo $Y_t = \beta_0 + \beta_1 X_t + u_t$, donde $Cov(u_t, u_s) \neq 0 \forall t \neq s$.

- ① Estimar $Y_t = \beta_0 + \beta_1 X_t + u_t$
- ② obtener $\hat{u}_t$.
- ③ Correr el modelo $\hat{u}_t = \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \dots + \rho_j \hat{u}_{t-j} + \epsilon_t$

$H_0: \rho_i = 0$ No autocorrelación

Estadístico de prueba: $t = \frac{\hat{\rho}}{se(\hat{\rho})} \sim t_{n-j}$



Soluciones:

MÍNIMOS CUADRADOS GENERALIZADOS FACTIBLES

Suponga que la autocorrelación es de la forma AR(2): $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + u_t \\ - \rho_1 Y_{t-1} &= \rho_1 \beta_0 + \rho_1 \beta_1 X_{t-1} + \rho_1 u_{t-1} \\ - \rho_2 Y_{t-2} &= \rho_2 \beta_0 + \rho_2 \beta_1 X_{t-2} + \rho_2 u_{t-2} \end{aligned}$$

$$(Y_t - \rho_1 Y_{t-1} - \rho_2 Y_{t-2}) = \beta_0(1 - \rho_1 - \rho_2) + \beta_1(X_t - \rho_1 X_{t-1} - \rho_2 X_{t-2}) + (u_t - \rho_1 u_{t-1} - \rho_2 u_{t-2})$$

Definimos: $\tilde{Y}_t = \tilde{\beta}_0 + \beta_1 \tilde{X}_t + \tilde{u}_t$

Ⓐ COCHRANNE - ORCUTT

1. Estimar el modelo $Y_t = \beta_0 + \beta_1 X_t + u_t$
2. Obtener $\hat{u}_t$.
3. Estimar $\hat{u}_t = \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \varepsilon_t$
4. Obtener $\hat{\rho}_1$ y $\hat{\rho}_2$
5. Construir $\tilde{Y}_t = \tilde{\beta}_0 + \beta_1 \tilde{X}_t + \tilde{u}_t$ usando $\hat{\rho}_1$ y $\hat{\rho}_2$.
6. Estimar $\tilde{Y}_t$.

Ⓑ PRAISE - WINSTEN

El hecho de rezagar el modelo hace perder una observación en Cochrane-Orcutt. Este método previene el problema.

Considere $Y_1 = \beta_0 + \beta_1 X_1 + u_1 \rightarrow$ La primera observación.

Para esta observación: I. $\text{Cov}(u_1, \varepsilon_t) = 0$
 II. $\text{Var}(u_1) = \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$, $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_t)$

Para homogeneizar la varianza:

$$\sqrt{1 - \rho^2} Y_1 = \sqrt{1 - \rho^2} \beta_0 + \beta_1 \sqrt{1 - \rho^2} X_1 + \sqrt{1 - \rho^2} u_1$$

Se corre el modelo con $\tilde{Y}_t$ en todas las observaciones y $\sqrt{1 - \rho^2} Y_1$ en la primera.

ERRORES ESTÁNDAR ROBUSTOS (para un AR(1))

$\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}$. La matriz Σ de un AR(1) es:

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho \sigma^2 & \dots & \rho^{T-1} \sigma^2 \\ \rho \sigma^2 & \sigma^2 & \dots & \rho^{T-2} \sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} \sigma^2 & \dots & \dots & \sigma^2 \end{bmatrix}$$

Es necesario estimar $\hat{\Sigma}$.

$$\sigma^2 = E[u_t u_t'] \Rightarrow \hat{\sigma}^2 = \frac{\sum u_t^2}{N - k - 1}$$

Obtenemos $\hat{\rho}$ de $u_t = \rho u_{t-1} + \varepsilon_t$.

Repaso
autocorrelación

Autocorrelación

$\text{cov}(u_t, u_s) \neq 0 \Rightarrow$ Para $y = \beta_0 + \beta_1 x_1 + u$: $\hat{\beta}$ insesgado, pero ineficiente.
 $\hookrightarrow \text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}$, $\Sigma = E[UU'] \neq \sigma^2 I_N$

En particular: AR(1): $u_t = \rho u_{t-1} + \varepsilon_t$, $E[\varepsilon_t] = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$
 $\text{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0 \forall j \neq 0$

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho^{T-1}\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho^{T-2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1}\sigma^2 & \dots & \dots & \sigma^2 \end{bmatrix}$$

Note que u_t sigue siendo homoscedástico.

$$\begin{aligned} \text{Var}(u_t) &= \text{Var}(\rho u_{t-1} + \varepsilon_t) = \rho^2 \text{Var}(u_{t-1}) + \sigma_\varepsilon^2 + 0 \\ (1 - \rho^2) \text{Var}(u_t) &= \sigma_\varepsilon^2 \\ \text{Var}(u_t) &= \frac{\sigma_\varepsilon^2}{1 - \rho^2} = \sigma^2 \end{aligned}$$

Solución

Solución

MCG Factibles

Estimar e.e. robustos a autocorrelación.

MCGF

MCG factibles

Para AR(1)

① AR(1): Transformación: Modelo cuasi-diferenciado

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + u_t \\ \rho y_{t-1} &= \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho u_{t-1} \end{aligned}$$

$$y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + (1 - \rho)u_t$$

$$\tilde{y}_t = \tilde{\beta}_0 + \beta_1 \tilde{x}_t + \varepsilon_t \rightarrow \text{Modelo no autocorrelacionado}$$

Detección

Note que $\tilde{\beta}_0 \neq \beta_0$: $\tilde{\beta}_0 = (1 - \rho)\beta_0 \Rightarrow \beta_0 = \tilde{\beta}_0 / (1 - \rho)$

Cochranne-Orcutt

① Cochranne - Orcutt (detección)

1. Estimar $y_t = \beta_0 + \beta_1 x_t + u_t$
2. Construye $\hat{u}_t$.
3. Correr el modelo $\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t$ (sin intercepto)
4. Obtener $\hat{\rho}$.
5. Se usa $\hat{\rho}$ para construir el modelo $\tilde{y}_t$.
6. $\tilde{y}_t$ se corre por MCO.

Para AR(2)

② AR(2): $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$. Modelo cuasi-diferenciado:

$$(y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2}) = (1 - \rho_1 - \rho_2)\beta_0 + \beta_1(x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2}) + \varepsilon_t$$

① Cochranne - Orcutt.

1. Igual.
2. Igual
3. Corro $\hat{u}_t$ sobre $\hat{u}_{t-1}$ y $\hat{u}_{t-2}$, obtengo $\hat{\rho}_1$ y $\hat{\rho}_2$.
4. =
5. Igual
6. Igual

El hecho de rezagar hace perder observaciones $\Rightarrow$ siguiente método.

Praise-Winsten

⑧ "Praise - Winsten".

Considere $Y_1 = \beta_0 + \beta_1 X_1 + u_1$ (la primera observación)
 $\tilde{Y}_t = \beta_0 + \beta_1 \tilde{X}_t + u_t$

Queremos incluir la primera observación.

- ① $\text{cov}(u_1, u_t) = 0 \quad \forall t$
 ② $\text{Var}(u_1) = \sigma^2 = \sigma_e^2 / (1 - \rho^2)$
 $\text{Var}(u_t) = \sigma_e^2$ } Heteroscedasticidad

Para homogeneizar la varianza:

$$\sqrt{1 - \rho^2} Y_1 = \sqrt{1 - \rho^2} \beta_0 + \sqrt{1 - \rho^2} \beta_1 X_1 + \sqrt{1 - \rho^2} u_1$$

Se corre el modelo con $\sqrt{1 - \rho^2} Y_1$ en la primera observación y $\tilde{Y}_t$ en el resto.

E.E Robustos

Errores estándar robustos. (a correlación serial de orden 1)

$\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}$. Considere la matriz Σ de un AR(1).

$$\hat{\Sigma}^1? \quad \sigma^2 = E[u_t u_t'] \Rightarrow \hat{\sigma}^2 = \frac{\sum_t \hat{u}_t^2}{T - K - 1}$$

Obtener $\hat{\rho}$ de la regresión de $\hat{u}_t$ sobre $\hat{u}_{t-1}$ sin constante por MCO.
 $\hookrightarrow$ Reemplazamos en Σ para obtener $\hat{\Sigma}$.

Endogeneidad

ENDOGENEIDAD

Considere $Y = X\beta + u \rightarrow E[u|X] = 0$ si X es exógena
 $E[u|X] \neq 0$ si X es endógena

Causas

Razones

Simultaneidad

① Causalidad simultánea:

$$(I) Y_0 = \beta_0 + \beta_1 Y_1 + \beta_2 X + u_0$$

$$(II) Y_1 = \alpha_0 + \alpha_1 Y_0 + \alpha_2 Z + u_1$$

$$\Rightarrow Y_0 = \beta_0 + \beta_1 \alpha_0 + \beta_1 \alpha_1 Y_0 + \beta_1 \alpha_2 Z + \beta_1 u_1 + \beta_2 X + u_0$$

Influye sobre Y_0 $\leftarrow$

Entonces $E[u_1 Y_0] \neq 0$ en (II) $\Rightarrow Y_0$ es endógena.

V. Omitida

② Variable omitida

$$\text{Salud} = \beta_0 + \beta_1 \text{Ejercicio} + \beta_2 \text{Hábitos} + u$$

Pero los hábitos no son observables.

Lo que tengo es $\text{Salud} = A_0 + A_1 \text{Ejercicio} + E$ donde E incluye los hábitos.

↳ Hay endogeneidad si $\text{Corr}(\text{Ejercicio}, \text{Hábitos}) \neq 0 \Rightarrow E[E | \text{Ejercicio}] \neq 0$.

Asumimos que $\text{Hábitos} = C_0 + C_1 \text{Ejercicio} + w$.

$$\Rightarrow \text{Salud} = \beta_0 + \beta_1 \text{Ejercicio} + \beta_2 (C_0 + C_1 \text{Ejercicio} + w) + u$$
$$\text{Salud} = \beta_0 + (\beta_1 + \beta_2 C_1) \text{Ejercicio} + \beta_2 C_0 + \beta_2 w + u$$

$$\Rightarrow A_1 = \beta_1 + \beta_2 C_1 \Rightarrow E[A_1] = \beta_1 \text{ si y sólo si } \beta_2 C_1 = 0.$$

$\beta_2 C_1$ es el sesgo.

→ β_1 : Efecto directo del ejercicio

→ $\beta_2 C_1$: Efecto indirecto del ejercicio a través de los hábitos saludables.

E. de medición

③ Error de medición

Considere $Y = X\beta + u$, pero X no se observa, sino que se observa un $\tilde{X} = X + E$.

$$\Rightarrow Y = \tilde{X}\beta + (\beta E + u) \text{ pero } \tilde{X} = f(E)$$

Sesgo de selección

④ Sesgo de auto-selección.

Variables Instrumentales

Variables instrumentales

$\text{Salud} = B_0 + B_1 \text{Ejercicio} + u$. Al estimar B_1 , estoy estimando un efecto causal. La variable instrumental trata de extraer una fracción de la varianza del ejercicio que es endógena.

$Z \longrightarrow$ Produce varianza en el ejercicio

$Z \not\longrightarrow u$, p.e: distancia al gimnasio. (no afecta los hábitos)

Econometría Avanzada - Clase 7

20-08-2014

$$Y = X\beta + u, \text{ donde } E[u|X] \neq 0$$

↳ Variable instrumental que (1) $E[Z'X] \neq 0 \rightarrow$ Relevancia
(2) $E[Z'u] = 0 \rightarrow$ Exogeneidad

¿El estimador está identificado?

Premultiplicando por Z' : $Z'Y = Z'X\beta + Z'u$ ○ por (2)

$$E[Z'Y] = \beta E[Z'X] + E[Z'u]$$
$$\Rightarrow \beta = E[Z'X]^{-1} E[Z'Y]$$

Considere $y = \beta_0 + \beta_1 x_1 + u$ donde x es endógena.

Método de momentos: $\hat{\beta} = \frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)} = \frac{\sum_i (y_i - \bar{y})(z_i - \bar{z})}{\sum_i (x_i - \bar{x})(z_i - \bar{z})}$ } Estimador de Wald

El estimador de Wald requiere que $E[Z'X]$ sea invertible y que se cumple el supuesto de exogeneidad del instrumento.

Mínimos cuadrados en dos etapasConsidere el modelo $y = \beta_0 + \beta_1 x + u$

$$\text{Etapa 1: } \hat{x} = \hat{\alpha}_0 + \hat{\alpha}_1 z \Rightarrow \hat{\alpha}_1 = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sum (z_i - \bar{z})^2}$$

$$\text{Etapa 2: } \hat{y} = \beta_0 + \beta_1 \hat{x} \Rightarrow \hat{\beta}_1 = \frac{\sum (\hat{x}_i - \bar{\hat{x}})(y_i - \bar{y})}{\sum (\hat{x}_i - \bar{\hat{x}})^2}$$

 $\bar{\hat{x}} = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{z} = \alpha_0 + \alpha_1 \bar{z}$. Reemplazando:

$$\hat{\beta}_1 = \frac{\sum (\hat{x}_i - \alpha_0 - \alpha_1 \bar{z})(y_i - \bar{y})}{\sum (\hat{x}_i - \alpha_0 - \alpha_1 \bar{z})^2} = \frac{\sum (\alpha_0 + \alpha_1 \bar{z} - \alpha_0 - \alpha_1 \bar{z})(y_i - \bar{y})}{\sum (\alpha_0 + \alpha_1 \bar{z} - \alpha_0 - \alpha_1 \bar{z})^2}$$

MC2E colapsa
a Wald si
 $k=1$

$$\hat{\beta}_1 = \frac{\alpha_1 \sum (z_i - \bar{z})(y_i - \bar{y})}{\alpha_1^2 \sum (z_i - \bar{z})^2} = \frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{\sum (x_i - \bar{x})(z_i - \bar{z})} \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sum (z_i - \bar{z})^2} = \frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{\sum (x_i - \bar{x})(z_i - \bar{z})}$$

Modelo GeneralConsidere $y = X\beta + u$, $E[u|X] = 0$ $Z_{N \times L}$: $L = \#$ de variables explicativas exógenas + V. Instrument. $\hookrightarrow VI \geq \#$ de v. explicativas endógenas (cond. de orden)

(1) $E[z'X] \neq 0$

(2) $E[z'u] = 0$

$$\text{Etapa 1: } \hat{X} = (\hat{\beta}' Z')' \Rightarrow \hat{\beta} = (Z'Z)^{-1} Z'X \Rightarrow \hat{X} = Z(Z'Z)^{-1} Z'X$$

$$\text{Etapa 2: } \hat{y} = \hat{X}\hat{\beta} \Rightarrow \hat{\beta}_{IV} = (\hat{X}'\hat{X})^{-1} \hat{X}'y$$

$$\hat{\beta}_{IV} = (X'Z(Z'Z)^{-1}Z'X)^{-1} (X'Z(Z'Z)^{-1}Z'X)'(Z[Z'Z]^{-1}Z'X)y$$

$$\hat{\beta}_{IV} = (X'Z(Z'Z)^{-1}Z'X)^{-1} (X'Z(Z'Z)^{-1}Z')y$$

 $\hookrightarrow$ Denota un efecto local y no promedio. (LATE y no ATE)Consistenciasupuestos necesarios: (1) $E[z'u] = 0$ (2) $Z'Z$ es invertible(3) $X'Z$ es de rango completo K . (Refinamiento de $E[z'X] \neq 0$).

$$\hat{\beta}_{IV} = (X'Z(Z'Z)^{-1}Z'X)^{-1} (X'Z(Z'Z)^{-1}Z')(X\beta + u)$$

$$\hat{\beta}_{IV} = (X'Z(Z'Z)^{-1}Z'X)^{-1} (X'Z(Z'Z)^{-1}Z'X)\beta + (X'Z(Z'Z)^{-1}Z'X)^{-1} (X'Z(Z'Z)^{-1}Z'u)$$

$$\Rightarrow \hat{\beta}_{IV} = \beta + \left[\frac{1}{N} \sum x_i' z_i \right] \left[\frac{1}{N} \sum z_i' z_i \right]^{-1} \left[\frac{1}{N} \sum z_i' u_i \right]$$

Por ley de grandes números:

$$\begin{array}{lcl} \frac{1}{N} \sum x_i' z_i & \xrightarrow{p} & E[X'Z] \\ \frac{1}{N} \sum z_i' z_i & \xrightarrow{p} & E[Z'Z] \\ \frac{1}{N} \sum z_i' u_i & \xrightarrow{p} & E[Z'U] \end{array}$$

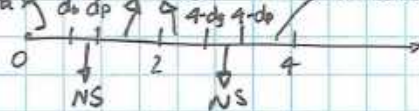
$$\hat{\beta}_1 = \text{Cov}(x, y)$$

Durbin Watson:

Autocorrelación positiva

No hay

Autocorrelación negativa



RESUMEN - CLASES 6 & 7

25-08-2014

ENDOGENEIDADConsiste en el incumplimiento del supuesto de independencia condicional. $E[U|X] \neq 0$.Razones

- (1) Simultaneidad: Considere
- $$(1) Y_0 = \beta_0 + \beta_1 Y_1 + \beta_2 X + u_0$$
- $$(2) Y_1 = A_0 + A_1 Y_0 + A_2 Z + u_1$$

Reemplazando (1) en (2):

$$Y_0 = \beta_0 + \beta_1 A_0 + \beta_1 A_1 Y_0 + \beta_1 A_2 Z + \beta_1 u_1 + u_0 + \beta_2 X$$

$$Y_0 = f(Z, X, u_1, u_0)$$

Influye sobre Y_0 .

- (2) Variable omitida: Considere $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, pero x_2 no es observable. Estimamos $y = \alpha_0 + \alpha_1 x_1 + \varepsilon$ donde ε comprende el efecto de x_2 .

$$\hookrightarrow \text{Si } \text{Cov}(x_1, x_2) \neq 0 \Rightarrow \text{Cov}(x_1, \varepsilon) \neq 0.$$

- (3) Error de medición: Considere $y = \beta_0 + \beta_1 x + u$, pero no tenemos x sino $\tilde{x} = x + \varepsilon$. $\Rightarrow y = \beta_0 + \beta_1 \tilde{x} + (u - \beta_1 \varepsilon)$. Sin embargo, $\tilde{x} = f(\varepsilon)$.

- (4) Sesgo de selección

Solución - Variable Instrumental

Condiciones de un buen instrumento

- (1) Relevancia: Produce varianza en la endógena. $E[Z'X] \neq 0$
 (2) Exogeneidad: Soluciona endogeneidad. $E[Z'u] = 0$.

Identificación del estimador de variables instrumentales

MC2E

Considere $y = X\beta + u$, $E[u|X] \neq 0$.
 Tenemos $Z_{L \times N}$ que cumple (1) y (2).

Etapas 1: $\hat{X} = Z(Z'Z)^{-1}Z'X$

Etapas 2: $\hat{Y} = \hat{X}\hat{\beta} = \hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}'Y$

$$\hat{\beta} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$$

$$\hat{X}'\hat{X} = X'Z(Z'Z)^{-1}Z'X$$

$$(\hat{X}'\hat{X})^{-1} = [X'Z(Z'Z)^{-1}Z'X]^{-1}$$

$$\Rightarrow \hat{\beta} = [X'Z(Z'Z)^{-1}Z'X]^{-1}[Z'(Z'Z)^{-1}Z'X]'Y$$

$$\Rightarrow \hat{\beta} = [X'Z(Z'Z)^{-1}Z'X]^{-1}[X'Z(Z'Z)^{-1}Z'Y]$$

$\hookrightarrow$ Está identificado.

Wald

Considere $y = X\beta + u$, $E[u|X] \neq 0$.

Premultiplico por Z' $\Rightarrow Z'y = Z'X\beta + Z'u$

$$E[Z'y] = E[Z'X\beta] + E[Z'u]$$

$$E[Z'y] = E[Z'X]\beta$$

$$\Rightarrow \hat{\beta} = E[Z'X]^{-1}E[Z'y]$$

Para $y = \beta_0 + \beta_1 x + u$, método de momentos:

$$\hat{\beta} = \frac{\text{Cov}(Z, y)}{\text{Cov}(Z, x)} = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})(x_i - \bar{x})}$$

$\hookrightarrow$ Requiere que $E[Z'X]$ sea invertible y que se cumpla el supuesto de exogeneidad del instrumento.

$\hookrightarrow$ Si $k=1$, $\hat{\beta}_{MC2E} = \hat{\beta}_{Wald}$

Consistencia del estimador de variables instrumentales

Supuestos necesarios: ① $E[z'u] = 0$ (Exogeneidad del instrumento)
 ② $z'z$ es invertible
 ③ $z'x$ es de rango completo K ($> E[z'x] \neq 0$)

$$\begin{aligned}\hat{\beta}_{IV} &= [X'Z(Z'Z)^{-1}Z'X]^{-1} [X'Z(Z'Z)^{-1}Z'Y] \\ \hat{\beta}_{IV} &= [X'Z(Z'Z)^{-1}Z'X]^{-1} [X'Z(Z'Z)^{-1}Z'X\beta] + [X'Z(Z'Z)^{-1}Z'X]^{-1} [X'Z(Z'Z)^{-1}Z'U] \\ \hat{\beta}_{IV} &= \beta + [X'Z(Z'Z)^{-1}Z'X]^{-1} [X'Z(Z'Z)^{-1}Z'U]\end{aligned}$$

$$\text{Reescribimos: } \hat{\beta}_{IV} = \beta + \left[\frac{1}{N} \sum (x_i' z_i) * \left(\frac{1}{N} \sum z_i' z_i \right)^{-1} * \left(\frac{1}{N} \sum z_i' x_i \right) \right]^{-1} \left[\frac{1}{N} \sum x_i' z_i * \left(\frac{1}{N} \sum z_i' z_i \right)^{-1} * \frac{1}{N} \sum z_i' u_i \right]$$

$$\begin{aligned}\text{Por WLLN: } \frac{1}{N} \sum (x_i' z_i) &\xrightarrow{p} E[X'Z] & \text{Por corolario 3.1} \\ \frac{1}{N} \sum (z_i' z_i) &\xrightarrow{p} E[Z'Z] & \left(\frac{1}{N} \sum z_i' z_i \right)^{-1} \xrightarrow{p} E[Z'Z]^{-1} \\ \frac{1}{N} \sum (z_i' x_i) &\xrightarrow{p} E[Z'X] \\ \frac{1}{N} \sum (z_i' u_i) &\xrightarrow{p} E[Z'U] = 0 \text{ por supuesto ①}\end{aligned}$$

$$\begin{aligned}\Rightarrow \text{plim } \hat{\beta}_{IV} &= \text{plim } \beta + [E[X'Z] * E[Z'Z]^{-1} * E[Z'X]]^{-1} [E[X'Z] * E[Z'Z]^{-1} * 0] \\ \Rightarrow \text{plim } \hat{\beta}_{IV} &= \beta\end{aligned}$$

Distribución asintótica del estimador de variables instrumentales

$$\text{Considere } \hat{\beta}_{IV} - \beta = \underbrace{\left[\frac{1}{N} \sum (x_i' z_i) * \left(\frac{1}{N} \sum z_i' z_i \right)^{-1} * \left(\frac{1}{N} \sum z_i' x_i \right) \right]^{-1}}_{a^{-1}} \left[\frac{1}{N} \sum x_i' z_i * \left(\frac{1}{N} \sum z_i' z_i \right)^{-1} * \frac{1}{N} \sum z_i' u_i \right]$$

$$\begin{aligned}\text{Por WLLN, corolario 3.1 y teorema de Slutsky } a^{-1} &\xrightarrow{p} [E(X'Z)E(Z'Z)^{-1}E(Z'X)]^{-1} \\ a^{-1} - A^{-1} &\xrightarrow{p} 0 \\ \Rightarrow a^{-1} - A^{-1} &= o_p(1)\end{aligned}$$

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) = a^{-1} \left[\left(\frac{1}{N} \sum x_i' z_i \right) \left(\frac{1}{N} \sum z_i' z_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum z_i' u_i \right) \right]$$

Considere $\frac{1}{\sqrt{N}} \sum (z_i' u_i)$. Es i.i.d por el supuesto ①
 Tiene varianza finita y valor esperado igual a cero.

$$\begin{aligned}\text{Por TLC: } \frac{1}{\sqrt{N}} \sum (z_i' u_i) &\xrightarrow{d} N(0, V) \quad V = \text{Var}(z_i' u_i) = E[z_i' u_i u_i' z_i] = \sigma^2 E[z_i' z_i] \\ &\Rightarrow \frac{1}{\sqrt{N}} \sum (z_i' u_i) = o_p(1) \Rightarrow \text{Tiene cota}\end{aligned}$$

Por WLLN, corolario 3.1 y teorema de Slutsky:

$$\begin{aligned}a^{-1} \left(\frac{1}{N} \sum x_i' z_i \right) \left(\frac{1}{N} \sum z_i' z_i \right)^{-1} &\xrightarrow{p} A^{-1} E[X'Z] E[Z'Z]^{-1} \\ \Rightarrow b - B &= o_p(1)\end{aligned}$$

$$o_p(1) * o_p(1) = o_p(1) = b * \frac{1}{\sqrt{N}} \sum (z_i' u_i) - B * \frac{1}{\sqrt{N}} \sum (z_i' u_i)$$

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) \Rightarrow \sqrt{N}(\hat{\beta}_{IV} - \beta) \xrightarrow{p} B \frac{1}{\sqrt{N}} \sum (z_i' u_i)$$

Asintóticamente equivalentes.

$$\begin{aligned}\text{Si } \frac{1}{\sqrt{N}} \sum z_i' u_i &\xrightarrow{d} N(0, V) \text{ por tma del mapa continuo } B^{-1} \frac{1}{\sqrt{N}} \sum (z_i' u_i) \xrightarrow{d} N(0, BVB) \\ \Rightarrow \sqrt{N}(\hat{\beta}_{IV} - \beta) &\rightarrow N(0, BVB)\end{aligned}$$

$$\begin{aligned}\text{Avar}(\sqrt{N}[\hat{\beta}_{IV} - \beta]) &= [E(X'Z)E(Z'Z)^{-1}E(Z'X)]^{-1} E[X'Z]E[Z'Z]^{-1}\sigma^2 E[Z'Z] [E[X'Z]E[Z'Z]^{-1}E(Z'X)]^{-1} \\ &= \sigma^2 A^{-1}\end{aligned}$$

$$\begin{aligned} \text{Avar}(\sqrt{N}[\hat{\beta}_{IV} - \beta]) &= \sigma^2 A^{-1} E[X'Z] E[Z'Z]^{-1} E[Z'Z] A^{-1} E[X'Z]' E[Z'Z]^{-1} \\ &= \sigma^2 A^{-1} E[X'Z] A^{-1} E[Z'X] E[Z'Z]^{-1} \\ A^{-1} &= [E(X'Z) E(Z'Z)^{-1} E(Z'X)]^{-1} \\ A^{-1} &= E(Z'X)^{-1} E(Z'Z) E(X'Z)^{-1} \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{Avar}(\sqrt{N}[\hat{\beta}_{IV} - \beta]) &= \sigma^2 E(Z'X)^{-1} E(Z'Z)^{-1} E(X'Z)^{-1} E(X'Z) E(Z'X)^{-1} E(Z'Z)^{-1} E(X'Z) E(X'Z)' E(Z'Z)^{-1} \\ &= \sigma^2 E(Z'X)^{-1} E(Z'Z)^{-1} E(Z'X)^{-1} \end{aligned}$$

$$\begin{aligned} B \vee B' &= A^{-1} E(X'Z) E(Z'Z)^{-1} \sigma^2 E(Z'Z)^{-1} E(Z'Z)^{-1} E(X'Z)' A^{-1} \\ &= E(Z'X)^{-1} E(Z'Z) E(X'Z)^{-1} E(X'Z) E(Z'Z)^{-1} \sigma^2 E(Z'Z)^{-1} E(Z'X) E(Z'X)^{-1} E(Z'Z) E(X'Z)^{-1} \\ &= \sigma^2 E(Z'X)^{-1} E(Z'Z) E(Z'Z)^{-1} E(Z'Z)^{-1} E(Z'Z) E(X'Z)^{-1} \end{aligned}$$

$$\begin{aligned} \text{Avar}(\sqrt{N}[\hat{\beta}_{IV} - \beta]) &= A^{-1} E(X'Z) E(Z'Z)^{-1} \sigma^2 E(Z'Z)^{-1} E(Z'Z)^{-1} E(Z'X) A^{-1} \\ &= \sigma^2 E(Z'X)^{-1} E(Z'Z) E(X'Z)^{-1} E(X'Z) E(Z'Z)^{-1} E(Z'Z)^{-1} E(Z'X) E(Z'X)^{-1} E(Z'Z) E(X'Z)^{-1} \\ &= \sigma^2 E(Z'X)^{-1} E(Z'Z) E(Z'Z)^{-1} E(Z'Z)^{-1} E(Z'Z) E(X'Z)^{-1} \\ &= \sigma^2 E(Z'X)^{-1} E(Z'Z)^{-1} E(X'Z)^{-1} \end{aligned}$$

$$\begin{aligned} B \vee B' &= [E(X'Z) E(Z'Z)^{-1} E(Z'X)]^{-1} E(X'Z) E(Z'Z)^{-1} \sigma^2 E(Z'Z)^{-1} E(Z'X) [E(Z'X)^{-1}] E(Z'Z) E(X'Z)^{-1} \\ &= E(Z'X)^{-1} E(Z'Z) E(X'Z)^{-1} E(X'Z) \sigma^2 E(X'Z)^{-1} \\ &= \sigma^2 E(Z'X)^{-1} E(Z'Z) E(X'Z)^{-1} = \sigma^2 [E(X'Z) E(Z'Z)^{-1} E(Z'X)]^{-1} \\ &= \sigma^2 A^{-1} \end{aligned}$$

$$\Rightarrow \sqrt{N}[\hat{\beta}_{IV} - \beta] \xrightarrow{d} N(0, \sigma^2 A^{-1})$$

$$\Rightarrow \hat{\beta}_{IV} \overset{\sim}{\sim} N\left(\beta, \frac{\sigma^2 A^{-1}}{N}\right)$$

	MCO	MC2E
Var. explicativa es exógena	Consistente	Consistente Ineficiente
Var. Explicativa es endógena	Inconsistente	Consistente

Prueba de HausmanPrueba de Hausman

Considere el modelo $Y = \beta_0 + \beta_1 Y_2 + \beta_2 X_1 + \beta_3 X_2 + u$

$$E(u|Y_2) \neq 0$$

$$\rightarrow Z = \begin{cases} E[Z'X] \neq 0 \\ E[Z'u] = 0 \end{cases}$$

Primera etapa $Y_2 = A_0 + A_1 X_1 + A_2 X_2 + A_3 Z + v \Rightarrow \begin{cases} Y_2 \text{ es endógena} \\ \text{si } v \text{ es endógena} \end{cases}$

$E(u|X_1, X_2, Z) = 0$

v es endógena si $E[u \cdot v] \neq 0$ (Hay correlación entre u y v)

Definimos $u = \beta v + \varepsilon$. $H_0: Y_2$ es exógena $\rightarrow \beta = 0$

Modelo $Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 X_1 + \beta_3 X_2 + \beta v + \varepsilon$

v no se observa, pero se puede obtener de $\hat{v}$.

- ① Construir primera etapa
- ② Correr Y_1 sobre $Y_2, X_1, X_2, \hat{v}$
- ③ Prueba estadística de significancia individual para $\hat{v}$.

Pruebas de validez del InstrumentoPruebas de validez de IV① Relevancia

Relevancia

En la primera etapa verifico si el instrumento es relevante.

H_0 : El/los instrumentos los irrelevantes

F (Idealmente > 10).

Exogeneidad

② Exogeneidad - Prueba de sobreidentificación.

Prueba de restricciones sobreidentificadas

u no se observa, pero se puede estimar por $\hat{u}$. Sin embargo, $\hat{u}$ sólo se estima de manera consistente si hay un instrumento válido.

$\rightarrow$ Sólo se puede hacer si el modelo está sobreidentificado.

Considere $Y = \beta_0 + \beta_1 X + u$, $E(u|X) \neq 0$, VI: Z_1, Z_2 .

- (1) Asumo que Z_1 es exógeno.
- (2) Estimo ecuación principal por MC2E usando Z_1 . Sólo si Z_1 es exógeno se puede demostrar que $\hat{u}$ es un estimador consistente de u .
- (3) Como $\hat{u}$ contra todas las variables exógenas incluidas VI.

H_0 : VI no están correlacionadas con $\hat{u}$

$\chi^2_q = NR^2$ donde q : # restricciones de sobreidentificación
= # VI - # Endógenas (VI sobrantes)

Si $\chi^2_q > \chi^2$, $RH_0 \Rightarrow$ Al menos uno de los instrumentos está correlacionado con u .

Desventajas de la estimación por VI

Desventajas de $\hat{\beta}_{IV}$

Es la metodología más sencilla a la hora de solucionar el problema de endogeneidad.

Instrumentos débiles

① Instrumentos débiles

Considere el modelo univariado con x endógena:

$$\hat{\beta}_{IV} = \frac{\text{Cov}(y, z)}{\text{Cov}(x, z)} = \frac{\text{Cov}(\beta_0 + \beta_1 x + u, z)}{\text{Cov}(x, z)} = \beta_1 \frac{\text{Cov}(x, z)}{\text{Cov}(x, z)} + \frac{\text{Cov}(z, u)}{\text{Cov}(x, z)}$$

$$\hat{\beta}_{IV} = \beta_1 + \frac{\text{Cov}(z, u)}{\text{Cov}(x, z)} = \beta_1 \Leftrightarrow \text{Cov}(z, u) = 0. \text{ De lo contrario, hay un sesgo.}$$

Si $\text{Cov}(x, z) \rightarrow 0 \Rightarrow$ Instrumento es débil $\Rightarrow$ Aumenta el sesgo.

Ineficiencia

② Errores estándar son mayores que los de MCO, y más aún si $\text{Cov}(x, z) \rightarrow 0$.

Interpretación del Estimador

③ Interpretación del estimador.

MCO: Estima efecto de x sobre $y \Rightarrow$ Efecto promedio de x sobre y

MCZE: Estima efecto de $\hat{x}$ sobre y , donde $\hat{x} : \Delta x$ que se debe a z . $\Rightarrow$ Efecto promedio LOCAL de x sobre y , donde LOCAL = submuestra de ind. que cambiaron su x debido a z .

RESUMEN - CLASE 8

Prueba de Hausman

Obj: Determinar si hay endogeneidad en el modelo.

① Estime la primera etapa: las endógenas en función de los instrumentos.

Endógenas $\{ Y_0 = Z\beta + v$
Segmento exógeno

Si Y es endógena $\Rightarrow v$ es endógena
 v es endógena si $E[u|v] \neq 0$.

② Estimar $\hat{v}$.

③ Correr el modelo $Y = x\beta + Y_0\alpha + \beta\hat{v} + u$

④ Realizar una prueba de significancia individual sobre $\hat{\beta}$.

$H_0: \beta = 0 \Rightarrow Y_0$ son exógenas

$H_1: \beta \neq 0 \Rightarrow Y_0$ son endógenas

RESUMEN - CLASE 8

Pruebas de validez del instrumento

I. RELEVANCIA

- 1) Estime la primera etapa: $Y_0 = Z\beta + v$
- 2) Realizar una prueba de significancia individual.

$$H_0: \beta = 0, Z \text{ no es relevante} \quad H_1: \beta \neq 0, Z \text{ es relevante}$$

En general: $F(\text{significancia conjunta}) > 10$.

II. EXOGENEIDAD (Prueba de sobreidentificación)

$\hat{u}$ sólo se estima de manera consistente si hay un instrumento válido $\Rightarrow$ sólo puede hacerse si el modelo está sobreidentificado.

Considere $Y = \beta_0 + \beta_1 X + u$, $E[u|X] \neq 0$, VI: X_1, X_2

- 1) Asumir que Z_1 es exógeno.
- 2) Estimar $Y = \beta_0 + \beta_1 X + u$ por MC2E usando Z_1 .
- 3) Estimar $\hat{u}$ contra todos los instrumentos.

$$H_0: \text{VI no están correlacionadas con } \hat{u}.$$

$$\chi^2_q = NR^2 \text{ donde } q: \# \text{VI} - \# \text{Endógenas}$$

Desventajas de $\hat{\beta}_{IV}$

- 1) Si el instrumento es débil, el sesgo aumenta.

$$\hat{\beta}_{IV} = \beta + \frac{\text{Cov}(Z, u)}{\text{Cov}(Z, X)} \rightarrow \infty \text{ as } \text{Cov}(Z, X) \rightarrow 0$$

- 2) Los errores estándar son ineficientes, y la ineficiencia aumenta si $\text{Cov}(X, Z) \rightarrow 0$.
- 3) Los $\hat{\beta}_{IV}$ sólo denotan efectos promedios locales.

Modelos de elección discretaTipos de
modelos de
elección
discreta

La variable dependiente es una variable categórica.

Ej 1: Elección binomial - Presencia o ausencia de una característica.

$$D_i = \begin{cases} 1 & \text{si es trabajador} \\ 0 & \text{si no es trabajador} \end{cases}; \quad D_i = f(X_{ik}, U_i)$$

Ej 2: Elección multinomial - Ayuda a explicar la escogencia de distintas alternativas tanto en función de los individuos como de las alternativas en sí.Ej 3: Elección ordenada - El orden de las categorías se importa.Máxima verosimilitudIntroducción
a MVLa función de densidad (pdf) de una variable aleatoria x con vector de parámetros θ se define $f(x|\theta)$.

↳ Expresión matemática del proceso generador de esos datos.

La distribución conjunta de una muestra i.i.d de tamaño N donde $f(x_i|\theta)$ es la pdf de x es:

$$f(x_1, x_2, \dots, x_N | \theta) = \prod_{i=1}^N f(x_i | \theta) = \mathcal{L}(\theta | x) \quad \left. \begin{array}{l} \text{Función de} \\ \text{verosimilitud} \end{array} \right\}$$

Probabilidad de observar esa muestra en particular

Ej: x es binomial, $x = \{0, 1\} \Rightarrow f(x|p) = p^x (1-p)^{1-x}$

$$\Rightarrow \mathcal{L}(p|x) = \prod_{i=1}^N p^{x_i} (1-p)^{1-x_i}$$

Introducción
al est.
de máx. ver.El estimador de máxima verosimilitud es $\hat{\theta}_{MV} = \arg\max \mathcal{L}(\theta|x)$

↳ El parámetro que maximiza la probabilidad de observar la muestra que tengo.

$$\hat{\theta}_{MV} = \arg\max \ln \mathcal{L}(\theta|x)$$

$$\ln \mathcal{L}(\theta|x) = \sum_{i=1}^N \ln f(x_i|\theta)$$

Distribución
binomialEj: Considere $x_1, x_2, \dots, x_N$ tal que $f(x_i|\theta) = f(x_i|p) = p^{x_i} (1-p)^{1-x_i}$

$$N = 100 \quad x_i = \begin{cases} \text{Se choca} & = 1 \\ \text{No se choca} & = 0 \end{cases}$$

$$\hat{p}_{MV} = \arg\max \left(\sum_{i=1}^{100} x_i \ln p + (N - \sum_{i=1}^{100} x_i) \ln(1-p) \right)$$

$$\text{CPO: } \frac{d\mathcal{L}}{d\hat{p}_{MV}} = \sum \frac{x_i}{p} + (N - \sum x_i) \left(-\frac{1}{1-p} \right) = 0$$

$$\frac{\sum x_i}{p} = (N - \sum x_i) \left(\frac{1}{1-p} \right)$$

$$(1-p) \sum x_i = p(N - \sum x_i)$$

$$\hat{p}_{MV} = \sum x_i / N$$

Distribución normal

Ej: $X_1, X_2, \dots, X_N$ donde:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

$$\ln f(x|\mu, \sigma^2) = -\frac{1}{2}\left[\ln(2\pi) + \ln(\sigma^2) + \frac{(x-\mu)^2}{\sigma^2}\right]$$

$$\mathcal{L}(\theta|x) = -\frac{1}{2}\left\{N \ln(2\pi) + N \ln(\sigma^2) + \sum_{i=1}^N \frac{(x_i - \mu)^2}{\sigma^2}\right\}$$

$$\frac{d\mathcal{L}}{d\mu} = -\frac{1}{2} \sum_{i=1}^N \frac{2(x_i - \mu)}{\sigma^2} = 0$$

$$\sum_{i=1}^N x_i = N\mu$$

$$\hat{\mu}_{MV} = \frac{\sum x_i}{N}$$

$$\frac{d\mathcal{L}}{d\sigma^2} = -\frac{1}{2}\left\{\frac{N}{\sigma^2} + (-1)\sigma^{-2} \sum_{i=1}^N (x_i - \mu)^2\right\} = 0$$

$$\frac{1}{\sigma^2} \left(N - \frac{1}{\sigma^2} \sum_{i=1}^N [x_i - \mu]^2 \right) = 0$$

$$\frac{1}{\sigma^4} \sum_{i=1}^N (x_i - \mu)^2 = \frac{N}{\sigma^2}$$

$$\frac{\sum (x_i - \mu)^2}{N} = \hat{\sigma}_{MV}^2$$

Distribución uniforme

Ej: Suponga una muestra $X_1, X_2, \dots, X_N$ con la siguiente pdf:

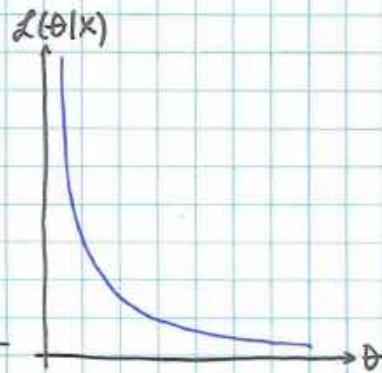
$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \forall x \in (0, \theta] \\ 0 & \text{e.o.c} \end{cases}$$

$$\mathcal{L}(\theta|x) = \prod_{i=1}^N \frac{1}{\theta} = \left(\frac{1}{\theta}\right)^N \text{ si } x \in (0, \theta]$$

$\mathcal{L}(\theta|x)$ va creciendo a medida que θ se hace más pequeño $\Rightarrow$ La gracia es que $\hat{\theta}_{MV}$ sea el θ más pequeño posible.

Sabemos que $0 < x < \theta \Rightarrow \theta \geq \max\{x_i\}$

$$\Rightarrow \hat{\theta}_{MV} = \max\{x_i\}$$



Propiedad de invarianza

① Propiedad de invarianza: Sea $h(\cdot)$ es una función uno a uno.

$$h(\hat{\theta}_{MV}) = h(\hat{\theta}_{MV})$$

P.e: $\hat{\sigma}^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum (x_i - \hat{\mu}_{MV})^2}{N}$

Teorema del límite inferior de Cramer-Rao

② Varianza de MV - Teorema del límite inferior de Cramer-Rao

$$\text{Var}(\hat{\theta}_{MV}) = E \left[\frac{d^2 \ln \mathcal{L}(\theta|X)}{d\theta d\theta'} \right]^{-1}$$

Matriz de información

Ej: Considere $\frac{d\mathcal{L}}{d\mu} = \frac{\sum (x_i - \mu)}{\sigma^2} = \frac{\sum x_i - N\mu}{\sigma^2} = \frac{\sum x_i}{\sigma^2} - \frac{N\mu}{\sigma^2}$

$$\frac{d^2 \mathcal{L}}{d\mu^2} = -\frac{N}{\sigma^2}$$

Usando el teorema del límite inferior de Cramer-Rao:

$$\text{Var}(\hat{\mu}_{MV}) = E \left[\frac{\sigma^2}{N} \right] = \frac{\hat{\sigma}_{MV}^2}{N}$$

RESUMEN - CLASE 10

MODELOS DE ELECCIÓN DISCRETA - Variable dependiente es dicótoma.

- I. Elección binomial
 - II. Elección multinomial
 - III. Elección categórica
- } El orden no importa
→ El orden sí importa

Método de estimación: Máxima verosimilitud

Asumimos una función de distribución para las variables: $f(X|\theta)$. Para una muestra $X_1, X_2, \dots, X_N$ la función de densidad conjunta es la probabilidad de observar esa muestra particular:

Dado que $X_1, X_2, \dots, X_N$ es i.i.d: $f(X_1, X_2, \dots, X_N|\theta) = \prod_{i=1}^N f(X_i|\theta) = \mathcal{L}(\theta|X)$
Función de verosimilitud

$\hat{\theta}_{MV} = \text{argmax} \{ \mathcal{L}(\theta|X) \}$. Es más fácil $\hat{\theta}_{MV} = \text{argmax} \{ \ln \mathcal{L}(\theta|X) \}$
↳ $\ln(\cdot)$ - Función monótonica $\Rightarrow$ Conserva máximos.

$$\hat{\theta}_{MV} = \text{argmax} \left\{ \ln \left[\prod_{i=1}^N f(X_i|\theta) \right] \right\} = \left\{ \sum_{i=1}^N \ln f(X_i|\theta) \right\}$$

Propiedades del estimador de MV

① Invarianza: Si $h(\theta)$ es una función uno a uno $\Rightarrow$

$$h(\hat{\theta}_{MV}) = \widehat{h(\theta_{MV})}$$

② Teorema del límite inferior de Cramer-Rao

$$\text{Var}(\hat{\theta}_{MV}) = E \left[\frac{d^2 \ln \mathcal{L}(\theta|X)}{d\theta d\theta'} \right]^{-1} \text{ donde } \frac{d^2 \ln \mathcal{L}(\theta|X)}{d\theta d\theta'} \text{ Matriz de información.}$$

Introduction

- ↳ "Unified approach to estimation" + The most efficient estimator of models that assume a distribution given X .
- But: $\rightarrow$ Not robust. Failure of normality = Inconsistent.
 - Specification of density - Somehow arbitrary.

Conditional
Maximum
Likelihood
EstimationPreliminaries & examples

- ↳ We estimate $\hat{\theta}$ in conditional distribution of y_i given $x_i \Rightarrow$ Conditional Maximum Likelihood Estimation.
- x_i can be random or nonrandom.

General Framework for Conditional MLE

$p_0(\vec{y}|\vec{x}) \rightarrow$ conditional density of y_i given $x_i = \vec{x}$, y, x dummies.
(True density) $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^K$ possible values of x_i & y_i (supports)

Assume that $\forall \mathcal{X} p_0(\cdot|\vec{x})$: Density w/ respect to a σ -finite measure
Noted $\nu(d\vec{y})$: can be chosen to make y discrete continuous or a mixture.

Kullback-Leibler
Information
Inequality

Conditional Kullback-Leibler information inequality: $\forall f(\cdot|\vec{x})$ no negativa

$$\int_{\mathcal{Y}} f(\vec{y}|\vec{x}) \nu(d\vec{y}) = 1 \quad \forall \vec{x} \in \mathcal{X} \quad (*)$$

$$\mathcal{K}(f|\vec{x}) = \int_{\mathcal{Y}} \log [p_0(\vec{y}|\vec{x}) / f(\vec{y}|\vec{x})] p_0(\vec{y}|\vec{x}) \nu(d\vec{y}) \geq 0 \quad \forall \vec{x} \in \mathcal{X}$$

$\downarrow$

Se minimiza si $f = p_0$ porque $\mathcal{K}(f|\vec{x}) = 0$ si $f = p_0$.

se aplica a un modelo paramétrico para $p_0(\cdot|\vec{x})$: $\{f(\cdot|\vec{x}, \vec{\theta}), \vec{\theta} \in \Theta \subset \mathbb{R}^p\}$
donde se satisface $(*)$

↳ Correctly specified model of conditional density $p_0(\cdot|\vec{x})$ if $\exists \vec{\theta}_0$
 $f(\cdot|\vec{x}, \vec{\theta}_0) = p_0(\cdot|\vec{x}) \quad \forall \vec{x} \in \mathcal{X}$.

$\forall \vec{x} \in \mathcal{X}, \mathcal{K}(f, \vec{x}) = E[\log(p_0(\vec{y}_i|\vec{x}_i)) | \vec{x}_i = \vec{x}] - E[\log(f(\vec{y}_i|\vec{x}_i)) | \vec{x}_i = \vec{x}]$
 $\Rightarrow$ If correctly specified model: $\mathcal{K}(f, \vec{x}) \geq 0$.

$$E[\log\{p_0(y_i|x_i)\} | x_i = \vec{x}] - E[\log\{f(y_i|\vec{x}_i)\} | \vec{x}_i = \vec{x}] \geq 0$$

$$E[\ell_i(\vec{\theta}_0) | \vec{x}_i] \geq E[\ell_i(\vec{\theta}) | \vec{x}_i], \quad \vec{\theta} \in \Theta$$

Donde $\ell_i(\vec{\theta}) = \ell(\underbrace{\vec{y}_i, \vec{x}_i}_{\text{Random}}, \vec{\theta}) = \log f(\vec{y}_i | \vec{x}_i; \vec{\theta})$

Tomando valor esperado y usando expectativas iteradas: $\vec{\theta}_0$ resuelve:

$$\max_{\vec{\theta} \in \Theta} E[\ell_i(\vec{\theta})] \xrightarrow{\text{nostrat}} \max_{\vec{\theta} \in \Theta} \frac{\sum \log f(y_i | x_i, \vec{\theta})}{N}$$

Consistency of CMLE

Consistency

Th 13.1 - Consistency of CMLE

Let $\{(x_i, y_i) : i = 1, 2, \dots, N\}$ be a random sample with $x_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$. $\mathcal{X} \in \mathbb{R}^K$, $\mathcal{Y} \in \mathbb{R}^P$. Let $\Theta \subset \mathbb{R}^p$ be the parameter set and denote the parametric model of conditional density as $\{f(\cdot | x; \theta)\}$.

Assume that: (a) $f(\cdot | x; \theta)$ is a true density w/ respect to measure $\nu(dy) \forall x, \theta$, so $(*)$ holds.

! (b) $\exists \theta_0 \in \Theta$ t.q. $p_0(\cdot | x) = f(\cdot | x; \theta_0)$, θ_0 is unique.

(c) Θ is a compact set ($\approx$ finite)

(d) $\forall \theta \in \Theta$, $\ell(\cdot, \theta)$ is a Borel measurable function on $\mathcal{Y} \times \mathcal{X}$ (doesn't need to be checked)

! (e) $\forall (y, x)$, $\ell(y, x, \cdot)$ is continuous on Θ

(f) $|\ell(w, \theta)| \leq b(w) \forall \theta \in \Theta$ & $E[b(w)] < \infty$

$\Rightarrow \exists$ a solution to $\max_{\theta} \frac{1}{N} \sum \log f(y_i | x_i; \theta)$

Main concern: Densities are correctly specified.

Asymptotic Normality

Asymptotic Normality

Assume that (a) $\theta_0 \in \Theta$, (b) $\ell_i(\theta)$ is twice differentiable in Θ .

$\Rightarrow$ score of log likelihood: $s_i(\theta) = \nabla_{\theta} \ell_i(\theta) = \left(\frac{\partial \ell_i}{\partial \theta_1}(\theta), \frac{\partial \ell_i}{\partial \theta_2}(\theta), \dots, \frac{\partial \ell_i}{\partial \theta_p}(\theta) \right)'$

Usually: $E[s_i(\theta) | x_i] = 0$.

Let the Hessian for obs. i be the $P \times P$ matrix of second partial derivatives of $\ell_i(\theta)$:

$$H_i(\theta) = \nabla_{\theta} s_i(\theta) = \nabla_{\theta}^2 \ell_i(\theta)$$

Let $A_0 = -E[H_i(\theta_0)] \rightarrow$ p.s.d if θ_0 is identified.

$$\Rightarrow \sqrt{N}(\hat{\theta} - \theta_0) \overset{d}{\sim} N(0, A_0^{-1} B_0 A_0^{-1}), B_0 = \text{Var}[s_i(\theta_0)] = A_0$$

Modelos de elección discreta

$$P(\text{Evento } j \text{ ocurra}) = F(X_{ij}, u_i | \theta)$$

↳ Cualitativa

↳ Parámetros

Modelos de
elección
binomialElección binomial

$$D_i = \begin{cases} 1 & \text{si trabaja} \\ 0 & \text{si no} \end{cases} \quad \text{Datos: } \{D_i, X_i\}_{i=1, \dots, N}$$

MPL

MCO: $D_i = X_i \beta + u_i$. Es consistente e insesgado si se cumple $E[u_i | X_i] = 0$

$$E[D_i | X_i] = X_i \beta + E[u_i | X_i] = X_i \beta. \quad D_i \text{ es binaria:}$$

$$E[D_i | X_i] = 1 \cdot \Pr(D_i = 1 | X_i \beta) + 0 \cdot [1 - \Pr(D_i = 1 | X_i \beta)]$$

$$E[D_i | X_i] = \Pr(D_i = 1 | X_i \beta) = X_i \beta$$

Probabilidad de éxito

$$\hat{\beta}_{MCO} : \Delta \Pr(D_i = 1 | X_i \beta) \text{ si } \Delta X_i = 1 \text{ (Puntos porcentuales).}$$

Problemas
del modelo
MPLInconvenientes: ① $\Pr(D_i = 1 | X_i \beta) = X_i \hat{\beta}$ no necesariamente $\in [0, 1]$.② $D_i = 1$ ocurre con probabilidad $X_i \beta \Rightarrow D_i = 0$ ocurre con probabilidad $1 - X_i \beta$.

$$\Rightarrow \text{Si } D_i = 1 : \hat{u}_i = 1 - X_i \beta \text{ con pr. } X_i \beta$$

$$\text{Si } D_i = 0 : \hat{u}_i = -X_i \beta \text{ con pr. } (1 - X_i \beta)$$

$$\Rightarrow \hat{u} \sim \text{Bernoulli}$$

$$\text{Var}(u_i) = p(1-p) = X_i \beta (1 - X_i \beta) = \text{Heteroscedástico}$$

↳ Solucionar por MCP, errores robustos.

Modelos de elección discreta no lineales

$$\text{Obj: } \Pr(\text{evento } j \text{ ocurra}) = F(X_{ij}, u_i | \theta), \quad 0 \leq F(\cdot) \leq 1.$$

Máxima verosimilitud

$$D_i = \begin{cases} 1 & \text{con } \Pr(D_i = 1 | X_i \beta) \\ 0 & \text{con } \Pr(D_i = 0 | X_i \beta) = 1 - \Pr(D_i = 1 | X_i \beta) \end{cases}$$

En una muestra de tamaño N :

$$\mathcal{L}(\theta | X) = \prod_{i=1}^N \Pr(D_i = 1 | X_i \beta)^{D_i} \cdot [1 - \Pr(D_i = 1 | X_i \beta)]^{1-D_i}$$

$$\ln \mathcal{L}(\theta | X) = \sum_{i=1}^N \left\{ D_i \ln [\Pr(D_i = 1 | X_i \beta)] - (1 - D_i) \ln [1 - \Pr(D_i = 1 | X_i \beta)] \right\}$$

$$\hat{\beta}_{ML} = \arg \max \ln \mathcal{L}(\beta | X)$$

Elección multinomial

Tres alternativas mutuamente excluyentes.

$$D_{1i} = \begin{cases} 1 & \text{si elige alt. 1} \\ 0 & \text{e.o.c} \end{cases}$$

$$D_{2i} = \begin{cases} 1 & \text{si elige alt. 2} \\ 0 & \text{e.o.c} \end{cases}$$

Datos: $\{D_{1i}, D_{2i}, X_{1i}, X_{2i}, X_{3i}\}_{i=1, \dots, N}$

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N \left\{ D_{1i} \ln [\Pr(D_{1i}=1|X\beta)] + D_{2i} \ln [\Pr(D_{2i}=1|X\beta)] + (1 - D_{1i} - D_{2i}) \ln [1 - \Pr(D_{1i}=1|X\beta) - \Pr(D_{2i}=1|X\beta)] \right\}$$

$$\ln \mathcal{L}(\theta|X) = \sum_{i: \text{con } D_1=1} \ln \Pr(D_{1i}=1|X\beta) + \sum_{i: \text{con } D_2=1} \ln \Pr(D_{2i}=1|X\beta) + \sum_{i: \text{con } D_3=1} \ln [1 - \Pr(D_{1i}=1|X\beta) - \Pr(D_{2i}=1|X\beta)]$$

Modelo de utilidad aleatoria

¿Qué es $\Pr(D_{1i}=1|X\beta)$? Definimos un modelo de utilidad aleatoria.

Binomial: $D_i = \begin{cases} 1 & \text{si escoge} \\ 0 & \text{e.o.c.} \end{cases}$ Asumimos que los agentes son racionales.

Regla de decisión: Maximización de utilidad

Si escoge 1 $\Rightarrow$ deriva $U_{1i} = X_{1i}\beta + \epsilon_{1i}$ (Lineales)
Si escoge 0 $\Rightarrow$ deriva $U_{0i} = X_{0i}\beta + \epsilon_{0i}$

Regla de decisión: Ej 1: $D_i = \begin{cases} 1 & \text{si } U_{1i} > U_{0i} \\ 0 & \text{e.o.c.} \end{cases}$



$$\Rightarrow \Pr(D_i=1|X\beta) = \Pr(U_{1i} > U_{0i}) = \Pr(X_{1i}\beta + \epsilon_{1i} > X_{0i}\beta + \epsilon_{0i}) = \Pr(\epsilon_{0i} < X_{1i}\beta - X_{0i}\beta + \epsilon_{1i})$$

$\approx F(\text{umbrales}), F(\cdot)$ densidad acumulada

No mutuamente excluyente.

Multinomial - No mutuamente excluyente.

Regla de decisión $D_{1i} = \begin{cases} 1 & \text{si } U_{1i} > 0 \\ 0 & \text{e.o.c.} \end{cases}$ $D_{2i} = \begin{cases} 1 & \text{si } U_{1i} > 0 \text{ y } U_{2i} > U_{1i} \\ 0 & \text{e.o.c.} \end{cases}$

1. ¿Cuál es el conjunto de alternativas factibles?

U_{1i}	U_{2i}	D_1	D_2
> 0	$> 0, U_{2i} > U_{1i}$	✓	✓
	$> 0, U_{2i} < U_{1i}$	✓	x
	< 0	✓	x
< 0	$> 0, U_{2i} > U_{1i}$	x	x
	$< 0, U_{2i} > U_{1i}$	x	x
	$< 0, U_{2i} < U_{1i}$	x	x

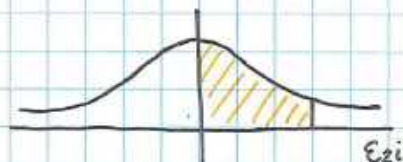
} 3 grupos de personas.

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N \left\{ D_{1i} \ln (\Pr(U_{1i} > 0)) + D_{1i} D_{2i} \{ \ln [\Pr(U_{1i} > 0)] + \ln [\Pr(U_{2i} > U_{1i})] \} + (1 - D_{1i} - D_{1i} D_{2i}) \{ 1 - [\Pr(U_{1i} > 0)] \} \right\}$$

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N \left\{ D_{1i} \ln (\Pr(U_{1i} > 0)) + (1 - D_{1i}) \ln (\Pr[U_{2i} < 0]) \right\}$$

$$= \sum_{i=1}^N \left\{ D_{1i} \{ \ln \Pr[U_{2i} > U_{1i} | U_{1i} > 0] \} + D_{1i} D_{2i} + (1 - D_{1i}) \ln \Pr[U_{2i} < U_{1i} | U_{1i} > 0] + (1 - D_{1i}) \ln (\Pr[U_{1i} < 0]) \right\}$$

$$\Pr(D_i = 1, D_2 = 1) = \Pr(U_{2i} > U_{1i} > 0) = \Pr(X_{2i}\beta + \varepsilon_{2i} > X_{1i}\beta + \varepsilon_{1i} > 0) \\ = \Pr(\varepsilon_{2i} > (X_{1i} - X_{2i})\beta + \varepsilon_{1i} > 0)$$



$$F(\varepsilon_{2i}^*) - F(0)$$

$$\Pr(D_i = 1, D_2 = 1) = \Pr(U_{1i} > 0, U_{1i} < U_{2i}) = \Pr(\varepsilon_{1i} > -X_{1i}\beta, \varepsilon_{2i} > X_{1i}\beta - X_{2i}\beta + \varepsilon_{1i}) \\ = \Pr(\varepsilon_{1i} > -X_{1i}\beta, \varepsilon_{2i} > X_{1i}\beta - X_{2i}\beta + \varepsilon_{1i}) \\ = \Pr(\varepsilon_{1i} > -X_{1i}\beta) * \Pr(\varepsilon_{2i} > X_{1i}\beta + \varepsilon_{1i})$$

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \text{MVN} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma \right)$$

CLASE 12 - ECONOMETRÍA AVANZADA

Modelo Logit

Logit

$$\Pr(D_i = 1 | X, \beta) = F(X, \beta) \in [0, 1]$$

$$\text{Modelo de utilidad aleatoria} = \begin{cases} U_{1i} = X_{1i}\beta + \varepsilon_{1i} \rightarrow D_i = 1 \\ U_{0i} = X_{0i}\beta + \varepsilon_{0i} \rightarrow D_i = 0 \end{cases}$$

Decisiones

$$D_i = \begin{cases} 1 & \text{si } U_{1i} > U_{0i} \\ 0 & \text{e.o.c.} \end{cases} \Rightarrow \Pr(D_i = 1 | X, \beta) = \Pr(U_{1i} > U_{0i}) \\ = \Pr(\varepsilon_{0i} < X_{1i}\beta - X_{0i}\beta + \varepsilon_{1i})$$

Se distribuye ① Logit
 $\varepsilon \sim \text{Valor ext. tipo 1}$
 ② Probit, $\varepsilon \sim N$

Funciones de V.E. tipo I

$$\text{Valor extremo tipo 1: } F(a) = \exp\{e^{-a}\} \\ f(a) = \exp\{-a - e^{-a}\}$$

Construcción función de densidad modelo logit.

$$\text{Considere } \Pr(D_i = 1 | X, \beta) = \Pr(\varepsilon_{0i} < X_{1i}\beta - X_{0i}\beta + \varepsilon_{1i}), \varepsilon_0 \sim \text{V.E. Tipo 1} \\ = E_{\varepsilon_0} [F(X_{1i}\beta - X_{0i}\beta + \varepsilon_{1i})]$$

$$E[X] = \sum_{i=1}^N x_i \underbrace{\Pr(x_i = X)}_{\text{pdf.}} = \int_{-\infty}^{\infty} F(X_{1i}\beta - X_{0i}\beta + \varepsilon_{1i}) \underbrace{f(\varepsilon_{1i})}_{\text{pdf.}} d\varepsilon_1$$

$$\Pr(D_i = 1 | X, \beta) = \int_{-\infty}^{\infty} \exp\{-e^{-(X_{1i}\beta - X_{0i}\beta + \varepsilon_{1i})}\} * \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} d\varepsilon_1 \\ = \int_{-\infty}^{\infty} \exp\{-e^{-(X_{1i}\beta - X_{0i}\beta + \varepsilon_{1i})} - \varepsilon_1 - e^{-\varepsilon_1}\} d\varepsilon_1 \\ = \int_{-\infty}^{\infty} \exp\{-\varepsilon_1 - e^{-\varepsilon_1}(1 + e^{X_{0i}\beta - X_{1i}\beta})\} d\varepsilon_1$$

$$\begin{aligned}
 \Pr(D_i=1|X, \beta) &= \int_{-\infty}^{\infty} \exp \left\{ -\varepsilon_i - \underbrace{e^{-\varepsilon_i} \left(1 + \frac{e^{x_0\beta}}{e^{x_1\beta}} \right)}_{e^\lambda} \right\} d\varepsilon_i \\
 &= \int_{-\infty}^{\infty} \exp \left\{ -\varepsilon_i - e^{\lambda - \varepsilon_i} \right\} d\varepsilon_i = \int_{-\infty}^{\infty} \exp \left\{ -\varepsilon_i - e^{-(\varepsilon_i - \lambda)} \right\} d\varepsilon_i \\
 &= \int_{-\infty}^{\infty} \exp \left\{ -(\varepsilon_i - \lambda) - e^{-(\varepsilon_i - \lambda)} + (-\lambda) \right\} d\varepsilon_i \\
 &= e^{-\lambda} \int_{-\infty}^{\infty} \exp \left\{ -(\varepsilon_i - \lambda) - e^{-(\varepsilon_i - \lambda)} \right\} d\varepsilon_i \\
 &= e^{-\lambda} \underbrace{\int_{-\infty}^{\infty} f(\varepsilon_i - \lambda) d\varepsilon_i}_{=1} = e^{-\lambda}
 \end{aligned}$$

$$\Pr(D_i=1|X, \beta) = \frac{1}{e^\lambda} = \frac{e^{x_1\beta}}{e^{x_0\beta} + e^{x_1\beta}}$$

Función para la binomial

$$\Pr(D_i=0|X, \beta) = \frac{e^{x_0\beta} + e^{x_1\beta} - e^{x_1\beta}}{e^{x_0\beta} + e^{x_1\beta}} = \frac{e^{x_0\beta}}{e^{x_0\beta} + e^{x_1\beta}}$$

En general:

Función para n alternativas

$$\Pr(D_{ij}=1|X, \beta) = \frac{e^{x_{ij}\beta}}{\sum_{k=1}^J e^{x_{ik}\beta}}$$

Reescalar el MUA

Identificación ① Escala

$$U_{1i} = x_{1i}\beta + \varepsilon_{1i}, \quad U_{0i} = x_{0i}\beta + \varepsilon_{0i} \Rightarrow p_i = \begin{cases} U_{1i} > U_{0i} \Rightarrow D_i=1 \\ D_i=0 \text{ e.o.c} \end{cases} = U_{1i} - U_{0i} > 0.$$

Considere $U_{11} > U_{10} \Rightarrow D_i=1$
 $\sigma U_{11} + k > \sigma U_{10} + k \Rightarrow D_i=1$ La decisión se mantiene.

No importan las magnitudes absolutas, sino la distancia relativa entre las utilidades.

↳ El modelo debe normalizarse contra una alternativa de base.

$$U_{1i}^* = U_{1i} - U_{0i} = (x_{1i} - x_{0i})\beta + (\varepsilon_{1i} - \varepsilon_{0i}) = x_{1i}^*\beta + \varepsilon_{1i}^*$$

Tomamos $U_{0i}^* = 0$

$$\Rightarrow \Pr(D_i=1|X, \beta) = \frac{e^{x_{1i}^*\beta}}{1 + e^{x_{1i}^*\beta}}$$

Estimabilidad de los coeficientes

② Coeficientes.

$$U_{1i} = \beta_{01} + \beta_{11} F_{1i} + \beta_{21} \text{edad} + \varepsilon_{1i}$$

$$U_{0i} = \beta_{00} + \beta_{10} F_{0i} + \beta_{20} \text{edad} + \varepsilon_{0i}$$

Varían a través de alternativas

Varían a través de individuos

$$U_{1i}^* = (\beta_{01} - \beta_{00}) + \beta_{11} F_{1i} - \beta_{10} F_{0i} + (\beta_{21} - \beta_{20}) \text{edad} + \varepsilon_{1i}^*$$

$$(2) \quad P(D_i=1|x,\beta) = \frac{\exp([\beta_{00} + \beta_{01}] + \beta_{11}i + (-\beta_{01}, F_{01}) + (\beta_{01} - \beta_{02}) \text{edad})}{1 + \text{lo mismo}}$$

$$\ln P(D_i=1|x,\beta) = \ln \left(\frac{\exp(w)}{1 + \exp(w)} \right)$$

Observe que $\beta_{01} - \beta_{00} = (\beta_{01} - \Delta) - (\beta_{00} - \Delta) \Rightarrow$ Si encontramos β_{01} y β_{00} que maximicen $\mathcal{L}(\theta|x)$, $\beta_{01} - \Delta$ y $\beta_{00} - \Delta$ también serán solución.

¿ β_{00} y β_{01} están identificados? No. Sólo $\beta_{01} - \beta_{00}$ está identificado.

Función de
verosimilitud
logit.

$$P(D_i=1|x,\beta) = F(x\beta) = \frac{e^{x\beta}}{1 + e^{x\beta}} \quad (\text{Binomial})$$

$$\ln \mathcal{L}(\beta|x) = \sum_i \{ D_i \ln F(x\beta) + (1 - D_i) \ln (1 - F(x\beta)) \}$$

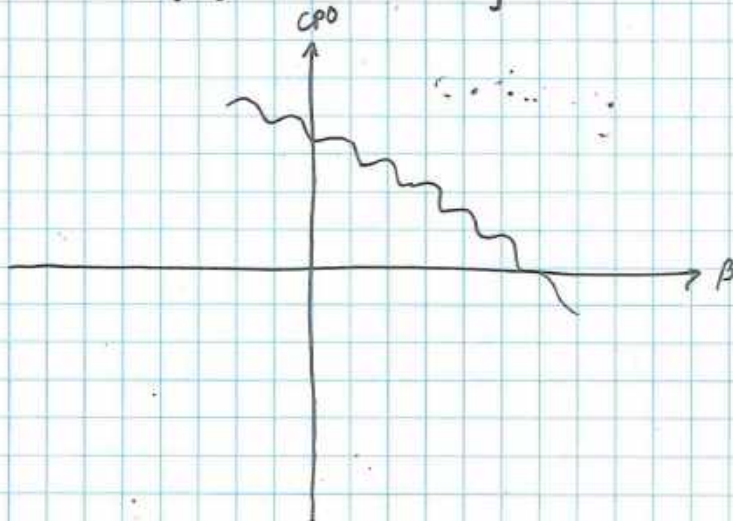
$$\frac{dF(x\beta)}{d\beta} = \frac{dF(x\beta)}{d(x\beta)} * \frac{d(x\beta)}{d\beta} = f(x\beta) * x_i$$

$$\frac{dF(x\beta)}{d(x\beta)} = \frac{e^{x\beta}(1 + e^{x\beta}) - e^{x\beta}e^{x\beta}}{(1 + e^{x\beta})^2} = \frac{e^{x\beta}}{(1 + e^{x\beta})^2} = F(x\beta)(1 - F(x\beta))$$

Sólo en el logit
↓
 $= f(x\beta)$

$$\begin{aligned} \text{CPO: } \frac{d \ln \mathcal{L}}{d\beta} &= \sum_i \left\{ D_i \frac{1}{F(x\beta)} + F(x\beta)[1 - F(x\beta)]x_i + \frac{(1 - D_i)}{1 - F(x\beta)} * (-F(x\beta)[1 - F(x\beta)])x_i \right\} \\ &= \sum_i \{ D_i(1 - F(x\beta))x_i + (1 - D_i)(F(x\beta)[1 - F(x\beta)])x_i \} \\ &= \sum_i \{ (D_i - D_i F(x\beta))x_i - x_i(F(x\beta) - D_i F(x\beta)) \} \\ &= \sum_i \{ x_i(D_i - D_i F(x\beta) + F(x\beta)D_i - F(x\beta)) \} \\ &= \sum_i \{ x_i(D_i - F(x\beta)) \} \end{aligned}$$

No linealidad
de la CPO.



Algoritmo maximización numérica

1. Conjetura $\beta = \beta_0$

2. $\ln \mathcal{L}(\beta_0 | X)$

3. $\frac{d \ln \mathcal{L}}{d \beta} \Big|_{\beta_0}$

4. $H = \frac{d^2 \ln \mathcal{L}}{d \beta_0 d \beta_0'}$

5. $\beta_1 = \beta_0 + \Delta$
 $\Delta = f(H)$

6. $\ln \mathcal{L}(\beta_k | X) - \ln \mathcal{L}(\beta_{k-1} | X) \rightarrow 0$

Maximización numérica [Búsqueda por grilla / Cuasi-Newton]

1. Se establece una conjetura de $\beta = \beta_0$ (elegida por econometrista)

2. Computar $\ln \mathcal{L}(\beta_0 | X)$

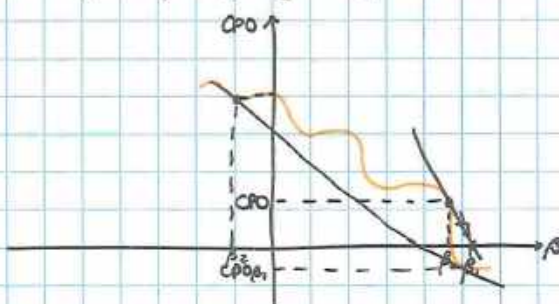
3. Hallar $\frac{d \ln \mathcal{L}}{d \beta} \Big|_{\beta_0}$

4. Calcular $H = \frac{d^2 \ln \mathcal{L}}{d \beta_0 d \beta_0'}$

5. $\beta_1 = \beta_0 + H(\Delta)$. Repetir (1) - (4) con β_1 .

6. La iteración se detiene cuando $\ln \mathcal{L}(\beta_k | X) - \ln \mathcal{L}(\beta_{k-1} | X)$ es muy pequeño:

$$(\hat{\beta}_k)_{MV}, \quad \begin{matrix} u_1 = x_1 \beta + \epsilon_1 \\ u_0 = 0 \end{matrix} \quad \left. \vphantom{\begin{matrix} u_1 \\ u_0 \end{matrix}} \right\} F(x_i \beta) = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}}$$



Interpretación de $(\hat{\beta}_k)_{MV}$

Si interpretamos $\hat{\beta}_{MV}$ directamente, muestra el cambio de u_1 con un cambio en x_1 , pero esto no nos interesa.

Efecto marginal

$$\text{Interés: } \frac{d \Pr(D_i = 1 | X, \beta)}{d x} = \frac{d}{d x} \left(\frac{e^{x \beta}}{1 + e^{x \beta}} \right)$$

$$\text{Efecto marginal} = \frac{e^{x \beta} (1 + e^{x \beta}) - (e^{x \beta})^2}{(1 + e^{x \beta})^2} \cdot \beta = \frac{\hat{\beta}_{MV} e^{x_i \hat{\beta}_{MV}}}{(1 + e^{x_i \hat{\beta}_{MV}})^2}$$

Ventajas del logit

Ventajas del modelo logit

1. La expresión de la probabilidad de éxito tiene solución de forma cerrada

2. $\ln \mathcal{L}(\cdot)$ del logit es globalmente cóncava $\Rightarrow \hat{\beta}_{MV}$ es único.

$$\sum_i [D_i - F(x_i \beta)] x_i = CPO = \sum_i D_i x_i - \sum_i x_i F(x_i \beta)$$

$$\frac{d CPO}{d \beta} = \sum_i x_i \left[D_i - \frac{d F(x_i \beta)}{d \beta} \right] = \sum_i x_i [D_i - f(x_i \beta) \cdot x_i]$$

$$= \sum_i x_i D_i - \sum_i x_i^2 f(x_i \beta) < 0$$

$$f(x_i \beta) = -x_i \beta - e^{-x_i \beta} \quad \begin{matrix} \sum_i x_i D_i < \sum_i x_i^2 f(x_i \beta) \\ \sum_i D_i < \sum_i x_i f(x_i \beta) \end{matrix}$$

$$\frac{d CPO}{d \beta} = - \sum_i x_i f(x_i \beta) x_i = - \sum_i x_i^2 f(x_i \beta) = - \sum_i x_i^2 F(x_i \beta) (1 - F(x_i \beta))$$

$$\frac{d CPO}{d \beta} < 0$$

Independencia de alternativas irrelevantes

Desventajas del modelo logit

Considere un modelo logit de J alternativas.

$$\left. \begin{aligned} \Pr(D_{ij}=1 | X, \theta) &= \frac{e^{x_{ij}\beta}}{1 + e^{x_{i1}\beta} + \dots + e^{x_{i(J-1)}\beta}} \\ \Pr(D_{ik}=1 | X, \theta) &= \frac{e^{x_{ik}\beta}}{1 + e^{x_{i1}\beta} + \dots + e^{x_{i(J-1)}\beta}} \end{aligned} \right\} \frac{\Pr(D_{ij}=1)}{\Pr(D_{ik}=1)} = \frac{e^{x_{ij}\beta}}{e^{x_{ik}\beta}} = f(x_{ij}, x_{ik})$$

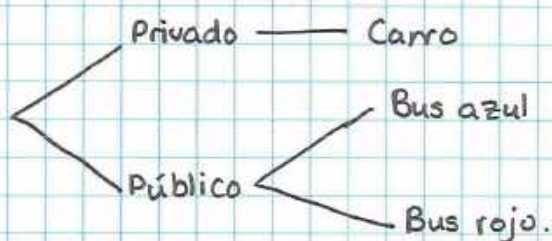
La probabilidad de elegir j con respecto a la probabilidad de elegir k sólo depende de estas dos alternativas.

Logit anidado

Possible solución: Logit anidado

La división de las categorías la hace el econometrista.

Carro Privado Bus azul público Bus rojo



$$U_c = \frac{e^{x_c\beta}}{e^{x_c\beta} + I(x_A\beta + x_R\beta)}$$

función de inclusión

$$U_c \text{ vs. } \max\{U_A, U_B\} \quad \max\{U_A, U_B\} \Rightarrow \Pr(A) = \frac{e^{x_A\beta}}{e^{x_A\beta} + e^{x_B\beta}}$$

PROBIT

Binomial

MODELO PROBIT

Probit binomial

considere: $U_1 = x_1\beta + \varepsilon_1, \varepsilon_1 \sim N(0, \sigma^2)$
 $U_0 = 0$

Regla de decisión: $D_i = \begin{cases} 1 & \text{si } U_1 > U_0 \\ 0 & \text{e.o.c.} \end{cases}$

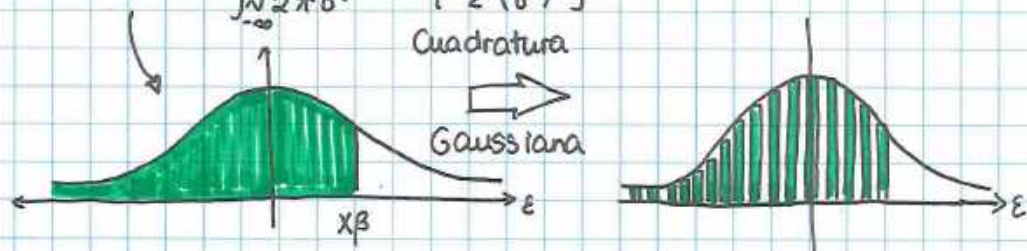
$$\ln \mathcal{L}(\theta | X) = \sum_{i=1}^N \left\{ D_i \ln \Pr(D_i=1 | X, \theta) + (1 - D_i) \ln [1 - \Pr(D_i=1 | X, \theta)] \right\}$$

simetría

Donde $\theta = (\beta, \sigma^2)'$. $\Pr(D_i=1 | X, \theta) = \Pr(\varepsilon_i > -x_i\beta) = \Pr(\varepsilon_i < x_i\beta)$

$$\Pr(D_i=1 | X, \theta) = \int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2\right\} d\varepsilon$$

Cuadratura Gaussiana



Normalizamos:

$$\begin{aligned} \Pr(D_i = 1 | X, \theta) &= \Pr\left(\frac{\varepsilon_i}{\sigma} < \frac{X_i \beta}{\sigma}\right), \quad \varepsilon_i \sim N(0, 1) \\ &= \int_{-\infty}^{\frac{X_i \beta}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\varepsilon^2} d\varepsilon, \quad \text{pero note que } \frac{X_i \beta}{\sigma} = \frac{X_i (\beta \sigma)}{\sigma^2} \end{aligned}$$

Normaliza/
uno de los
parámetros
($\sigma = 1$)

Por lo tanto, β y σ no están identificados de manera única. En consecuencia, es necesario normalizar uno de los parámetros. Convencionalmente, se asume $\sigma = 1$. De esta forma, se estima $\hat{\beta}_{NLV}$ por métodos numéricos.

Multinomial

Probit multinomial

Considere

$$\begin{aligned} u_1 &= X_1 \beta + \varepsilon_1 \\ u_2 &= X_2 \beta + \varepsilon_2 \\ u_0 &= 0 \end{aligned} \quad \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim MVN \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right\}$$

Note que si $\sigma_{12} > 0 \Rightarrow \varepsilon_1$ y ε_2 se mueven en la misma dirección.
 $\Rightarrow$ "Son del mismo grupo"

Si $\sigma_{12} < 0 \Rightarrow \varepsilon_1$ y ε_2 se mueven en direcciones opuestas.
 $\Rightarrow$ "Son de grupos opuestos".

Regla de decisión: Max u . $\theta = (\beta, \sigma_1^2, \sigma_2^2, \sigma_{12})$

$$\Pr(D_1 = 1 | X, \theta) = \Pr(u_1 > 0, u_1 > u_2)$$

$$\Pr(D_1 = 1 | X, \theta) = \Pr(X_1 \beta + \varepsilon_1 > 0, X_1 \beta + \varepsilon_1 > X_2 \beta + \varepsilon_2)$$

Normalizamos, dado que ε_1 y ε_2 no son independientes:

$$\begin{aligned} \Pr(D_1 = 1 | X, \theta) &= \Pr\left(\frac{X_1 \beta + \varepsilon_1}{\sigma_1} > 0, \frac{X_1 \beta + \varepsilon_1}{\sigma_1} > \frac{X_2 \beta + \varepsilon_2}{\sigma_2}\right) \\ &= \Pr\left(\frac{\varepsilon_1}{\sigma_1} > -\frac{X_1 \beta}{\sigma_1}, \frac{\varepsilon_2}{\sigma_2} < \frac{X_1 \beta - X_2 \beta + \varepsilon_1}{\sigma_2}\right) \\ &= \int_{-\frac{X_1 \beta}{\sigma_1}}^{\infty} \int_{-\infty}^{\frac{X_1 \beta - X_2 \beta + \varepsilon_1}{\sigma_2}} f(\varepsilon_1, \varepsilon_2 | \Sigma) d\varepsilon_2 d\varepsilon_1 \end{aligned}$$

pdf normal multivariada.

Tenemos β/σ_1 , β/σ_2 . Si asumimos $\sigma_1 = 1 \Rightarrow$ Podemos encontrar β .
Si tenemos $\beta \Rightarrow$ Podemos encontrar σ_2 .

LOGIT MULTINOMIAL

Considerare $y_i = \begin{cases} u_0 = x_0\beta + \varepsilon_0, & 0 \text{ si } u_0 > u_1, u_0 > u_2 \\ u_1 = x_1\beta + \varepsilon_1, & 1 \text{ si } u_1 > u_2, u_1 > u_0 \\ u_2 = x_2\beta + \varepsilon_2, & 2 \text{ si } u_2 > u_1, u_2 > u_0 \end{cases}$ $\lim_{x \rightarrow -\infty} -\varepsilon_1 = \lim_{x \rightarrow -\infty} e^{-\varepsilon_2}$

$$f(\varepsilon_0, \varepsilon_1, \varepsilon_2) = \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} * \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} * \exp\{-\varepsilon_2 - e^{-\varepsilon_2}\}$$

$$\begin{aligned} \Pr(y_i = 1 | x, \beta) &= \Pr(u_1 > u_0, u_2 < u_1) = \Pr(\varepsilon_1 > x_0\beta - x_1\beta + \varepsilon_0, \varepsilon_2 < x_2\beta - x_1\beta + \varepsilon_1) \\ &= \Pr(\varepsilon_1 > x_0\beta - x_1\beta + \varepsilon_0) * \Pr(\varepsilon_2 < x_2\beta - x_1\beta + \varepsilon_1) \\ &= \Pr(\varepsilon_1 > x_0\beta - x_1\beta + \varepsilon_0) * \Pr(\varepsilon_1 > x_2\beta - x_1\beta + \varepsilon_2) \end{aligned}$$

$$\Pr(y_i = 1 | x, \beta) = \int_{-\infty}^{\infty} \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} \int_{-\infty}^{x_1\beta - x_2\beta + \varepsilon_1} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} \exp\{-\varepsilon_2 - e^{-\varepsilon_2}\} d\varepsilon_2 d\varepsilon_1 d\varepsilon_0$$

$$\Pr(y_i = 1) = \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} d\varepsilon_0 * \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} \int_{-\infty}^{x_1\beta - x_2\beta + \varepsilon_1} \exp\{-\varepsilon_2 - e^{-\varepsilon_2}\} d\varepsilon_2 d\varepsilon_1$$

$$= \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} F(x_1\beta - x_2\beta + \varepsilon_1) d\varepsilon_1 d\varepsilon_0$$

$$= \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} \exp\{-e^{x_2\beta - x_1\beta - \varepsilon_1}\} d\varepsilon_1 d\varepsilon_0$$

$$= \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} \exp\{-\varepsilon_1 - e^{-\varepsilon_1} (1 + e^{x_2\beta - x_1\beta})\} d\varepsilon_1 d\varepsilon_0$$

$$= \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} \exp\{-\varepsilon_1 - e^{-(\varepsilon_1 - \lambda_2)} - \lambda_2 + \lambda_2\} d\varepsilon_1 d\varepsilon_0$$

$$= \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} \exp\{-(\varepsilon_1 - \lambda_2) - e^{-(\varepsilon_1 - \lambda_2)} - \lambda_2\} d\varepsilon_1 d\varepsilon_0$$

$$= e^{-\lambda_2} \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} \int_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} f(\varepsilon_1 - \lambda_2) d\varepsilon_1 d\varepsilon_0$$

$$= e^{-\lambda_2} \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} F(\varepsilon_1 - \lambda_2) \Big|_{x_0\beta - x_1\beta + \varepsilon_0}^{\infty} d\varepsilon_0$$

$$= e^{-\lambda_2} \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0}\} F(x_0\beta - x_1\beta + \varepsilon_0) d\varepsilon_0$$

$$= e^{-\lambda_2} \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-\varepsilon_0} (1 + e^{x_1\beta - x_0\beta})\} d\varepsilon_0$$

$$= e^{-\lambda_2} \int_{-\infty}^{\infty} \exp\{-\varepsilon_0 - e^{-(\varepsilon_0 - \lambda_0)} \frac{e^{\lambda_0}}{1 + \lambda_0 - \lambda_0}\} d\varepsilon_0$$

$$= e^{-\lambda_2} \int_{-\infty}^{\infty} \exp\{-(\varepsilon_0 - \lambda_0) - e^{-(\varepsilon_0 - \lambda_0)} - \lambda_0\} d\varepsilon_0$$

$$= e^{-\lambda_2} \int_{-\infty}^{\infty} f(\varepsilon_0 - \lambda_0) d\varepsilon_0 = e^{-\lambda_2} e^{-\lambda_0} = e^{-(\lambda_2 + \lambda_0)} = \frac{1}{(1 + e^{x_2\beta - x_1\beta})(1 + e^{x_1\beta - x_0\beta})}$$

$$= \frac{e^{x_1\beta} * e^{x_0\beta}}{(e^{x_0\beta} + e^{x_1\beta}) * (e^{x_1\beta} + e^{x_2\beta})} = \frac{e^{x_1\beta + x_0\beta}}{e^{x_0\beta + x_1\beta} + e^{x_1\beta + x_2\beta} + e^{x_0\beta + x_2\beta} + e^{x_1\beta + x_0\beta}} =$$

REV3AR

$$\begin{aligned} e^{\lambda_1} &= 1 + e^{x_2\beta - x_1\beta} \\ e^{\lambda_0} &= 1 + e^{x_1\beta - x_0\beta} \end{aligned}$$

$$\begin{aligned} e^{\lambda_2} &= \frac{e^{x_1\beta} + e^{x_2\beta}}{e^{x_1\beta}} \\ e^{\lambda_0} &= \frac{e^{x_0\beta} + e^{x_1\beta}}{e^{x_0\beta}} \end{aligned}$$

COMPLEMENTARIA

12-09-2014

Considere

$$y_i = \begin{cases} 1 & \text{si } u_1 = x_{i1}\beta + z_i\delta_1 + \varepsilon_1 \\ 2 & \text{si } u_2 = x_{i2}\beta + z_i\delta_2 + \varepsilon_2 \\ 3 & \text{si } u_3 = x_{i3}\beta + z_i\delta_3 + \varepsilon_3 \end{cases}$$

Base de datos

id	x ₁	x ₂	x ₃	z
1	x	x	x	M
1	x	x	x	F
2	x	x	x	M
2	x	x	x	F
3	x	x	x	M
3	x	x	x	F

Normalizamos:

$$u_1 - u_3 = (x_{i1} - x_{i3})\beta + z_i(\delta_1 - \delta_3) + \varepsilon_1 - \varepsilon_3$$

$$u_2 - u_3 = (x_{i2} - x_{i3})\beta + z_i(\delta_2 - \delta_3) + \varepsilon_2 - \varepsilon_3$$

$$u_3 = 0$$

Varía entre alternativas
⇒ β se puede estimar

$$\delta_1 \text{ y } \delta_2$$

no se pueden

estimar individualmente → sólo varía entre individuos.

PREGUNTAS

- ¿Qué es mlogit?
- ¿Qué es mfx?
- ¿Qué hace reshape?

LOGIT =

Multinomial ⇒ $P(y_i = 1) = \frac{e^{x_i\beta_1}}{1 + \sum_{k=1}^{J-1} e^{x_i\beta_k}}$

Condicional = $P(y_i = 1) = \frac{e^{x_i\beta}}{1 + \sum_{k=1}^{J-1} e^{x_i\beta_k}}$

! Las probabilidades relativas son importantes porque nos permite ver qué tanto más probable es un evento que otro.

CLASE 13

15-09-2014

MODELO PROBIT

$$\begin{aligned} u_1 &= x_1\beta + \varepsilon_1 \\ u_2 &= x_2\beta + \varepsilon_2 \\ u_3 &= 0 \end{aligned} ; (\varepsilon_1, \varepsilon_2) \sim MVN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right)$$

Si regla de decisión: max (1)

$$Pr(D_{1i} = 1 | X, \theta) = Pr(u_1 > 0, u_1 > u_2) = Pr(\varepsilon_1 > -x_1\beta, \varepsilon_2 < x_1\beta - x_2\beta + \varepsilon_1)$$

Hay correlación = σ_{12}

$$Pr(D_{1i} = 1 | X, \theta) = \int_{-x_1\beta}^{\infty} \int_{-\infty}^{x_1\beta - x_2\beta + \varepsilon_1} f(\varepsilon_1, \varepsilon_2 | \Sigma) d\varepsilon_2 d\varepsilon_1$$

$f(\varepsilon_1, \varepsilon_2 | \Sigma)$ = Función de densidad normal bivariada con matriz var-cov Σ

Simulaciones - Para solución numérica

Todos los tipos de simulaciones buscan maximizar la función de verosimilitud. Para ello:

1. Se establece una conjetura $\theta_0 = (\beta_0, \sigma_{12}^0, \sigma_{22}^0)$.

2. Se computa $\ln \mathcal{L}(\theta_0 | \cdot)$
3. Se halla la condición de primer orden en θ_0 .
4. Se establece el hessiano en θ_0
 $\hookrightarrow$ Obtiene $\theta_1 = \theta_0 + \Delta(\text{Hess})$

$$\ln \mathcal{L}(\theta | X) = \sum_{i=1}^N \sum_{j=1}^J D_{ij} \ln \Pr(D_{ij} = 1 | X, \theta)$$

Simulador de frecuencias (Montecarlo)

$$\begin{aligned} U_{1i} &= X_{1i} \beta + \varepsilon_{1i} & X_{1i} \text{ y } X_{2i} \text{ son conocidos } \forall \text{ individuo, y ya} \\ U_{2i} &= X_{2i} \beta + \varepsilon_{2i} & \text{tenemos conjeturas sobre los } \beta. \\ U_{3i} &= 0 & \hookrightarrow \text{Para conocer } U_{1i}, \text{ hay que encontrar } \varepsilon_{1i} \text{ y } \varepsilon_{2i}. \end{aligned}$$

$\forall i$: 1. Sacar $\{\varepsilon_1^d, \varepsilon_2^d\}_{d=1, \dots, D}$ donde $D = \#$ simulaciones por individuo. de su correspondiente distribución de probabilidad:

$$\text{Normal}(0, \Sigma).$$

$$\begin{aligned} 2. \text{ Para } \forall d, \text{ reemplazamos: } & U_1^d = X_1 \beta + \varepsilon_1^d \\ & U_2^d = X_2 \beta + \varepsilon_2^d \quad d=1, 2, \dots, D. \\ & U_3^d = 0 \end{aligned}$$

$$\text{Definimos } I[\cdot] = \begin{cases} 1 & \text{si la condición } \cdot \text{ se cumple} \\ 0 & \text{e.o.c} \end{cases}$$

$$\hat{\Pr}(D_{ij}=1 | X, \theta)_{FS} = \frac{1}{D} \sum_{d=1}^D I[U_{1j}^d > U_{ik}^d, \forall k \neq j]$$

3. Se reemplaza en la función de verosimilitud la opción que el individuo efectivamente escogió.

$$\begin{aligned} \text{p.e: } i=1 & \rightarrow \text{Alt 2} \\ i=2 & \rightarrow \text{Alt 3} \Rightarrow \ln \mathcal{L} = \ln \Pr(D_{12}=1) + \ln \Pr(D_{23}=1) + \ln \Pr(D_{31}=1) \\ i=3 & \rightarrow \text{Alt 1} \end{aligned}$$

Desventajas

$$\hat{\Pr}_{FS} \in \{0, \frac{1}{D}, \frac{2}{D}, \dots, 1\}$$

① $\hat{\Pr}_{FS}$ puede ser cero $\Rightarrow \ln 0$? $\ln \mathcal{L}(\cdot)$ no está definido.

② En la búsqueda del máximo si necesito:
 $\frac{\partial \Pr(D_{ij}=1 | \cdot)}{\partial \theta}$



Perd: $d\Pr(D_{ij}=1 | \cdot) / d\theta = 0$ en casi todas partes.

③ Imprecisión.

Simulador GHK

$$\begin{cases} u_1 = x_1 \beta + \varepsilon_1 \\ u_2 = x_2 \beta + \varepsilon_2 \\ u_3 = 0 \end{cases} \quad \left\{ \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right] \right.$$

$$\Pr(D_{ii} = 1 | \cdot) = \Pr(\varepsilon_1 > -x_1 \beta, \varepsilon_2 < x_1 \beta - x_2 \beta + \varepsilon_1).$$

El problema es que los eventos son dependientes. Este simulador busca escribir los ε como unos η que se distribuyan normalmente y que **sí** sean independientes.

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c_{12} & c_{22} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \quad \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right]$$

$\varepsilon = C * \eta$ donde $\Sigma = CC'$
 ↳ Descomposición de Cholesky → Triangular sup

$$\text{En este caso: } \begin{pmatrix} 1 & 0 \\ c_{12} & c_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \sigma_{12} & \sqrt{\sigma_{22} - \sigma_{12}^2} \end{pmatrix}$$

$$\text{Según el modelo: } \begin{aligned} \varepsilon_1 &= \eta_1 \\ \varepsilon_2 &= \sigma_{12} \eta_1 + \sqrt{\sigma_{22} - \sigma_{12}^2} \eta_2 \end{aligned}$$

Comprobaremos la igualdad:

$$\begin{aligned} 1 &= \text{Var}(\varepsilon_1) = \text{Var}(\eta_1) = 1 \quad \checkmark \\ \sigma_{22} &= \text{Var}(\varepsilon_2) = \text{Var}(\sigma_{12} \eta_1 + \sqrt{\sigma_{22} - \sigma_{12}^2} \eta_2) \\ &= \sigma_{12}^2 \text{Var}(\eta_1) + (\sigma_{22} - \sigma_{12}^2) \text{Var}(\eta_2) + 2 \text{Cov}(\eta_1, \eta_2) \dots \\ &= \sigma_{12}^2 + \sigma_{22} - \sigma_{12}^2 \\ &= \sigma_{22} \quad \checkmark \\ \sigma_{12} &= \text{Cov}(\varepsilon_1, \varepsilon_2) = \text{Cov}(\eta_1, \sigma_{12} \eta_1 + \sqrt{\sigma_{22} - \sigma_{12}^2} \eta_2) \\ &= \sigma_{12} \text{Var}(\eta_1) + \sqrt{\sigma_{22} - \sigma_{12}^2} \text{Cov}(\eta_1, \eta_2) \\ &= \sigma_{12} \quad \checkmark \quad \text{El modelo es acertado.} \end{aligned}$$

Reescribimos el modelo con los η :

$$\begin{aligned} u_1 &= x_1 \beta + \eta_1 \\ u_2 &= x_2 \beta + c_{12} \eta_1 + c_{22} \eta_2 \\ u_3 &= 0 \end{aligned}$$

Debe cumplir $\eta_1 < x_1 \beta$

$$\Rightarrow \Pr(D_{ii} = 1 | \cdot) = \Pr(\eta_1 > -x_1 \beta, \eta_2 < \frac{x_1 \beta + \eta_1 - x_2 \beta - c_{12} \eta_1}{c_{22}})$$

$$\text{Por regla de Bayes: } P(A|B) = \frac{P(A, B)}{P(B)}$$

$$\Pr(D_{ii} = 1 | \cdot) = \Pr(\eta_1 > -x_1 \beta) * \Pr(\eta_2 < \frac{x_1 \beta + \eta_1 - x_2 \beta - c_{12} \eta_1}{c_{22}} | \eta_1 > -x_1 \beta)$$

Linealidad de $u_1, u_2, u_3 \Rightarrow$ Permite escribir $\Pr(\varepsilon_1 > -x_1 \beta, \varepsilon_2 < x_1 \beta - x_2 \beta + \varepsilon_1)$

$$\begin{aligned} \text{4 alt: } \Pr(D_{ii} = 1 | \cdot) &= \Pr(\eta_1 > -x_1 \beta) * \Pr(\eta_2 < B_1(\eta_1) | \eta_1 > -x_1 \beta) \\ &\quad * \Pr(\eta_3 < B_2(\eta_1, \eta_2) | \eta_1 > -x_1 \beta, \eta_2 < B_1(\eta_1)) \end{aligned}$$

CLASE 15

Simulador Kernel - Permite no linealidad en las utilidades.

Considere
$$\begin{aligned} u_1 &= f(x_1, \varepsilon_1 | \theta) \\ u_2 &= f(x_2, \varepsilon_2 | \theta) \\ &\vdots \\ u_J &= f(x_J, \varepsilon_J | \theta) \end{aligned} \quad \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_J \end{pmatrix} \sim \text{MVN}(0, \Sigma)$$

Objetivo: Volver a un logit, que es lo más simple en términos de no linealidad.

⇒ Se estima un modelo aproximado:

$$\begin{aligned} u_1^* &= u_1 + \rho \varepsilon_1 \\ &\vdots \\ u_J^* &= u_J + \rho \varepsilon_J \end{aligned} \quad \left. \vphantom{\begin{aligned} u_1^* &= u_1 + \rho \varepsilon_1 \\ &\vdots \\ u_J^* &= u_J + \rho \varepsilon_J \end{aligned}} \right\} \begin{array}{l} \text{Donde } \varepsilon_i: \text{valor extremo tipo I} \\ \rho: \text{Factor de escala (escogido por el economista).} \end{array}$$

$$\Pr(D_i^* = 1 | \cdot) = \frac{e^{f(x_i, \varepsilon_i | \theta)/\rho}}{\sum_{j=1}^J e^{f(x_j, \varepsilon_j | \theta)/\rho}} \quad \begin{array}{l} \rho \rightarrow 0 \text{ es deseable porque} \\ \Rightarrow u_i^* \rightarrow u_i \end{array}$$

Sin embargo, ε_i y ε_j no son conocidos. Asumimos que se distribuyen normalmente. Encontramos la integral de ese valor esperado por Montecarlo.

→ Sacar $\{\varepsilon_1^d, \varepsilon_2^d, \dots, \varepsilon_J^d\}_{d=1, \dots, D}$ de su distribución ($\text{MVN}(0, \Sigma)$)

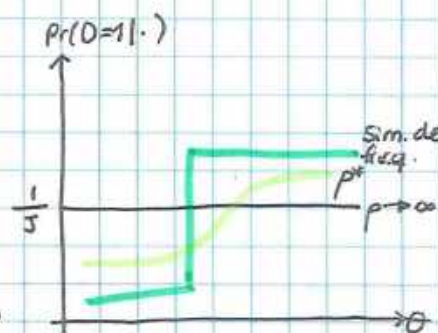
$$\Rightarrow \hat{\Pr}(D_i = 1 | \cdot) = \frac{1}{D} \sum_{d=1}^D \frac{e^{f(x_i, \varepsilon_i^d | \theta)/\rho}}{\sum_{j=1}^J e^{f(x_j, \varepsilon_j^d | \theta)/\rho}}$$

¿Cómo elegir ρ ?

(1) Si $\rho \rightarrow 0 \Rightarrow u_i^* \rightarrow u_i$.

(2) Si $\rho \rightarrow 0 \Rightarrow \Pr(D=1 | \cdot) \rightarrow \frac{1}{J}$

queremos un ρ que sea tan pequeño como sea posible, pero que no sea idéntico al simulador de frecuencias porque tiene muchas porciones planas.



Se empieza con ρ muy pequeña, y se va aumentando para que logre correr.

Desventajas

1. Es más inexacto que el GHK
2. Es difícil elegir un ρ .

Probit bivariado

Trabajar sí/No

Casarse sí/No

$$\text{Probit binomial} \begin{cases} U_T = X_T \beta_T + \varepsilon_T \\ U_C = 0 \end{cases} \quad \begin{matrix} \xleftarrow{\varepsilon_T} \\ \xrightarrow{\varepsilon_C} \end{matrix} \quad \begin{cases} U_C = X_C \beta_C + \varepsilon_C \\ U_{NC} = 0 \end{cases}$$

donde $\begin{pmatrix} \varepsilon_T \\ \varepsilon_C \end{pmatrix} \sim \text{MVN} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$

$$\begin{aligned} \Pr(D_T = 1, D_C = 1 | \theta, X) &= \Pr(X_T \beta_T + \varepsilon_T > 0, X_C \beta_C + \varepsilon_C > 0) \\ &= \Pr(\varepsilon_T > -X_T \beta_T, \varepsilon_C > -X_C \beta_C) \\ &= \int_{-X_T \beta_T}^{\infty} \int_{-X_C \beta_C}^{\infty} \varphi_2(\varepsilon_T, \varepsilon_C | \rho) d\varepsilon_C d\varepsilon_T \end{aligned}$$

Normal bivariada

Nota

Tanto el simulador de frecuencias como el Kernel requiere sacar $\{\varepsilon_1^d, \dots, \varepsilon_T^d\}_{d=1, \dots, D}$ de $\text{MVN}(0, \Sigma)$.

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c_{12} & c_{22} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \rightarrow \text{Se extrae de randn(\cdot) en Excel.}$$

$\downarrow$ $\downarrow$ $\downarrow$
 $\text{MVN}(0, \Sigma)$ $C\Sigma = CC'$ $\text{MVN}(0, I)$

EJERCICIO

① Decisiones:

	Trabajar	Trabajar 1/2	Sub 1	Sub 2	Sub 3
1.	✓	×	×	×	×
2.	✓	×	✓	×	×
3.	✓	×	✓	✓	×
4.	✓	×	✓	✓	✓
5.	✓	×	×	✓	×
6.	✓	×	×	×	✓
7.	✓	×	×	✓	✓
8.	✓	×	✓	×	✓

+ 8 si no trabaja. = 16 alternativas.
+ 8 si trabaja 1/2 tiempo $\Rightarrow$ 24 alternativas

② Regla de decisión:

Trabajo si $U_{TC} > U_{MT} > U_{NT}$
(TC)

Trabajo: decisión k, Subsidio: decisión j.

$$U_k = (\alpha X_k + \varepsilon_k) H_k + w H_k + B_j P_j - \beta_H H_k^2 - \beta_Y (w H_k + B_j P_j)^2 - \lambda (\bar{\varphi}_T X_k + \varepsilon_T) P_j - (1 - \lambda) \max \{ (\bar{\varphi}_T X_k + \varepsilon_T) P_j \}$$

Elige la op. T y la op. S si $U_{T,S} > U_{J,R} \quad \forall J \neq T, S \neq R$

Resumen MV

① 24 alternativas

② $D_{ij} = 1$ si $U_{ij} > U_{ik} \quad \forall j \neq k$.

$$\textcircled{3} \ln \mathcal{L} = \sum_{i=1}^N \sum_{j=1}^{24} D_{ij} \ln \Pr(D_{ij} = 1 | x, \theta)$$

⑤ No se puede despejar para usar GHK
 $\Rightarrow$ Debe ser Kernel.

RESUMEN - MÁXIMA VEROSIMILITUD

Modelos de elección discreta: La variable dependiente no es continua.

$\hookrightarrow$ Se asume una función de distribución para las variables $f(X|\theta)$.

$\hookrightarrow$ Para una muestra $X_1, X_2, \dots, X_N$, la función de densidad conjunta es la probabilidad de observar esa muestra en particular.

Dado que $X_1, X_2, \dots, X_N$ es i.i.d: $f(X_1, X_2, \dots, X_N | \theta) = \prod_{i=1}^N f(X_i | \theta) = \mathcal{L}(\theta | X)$

$$\hat{\theta}_{MV} = \operatorname{argmax} \{ \mathcal{L}(\theta | X) \} = \operatorname{argmax} \{ \ln \mathcal{L}(\theta | X) \}$$

Propiedades del estimador de MV

1. Invarianza: Si $h(\theta)$ es una función uno a uno $\Rightarrow h(\hat{\theta}_{MV}) = \hat{h(\theta)}$

2. Teorema del límite inferior de Cramer-Rao: $\operatorname{Var}(\hat{\theta}_{MV}) = -E \left[\frac{d^2 \ln \mathcal{L}(\theta | X)}{d\theta d\theta'} \right]^{-1}$

Matriz de Información

MODELO DE PROBABILIDAD LINEAL

Se estima $D_i = X_i \beta + u_i$

$\hookrightarrow$ Consistente e insesgado si $E[u_i | X_i] = 0$.

$$\begin{aligned} E[D_i | X_i] &= E[X_i \beta + u_i | X_i] \\ E[D_i | X_i] &= X_i \beta + E[u_i | X_i] = X_i \beta \end{aligned}$$

Características de $\hat{\beta}_{MV}$

- ① Invarianza
- ② Consistencia
- ③ Asintóticamente normal
- ④ Alcanza la cota de Cramer-Rao.

Desventajas

① $\hat{\Pr}(D_i = 1 | X_i \beta) = X_i \hat{\beta}$ no necesariamente pertenece al intervalo $(0, 1)$

② El error se distribuye Bernoulli $\Rightarrow \operatorname{Var}(u_i) = X_i \beta (1 - X_i \beta) \Rightarrow$ Heteroscedástico

Exp: $D_i = 1$ ocurre con probabilidad $X_i \beta$, y $D_i = 0$ ocurre con probabilidad $1 - X_i \beta$.

$$\Rightarrow D_i = 1 \Rightarrow \hat{u}_i = 1 - X_i \beta$$

$$D_i = 0 \Rightarrow \hat{u}_i = -X_i \beta$$

$\hookrightarrow$ Se puede solucionar por mínimos cuadrados ponderados.

MODELOS DE ELECCIÓN DISCRETA NO LINEALES

Construcción de funciones de verosimilitud

En una muestra de tamaño N : $\mathcal{L}(\theta | X) = \prod_{i=1}^N \Pr(D_i = 1 | X_i \beta)^{D_i} (1 - \Pr(D_i = 1 | X_i \beta))^{1-D_i}$

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N D_i \ln \Pr(D_i=1|X\beta) + (1-D_i) \ln(1-\Pr(D_i=1|X\beta)) \Rightarrow \hat{\beta}_{ML} = \arg\max(\ln \mathcal{L})$$

Elección multinomial.

Mutualmente excluyente: Suponga el siguiente modelo.

$$\left. \begin{array}{l} u_1 = x_1 \beta_1 + \varepsilon_1 \\ u_2 = x_2 \beta_2 + \varepsilon_2 \\ \vdots \\ u_K = x_K \beta_K + \varepsilon_K \end{array} \right\} \text{Criterio: Maximización de utilidad.}$$

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N \sum_{j=1}^K (1 - \sum_{s \neq j} D_{is}) D_{ij} \ln \Pr(D_{ij}=1|X\beta)$$

No mutuamente excluyente: Suponga el siguiente modelo.

$$D_{1i} = \begin{cases} u_{1i} > 0 & \Rightarrow 1 \\ \text{e.o.c.} & \end{cases} \quad D_{2i} = \begin{cases} u_{2i} > u_{1i} & \Rightarrow 1 \\ \text{e.o.c.} & \end{cases} \quad \begin{array}{c} \text{Posibles casos} \\ D_{1i} \quad D_{2i} \\ 1 \quad 1 \\ 1 \quad 0 \\ 0 \quad 0 \end{array}$$

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N D_{1i} D_{2i} \ln \Pr(D_{1i}=1, D_{2i}=1) + D_{1i}(1-D_{2i}) \ln \Pr(D_{1i}=1, D_{2i}=0) + (1-D_{1i}) \ln \Pr(D_{1i}=0)$$

Modelos de utilidad aleatoria $\rightarrow$ Para definir la forma de $\Pr(D_{ij}=1|X\beta)$.

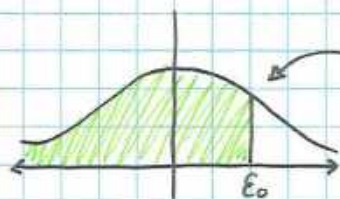
Si escoge 1 $\Rightarrow$ Deriva $u_1 = x_1 \beta_1 + \varepsilon_1$
Si escoge 0 $\Rightarrow$ Deriva $u_0 = x_0 \beta_0 + \varepsilon_0$ } Criterio: Maximización de utilidad

$$D_i = \begin{cases} 1 & \text{si } u_{1i} > u_{0i} \\ 0 & \text{e.o.c.} \end{cases} \Rightarrow \Pr(D_{1i}=1|X\beta) = \Pr(u_{1i} > u_{0i})$$

$$= \Pr(x_1 \beta + \varepsilon_1 > x_0 \beta + \varepsilon_0)$$

$$= \Pr(\varepsilon_1 > x_0 \beta - x_1 \beta + \varepsilon_0)$$

$$\Pr(D_{1i}=1|X\beta) = \Pr(\varepsilon_0 < x_1 \beta - x_0 \beta + \varepsilon_1)$$



MODELO LOGIT

Considere un modelo de probabilidad aleatoria: $u_1 = x_1 \beta + \varepsilon_1$ si escoge 1, $u_0 = x_0 \beta + \varepsilon_0$ si escoge cero. $\Rightarrow \Pr(D_{1i}=1|X\beta) = \Pr(\varepsilon_0 < x_1 \beta - x_0 \beta + \varepsilon_1)$

$\hookrightarrow$ Distribución V.E. tipo I.

Distribución acumulada para V.E. tipo I $F(a) = \exp(-e^{-a})$ Función de densidad para V.E. tipo I $f(a) = \exp(-a - e^{-a})$

Derivación - Expresión para $\Pr(D_i=1|X\beta)$ - Caso bivariado.

Considere $\Pr(D_{1i}=1|X\beta) = \Pr(\varepsilon_0 < x_1 \beta - x_0 \beta + \varepsilon_1)$, $\varepsilon_0 \sim$ V.E. tipo I.

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_1 \beta - x_0 \beta + \varepsilon_1} f(\varepsilon_1) f(\varepsilon_0) d\varepsilon_0 d\varepsilon_1 = \int_{-\infty}^{\infty} f(\varepsilon_1) \int_{-\infty}^{x_1 \beta - x_0 \beta + \varepsilon_1} f(\varepsilon_0) d\varepsilon_0 d\varepsilon_1$$

$$= \int_{-\infty}^{\infty} F(x_1 \beta - x_0 \beta + \varepsilon_1) f(\varepsilon_1) d\varepsilon_1 = \int_{-\infty}^{\infty} \exp\{-e^{x_0 \beta - x_1 \beta - \varepsilon_1}\} \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} d\varepsilon_1$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} \exp\{-\varepsilon_1 - e^{\varepsilon_1} \underbrace{(e^{x_1\beta + x_0\beta} + 1)}_{e^\lambda}\} d\varepsilon_1 = \int_{-\infty}^{\infty} \exp\{-\varepsilon_1 - e^{-(\varepsilon_1 - \lambda)}\} d\varepsilon_1 \\
 &= \int_{-\infty}^{\infty} \exp\{-(\varepsilon_1 - \lambda) - e^{-(\varepsilon_1 - \lambda)} - \lambda\} d\varepsilon_1 = \int_{-\infty}^{\infty} e^{-\lambda} f(\varepsilon_1 - \lambda) d\varepsilon_1 = e^{-\lambda} \int_{-\infty}^{\infty} f(\varepsilon_1 - \lambda) d\varepsilon_1 \\
 \Pr(\varepsilon_0 < x_1\beta - x_0\beta + \varepsilon_1) &= \frac{1}{e^\lambda} = \frac{1}{\left(\frac{e^{x_1\beta}}{e^{x_0\beta}} + 1\right)} = \frac{1}{\left(\frac{e^{x_1\beta} + e^{x_0\beta}}{e^{x_0\beta}}\right)} = \frac{e^{x_0\beta}}{e^{x_1\beta} + e^{x_0\beta}}
 \end{aligned}$$

Derivación - Expresión de $\Pr(D_{1i}=1 | x, \beta)$ - caso multivariado

Considere el modelo de utilidad aleatoria: $\left. \begin{aligned} u_{1i} &= x_{1i}\beta + \varepsilon_{1i} \\ u_{2i} &= x_{2i}\beta + \varepsilon_{2i} \\ u_{3i} &= x_{3i}\beta + \varepsilon_{3i} \end{aligned} \right\} \text{Regla de decisión: } \text{Max } u.$

$$\begin{aligned}
 \Pr(D_{1i}=1 | x, \beta) &= \Pr(u_{1i} > u_{2i}, u_{1i} > u_{3i}) = \Pr(x_{1i}\beta + \varepsilon_{1i} > x_{2i}\beta + \varepsilon_{2i}, x_{1i}\beta + \varepsilon_{1i} > x_{3i}\beta + \varepsilon_{3i}) \\
 &= \Pr(\varepsilon_{2i} < x_{1i}\beta - x_{2i}\beta + \varepsilon_{1i}, \varepsilon_{3i} < x_{1i}\beta - x_{3i}\beta + \varepsilon_{1i})
 \end{aligned}$$

$$\Pr(D_{1i}=1 | x, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_{1i}\beta - x_{2i}\beta + \varepsilon_{1i}} \int_{-\infty}^{x_{1i}\beta - x_{3i}\beta + \varepsilon_{1i}} f(\varepsilon_1) f(\varepsilon_2) f(\varepsilon_3) d\varepsilon_3 d\varepsilon_2 d\varepsilon_1$$

$$\begin{aligned}
 \Pr(D_{1i}=1 | x, \beta) &= \int_{-\infty}^{\infty} f(\varepsilon_1) F(x_{1i}\beta - x_{2i}\beta + \varepsilon_1) F(x_{1i}\beta - x_{3i}\beta + \varepsilon_1) d\varepsilon_1 \\
 &= \int_{-\infty}^{\infty} \exp\{-\varepsilon_1 - e^{-\varepsilon_1}\} \exp\{-e^{-(x_{1i}\beta - x_{2i}\beta + \varepsilon_1)}\} \exp\{-e^{-(x_{1i}\beta - x_{3i}\beta + \varepsilon_1)}\} d\varepsilon_1 \\
 &= \int_{-\infty}^{\infty} \exp\{-\varepsilon_1 - e^{-\varepsilon_1} \underbrace{(1 + e^{-x_{1i}\beta + x_{2i}\beta} + e^{-x_{1i}\beta + x_{3i}\beta})}_{e^\lambda}\} d\varepsilon_1 \\
 &= \int_{-\infty}^{\infty} \exp\{-(\varepsilon_1 - \lambda) - e^{-(\varepsilon_1 - \lambda)} - \lambda\} d\varepsilon_1
 \end{aligned}$$

$$\Pr(D_{1i}=1 | x, \beta) = e^{-\lambda} \int_{-\infty}^{\infty} f(\varepsilon_1 - \lambda) d\varepsilon_1 = e^{-\lambda} = \frac{1}{e^\lambda} = \frac{1}{\left(\frac{e^{x_1\beta} + e^{x_2\beta} + e^{x_3\beta}}{e^{x_1\beta}}\right)}$$

$$\Pr(D_{1i}=1 | x, \beta) = \frac{e^{x_1\beta}}{e^{x_1\beta} + e^{x_2\beta} + e^{x_3\beta}} \quad \text{En general: } \Pr(D_{ij}=1 | x, \beta) = \frac{e^{x_j\beta}}{\sum_{i=1}^J e^{x_i\beta}}$$

Identificación

① Escala

Considere $u_{1i} > u_{0i} \Rightarrow \sigma u_{1i} + k > \sigma u_{0i} + k \Rightarrow$ La decisión se mantiene.

$$\Rightarrow D_{1i}=1 \quad \Rightarrow D_{1i}=1$$

Es posible normalizar una utilidad. En ese caso:

$$\Pr(D_{1i}=1 | x, \beta) = \frac{e^{x_1\beta}}{1 + e^{x_1\beta}}$$

② Coeficientes

$$\begin{aligned}
 \text{Si } \left. \begin{aligned} u_{1i} &= \beta_{01} + \beta_{11} F_{1i} + \beta_{12} \text{edad} + \varepsilon_{1i} \\ u_{0i} &= \beta_{00} + \beta_{01} F_{0i} + \beta_{02} \text{edad} + \varepsilon_{0i} \end{aligned} \right\} u_{1i}^* &= (\beta_{01} - \beta_{00}) + \beta_{11} F_{1i} - \beta_{01} F_{0i} + (\beta_{12} - \beta_{02}) \text{edad} + (\varepsilon_{1i} - \varepsilon_{0i}) \\
 &\text{Varían a través de alternativas} \quad \text{Varían a través de individuos}
 \end{aligned}$$

No están identificados

Función de verosimilitud

$$F(x\beta) = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}} \Rightarrow \ln \mathcal{L}(\theta|X) = \sum_{i=1}^N \left\{ \ln e^{x_i\beta} - \ln(1 + e^{x_i\beta}) \right\} D_i + \left\{ \ln \left[\frac{1}{1 + e^{x_i\beta}} \right] \right\} (1 - D_i)$$

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N \left\{ \ln 1 - \ln(1 + e^{x_i\beta}) \right\} (1 - D_i) + \left\{ x_i\beta - \ln(1 + e^{x_i\beta}) \right\} D_i$$

CPO: $\frac{d \ln \mathcal{L}}{d\beta} = \frac{d}{d\beta} \left\{ \sum_{i=1}^N D_i \ln F(x\beta) + (1 - D_i) \ln (1 - F(x\beta)) \right\}$

$$= \sum_{i=1}^N \frac{D_i}{F(x\beta)} * \frac{dF(x\beta)}{d\beta} + \frac{(1 - D_i)}{1 - F(x\beta)} (-1) \frac{dF(x\beta)}{d\beta} = 0$$

$$\frac{dF(x\beta)}{d\beta} = \frac{d}{d\beta} \left(\frac{e^{x_i\beta}}{1 + e^{x_i\beta}} \right) = \frac{dF(x\beta)}{d(x\beta)} * \frac{d(x\beta)}{d\beta} = \frac{e^{x_i\beta}(1 + e^{x_i\beta}) - e^{x_i\beta}e^{x_i\beta}}{(1 + e^{x_i\beta})^2} * x_i$$

$$\frac{dF(x\beta)}{d\beta} = \frac{e^{x_i\beta}}{(1 + e^{x_i\beta})^2} * x_i, \Rightarrow \frac{dF(x\beta)}{d\beta} * \frac{1}{F(x\beta)} = \frac{e^{x_i\beta} x_i}{(1 + e^{x_i\beta})^2} * \frac{(1 + e^{x_i\beta})}{e^{x_i\beta}} = f(x\beta) / F(x\beta) * x_i$$

$F(x\beta)(1 - F(x\beta)) = f(x\beta)$

$$\frac{d \ln \mathcal{L}}{d\beta} = \sum_{i=1}^N \frac{D_i f(x\beta) * x_{1i}}{F(x\beta)} - \frac{f(x\beta) * x_{1i}}{1 - F(x\beta)} + \frac{D_i f(x\beta) * x_{1i}}{1 - F(x\beta)} = 0$$

$$= \sum_{i=1}^N D_i (1 - F(x\beta)) x_{1i} - F(x\beta) x_{1i} + D_i (F(x\beta)) x_{1i} = 0$$

$$= \sum_{i=1}^N x_{1i} (D_i - D_i F(x\beta) - F(x\beta) + D_i F(x\beta)) = 0$$

$$= \sum_{i=1}^N x_{1i} (D_i - F(x\beta)) = 0 \Rightarrow \text{CPO no es lineal en } \beta$$

Algoritmo de maximización numérica

- ① Se establece una conjetura $\beta = \beta_0$ (elegida por el econometrista.)
- ② Computar $\ln \mathcal{L}(\beta_0|X)$.
- ③ Encontrar $d \ln \mathcal{L} / d\beta |_{\beta = \beta_0}$.
- ④ Calcular el hessiano $H = [d^2 \ln \mathcal{L} / d\beta d\beta']$.
- ⑤ Establecer $\beta_1 = \beta_0 + \Delta$ donde $\Delta = f(H)$. Repetir pasos (2)-(4) con β_1 .
- ⑥ La iteración se detiene cuando $\ln \mathcal{L}(\beta_k|X) - \ln \mathcal{L}(\beta_{k-1}|X) \rightarrow 0$.

Interpretación de $\hat{\beta}_{MNL}$

Al interpretar $\hat{\beta}_{MNL}$, hablamos del cambio en la utilidad con un cambio en X . Sin embargo, la utilidad es un constructo del econometrista y no tiene relevancia.

Efecto marginal de $\hat{\beta}_{MNL}$ sobre D_i : $\frac{d \Pr(D_i = 1|X, \beta)}{dx} = \frac{d}{dx} \left(\frac{e^{x\beta}}{1 + e^{x\beta}} \right)$

$$= \frac{dF(x\beta)}{d(x\beta)} * \frac{d(x\beta)}{dx}$$

$$= F(x\beta)(1 - F(x\beta)) \beta$$

$$\frac{d \Pr(D_i = 1|X, \beta)}{dx} = \frac{\beta e^{x\beta}}{(1 + e^{x\beta})^2}$$

Ventajas del modelo logit.

I. La expresión de la probabilidad de éxito tiene solución cerrada.

II. $\ln \mathcal{L}(\cdot)$ es globalmente cóncava $\Rightarrow \hat{\beta}_{ML}$ es único.

$\hookrightarrow$ Considere $\sum_{i=1}^N X_i (D_i - F(X_i\beta)) = \sum_{i=1}^N X_i D_i - \sum_{i=1}^N X_i F(X_i\beta)$

$$\frac{dCPO}{d\beta} = \sum_{i=1}^N -X_i \left(\frac{dF(X_i\beta)}{d\beta} \right) = \sum_{i=1}^N -X_i * \left(\frac{e^{X_i\beta} X_i}{(1+e^{X_i\beta})^2} \right)$$

$$\frac{dCPO}{d\beta} = \sum_{i=1}^N \underbrace{-f(X_i\beta)}_{>0} \underbrace{X_i^2}_{>0} < 0 \Rightarrow \text{Globalmente cóncava.}$$

Desventajas del modelo logit - Independencia de alternativas irrelevantes

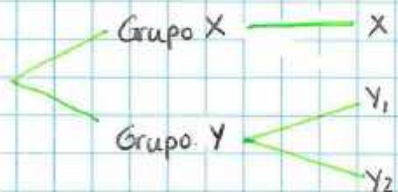
Considere un modelo con J alternativas.

$$\left. \begin{aligned} \Pr(D_{ij}=1|X, \theta) &= \frac{e^{X_j\beta}}{1 + e^{X_1\beta} + \dots + e^{X_{J-1}\beta}} \\ \Pr(D_{ik}=1|X, \theta) &= \frac{e^{X_k\beta}}{1 + e^{X_1\beta} + \dots + e^{X_{J-1}\beta}} \end{aligned} \right\} \frac{\Pr(D_{ij}=1)}{\Pr(D_{ik}=1)} = \frac{e^{X_j\beta}}{e^{X_k\beta}} = f(X_k, X_j)$$

La probabilidad de elegir j con respecto a la probabilidad de elegir k sólo depende de estas dos alternativas.

Possible solution: logit anidado

Considere las categorías X, Y_1, Y_2 . Se clasifican en grupos



Decisión maximizar U_X vs. $\max(U_{Y_1}, U_{Y_2})$

$$\max(U_{Y_1}, U_{Y_2}) \Rightarrow \Pr(\cdot) = \frac{e^{X_{Y_1}\beta}}{e^{X_{Y_1}\beta} + e^{X_{Y_2}\beta}}$$

$$U_X = \frac{e^{X\beta}}{e^{X\beta} + I(Y_1\beta + Y_2\beta)} \quad \left. \vphantom{U_X} \right\} \begin{array}{l} \text{Función de inclusión} \\ I(\cdot) \end{array}$$

MODELO PROBIT

Probit binomial

Considere $U_1 = X_1\beta + \varepsilon_1$, $\varepsilon_1 \sim N(0, \sigma^2)$
 $U_0 = 0$

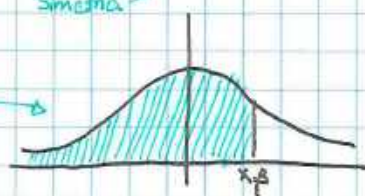
Regla de decisión $D_i = \begin{cases} 1 & \text{si } U_1 > 0 \\ 0 & \text{e.o.c} \end{cases}$

$$\ln \mathcal{L}(\theta|X) = \sum_{i=1}^N D_i \ln \Pr(U_1 > 0) + (1 - D_i) \ln (1 - \Pr(U_1 > 0)) \Rightarrow \theta = (\beta, \sigma^2)'$$

$$\Pr(D_i = 1|X, \theta) = \Pr(U_1 > 0) = \Pr(X_1\beta + \varepsilon_1 > 0) = \Pr(\varepsilon_1 > -X_1\beta) = \Pr(\varepsilon_1 < X_1\beta)$$

$$\Pr(\varepsilon_1 > -X_1\beta) = \int_{-X_1\beta}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{\varepsilon_1}{\sigma} \right)^2 \right\} d\varepsilon_1$$

$$\Pr(D_i = 1|X, \theta) = \int_{-\infty}^{X_1\beta} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{\varepsilon_1}{\sigma} \right)^2 \right\} d\varepsilon_1$$



Si normalizamos para que $\varepsilon_i \sim N(0, 1)$:

$$Pr(D_i = 1 | X, \theta) = Pr\left(\frac{\varepsilon_i}{\sigma} < \frac{X_i \beta}{\sigma}\right) = \int_{-\infty}^{\frac{X_i \beta}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\varepsilon_i^2} d\varepsilon_i$$

Es necesario fijar uno de los parámetros. Por convención, se asume $\sigma = 1$.

No obstante, en el proceso de normalización tenemos que:

$$\frac{X_i \beta(\beta)}{\sigma(\beta)} = \frac{X_i \beta}{\sigma}$$

$\Rightarrow \beta$ y σ no están identificados de manera única.

Probit multinomial

Considere $u_0 = 0$
 $u_1 = X_1 \beta + \varepsilon_1$
 $u_2 = X_2 \beta + \varepsilon_2$

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right]$$

En este caso, $\theta = (\beta, \sigma_{11}, \sigma_{12}, \sigma_{22})$ $D_i = k$ si $u_k > u_j \forall k \neq j$.

$$Pr(D_i = 1 | X, \theta) = Pr(u_1 > 0, u_1 > u_2) = Pr(\varepsilon_1 > -X_1 \beta, \varepsilon_2 > -X_2 \beta + X_1 \beta + \varepsilon_1)$$

$$Pr(D_i = 1 | X, \theta) = \int_{-\infty}^{X_1 \beta} \int_{-X_2 \beta + X_1 \beta + \varepsilon_1}^{\infty} \underbrace{f(\varepsilon_1, \varepsilon_2 | \Sigma)}_{\text{pdf normal multivariada}} d\varepsilon_2 d\varepsilon_1$$

Si normalizamos: $Pr(D_i = 1 | X, \theta) = Pr\left(\frac{\varepsilon_1}{\sigma_1} > -\frac{X_1 \beta}{\sigma_1}, \frac{\varepsilon_2}{\sigma_2} > \frac{X_1 \beta - X_2 \beta + \varepsilon_1}{\sigma_2}\right)$

Aún es necesario fijar un sólo parámetro.

SIMULADORES

Objetivo: Simular las integrales que no podemos solucionar de forma analítica.

Simulador de frecuencias (Montecarlo)

Vi 1. Sacar $\{\varepsilon_1^d, \varepsilon_2^d\}_{d=1, \dots, D}$, donde $D = \#$ de simulaciones por individuo, de su correspondiente distribución de probabilidad.

2. $\forall d$ definimos $u_1^d = X_1 \beta + \varepsilon_1^d$, $u_2^d = X_2 \beta + \varepsilon_2^d$, $u_3^d = 0$

MODELOS LINEALES DE PANEL

Definición de panel: Mismos Individuos a lo largo del tiempo

Considere el modelo $y_{it} = X_{it}\beta + \varepsilon_{it}$
 $\rightarrow y_{it}$ es continua: $i = \{1, \dots, N\}$, $t = \{1, \dots, T\}$

Mismos individuos a lo largo del tiempo.

Tipos de panel

TIPOS DE PANEL

Según el tamaño de N o T

Micropanels

$N \gg T$

Análisis basado en $N \rightarrow \infty$, T fijo (Consistencia)

Macropanel

$T \gg N$

Análisis basado en técnicas de series de tiempo

Balanceados

El mismo número de períodos para cada i

No balanceados

Algunas unidades de sección cruzada tienen más períodos de tiempo.

Ventajas de datos panel

- 1) A. Dinámico
- 2) Het. Indiv.
- 3) Mejora ef.

Ventajas

- ① Permite análisis de procesos dinámicos.
- ② Permite controlar por heterogeneidad individual / de sección cruzada / de corte transversal que es fija a lo largo del tiempo.
- ③ Es más informativo, busca más variabilidad y tiende a mejorar la eficiencia.

Problemas de datos panel

- 1) Heterosced. por problemas de escala

Problemas

- ① Heteroscedasticidad (si es estimado por MCO $\Rightarrow$ Errores estándar incorrectos). Causa: Escala de las observaciones y_{it} .

E.g: $PIB_{it} = \text{Det}_{it}\beta + \varepsilon_{it}$ EEUE vs. Etiopía
 $PIB_{EEUE,t} \gg PIB_{Etiopía,t}$

E.g: Flujos del comercio: $Tarifa_{it} = \text{Tarifas}_{it}\beta + \varepsilon_{it}$ EEUE vs. Colombia
 $SCE = SCE_{EEUE} + SCE_{Colombia}$, $SCE_{EEUE} \gg SCE_{Colombia}$
 $\min SCE \rightarrow \hat{\beta}$ tiene más en cuenta a SCE_{EEUE}

Sols. A Het.

(a) Transformar el modelo a logs

Posibles soluciones

- (a) Transformar y_{it} en logaritmos para balancear la escala.
 Limitación: $y_{it} = 0$ o $y_{it} < 0 \Rightarrow$ Se pierden las observaciones.

(b) MCP

- (b) Dividir por un factor de escala (p.e: z_i)

$$\frac{y_{it}}{z_i} = \frac{X_{it}}{z_i}\beta + \frac{\varepsilon_{it}}{z_i} \quad \text{s.t.} \quad \text{var}\left(\frac{\varepsilon_{it}}{z_i}\right) = \sigma_\varepsilon^2 \quad \text{MCP}$$

E.g: $\frac{\text{Flujo del com}_{it}}{\text{Producción}_i} = \frac{\text{Tarifas}_{it}}{\text{Producción}_i}\beta + \frac{\varepsilon_{it}}{\text{Producción}_i}$
 Donde producción_i = El promedio de producción de las firmas a lo largo del tiempo para la unidad i .

MCGF

¿Cómo definir z_i ? (MCGF)

- (I) Estime el modelo original por MCO y obtenga $\hat{e}_i$.
- (II) Estimar una regresión auxiliar de la siguiente forma:

$$\hat{e}_{it} = \theta_0 + \theta_1 z_i + \theta_2 z_i^2 + \theta_3 \sqrt{z_i} + w_{it}$$
- (III) Realizar una prueba de significancia individual para identificar cuál forma funcional de z_i es relevante.
- (IV) Dividimos el modelo por esa forma funcional

(c) E.E. Robustos

- (c) Estimar con errores estándares robustos: Corrige los errores estándares usando estimadores tipo "Sandwich" (Huber - White)

↳ Errores estándar válidos para efectuar pruebas de hipótesis.

2) Corr. transvers.

(2) Correlación de corte transversal

- (I) $E(e_{it}, e_{is}) \neq 0 \quad \forall t \neq s$: Por ser el mismo individuo.
- (II) $E(e_{it}, e_{jt}) \neq 0 \quad \forall i \neq j$: Por choques agregados o estacionales. P.e: Ola invernal

↳ Posible solución: Considere $y_{it} = x_{it}\beta + e_{it}$. Se introduce un efecto fijo de tiempo. $y_{it} = x_{it}\beta + \nu_t + e_{it}$.

(3) Heterogeneidad

(3) Heterogeneidad de parámetros

} Ignorar heterogeneidad ind.
Causa correlación serial

(4) Corr. serial

(4) Correlación serial

E.g: Considere una función de producción de tipo Cobb-Douglas.

$$y_{it} = A_i x_{it}^{\beta_i} e_{it} \Rightarrow \ln y_{it} = \ln A_i + \beta_i \ln x_{it} + \ln e_{it}$$

$$y_{it} = \beta_{0i} + \beta_{1i} x_{it} + e_{it}$$

Si imponemos un β_0 común: $y_{it} = \beta_0 + \beta_{1i} x_{it} + (e_{it} - \beta_{0i} + \beta_{0i})$

$$y_{it} = \beta_0 + \beta_{1i} x_{it} + w_{it} \quad \mu_i$$

$w_{it} = \mu_{it} + e_{it}$. Se puede demostrar que: $E[w_{it} w_{is}] = \sigma_{\mu}^2 = E[\mu_i^2]$
Si asumimos $E[\mu_i e_{it}] = 0$ y $E[e_{it} e_{is}] = 0 \quad \forall t \neq s$. $E[e_{it}^2] = \sigma_e^2$
 $= \rho_{\mu}$ $= \theta$

RESUMEN - CLASE 1

DATOS PANEL

Def: Base de datos que sigue a los mismos individuos a lo largo del tiempo.
 $y_{it}, i = \{1, \dots, N\}$
 $t = \{1, \dots, T\}$

Desventajas

Tipos de panel

Según N & T

Micropanel
 $N \gg T$

Macropanel
 $T \gg N$

Según balance

Balanceado
 $N = T$

No balanceado
 $N \neq T$

Ventajas

- Permite análisis dinámico
- Controlar por heteroscedasticidad indiv.
- Mejora eficiencia

Heteroscedasticidad

→ Soluciones:

- Logaritmos
- MCP
- MCGF
- E.E robustos

Corr. Transversal : $E(e_{it} e_{is}) \neq 0, E(e_{it} e_{jt}) \neq 0$

Heterogeneidad de Parámetros → Correlación serial

MÉTODOS DE ESTIMACIÓN PANEL

El método a plantear depende de los supuestos que se planteen.
Sea el modelo:

$$y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$$

Efecto fijo que no dependiera de t

Consideramos $E[x_{it}\varepsilon_{it}] = 0 \rightarrow$ No hay endogeneidad debida a ε_{it} y x_{it} de manera contemporánea

≠ Métodos según supuestos respecto a $E[x_{it}\mu_i]$

¿Qué asumimos respecto a $E[x_{it}\mu_i]$?

- ↳ Si $E[x_{it}\mu_i] = 0 \Rightarrow E[x_{it}w_{it}] = 0 \Rightarrow$ No hay endogeneidad (Efectos aleatorios)
- ↳ Si $E[x_{it}\mu_i] \neq 0 \Rightarrow E[x_{it}w_{it}] \neq 0 \Rightarrow$ Hay endogeneidad (Efectos fijos)

Si $E[x_{it}\mu_i] \neq 0$

Métodos de estimación con efectos fijos.

Dummy $\forall i$

- (1) Incluir variables dummy para cada i (LSDV)
 - ↳ Problema: Si T es pequeño, no habría información suficiente para estimar cada constante i .
 - ↳ Si K se hace muy grande, se reducen los grados de libertad disponible, y por tanto, la potencia de las P.H.

Primera diferencia

(2) Primera diferencia (FD)

Considere un modelo con t y $t-1$.

$$y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}, \quad y_{it-1} = x_{it-1}\beta + \mu_i + \varepsilon_{it-1}$$

Se toma la primera diferencia:

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (\mu_i - \mu_i) + (\varepsilon_{it} - \varepsilon_{it-1})$$

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (\varepsilon_{it} - \varepsilon_{it-1})$$

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta \varepsilon_{it}$$

Aplicamos MCO a las variables transformadas:

$$\hat{\beta}_{FD} = (\Delta X' \Delta X)^{-1} \Delta X' \Delta Y$$

Para que $\hat{\beta}_{FD}$ sea consistente, necesitamos que:

$$E[\Delta x_{it} \Delta \varepsilon_{it}] = E[(x_{it} - x_{it-1})(\varepsilon_{it} - \varepsilon_{it-1})] = 0$$

Puede generar endogeneidad

Esto se cumple si:

- $E[x_{it}\varepsilon_{it}] = 0$
 - $E[x_{it-1}\varepsilon_{it-1}] = 0$
 - $E[x_{it-1}\varepsilon_{it}] = 0 \rightarrow$ Plausible
 - $E[x_{it}\varepsilon_{it-1}] = 0 \rightarrow$ Supuesto FUERTE.
- } contemporáneos
} Supuesto clásico

$\Rightarrow \hat{\beta}_{FD}$ probablemente sea inconsistente.

Primera diferencia con VI

(3) Primera diferencia con variables instrumentales (FD-IV)

Asumamos que $E[x_{it}\varepsilon_{it-1}] = 0 \Rightarrow \hat{\beta}_{FD}$ inconsistente.

Zit resuelve endogeneidad en el modelo Δy_{it} .

Buscamos un instrumento Zit que cumpla con

- (i) $E[Z_{it}(X_{it} - X_{it-1})] \neq 0 \Rightarrow$ Zit debe ser **relevante**
- (ii) $E[Z_{it}(E_{it} - E_{it-1})] = 0 \Rightarrow$ Zit debe ser **exógeno**

Buscamos resolver endogeneidad del modelo $\Delta y_{it} = \Delta X_{it}\beta + \Delta E_{it}$

Posibles instrumentos $\begin{cases} X_{it-1} \\ X_{it-1} - X_{it-2} \end{cases}$

Xit: Para que sea exógeno:

$$E[X_{it-1}(E_{it} - E_{it-1})] = 0 \quad \begin{cases} E[X_{it-1} E_{it}] = 0 & \text{Plausible} \\ E[X_{it-1} E_{it-1}] = 0 & \text{Clásico} \end{cases}$$

Zit impone exigencias adicionales para eliminar endo.

Para que sea relevante:

$$E[X_{it}(X_{it} - X_{it-1})] \neq 0 \quad \begin{cases} E[X_{it}^2] \neq 0 \\ E[X_{it}X_{it-1}] \neq 0 \\ E[X_{it}^2] - E[X_{it}X_{it-1}] \neq 0 \end{cases}$$

↳ Tiene que haber variabilidad

Efectos fijos Within

④ Restar la media - Efectos fijos Within (FE - Within)

Considere: $y_{it} = X_{it}\beta + \mu_i + E_{it}$ donde $\bar{y}_i = \frac{\sum_{t=1}^T y_{it}}{T}$

$$- \bar{y}_i = \bar{X}_i\beta + \mu_i + \bar{E}_i$$

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + (E_{it} - \bar{E}_i)$$

$$\tilde{y}_{it} = \tilde{X}_{it}\beta + \tilde{E}_{it}$$

Necesitamos que $X_{it} - \bar{X}_i \neq 0 \Rightarrow$ Que haya variabilidad de X y en el tiempo, y que no sea igual $\forall i$, porque si no habría multicolinealidad perfecta con la constante.

Estimamos β por MCO: $\hat{\beta} = (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{y}$

Para que sea consistente, se necesita que: $E[\tilde{X}_{it}\tilde{E}_{it}] = 0$
 $E[(X_{it} - \bar{X}_i)(E_{it} - \bar{E}_i)] = 0$

Esto requiere que $E[X_{it}\bar{E}_i] = 0 = \frac{1}{T} E[X_{it}(\sum_{s=1}^T E_{is})]$
 $= \frac{1}{T} \sum_{s=1}^T E[X_{it}E_{is}]$

Hay exogeneidad estricta si $\forall t \neq s$
 $E[X_{it}E_{is}] = 0$

$$E[X_{it}E_{is}] = 0 \quad \begin{cases} s < t & \text{(Supuesto fuerte)} \\ s = t & \text{(Clásico, contemporáneo)} \\ s > t & \text{(Plausible)} \end{cases} \quad \begin{matrix} \textcircled{A} \\ \textcircled{B} \\ \textcircled{C} \end{matrix} \quad (*)$$

predeterminadas $E[X_{it}E_{is}] = 0$
 $\Rightarrow t \leq s$

Si se cumplen A, B, C se dice que hay exogeneidad estricta $\Rightarrow$ FE-Within provee estimadores consistentes

Si se cumplen B y C se dice que hay exogeneidad predeterminada o secuencial $\Rightarrow$ FE-Within provee estimadores inconsistentes.

Incluso si intentamos instrumentar $(X_{it} - \bar{X}_i)$ con X_{it-1} , se requeriría que $E[X_{it-1}(E_{it} - \bar{E}_i)] = 0$, y esto plantea el mismo problema (*).

Modelos de estimación con efectos aleatorios

Efectos aleatorios

Considere el modelo $y_{it} = X_{it}\beta + \mu_i + E_{it}$ donde $E[X_{it}\mu_i] = 0$ y $E[X_{it}E_{it}] = 0$.
 ↳ Si estimamos por OLS (Pooled OLS) $\Rightarrow \hat{\beta}$ insesgado, consistente pero ineficiente.

$\Rightarrow$ MCO al modelo

Genera
heteroscedas-
tidad

Porque $E[w_{it} w_{is}] = \sigma_\mu^2 \Rightarrow$ Hay correlación serial.

Soluciones

④ Mínimos Cuadrados generalizados

Tenemos que $E[w_{it}^2] = E[(\mu_i + \varepsilon_{it})^2] = \sigma_\mu^2 + \sigma_\varepsilon^2$ si $E[\varepsilon_{it} \mu_i] = 0$

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\mu^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 \\ \sigma_\mu^2 & \ddots & & \sigma_\mu^2 \\ \vdots & & \ddots & \vdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \dots & \sigma_\varepsilon^2 + \sigma_\mu^2 \end{bmatrix}_{T \times T} = \sigma_\varepsilon^2 I_T + \sigma_\mu^2 j_T j_T' = \sigma_\varepsilon^2 I_T + \sigma_\mu^2 J_T$$

donde $J_T = \frac{1}{T} j_T j_T'$

Solución por
MCG

Matriz de varianza por individuo.

$$\text{Note que Var-Cov} = \Omega = \begin{bmatrix} \Sigma & 0 & \dots & 0 \\ 0 & \Sigma & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma \end{bmatrix} = I_N \otimes \Sigma$$

El estimador de efectos aleatorios se obtiene como:

$$\hat{\beta}_{RE} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y \quad \text{Para transformar el modelo, necesitamos}$$

$$\Omega^{-1} = \Omega^{-1/2} \Omega^{-1/2} \rightarrow \Omega^{-1/2} = [I_N \otimes \Sigma]^{-1/2}$$

RESUMEN - CLASE 2

Métodos de estimación en datos panel

Considere el modelo
 $y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$

I. Modelos de efectos aleatorios

En este tipo de modelos, $E[x_{it} \mu_i] = 0 \rightarrow E[x_{it} w_{it}] = 0$ No Hay endogeneidad.

Método de estimación: Pooled OLS

$\rightarrow$ Estimamos $y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$ $\left\{ \begin{array}{l} \hat{\beta} \text{ es: } \rightarrow \text{Insesgado} \\ \rightarrow \text{Consistente} \\ \rightarrow \text{Ineficiente} \end{array} \right.$
 $\rightarrow$ Se corrige por MCGF o EE Robustos

Fuente de ineficiencia: Heteroscedasticidad - $E[w_{it} w_{is}] = \sigma_\mu^2$, $E[w_{it}^2] = \sigma_\varepsilon^2 + \sigma_\mu^2$

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\mu^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\varepsilon^2 + \sigma_\mu^2 & \dots & \sigma_\mu^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \dots & \sigma_\varepsilon^2 + \sigma_\mu^2 \end{bmatrix} \rightarrow \text{Var-Cov} = I_N \otimes \Sigma$$

Varianza por i :

II. Modelos de efectos fijos

En este tipo de modelos, $E[x_{it} \mu_i] \neq 0 \rightarrow E[x_{it} w_{it}] \neq 0 \Rightarrow$ Hay endogeneidad

A. Incluir v. dummy β_i (LDSV) $\rightarrow$ Pero T tiene que ser muy grande para poder estimar β_0
 $\rightarrow$ Si $K \rightarrow \infty \Rightarrow$ Grados de libertad $\rightarrow 0 \Rightarrow$ Invalidación P.H.

B. Primera diferencia (FD) $y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (\mu_i - \mu_i) + (\varepsilon_{it} - \varepsilon_{it-1})$
 $\Delta y_{it} = \Delta x_{it}\beta + \Delta \varepsilon_{it}$

Estimamos Δy_{it} por MCO: $\hat{\beta}_{FD} = (\Delta x_{it}' \Delta x_{it})^{-1} \Delta x_{it}' \Delta y_{it}$

$\rightarrow$ Es consistente si $E[\Delta x_{it} \Delta \varepsilon_{it}] = 0$. Requiere:

$$E[(x_{it} - x_{it-1})(\varepsilon_{it} - \varepsilon_{it-1})] = E[\underbrace{x_{it}\varepsilon_{it}}_{=0 \text{ por sup. clásico}} - \underbrace{x_{it-1}\varepsilon_{it}}_{=0 \text{ sup. débil}} - \underbrace{x_{it}\varepsilon_{it-1}}_{=0 \text{ sup. fuerte}} + \underbrace{x_{it-1}\varepsilon_{it-1}}_{=0 \text{ por sup. clásico}}]$$

C. Primera diferencia con variables instrumentales (FD-IV)

Se calcula el modelo con primeras diferencias instrumentando Δx_{it} . El instrumento z_{it} debe ser:

(i) Relevante $E[z_{it} \Delta x_{it}] \neq 0$
 (ii) Exógeno $E[z_{it} \Delta \varepsilon_{it}] = 0$

P.e: $x_{it-1} - x_{it-2}$: ¿Es relevante? $E[(x_{it-1} - x_{it-2})(x_{it} - x_{it-1})]$
 $= E[x_{it}x_{it-1} - x_{it}x_{it-2} - \underbrace{x_{it-1}^2}_{\text{req. variabilidad}} + x_{it-1}x_{it-2}] \neq 0$

¿Es exógeno? $E[(x_{it-1} - x_{it-2})(\varepsilon_{it} - \varepsilon_{it-1})] = E[\underbrace{x_{it-1}\varepsilon_{it}}_{=0 \text{ sup. débil}} - \underbrace{x_{it-2}\varepsilon_{it}}_{=0 \text{ sup. débil}} - \underbrace{x_{it-1}\varepsilon_{it-1}}_{=0 \text{ sup. clásico}} + \underbrace{x_{it-2}\varepsilon_{it-1}}_{=0 \text{ sup. débil}}]$

Depende mucho del instrumento. Requiere de mín. 3 períodos para usar el instr.

D. Efectos fijos - within (FE-within) Restamos la media, $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\mu_i - \mu_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$y_{it}^* = x_{it}^*\beta + \varepsilon_{it}^* \Rightarrow \text{Se estima por MCO. } \hat{\beta}_{FE-W} = (x_{it}^{*'} x_{it}^*)^{-1} x_{it}^{*'} y_{it}^*$$

Para que sea consistente: $E[x_{it}^* \varepsilon_{it}^*] = 0$
 $= E[(x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)]$
 $= E[\underbrace{x_{it}\varepsilon_{it}}_{\text{sup. clásico}} - \underbrace{\bar{x}_i\varepsilon_{it}}_{\text{sup. Fuerte}} - \underbrace{x_{it}\bar{\varepsilon}_i}_{=0} + \underbrace{\bar{x}_i\bar{\varepsilon}_i}_{=0}]$

Considere $E[\bar{x}_i \varepsilon_{it}] = \frac{1}{T} E[(\sum_{s=1}^T x_{is}) \varepsilon_{it}] = \frac{1}{T} \sum_{s=1}^T E[x_{is} \varepsilon_{it}]$, $\forall s \neq t$.
 $E[x_{it} \bar{\varepsilon}_i] = \frac{1}{T} E[(\sum_{s=1}^T x_{is}) \varepsilon_{it}] = \frac{1}{T} \sum_{s=1}^T E[x_{is} \varepsilon_{it}]$

$$E[x_{is} \varepsilon_{it}] = 0 \begin{cases} \text{Si } t > s \Rightarrow \text{Plausible} \\ \text{Si } t = s \Rightarrow \text{Clásico} \\ \text{Si } s > t \Rightarrow \text{sup. fuerte} \end{cases} \quad E[x_{it} \varepsilon_{is}] = 0 \begin{cases} \text{Si } s > t \Rightarrow \text{Plausible} \\ \text{Si } t = s \Rightarrow \text{Clásico} \\ \text{Si } s < t \Rightarrow \text{sup. fuerte} \end{cases}$$

Exogeneidad

- Estricta** Si $\forall t \neq s \quad E[x_{is} \varepsilon_{it}] = 0 \ \& \ E[x_{it} \varepsilon_{is}] = 0$
- Secuencial / Predeterminada** Si $\forall t \leq s \quad E[x_{it} \varepsilon_{is}] = 0$
 Si $\forall s \leq t \quad E[x_{is} \varepsilon_{it}] = 0$

Soluciones a
AC en RE

MCG

$$\Sigma^{-1/2} = f(\theta)$$

$$\theta = g(\sigma_e^2, \sigma_\mu^2)$$

Transformación

→
Similar a FE-
withinModelo
transformado

Consistente

↔
Ex. estricta

MCGF

Estimar σ_e^2, σ_μ^2 Soluciones

(A) Mínimos Cuadrados Generalizados

Para transformar los datos, requerimos $\Sigma^{-1/2}$. Esta matriz tiene la forma:

$$\Sigma^{-1/2} = \begin{bmatrix} 1 - \frac{\theta}{T} & -\frac{\theta}{T} & \dots & -\frac{\theta}{T} \\ -\frac{\theta}{T} & 1 - \frac{\theta}{T} & \dots & -\frac{\theta}{T} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\theta}{T} & -\frac{\theta}{T} & \dots & 1 - \frac{\theta}{T} \end{bmatrix}_{T \times T} \quad \text{Donde}$$

$$\theta = 1 - \left(\frac{\sigma_e^2}{\sigma_e^2 + T\sigma_\mu^2} \right)^{1/2}$$

$$\Sigma^{-1/2} = I_T - \frac{\theta}{T} \mathbf{j}_T \mathbf{j}_T' = I_T - \theta \mathbf{J}_T \quad \text{donde } \mathbf{J}_T = \frac{1}{T} \mathbf{j}_T \mathbf{j}_T'$$

Recuerde que $y = X\beta + w$. Transformaremos el modelo de la siguiente manera: $y_{it}^* = \Sigma^{-1/2} y_i$ donde el vector $y_i = (y_{i1}, y_{i2}, \dots, y_{iT})'_{T \times 1}$.

Para el primer componente de la multiplicación:

$$\begin{aligned} y_{i1} \left(1 - \frac{\theta}{T}\right) - \frac{\theta}{T} y_{i2} - \dots - \frac{\theta}{T} y_{iT} \\ = y_{i1} - \frac{\theta}{T} y_{i1} - \frac{\theta}{T} y_{i2} - \dots - \frac{\theta}{T} y_{iT} \\ = y_{i1} - \frac{\theta}{T} \sum_{i=1}^T y_{it} = y_{i1} - \theta \bar{y}_i \end{aligned} \quad y_{it}^* = \begin{bmatrix} y_{i1} - \theta \bar{y}_i \\ y_{i2} - \theta \bar{y}_i \\ \vdots \\ y_{iT} - \theta \bar{y}_i \end{bmatrix}$$

De la misma manera, se hace para x_{it} y w_{it} .

$$w_{it}^* = \Sigma^{-1/2} w_i \Rightarrow E[w_{it}^* w_{it}^{*'}] = \Sigma^{-1/2} \Sigma \Sigma^{-1/2} = \sigma_e^2 I_T$$

El modelo con los datos transformados será:

$$(y_{it} - \theta \bar{y}_i) = (x_{it} - \theta \bar{x}_i) \beta + (w_{it} - \theta \bar{w}_i)$$

Para que sea consistente: $E[(x_{it} - \theta \bar{x}_i)(w_{it} - \theta \bar{w}_i)] = 0$ ↳ Requiere $E[x_{it} \bar{w}_i] \Rightarrow E[x_{it} \bar{e}_i] = 0 \Rightarrow$ Exogeneidad estr.

(B) Mínimos Cuadrados generalizados factibles.

Estimamos σ_e^2 y σ_μ^2 .Procedimiento

- I. Estimar $y_{it} = x_{it}\beta + w_{it}$ por POLS. $\rightarrow \hat{\beta}_{\text{POLS}}$
- II. Obtenemos $\hat{w}_{it} = y_{it} - x_{it}\hat{\beta}_{\text{POLS}}$.
- III. Obtenemos $\hat{\sigma}_w^2 = \hat{\sigma}_e^2 + \hat{\sigma}_\mu^2$.

$$\hat{\sigma}_w^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T \hat{w}_{it}^2}{NT - K}$$

IV. Obtenemos $\hat{\sigma}_{\mu}^2$ y $\hat{\sigma}_{\varepsilon}^2$

(A) Para $\hat{\sigma}_{\mu}^2$: Sabemos que $\sigma_{\mu}^2 = E[w_{it} w_{is}]$

$$\hat{\sigma}_{\mu}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \hat{w}_{it} \hat{w}_{is}}{\left\{ \frac{NT(T-1)}{2} - K \right\}}$$

(B) Para $\hat{\sigma}_{\varepsilon}^2 = \hat{\sigma}_w^2 - \hat{\sigma}_{\mu}^2$.

V. Construir y_i^* y x_i^* o construir $\hat{\Sigma}^{-1}$ para obtener

$\hat{\beta}_{RE}$ { Matricial
Ponderando las medias.

$\hat{\beta}_{RE}$

Valores de θ

$$\theta = 1 - \left(\frac{\sigma_{\varepsilon}^2}{\sigma_{\varepsilon}^2 + T\sigma_{\mu}^2} \right)^{-1/2}$$

→ Si $\sigma_{\mu}^2 \rightarrow 0 \Rightarrow \theta \rightarrow 0$. Se aproxima a POLS

→ Si $\sigma_{\mu}^2 / \sigma_{\varepsilon}^2 \rightarrow \infty \Rightarrow \theta \rightarrow 1$. Se aproxima a FE-Within.

→ Si $\theta \in (0, 1) \Rightarrow$ Random effects.

Intervalo de θ

Modelos de Panel con Endogeneidad

MODELOS DE PANEL CON ENDOGENIDAD

Suponemos ahora que $y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$, donde $E[x_{it}\varepsilon_{it}] \neq 0$. Dado que tenemos efectos fijos, considere:

FD-IV vs. FE-Within IV

(a) FD-IV: $y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (\varepsilon_{it} - \varepsilon_{it-1})$. Tomamos un instrumento z_{it} .

→ Requerimos $E[z_{it}(\varepsilon_{it} - \varepsilon_{it-1})] = 0 \Leftrightarrow E[z_{it}\varepsilon_{it-1}] = 0$ {Sup. Fuerte. Esto elimina x_{it-1} como instrumento.

Sin ex. estricta:
→ FD-IV: Cons
→ FE-W: Incons

(b) FE-Within-IV: $y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$. Sea z_{it} un instrumento.

→ Requerimos $E[z_{it}(\varepsilon_{it} - \bar{\varepsilon}_i)] = 0 \Leftrightarrow E[z_{it}\bar{\varepsilon}_i] = 0$ {Sup. Fuerte

Si $T \rightarrow \infty$:
FD-IV > FEW-IV

Sea z_{it-1} un instrumento. $E[z_{it}(\varepsilon_{it} - \bar{\varepsilon}_i)] = 0 \Leftrightarrow E[z_{it-1}\bar{\varepsilon}_i] = 0$ {Sup. Fuerte
⇒ El instrumento debe ser estrictamente exógeno.

Si se viola el supuesto de exogeneidad estricta, FD-IV sigue siendo consistente, mientras que FE-Within no lo es. Si $T \rightarrow \infty$, FD-IV cobra importancia sobre FE-Within-IV.

Pruebas de Hausman

PRUEBAS DE HAUSMAN

Exogeneidad estricta

(1) Exogeneidad estricta

Considere el modelo $y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Asumamos que

(A) $E[x_{it}\varepsilon_{it}] = 0 \quad \forall t$
(B) $E[x_{it}\varepsilon_{is}] = 0 \quad \forall t > s$ } Si se cumplen (A) y (B) ⇒ Exogeneidad secuencial

$H_0: E[x_{it}\varepsilon_{is}] = 0 \quad \forall s < t \Rightarrow$ Hay exogeneidad estricta

$H_a: E[x_{it}\varepsilon_{is}] \neq 0 \quad \forall s < t \Rightarrow$ Hay exogeneidad secuencial

Planteemos dos estimadores y compárelos. (FD vs. FD-IV)

FD: $y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (\varepsilon_{it} - \varepsilon_{it-1})$. $\hat{\beta}_{FD}$ es consistente si:
 $E[x_{it}\varepsilon_{it}] = 0$ (A) $E[x_{it-1}\varepsilon_{it}] = 0$ (A) $E[x_{it-1}\varepsilon_{it}] = 0$ (B) $E[x_{it}\varepsilon_{it-1}] = ?$

Compara FD vs FD-IV

Sin ex. estricta:

FD: Incons

FD-IV: Cons

Ho: Ex. estricta

Bajo H_0 , $\hat{\beta}_{FD}$ es consistente

FD-IV: $y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (\varepsilon_{it} - \varepsilon_{it-1})$. Sea z_{it} un instrumento, pero requiere $E[z_{it}\varepsilon_{it-1}] = 0$. Si $z_{it-1} = x_{it-1}$: $E[x_{it-1}\varepsilon_{it-1}] = 0$ (A) y $E[x_{it-1}\varepsilon_{it}] = 0$ (B).

Bajo H_0 y H_a , $\hat{\beta}_{FD-IV}$ es consistente $\Rightarrow$ (A) y (B) son condiciones suficientes para exogeneidad estricta.

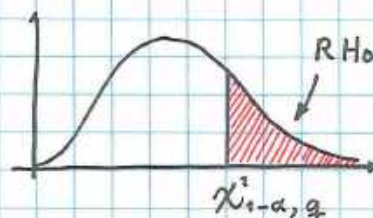
Estadístico de prueba

Criterio de Wald Sea $\Delta\hat{\beta} = \hat{\beta}_{FD-IV} - \hat{\beta}_{FD}$.

$$H = (\Delta\hat{\beta})' (\text{var}\{\Delta\hat{\beta}\})^{-1} (\Delta\hat{\beta}) \sim \chi^2_q$$

q : # de variables explicativas, # β 's sin β_0

Región de rechazo



FE vs. RE

② Efectos fijos vs. Efectos aleatorios

Compara

POLS - GLS vs.

FE - Within

Considere $y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Asumamos que se cumple exogeneidad estricta.

$H_0: E[x_{it}\mu_i] = 0 \Rightarrow$ Efectos aleatorios - MCG

$H_a: E[x_{it}\mu_i] \neq 0 \Rightarrow$ Efectos fijos - Within

FE-Within: $(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$. $\hat{\beta}_{FE-Within}$ es consistente si $E[(x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)] = 0$. Esto se cumple por supuesto. Por lo tanto $\hat{\beta}_{FE-w}$ es consistente bajo H_0 y H_a , dado que hay exogeneidad estricta.

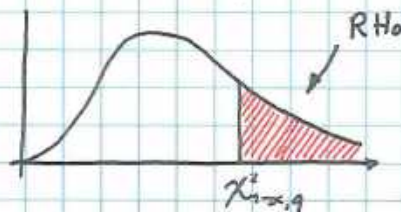
RE-GLS: $(y_{it} - \theta \bar{y}_i) = (x_{it} - \theta \bar{x}_i)\beta + \underbrace{(1-\theta)\mu_i + (\varepsilon_{it} - \theta \bar{\varepsilon}_i)}_{w_{it} - \theta \bar{w}_i}$. $\hat{\beta}_{RE-GLS}$ es consistente si $E[(x_{it} - \theta \bar{x}_i)(w_{it} - \theta \bar{w}_i)] = E[(x_{it} - \theta \bar{x}_i)[(1-\theta)\mu_i + (\varepsilon_{it} - \theta \bar{\varepsilon}_i)]]$. Bajo H_0 , no hay correlación debido a μ_i , y por exogeneidad estricta tampoco hay correlación debido a ε_{it} . Bajo H_0 , $\hat{\beta}_{RE-GLS}$ es consistente. Bajo H_a , es inconsistente.

Estadístico de prueba

$$H = (\Delta\hat{\beta})' (\text{var}\{\Delta\hat{\beta}\})^{-1} (\Delta\hat{\beta}) \sim \chi^2_q$$

$$\Delta\hat{\beta} = \hat{\beta}_{FE} - \hat{\beta}_{RE}$$

Región de rechazo



Hausman demostró que: $\text{Var}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) = \text{Var}(\hat{\beta}_{FE}) - \text{Var}(\hat{\beta}_{RE})$ porque $\text{cov}(\hat{\beta}_{FE}, \hat{\beta}_{RE}) = \text{Var}(\hat{\beta}_{RE})$ dado que RE es el estimador eficiente.

$\hookrightarrow$ Si H no es positiva definida: Puede deberse a que la diferencia en los coeficientes es muy pequeña $\Rightarrow$ Se favorece RE.

MODELOS DE PANEL DINÁMICO

Considere el modelo $y_{it} = y_{it-1}\delta + X_{it}\beta + \mu_i + \varepsilon_{it}$, donde la variable dependiente rezagada ahora es una variable explicativa.

Ejemplo: $y_{it} - y_{it-1} = \underbrace{y_{it-1}}_{\text{Crecimiento}} \alpha + \underbrace{X_{it}\beta}_{\text{PIB}_{t-1}} + \mu_i + \varepsilon_{it}$. Reordenando:

$$y_{it} = y_{it-1}\gamma + X_{it}\beta + \mu_i + \varepsilon_{it} \text{ donde } \gamma = 1 + \alpha$$

Características

1) Endogeneidad inherente

① y_{it} y y_{it-1} dependen de μ_i , y por tanto y_{it-1} está correlacionado con w_{it} .

2) $\neq$ Ex. estricta

② Se viola el supuesto de exogeneidad estricta.

$$\text{Considere } y_{it-1} = y_{it-2}\delta + X_{it-1}\beta + \mu_i + \varepsilon_{it-1}$$

$$\rightarrow y_{it-1} = y_{it-3}\delta^2 + (X_{it-1} + \delta X_{it-2})\beta + (1 + \delta)\mu_i + \varepsilon_{it-1} + \delta\varepsilon_{it-2}$$

y_{it-1} depende de $\varepsilon_{it-1}, \varepsilon_{it-2}, \dots, \varepsilon_{i1}$. Por lo tanto, $E[y_{it-1}\varepsilon_{is}] \neq 0 \forall s < t$.

Soluciones

Posibles soluciones

POLS: Pero incon
RE: Incons.

↳ POLS: Los estimadores son inconsistentes

↳ RE: Requiere $E[y_{it-1}\mu_i] = 0$, pero esto no se cumple. $\Rightarrow$ Los estimadores son inconsistentes.

COMPLEMENTARIA

Panel desbalanceado $\rightarrow$ Es un problema si está asociado a las variables explicativas del modelo.

Para estimar panel: Base debe estar en long.

Pooled OLS: $\rightarrow$ Asume que el individuo en $t \neq$ Individuo en $t+1$

RESUMEN - CLASES 3 & 4

Métodos de estimación en datos panel - Efectos aleatorios

Tenemos que para $y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$, $E[x_{it}\mu_i] = 0 \Rightarrow E[x_{it}\varepsilon_{it}] = 0$

Método de estimación: POLS $\rightarrow$ MCO sobre el modelo $\rightarrow \hat{\beta}_{POLS}$ es ineficiente por autocorrelación

Σ_i (matriz de varianza por individuo) $= \sigma_\varepsilon^2 I_T + \sigma_\mu^2 j_T j_T' = \sigma_\varepsilon^2 I_T + T\sigma_\mu^2 J_T$, $J_T = \frac{1}{T} j_T j_T'$

Para transformar el modelo, se necesita $\Sigma_i^{-1/2} = I_T - \theta J_T$, donde $\theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2}}$
 $\hookrightarrow$ MCGF

Se transforma el modelo $\Sigma_i^{-1/2} y_{it} = \Sigma_i^{-1/2} x_{it} \beta + \Sigma_i^{-1/2} \mu_i + \Sigma_i^{-1/2} \varepsilon_{it}$
 $y_{it}^* = x_{it}^* \beta + w_{it}^*$

$y_{it}^* = [y_{it} - \bar{y}_i, y_{it} - \bar{y}_i, \dots, y_{it} - \bar{y}_i]'$ $\Rightarrow$ Similar a FE-Within en efectos fijos.
 Tarea: Demostrar que $E[w_{it}^* w_{it}^{*'}] = \sigma_\varepsilon^2 I_T$.

Para que sea consistente: $E[(x_{it} - \bar{x}_i)(w_{it} - \bar{w}_i)] = 0$
 $\rightarrow E[x_{it} \bar{w}_i] = 0 \rightarrow E[x_{it} \bar{\varepsilon}_i] = 0 \Rightarrow$ Requiere ex. estricta

MCGF
Procedimiento

1. Estimar $y_{it} = x_{it}\beta + w_{it}$ y obtener $\hat{\beta}_{POLS}$.
2. Obtener $\hat{w}_{it} = y_{it} - x_{it}\hat{\beta}_{POLS}$
3. Obtener $\hat{\sigma}_w^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T \hat{w}_{it}^2}{NT - K}$
4. Obtener $\hat{\sigma}_\mu^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T \sum_{s=R+1}^T \hat{w}_{it} \hat{w}_{is}}{\left\{ \frac{NT(T-1)}{2} - K \right\}}$
5. Obtener $\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_w^2 - \hat{\sigma}_\mu^2$
6. Construir y_{it}^* , x_{it}^* o $\Sigma_i^{-1/2}$ para estimar el modelo corregido.

Modelos de panel con endogeneidad

Asuma que tiene un modelo con efectos fijos ($E[x_{it}\mu_i] \neq 0$) y $E[x_{it}\varepsilon_{it}] \neq 0$. Es necesario instrumentar

FD-IV: Los instrumentos que sean rezagos o combinaciones lineales de x_{it} se eliminan como posibilidades.

$\hookrightarrow$ Sin exogeneidad estricta, sigue siendo consistente.

$\hookrightarrow$ Si $T \rightarrow \infty$, es un método muy favorable.

FE-Within-IV: Requerimos que $E[z_{it}\bar{\varepsilon}_i] = 0 \Rightarrow$ Debe haber exogeneidad estricta en el modelo para que los estimadores sean consistentes.

Pruebas de Hausman

I. Exogeneidad estricta

Obj: Determinar si hay exogeneidad estricta en el modelo comparando los $\hat{\beta}$ de FD con los de FD-IV.

$H_0: E[X_{it} \varepsilon_{it}] = 0 \quad \forall \quad s < t$

- Hay exogeneidad estricta
- FD otorga estimadores consistentes
- FD-IV otorga estimadores consistentes

$H_1: E[X_{it} \varepsilon_{it}] \neq 0 \quad \forall \quad s < t$

- Hay exogeneidad secuencial
- FD otorga estimadores inconsistentes
- FD-IV otorga estimadores consistentes.

Estadístico de prueba $H = (\Delta \hat{\beta})' (\text{var} [\Delta \hat{\beta}])^{-1} (\Delta \hat{\beta}) \sim \chi^2_q$
 $\Delta \hat{\beta} = \hat{\beta}_{FD-IV} - \hat{\beta}_{FD}$, $q = \#$ de variables explicativas.

II. Efectos fijos vs. Efectos aleatorios

Obj: Determinar si un modelo tiene efectos fijos y aleatorios comparando los $\hat{\beta}$ de POLS - GLS y FE-Within.

$H_0: E[X_{it} \mu_i] = 0$

- Hay efectos aleatorios
- POLS - GLS arroja estimadores consistentes
- FE-Within arroja estimadores consistentes

$H_1: E[X_{it} \mu_i] \neq 0$

- Hay efectos fijos
- POLS - GLS arroja estimadores inconsistentes
- FE-Within arroja estimadores consistentes

Estadístico de prueba $H = (\Delta \hat{\beta})' (\text{var} [\Delta \hat{\beta}])^{-1} (\Delta \hat{\beta}) \sim \chi^2_q$
 $\Delta \hat{\beta} = \hat{\beta}_{FE} - \hat{\beta}_{RE}$, donde $\text{var} [\Delta \hat{\beta}] = \text{var} (\hat{\beta}_{FE}) - \text{var} (\hat{\beta}_{RE})$
↳ Si H no es p.d → Se favorece RE

MODELOS DE PANEL DINÁMICOS

Problemas inherentes al panel dinámico

$$y_{it} \begin{cases} \mu_i \\ \varepsilon_{it} \end{cases}$$

Considere $y_{it} = y_{it-1} \delta + x_{it} \beta + w_{it}$. Hay dos problemas inherentes al modelo

- ① y_{it-1} se correlaciona con w_{it} por μ_i .
- ② No hay exogeneidad estricta.

Métodos de efectos fijos

Efectos fijos

FE-Within

Soluciona $\text{Cov}(y_{it}, \mu_i) \neq 0$

Considere $(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i) \beta + (w_{it} - \bar{w}_i) = (x_{it} - \bar{x}_i) \beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$
 $\rightarrow$ Resolvemos el problema ①

Para que sea consistente: $E[y_{it-1} \varepsilon_{it}] = 0$
 $E[(y_{it-1} - \bar{y}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)] = 0$

No cumple exogeneidad estricta

Recordemos que $\bar{y}_i = \frac{1}{T-1} \sum_{t=2}^T y_{it}$, $\bar{\varepsilon}_i = \frac{1}{T-1} \sum_{t=2}^T \varepsilon_{it}$ } Requiere exogeneidad estricta
 Por el rezago, se pierde la primera observación

Sin embargo, este supuesto (exogeneidad estricta) se viola.

Macropanel: $\hat{\beta}$ consistente

- $\rightarrow$ Note que si tuviéramos un macro-panel, $T \rightarrow \infty \Rightarrow \bar{y}_i, \bar{\varepsilon}_i \rightarrow 0$
 $\Rightarrow \hat{\beta}$ son consistentes
- $\rightarrow$ En micro-panel, $\hat{\beta}$ son inconsistentes.

Primeras diferencias

FD

Considere $y_{it} - y_{it-1} = (y_{it-1} - y_{it-2}) \delta + (x_{it} - x_{it-1}) \beta + (\varepsilon_{it} - \varepsilon_{it-1})$
 $\Delta y_{it} = \Delta y_{it-1} \delta + \Delta x_{it} \beta + \Delta \varepsilon_{it}$

Soluciona $\text{Cov}(y_{it}, \mu_i) \neq 0$

$\rightarrow$ Resolvemos el problema ①

Para que sea consistente: $E[\Delta y_{it-1} \Delta \varepsilon_{it}] = 0$
 $E[(y_{it-1} - y_{it-2})(\varepsilon_{it} - \varepsilon_{it-1})] = 0$
 $\rightarrow E[y_{it-1} \varepsilon_{it-1}] = 0 \rightarrow$ No se cumple
 $\rightarrow E[y_{it-1} \varepsilon_{it}] = 0 \rightarrow$ Razonable
 $\rightarrow E[y_{it-2} \varepsilon_{it}] = 0 \rightarrow$ Razonable
 $\rightarrow E[y_{it-2} \varepsilon_{it-1}] = 0 \rightarrow$ Razonable

$\hat{\beta}$ inconsistente
 menos grave que en FE

Por tanto $\hat{\beta}_{FD}$ es inconsistente, pero es menos grave que en FE.

FD-IV

FD-IV (Modelo Anderson - Hsiao)

Soluciona $\text{Cov}(y_{it}, \varepsilon_{it}) = 0$

Busquemos instrumentos para $y_{it-1} - y_{it-2}$. $\begin{cases} y_{it-2}, & T > 3 \\ y_{it-2} - y_{it-3}, & T > 4 \end{cases}$

Para que sea consistente: $E[y_{it-2} (\varepsilon_{it} - \varepsilon_{it-1})] = 0$
 $\rightarrow E[y_{it-2} \varepsilon_{it}] = 0 \rightarrow$ Razonable
 $\rightarrow E[y_{it-2} \varepsilon_{it-1}] = 0 \rightarrow$ Razonable

$\hat{\beta}$ consistente

Por tanto, $\hat{\beta}_{FD-IV}$ es consistente.

Arellano-Bond

A-H: Ineficiente "desperdicia información".

Método generalizado de momentos

↳ Aprovecha toda la muestra

Método Arellano-Bond (1991)

Anderson-Hsiao para $T > 3$ genera estimadores consistentes pero ineficientes ("se desperdicia información".)

↳ En $T = 3$, sólo puedo utilizar y_{i1} como instrumento.

↳ En $T = 4$, puedo usar y_{i1} y y_{i2} como instrumentos.

↳ En $T = T$, puedo usar $y_{i1}, y_{i2}, \dots, y_{i(T-2)}$ como instrumentos

Se pueden demostrar que se tienen $m = (T-2)(T-1)/2$ condiciones de momentos que se pueden usar para estimar los β de manera eficiente. Estos momentos son de la forma $E[y_{it-s} \Delta \epsilon_{it}] = 0 \quad \forall s \geq 2$

→ Método de momentos generalizado $t = 3, \dots, T$

PANEL DESBALANCEADO

Definición
Panel desbalanceado

Considere $y_{it} = X_{it}\beta + \mu_i + \epsilon_{it}$, y sea S_{it} un indicador de selección tal que:

$$S_{it} = \begin{cases} 1 & \text{si } (X_{it}, y_{it}) \text{ se observan} \\ 0 & \text{e.o.c.} \end{cases}$$

Efectos fijos

Método de efectos fijos

$$\ddot{X}_{it} = X_{it} - \frac{1}{T_i} \sum_{t=1}^T S_{it} X_{it}, \quad \ddot{\epsilon}_{it} = \epsilon_{it} - \frac{1}{T_i} \sum_{t=1}^T S_{it} \epsilon_{it} \quad \text{con } T_i = \sum_{t=1}^T S_{it}$$

Consistencia requiere

Exogeneidad estricta

Aparición en panel aleatoria

Para que sea consistente: $E[\ddot{X}_{it} \ddot{\epsilon}_{it}] = 0$. Requiere

↳ Exogeneidad estricta: $E[X_{it} \epsilon_{ip}] = 0 \quad \forall p \neq t$

↳ $E[X_{it} S_{ip}] = 0 \quad \forall t \neq p \rightarrow$ No hay correlación entre aparecer en el panel y las características del individuo $\Leftrightarrow$ Aparecer en el panel es aleatorio.

Si $E[X_{it} S_{ip}] \neq 0$ (muy probable), FE-Within es inconsistente.

Efectos aleatorios

Método de efectos aleatorios

$$\ddot{X}_{it} = X_{it} - \frac{\theta_i}{T_i} \sum_{t=1}^T X_{it} S_{it}, \quad \ddot{w}_{it} = w_{it} - \frac{\theta_i}{T_i} \sum_{t=1}^T w_{it} S_{it}$$

$$\text{Donde } \theta_i = 1 - \sqrt{\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + T_i \sigma_\mu^2}} \quad \Omega \neq I_N \otimes \Sigma$$

Consistencia requiere

Exogeneidad estricta

Para que sea consistente $E[\ddot{X}_{it} \ddot{w}_{it}] = 0$. Esto requiere:

↳ Exogeneidad estricta

↳ $E[X_{it} S_{ip}] = 0 \quad \forall p \neq t$

↳ $E[S_{it} \mu_i] = 0$

Si $E[X_{it} S_{ip}] \neq 0$ y/o $E[S_{it} \mu_i] \neq 0$, $\hat{\beta}_{RE}$ es inconsistente.

FD: No se puede aplicar.

MODELOS NO LINEALES DE PANEL

Definición
Modelos no
lineales

Considere $u_{it} = X_{it}\beta + \mu_i + \varepsilon_{it}$, donde la decisión observable es:

$$d_{it} = \begin{cases} 1 & \text{si } u_{it} > 0 \\ 0 & \text{e.o.c} \end{cases}$$

Caso A: $\mu_i =$
 $\varepsilon_{it} \varepsilon_{is} = 0$

Pool con
logit o
probit

CASO A: No hay efectos individuales ni correlación serial en ε_{it}

Note que no hay efecto individual (μ_i) ni correlación serial $\Rightarrow$ Podemos estimar el modelo por medio de un pool de datos $\Rightarrow$ NT observaciones.

Probit $\rightarrow$ Si $\varepsilon_{it} \sim N(0, \sigma^2)$ $\rightarrow F(X_{it}\beta) = \Phi(X_{it}\beta)$
Logit $\rightarrow$ Si $\varepsilon_{it} \sim \text{V.E tipo I}$ $\rightarrow F(X_{it}\beta) = \Lambda(X_{it}\beta)$ } $\hat{\beta}$ consistente

La función de verosimilitud parcial es:

$$\ln \mathcal{L} = \sum_{i=1}^n \sum_{t=1}^T d_{it} \ln F(X_{it}\beta) + (1-d_{it}) \ln (1-F(X_{it}\beta))$$

Caso B: $\mu_i,$
 $\varepsilon_{it} \varepsilon_{is} = 0$

CASO B: Efectos fijos sin correlación serial en ε_{it}

Recuerde que en efectos fijos no asumimos una distribución paramétrica para μ_i . Note que existe correlación en w_{it} :

Decisiones cor-
relacionadas
por ser del
mismo individuo
 $\Rightarrow$ Probabilidad
conjunta
pero no son ex.
Independientes

$\text{Cov}(w_{it}, w_{is}) = \text{Cov}(\mu_i + \varepsilon_{it}, \mu_i + \varepsilon_{is}) = \text{Var}(\mu_i) = \sigma_\mu^2$, incluso si $\text{Cov}(\mu_i, \varepsilon_{it}) = 0$ y $\text{Cov}(\varepsilon_{it}, \varepsilon_{is}) = 0$.

$\rightarrow$ Las decisiones están correlacionadas a través del tiempo por ser un mismo individuo.

Sea $P(d_i)$ la probabilidad de que un individuo haya elegido una secuencia de alternativas, donde $d_i = (d_{i1}, d_{i2}, \dots, d_{iT})$. Es decir, $P(d_i)$ es una probabilidad conjunta, pero d_{i1} y d_{i2} no son eventos independientes. $P(d_i) \neq P(d_{i1}) * P(d_{i2}) * \dots * P(d_{iT})$

Condicionar
por μ_i

Una alternativa es condicionar por μ_i :

$$P(d_i | \mu, X) = P(d_{i1} | \mu, X) P(d_{i2} | \mu, X) \dots P(d_{iT} | \mu, X)$$

$$P(d_{it} = 1 | X, \mu) = P(X_{it}\beta + \mu_i + \varepsilon_{it} > 0 | \mu, X) \\ = P(\varepsilon_{it} > -X_{it}\beta - \mu_i | \mu, X)$$

Por sup. de simetría $= F(X_{it}\beta + \mu_i)$

$$P(d_{it} = 0 | X, \mu) = 1 - F(X_{it}\beta + \mu_i)$$

Función de
verosimilitud
condicional
Aún depende
de μ_i .

La función de verosimilitud condicional es:

$$\mathcal{L}_c = \sum_{i=1}^n l_{ic}(\mu_i, \beta)$$

$$\text{Donde } l_{ic}(\mu_i, \beta) = \ln P(d_i | \mu_i) = \ln \left\{ \prod_{t=1}^T F(X_{it}\beta + \mu_i)^{d_{it}} (1-F(X_{it}\beta + \mu_i))^{1-d_{it}} \right\}$$

Note que μ_i aún está dentro de F .

Suponga que $F(x_i\beta + \mu_i) = \Phi(x_i\beta + \mu_i)$ (Modelo probit.)

Usualmente, este modelo considera μ_i como parámetros a estimar, incluyendo dummies.

Problemas

Si $N \rightarrow \infty$, est.
 $\mu_i \neq$ consistente
 $\Rightarrow \hat{\beta}$ inconsistente
 por no linealidad
 Perder obs. que
 escogieron una
 sola opción sist.

Soluciones

Condicionar
 logit por un
 estadístico
 mínimo suficiente
 para μ_i

Problemas (Aplican también a logit)

- ↳ Problema de parámetros incidentales: Si $N \rightarrow \infty$, entonces μ_i no se puede estimar consistentemente
- ↳ La inconsistencia de μ_i se transmite a β por la no linealidad del modelo.
- ↳ Si un individuo escogió $(0, 0, \dots, 0) \Rightarrow \hat{\beta}_{MLE} \rightarrow -\infty$, y si escogió $(1, 1, \dots, 1) \Rightarrow \hat{\beta}_{MLE} \rightarrow \infty$.

Soluciones

Logit condicional: Recuerde que requerimos condicionar por μ_i para obtener independencia condicional. Chamberlain planteó que es posible proponer un estadístico mínimo suficiente para μ_i tal que podríamos condicionar la probabilidad por ese estadístico. Chamberlain sugirió:

$$\sum_{t=1}^T d_{it}$$

Considere un modelo con dos períodos, y sea:

$$P(d_{i1} = j, d_{i2} = k \mid \sum_{t=1}^2 d_{it} = j+k, \mu) = \frac{P(d_{i1} = j, d_{i2} = k \mid \mu)}{P(\sum_{t=1}^2 d_{it} = j+k \mid \mu)}$$

Considere $d_{i1} = 0, d_{i2} = 1$.

$$P(d_{i1} = 0, d_{i2} = 1 \mid \sum_{t=1}^2 d_{it} = 1, \mu) = \frac{P(d_{i1} = 0, d_{i2} = 1 \mid \mu)}{P(\sum_{t=1}^2 d_{it} = 1 \mid \mu)} = P(0, 1 \mid \mu) + P(1, 0 \mid \mu)$$

Donde $P(d_{i1} = 0, d_{i2} = 1 \mid \mu) = P(d_{i1} = 0 \mid \mu) * P(d_{i2} = 1 \mid \mu)$

Dado que tenemos un logit:

$$P(d_{i1} = 0 \mid \mu) = \frac{1}{1 + \exp(x_{i1}\beta + \mu_i)}, \quad P(d_{i2} = 1 \mid \mu) = \frac{\exp(x_{i2}\beta + \mu_i)}{1 + \exp(x_{i2}\beta + \mu_i)}$$

$$\Rightarrow P(d_{i1} = 0, d_{i2} = 1 \mid \mu) = \frac{\exp(x_{i2}\beta + \mu_i)}{(1 + \exp(x_{i1}\beta + \mu_i))(1 + \exp(x_{i2}\beta + \mu_i))}$$

$$P(d_{i1} = 1, d_{i2} = 0 \mid \mu) = \frac{\exp(x_{i1}\beta + \mu_i)}{(1 + \exp(x_{i1}\beta + \mu_i))(1 + \exp(x_{i2}\beta + \mu_i))}$$

El denominador será:

El numerador:

$$\frac{\exp(x_{i1}\beta + \mu_i)(\exp(x_{i2}\beta + \mu_i))}{(1 + \exp(x_{i1}\beta + \mu_i))(1 + \exp(x_{i2}\beta + \mu_i))} \quad \frac{\exp(x_{i2}\beta + \mu_i)}{(1 + \exp(x_{i1}\beta + \mu_i))(1 + \exp(x_{i2}\beta + \mu_i))}$$

$$P(d_{i1} = 0, d_{i2} = 1 \mid \sum_{t=1}^2 d_{it} = 1, \mu) = \frac{\exp(\mu_i) \exp(x_{i2}\beta)}{\exp(\mu_i) [\exp(x_{i1}\beta) + \exp(x_{i2}\beta)]}$$

$$P(d_{i1}=0, d_{i2}=1 | \sum d_{it}=1, \mu) = \frac{\exp(x_{i2}\beta)}{\exp(x_{i1}\beta) + \exp(x_{i2}\beta)}$$

Por lo tanto:

$$P(d_{i1}=1, d_{i2}=0 | \sum d_{it}=1, \mu) = \frac{\exp(x_{i1}\beta)}{\exp(x_{i1}\beta) + \exp(x_{i2}\beta)}$$

Pero si:

$$P(d_{i1}=1, d_{i2}=1 | \sum d_{it}=1, \mu) = \frac{P(0,0|\mu)}{P(1,1|\mu)} = 1 \quad \text{Lo mismo ocurre para } P(0,0 | \sum d_{it}=1)$$

β consistente pero ineficiente

Para estas probabilidades, no aparece β : por ende, $\hat{\beta}$ es consistente pero ineficiente.

Función de verosimilitud condicional

La función de verosimilitud condicional tiene la siguiente forma:

$$\mathcal{L}_c = \prod_{i=1}^N P(d_{i1}, d_{i2}, \dots, d_{iT} | \sum_{t=1}^T d_{it} = 1)$$

$$\ln \mathcal{L}_c = \sum_{i=1}^N \ln P(d_{i1}, d_{i2}, \dots, d_{iT} | \sum_{t=1}^T d_{it} = 1)$$

Test de Hausman para EF

Se puede plantear un test de Hausman para determinar si hay o no efectos fijos.

Test de Hausman para efectos fijos

Obj: Comparar los estimadores de pool y Chamberlain

H₀: $\mu = 0$ → POLS es consistente y eficiente
→ Chamberlain es consistente, pero ineficiente

H₁: $\mu \neq 0$ → POLS es inconsistente
→ Chamberlain es consistente

Efectos aleatorios → Asumir distribución paramétrica para μ_i

EFFECTOS ALEATORIOS

Asumimos una distribución paramétrica para μ .

$$\mu_i \sim \text{i.i.d } N(0, \sigma_\mu^2) \\ \varepsilon_{it} \sim \text{i.i.d } N(0, \sigma_\varepsilon^2)$$

Caso C: Probit
sin ε_{it} es $\neq 0$

Caso C: Probit sin correlación serial en ε_{it}

Condicionar por μ_i

Debido a que μ_i está en w_{it} $\Rightarrow \text{Cov}(w_{it}, w_{is}) = \sigma_\mu^2 \neq 0$. Hay correlación en las decisiones: $P(d_i) \neq P(d_{i1}) \cdot P(d_{i2}) \cdots P(d_{iT})$
Por lo tanto: $P(d_i | \mu) = P(d_{i1} | \mu) \cdot \dots \cdot P(d_{iT} | \mu)$

Calcular función de verosimilitud con pr de dist. de μ_i

$$\text{Donde } P(d_i | \mu) = \prod_{t=1}^T F(x_{it}\beta + \mu_i)^{d_{it}} \cdot [1 - F(x_{it}\beta + \mu_i)]^{1-d_{it}}$$

Note que yo he asumido una distribución para μ_i , y por lo tanto, la función de verosimilitud debe construirse con las probabilidades no

condicionales.

"Integrar afuera"

$P(d_i) = \int$ conjunta

Por lo tanto: $P(d_i) = \int_{-\infty}^{\infty} P(d_i | \mu_i) f(\mu_i) d\mu_i$

Esta función proviene de $P(d_i | \mu_i) = \frac{P(d_i, \mu_i)}{f(\mu_i)} \Rightarrow P(d_i, \mu_i) = P(d_i | \mu_i) f(\mu_i)$

Al integrar afuera la expresión $P(d_i, \mu_i)$ con respecto a $\mu_i \rightarrow P(d_i)$.

función de verosimilitud

La función de verosimilitud toma la forma:

$$\begin{aligned} \ln \mathcal{L}(\beta, \sigma_\mu^2, \sigma_\varepsilon^2) &= \sum_{i=1}^N \ln P(d_i) \\ &= \sum_{i=1}^N \ln \int_{-\infty}^{\infty} P(d_i | \mu_i) f(\mu_i) d\mu_i \end{aligned}$$

Dado que estamos en Probit: $F(X_{it}\beta + \mu_i) = \Phi(X_{it}\beta + \mu_i)$.

Métodos de est.

$\hat{\beta}$: $\rightarrow$ Simular

$\rightarrow$ MV Simulada

$\rightarrow$ Cuadratura

Coarussiana

Métodos de estimación para obtener $\hat{\beta}$

$\rightarrow$ Simulaciones para probit

$\rightarrow$ Máxima verosimilitud simulada

$\rightarrow$ Cuadratura Gauss-Hermite (sugerida por Butler y Moffit, 1982)

Caso D: $E_{it}E_{it} \neq 0$

Caso D: Probit con correlación serial en E_{it}

Decisiones no son independientes, ni cond. por μ_i

Considere el modelo $u_{it} = X_{it}\beta + \varepsilon_{it}$ donde $E_{it} = \rho E_{it-1} + \eta_{it}$ con $\eta_{it} \sim N(0, \sigma_\eta^2)$.

$\rightarrow$ Las decisiones no son independientes a lo largo del tiempo, incluso al condicionar por μ_i .

Considere u_{it} en $t=2$: $u_{i2} = X_{i2}\beta + \mu_i + \varepsilon_{i2}$

$$u_{i2} = X_{i2}\beta + \mu_i + \rho \varepsilon_{i1} + \eta_{i2}$$

Condicionar en decisiones pasadas

Dado que las decisiones pasadas generan la correlación: La probabilidad de una secuencia $P(d_i | \mu_i) = P(d_{i1}, \dots, d_{iT} | \mu_i)$, debe condicionar por las decisiones pasadas:

$$P(d_i | \mu_i) = P(d_{i1} | \mu_i) * P(d_{i2} | \mu_i, d_{i1}) * \dots * P(d_{iT} | \mu_i, d_{i1}, \dots, d_{iT-1})$$

Sin E_{it} :

Supongamos que no hay μ_i . Considere dos periodos y la siguiente secuencia $d_{i1}=1$ y $d_{i2}=1$

$$\begin{aligned} P(d_{i1}=1, d_{i2}=1) &= P(d_{i1}=1) * P(d_{i2}=1 | d_{i1}=1) \\ P(1, 1) &= P(1) * P(1|1) \end{aligned}$$

$$P(d_{i1}=1) = P(u_{i1} > 0) = P(X_{i1}\beta + \varepsilon_{i1} > 0) = F(X_{i1}\beta)$$

$$\begin{aligned} P(d_{i2}=1 | d_{i1}=1) &= P(u_{i2} > 0 | u_{i1} > 0) = P(X_{i2}\beta + \varepsilon_{i2} > 0 | X_{i1}\beta + \varepsilon_{i1} > 0) \\ &= P(\varepsilon_{i2} > -X_{i2}\beta | \varepsilon_{i1} > -X_{i1}\beta) \\ &= P(\rho \varepsilon_{i1} + \eta_{i2} > -X_{i2}\beta | \varepsilon_{i1} > -X_{i1}\beta) \\ &= P(\eta_{i2} > -X_{i2}\beta - \rho \varepsilon_{i1} | \varepsilon_{i1} > -X_{i1}\beta) \end{aligned}$$

$$= \int_{-X_{i1}\beta}^{\infty} P(\eta_{i2} > -X_{i2}\beta - \rho \varepsilon_{i1}) f(\varepsilon_{i1} | \varepsilon_{i1} > -X_{i1}\beta) d\varepsilon_{i1}$$

Recuerde que $f(\varepsilon_{it} | \varepsilon_{it} > -x_{it}\beta) = \frac{f(\varepsilon_{it})}{P(\varepsilon_{it} > -x_{it}\beta)} = \frac{f(\varepsilon_{it})}{F(x_{it}\beta)}$

Estas probabilidades se pueden simular con GHK. Recuerde que, con más periodos, las funciones son más complejas.

Con EF:

Matriz de
varianza
de w_t

Ahora, si $\rho \neq 0$. Recuerde que $\varepsilon_{it} \sim \text{i.i.d } N(0, \sigma_\varepsilon^2)$ y $\mu_{it} \sim \text{i.i.d } N(0, \sigma_\mu^2)$. Debemos construir la matriz de varianza-covarianza de w_{it} ; para ello, requerimos que ε_{it} siga un proceso estacionario.

Considere $\varepsilon_{it} = \rho \varepsilon_{it-1} + \eta_{it}$

$$\text{Var}(\varepsilon_{it}) = \text{Var}(\rho \varepsilon_{it-1} + \eta_{it}) = \rho^2 \text{Var}(\varepsilon_{it-1}) + \text{Var}(\eta_{it})$$

$$(1 - \rho^2) \text{Var}(\varepsilon_{it}) = \text{Var}(\eta_{it}) \quad \text{si } \text{Var}(\varepsilon_{it}) = \text{Var}(\varepsilon_{it-1})$$

$$\text{Var}(\varepsilon_{it}) = \frac{\text{Var}(\eta_{it})}{1 - \rho^2}$$

$$\begin{aligned} \text{Para } t=1, s=2: \text{Cov}(w_{i1}, w_{i2}) &= \text{Cov}(\mu_i + \varepsilon_{i1}, \mu_i + \rho \varepsilon_{i1} + \eta_{i2}) \\ &= \text{Cov}(\mu_i + \varepsilon_{i1}, \mu_i, \rho \varepsilon_{i1} + \eta_{i2}) \\ &= \sigma_\mu^2 + \rho \text{Var}(\varepsilon_{i1}) \\ &= \sigma_\mu^2 + \frac{\rho \sigma_\eta^2}{1 - \rho^2} \end{aligned}$$

$$\begin{aligned} \text{Para } t=1, s=3: \text{Cov}(w_{i1}, w_{i3}) &= \text{Cov}(\mu_i + \varepsilon_{i1}, \mu_i + \rho \varepsilon_{i2} + \eta_{i3}) \\ &= \sigma_\mu^2 + \rho \text{Cov}(\varepsilon_{i1}, \varepsilon_{i2}) = \sigma_\mu^2 + \frac{\rho^2 \sigma_\eta^2}{1 - \rho^2} \end{aligned}$$

$$\Sigma = \begin{bmatrix} \sigma_\mu^2 + \frac{\sigma_\eta^2}{1 - \rho^2} & \dots & \sigma_\mu^2 + \frac{\rho^{T-1} \sigma_\eta^2}{1 - \rho^2} \\ \vdots & \ddots & \vdots \\ \sigma_\mu^2 + \frac{\rho^{T-1} \sigma_\eta^2}{1 - \rho^2} & \dots & \sigma_\mu^2 + \frac{\sigma_\eta^2}{1 - \rho^2} \end{bmatrix}$$

Se requiere la descomposición de Cholesky ($\Sigma = TT'$) para construir la matriz de variables transformadas se puede usar β 's por GHK

CENSURAMIENTO

Obs > a
se registran
como a
Problema
de y_i^*

↳ Común en los datos de encuestas (problema de los valores de la variable de interés) cuando las observaciones se registran con un valor (por encima o por debajo de un límite y el resto de las observaciones corresponden a las observaciones específicas de la variable)

⇒ EIH (1990s): Se registraba como 999,999 a todas las observaciones ≥ 1 millón de pesos de salario ("top coding")

Considere el modelo $y_i^* = x_i\beta + \varepsilon_i$ donde y_i^* es una variable latente:

$$y_i^* \sim N(\mu, \sigma^2) \quad \varepsilon_i \sim N(0, \sigma^2)$$

↳ $x_i\beta$

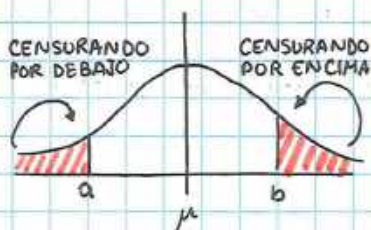
$y_i^* \rightarrow$ Verdadero

$y_i \rightarrow$ Observado
↳ Discreta
en a
↳ Continua
en lo demás

Tenemos que y_i es observable. Esta variable tiene una forma híbrida; una parte discreta y una parte continua

→ Discreta: En el segmento no observado $\rightarrow$ Cuando las observaciones se apilan en el valor de censura.

→ Continua: Cuando las observaciones se registran de manera verdadera



Ejemplos

Censurando por debajo

$$f(y_i | x_i) = \begin{cases} y_i = 0 & \text{si } y_i^* \leq 0 \\ y_i = y_i^* & \text{si } y_i^* > 0 \end{cases}$$

$$P(y_i = 0 | x_i) = P(y_i^* \leq 0 | x_i) = \Phi\left(\frac{0 - \mu}{\sigma}\right) = \Phi\left(\frac{-x_i\beta}{\sigma}\right) = 1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)$$

$$f(y_i | x_i) = \begin{cases} \Phi\left(\frac{-x_i\beta}{\sigma}\right) & \text{si } y_i^* \leq 0 \\ f(y_i^* | x_i) & \text{si } y_i^* > 0 \end{cases}$$

Función de
densidad si
está censurado
por debajo

En general:

$$f(y_i | x_i) = \begin{cases} \Phi\left(\frac{a - x_i\beta}{\sigma}\right) & \text{si } y_i^* \leq 0 \\ f(y_i^* | x_i) & \text{si } y_i^* > 0 \end{cases}$$

Censurando por arriba

$$f(y_i | x_i) = \begin{cases} f(y_i^* | x_i) & \text{si } y_i^* \leq b \\ 1 - \Phi\left(\frac{b - x_i\beta}{\sigma}\right) & \text{si } y_i^* > b \end{cases}$$

Solución:
Tobit

Modelo Tobit

Asumamos que hay censuramiento por debajo en cero.

$$y_i = \begin{cases} y_i = 0 & \text{si } y_i^* \leq 0 \\ y_i = y_i^* & \text{si } y_i^* > 0 \end{cases} \quad y_i^* \sim N(x_i\beta, \sigma^2)$$

Usamos MV para obtener estimadores consistentes de β .

Usa MV y la función de densidad de y_i truncada para est. β .

La función de log-verosimilitud es:

$$\ln \mathcal{L} = \sum_{y_i=0} \ln [1 - \Phi(\frac{x_i\beta}{\sigma})] + \sum_{y_i>0} \ln f(y_i^*|x_i)$$

$$\ln \mathcal{L} = \sum_{y_i=0} \ln [1 - \Phi(\frac{x_i\beta}{\sigma})] + \sum_{y_i>0} -\frac{1}{2} (\ln 2\pi + \ln \sigma^2 + \frac{(y_i - x_i\beta)^2}{\sigma^2})$$

Utilizamos métodos numéricos para obtener $\hat{\beta}$. Podemos usar LR para pruebas de hipótesis, se puede usar R^2 (pseudo) como una medida alternativa de bondad de ajuste.

Ejemplo: Censurando por arriba y por debajo

$$y_i = \begin{cases} y_i = 0 & \text{si } y_i^* \leq 0 \\ y_i = y_i^* & \text{si } y_i^* \in (0, 4) \\ y_i = 4 & \text{si } y_i^* \geq 4 \end{cases}$$

$$f(y_i|x_i) = \begin{cases} P(y_i=0) = P(y_i^* \leq 0|x_i) = 1 - \Phi(x_i\beta/\sigma) \\ P(y_i=y_i^*|x_i) = f(y_i^*|x_i) \\ P(y_i=4) = P(y_i^* \geq 4|x_i) = 1 - \Phi(\frac{4-x_i\beta}{\sigma}) \end{cases}$$

$$\ln \mathcal{L} = \sum_{y_i=0} \ln [1 - \Phi(\frac{x_i\beta}{\sigma})] + \sum_{y_i \in (0,4)} -\frac{1}{2} (\ln 2\pi + \ln \sigma^2 + \frac{(y_i - x_i\beta)^2}{\sigma^2}) + \sum_{y_i=4} \ln [1 - \Phi(\frac{4-x_i\beta}{\sigma})]$$

Media de una v. censurada

Media de una variable censurada (censurando por debajo)

$$y_i = \begin{cases} y_i = a & \text{si } y_i^* \leq a \\ y_i = y_i^* & \text{si } y_i^* > a \end{cases} \quad \left\| \begin{aligned} E(y_i) &= P(y_i=a) E(y_i|y_i=a) + \\ &P(y_i > a) E(y_i|y_i > a) \end{aligned} \right.$$

$$E[y_i] = (1 - \Phi[\frac{a-x_i\beta}{\sigma}] + (-1))a + (1 - \Phi[\frac{a-x_i\beta}{\sigma}]) E(y_i^*|y_i^* > a)$$

Valor esperado de una variable truncada

TRUNCAMIENTO

Falta de información de ciertos individuos

Similar a censuramiento, pero no tenemos acumulación de la distribución en un punto, sino que hay falta de información.

La muestra es un subconjunto de la población; es decir, las variables están disponibles para un subconjunto de características de la población

P.e: ELCA: Representativa de hogares de estratos de 1-4

Banco Mundial: Encuestas a firmas con más de 10 empleados

Considere $y_i = x_i\beta + \varepsilon_i$, pero tenemos una submuestra de la población completa de interés donde $y_i / y_i > a$, donde a es el punto de truncamiento.

Distribución Truncada

Por lo tanto, tenemos una distribución truncada.

Función de densidad de una variable truncada

Tma 1: Si una variable aleatoria continua y tiene una función de densidad de probabilidad $f(y)$ y ' a ' es una constante, la función de densidad de la variable truncada es:

$$f(y|y > a) = \frac{f(y)}{P(y > a)} \quad \left. \begin{array}{l} \text{Reescalar la v. para} \\ \text{que el área bajo la curva} \\ \text{sea} = 1. \end{array} \right\}$$

$$\text{Si } y \sim N(\mu, \sigma^2) \Rightarrow f(y|y > a) = \frac{\frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)} = \frac{\frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right)}{1 - \Phi(\alpha)}$$

Dos primeros momentos de una variable truncada

Tma 2: Los dos primeros momentos de la variable truncada son:

$$E[y|y > a] = \int_a^{\infty} y f(y|y > a) dy = \mu + \sigma \lambda(\alpha)$$

$$\text{Donde } \alpha = \frac{a-\mu}{\sigma} \quad y \quad \lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \left. \begin{array}{l} \text{Inverse Mills} \\ \text{Ratio} \end{array} \right\}$$

$$\text{Var}(y|y > a) = \sigma^2(1 - \lambda(\alpha)) \quad \text{donde } \lambda(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$$

Modelo clásico de truncamiento

Modelo clásico de truncamiento

Considere $y_i = x_i \beta + \varepsilon_i$ donde $\varepsilon_i | x_i \sim N(0, \sigma^2)$, $y_i | x_i \sim N(\mu_i, \sigma^2)$ donde $\mu_i = x_i \beta$.

Sea y_i^* el salario de reserva (no observado). sólo tenemos $y_i > y_i^*$.
Si $y_i > y_i^* \Rightarrow x_i \beta + \varepsilon_i > y_i^* \Rightarrow \varepsilon_i > y_i^* - x_i \beta$

$$\begin{aligned} \text{Entonces } E[y_i | y_i > y_i^*] &= E[x_i \beta | y_i^* - x_i \beta < \varepsilon_i] \\ &= E[x_i \beta | \varepsilon_i > y_i^* - x_i \beta] \\ &= x_i \beta + E[\varepsilon_i > y_i^* - x_i \beta] \end{aligned}$$

$$\text{Si } y_i^* = a: E[y_i | y_i > a] = x_i \beta + E[\varepsilon_i > a - x_i \beta]$$

$$E[y_i | y_i > a] = x_i \beta + \sigma E\left[\frac{\varepsilon_i}{\sigma} > \frac{a - x_i \beta}{\sigma}\right]$$

Aplicamos el teorema de la variable truncada:

$$E[y_i | y_i > a] = x_i \beta + \sigma \frac{\phi\left(\frac{a - x_i \beta}{\sigma}\right)}{1 - \Phi\left(\frac{a - x_i \beta}{\sigma}\right)} = x_i \beta + \sigma \lambda(\alpha)$$

$$\begin{aligned} \text{Considere } v_i \text{ como el error de predicción. } &= y_i | y_i > a - E[y_i | y_i > a] \\ y_i | y_i > a &= E[y_i | y_i > a] + v_i \\ y_i | y_i > a &= x_i \beta + \sigma \lambda(\alpha) + v_i \end{aligned}$$

Si se omite el término $\lambda(\alpha)$, éste va al término de error y generará endogeneidad. Este problema se puede solucionar usando el supuesto de normalidad y la información del límite en " a ".

El log de una función de verosimilitud es:

$$\ln L = \sum_{i=1}^n \left\{ \ln \left[\frac{1}{\sigma} \phi\left(\frac{x_i \beta + y_i}{\sigma}\right) \right] - \ln \left[1 - \Phi\left(\frac{a - x_i \beta}{\sigma}\right) \right] \right\}$$

$\max \ln L \Rightarrow \hat{\beta}$ consistente

SESGO DE SELECCIÓN

Muestra no aleatoria de la población donde hay truncamiento por un conjunto de variables económicas (especialmente en y).

Considere el modelo de interés $y_i = X_i\beta + \varepsilon_i$, donde X_i es exógena.

Considere que la selección está determinada por $Z_i^* = w_i\gamma + u_i$, donde

$$Z_i = \begin{cases} 1 & \text{si } Z_i^* > 0 \\ 0 & \text{si } Z_i^* \leq 0 \end{cases} \quad \text{Suponemos que observamos } w_i, \text{ que contiene } X_i \text{ y otras variables que determinan la selección.}$$

RESUMEN - PANEL DINÁMICO

Definición: $y_{it} = y_{it-1}\gamma + X_{it}\beta + \mu_i + \varepsilon_{it}$

Problemas: $y_{it-1} \begin{cases} \text{Correlacionado con } \mu_i \text{ (efecto fijo de individuo)} \\ y_{it} = f(y_{it-1}) = f(\varepsilon_{it-1}) \Rightarrow \text{Cov}(y_{it}, \varepsilon_{it-1}) \neq 0 \rightarrow \text{Se incumple exogeneidad estricta} \end{cases}$

MODELO	DESCRIPCIÓN	$\text{Cov}(y_{it}, \mu_i) \neq 0$	$\text{Cov}(y_{it}, \varepsilon_{it}) \neq 0$
FE-Within	Estimamos $(y_{it} - \bar{y}_i) = (X_{it} - \bar{X}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$ Para que sea consistente: $E[(y_{it} - \bar{y}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)] = 0$ $\hookrightarrow$ <u>Micro-panel</u> : $\hat{\beta}$ inconsistentes $\hookrightarrow$ <u>Macro-panel</u> : $\bar{y}_i, \bar{\varepsilon}_i \rightarrow 0 \Rightarrow \hat{\beta}$ consistentes	Solucionado restando la media	No se cumple
Primeras diferencias	Estimamos: $y_{it} - y_{it-1} = (y_{it-1} - y_{it-2})\gamma + (X_{it} - X_{it-1})\beta + (\varepsilon_{it} - \varepsilon_{it-1})$ Consistencia requiere $E[(y_{it-1} - y_{it-2})(\varepsilon_{it} - \varepsilon_{it-1})] = 0$ $\hookrightarrow E[y_{it-1} \varepsilon_{it-1}] = 0$ no se cumple $\hookrightarrow \hat{\beta}$ inconsistentes, menos grave que en FE-Within	Solucionado con diferencia	No se cumple
FD-IV (Anderson-Hsiao)	Estimamos el modelo de primeras diferencias instrumentando Δy_{it-1} con y_{it-2} o Δy_{it-2} Consistencia requiere $E[y_{it-2}(\varepsilon_{it} - \varepsilon_{it-1})] = 0$ Se cumple $\hookrightarrow \hat{\beta}$ consistente	Solucionado con diferencia	Solucionado con VI
Método Anellano Bond	FD-IV, consistente pero ineficiente porque no aprovecha toda la muestra (si $T > 3$) $\hookrightarrow$ En $T=3$, sólo y_{it} es instrumento $\hookrightarrow$ En $T=4$, y_{it}, y_{it-1} son instrumentos $\hookrightarrow$ En T , $y_{it}, y_{it-1}, y_{it-2}, \dots, y_{it-T+2}$ son instrumentos $\Rightarrow \exists m = \frac{(T-1)(T-2)}{2}$ condiciones de momentos para generar $\hat{\beta}$ (Método generalizado de momentos)	Solucionado con diferencia	Solucionado con IV

RESUMEN - PANEL DESBALANCEADO

Considere el modelo $y_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. S_{it} es un indicador de selección, tal que:

$$S_{it} = \begin{cases} 1 & \text{si } (x_{it}, y_{it}) \text{ se observan} \\ 0 & \text{en otro caso} \end{cases}$$

MODELO	DESCRIPCIÓN	REQUISITOS CONSISTENCIA
Efectos fijos	Estimamos el modelo $\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{\mu}_i + \ddot{\varepsilon}_{it}$ Donde $\ddot{x}_{it} = x_{it} - \frac{1}{T_i} \sum_{t=1}^{T_i} S_{it} x_{it}$ $\ddot{y}_{it} = y_{it} - \frac{1}{T_i} \sum_{t=1}^{T_i} S_{it} y_{it}$	$E[\ddot{x}_{it}\ddot{\varepsilon}_{it}] = 0$. Para que esto se cumpla: ↳ Exogeneidad estricta ↳ $E[x_{it}S_{it}] \neq 0 \forall t \neq p$: Implica que la aparición en el panel es aleatoria y no depende de las características de los individuos.
Efectos aleatorios	Estimamos el modelo $\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{\varepsilon}_{it}$ Donde $\ddot{x}_{it} = x_{it} - \frac{\theta_i}{T_i} \sum_{t=1}^{T_i} x_{it} S_{it}$ $\ddot{y}_{it} = y_{it} - \frac{\theta_i}{T_i} \sum_{t=1}^{T_i} y_{it} S_{it}$ $\theta_i = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i \sigma_\mu^2}}$	$E[\ddot{x}_{it}\ddot{\varepsilon}_{it}] = 0$. Para que esto se cumpla: ↳ Exogeneidad estricta ↳ $E[x_{it}S_{it}] = 0 \forall t \neq p$: Aparición en el panel es aleatoria. ↳ $E[S_{it}\mu_i] = 0$

RESUMEN - MODELOS NO LINEALES DE PANEL

Considere el modelo $u_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$, donde la decisión observable es:

$$d_{it} = \begin{cases} 1 & \text{si } u_{it} > 0 \\ 0 & \text{o.e.c.} \end{cases}$$

Caso A: Sin efectos individuales, sin correlación serial en ε_{it}

Es posible obtener estimadores consistentes a través de un pool de datos. (obs = NT)

Probit: $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2) \Rightarrow F(x_{it}\beta) = \Phi(x_{it}\beta)$

Logit: $\varepsilon_{it} \sim \text{V.E. tipo I} \Rightarrow F(x_{it}\beta) = \Delta(x_{it}\beta)$

Función de verosimilitud parcial:

$$\ln \mathcal{L} = \sum_{i=1}^N \sum_{t=1}^T d_{it} \ln F(x_{it}\beta) + (1 - d_{it}) \ln (1 - F(x_{it}\beta))$$

Caso B: Efectos fijos, sin correlación serial en ε_{it}

⊛ No asumimos una distribución para μ_i .

Las decisiones están correlacionadas a través del tiempo: $\text{Cov}(w_{it}, w_{is}) = \sigma_\mu^2$

↳ Causa: Es el mismo individuo el que toma las decisiones.

⇒ Para una secuencia de decisiones $(d_{i1}, d_{i2}, \dots, d_{iT})$:

$$P(d_{i1}, d_{i2}, \dots, d_{iT}) \neq P(d_{i1}) * P(d_{i2}) * \dots * P(d_{iT})$$

Solución: Condicionar por μ_i . $\Rightarrow P(d_{i1}, d_{i2}, \dots, d_{iT} | \mu_i, X) = P(d_{i1} | \mu_i, X) * \dots * P(d_{iT} | \mu_i, X)$
 Donde $P(d_{it} | \mu_i, X) = \prod_{t=1}^T P(d_{it} = 1 | \mu_i, X)^{d_{it}} P(d_{it} = 0 | \mu_i, X)^{1-d_{it}}$

$$P(d_{it} = 1 | \mu_i, X) = P(u_{it} = 1 | \mu_i, X) = P(x_{it}\beta + \mu_i + \varepsilon_{it} > 0 | \mu_i, X) \\ = P(\varepsilon_{it} > -x_{it}\beta - \mu_i | \mu_i, X) = F(x_{it}\beta + \mu_i)$$

Sup. simetría

$$\Rightarrow P(d_{it} = 0 | \mu_i, X) = 1 - F(x_{it}\beta + \mu_i)$$

La función de verosimilitud condicional es:

$$\mathcal{L}_c = \sum_{i=1}^N l_{ic}(\mu_i, \beta)$$

Donde $l_{ic}(\mu_i, \beta) = \ln P(d_{it} | \mu_i, X) = \sum_{t=1}^T d_{it} \ln F(x_{it}\beta + \mu_i) + (1-d_{it}) \ln (1 - F(x_{it}\beta + \mu_i))$
 ↳ μ_i : Un parámetro a estimar.

Problemas

- ↳ Si $N \rightarrow \infty$, μ_i no se puede estimar consistentemente $\Rightarrow \hat{\beta}$ tampoco es consistente, por no linealidad
- ↳ Si $d_{it} = (0, 0, \dots, 0) \Rightarrow \hat{\beta} \rightarrow -\infty$
- Si $d_{it} = (1, 1, \dots, 1) \Rightarrow \hat{\beta} \rightarrow \infty$

Soluciones

- ↳ Logit condicional: Chamberlain propone estadístico mínimo suficiente para μ_i y condicionar la probabilidad a ese estadístico. Estadístico propuesto: $\sum_{t=1}^T d_{it}$

Para un modelo de dos periodos:

$$P(d_{i1} = j, d_{i2} = k | \sum_{t=1}^2 d_{it} = j+k, \mu_i, X) = \frac{P(d_{i1} = j, d_{i2} = k | \mu_i, X)}{P(\sum_{t=1}^2 d_{it} = j+k | \mu_i, X)}$$

Sin embargo, si $d_{i1} = d_{i2} = j \Rightarrow P(d_{i1} = j, d_{i2} = k | \sum_{t=1}^2 d_{it} = j+k, \mu_i, X) = 1$
 ↳ También se pierden las observaciones para las que $d_{it} = j \forall j, t \in (1, \dots, T)$
 $\Rightarrow \hat{\beta}$ consistente, pero ineficiente.

función de verosimilitud condicional

$$\ln \mathcal{L}_c = \sum_{i=1}^N \ln P(d_{i1}, d_{i2}, \dots, d_{iT} | \sum_{t=1}^T d_{it} = 1, \mu_i, X)$$

Es posible hacer un test de Hausman para efectos fijos en modelos no lineales, comparando POLS y Chamberlain.

Caso C: Efectos aleatorios, sin correlación serial en ε_{it}

Asumimos una distribución paramétrica para $\mu_i \sim N(0, \sigma_\mu^2)$. Considere el modelo $u_{it} = x_{it}\beta + \mu_i + \varepsilon_{it} \Rightarrow \text{Cov}(u_{it}, u_{is}) = \sigma_\mu^2 \neq 0$. Por ende, para una secuencia de decisiones $(d_{i1}, \dots, d_{iT})$, $P(d_{i1}, \dots, d_{iT}) \neq P(d_{i1}) * \dots * P(d_{iT})$.

Solución: Condicionar por $\mu_i \Rightarrow P(d_{i1}, \dots, d_{iT} | \mu_i, X) = P(d_{i1} | \mu_i, X) * \dots * P(d_{iT} | \mu_i, X)$
 Donde $P(d_{it} | \mu_i, X) = \prod_{t=1}^T P(d_{it} = 1 | \mu_i, X)^{d_{it}} P(d_{it} = 0 | \mu_i, X)^{1-d_{it}}$

↳ Dado que se asumió una función de distribución para μ_i , la función de verosimilitud se construye teniendo en cuenta la distribución de μ_i .

$$P(d_i) = \int_{-\infty}^{\infty} P(d_i, \mu_i) d\mu_i = \int_{-\infty}^{\infty} \frac{P(d_i | \mu_i, X)}{f(\mu_i)} d\mu_i$$

Regla de Bayes

La función de verosimilitud toma la forma:

$$\ln \mathcal{L}(\beta, \sigma_\mu^2, \sigma_\varepsilon^2) = \sum_{i=1}^N \ln P(d_i) = \sum_{i=1}^N \ln \int_{-\infty}^{\infty} \frac{P(d_i | \mu_i, x)}{f(\mu_i)} d\mu_i$$

Esto puede estimarse por medio de

- Máxima verosimilitud simulada
- Cuadratura gaussiana
- Simulaciones para probit

Caso D: Sin efectos fijos, con correlación serial en ε_{it}

Considere el modelo $u_{it} = X_{it}\beta + \mu_i + \varepsilon_{it}$ con $\varepsilon_{it} = \rho \varepsilon_{it-1} + \eta_{it}$ con $\eta_{it} \sim N(0, \sigma_\eta^2)$. Las decisiones en el período t dependen de ε_{it-k} . $\Rightarrow$ Probabilidades condicionadas en μ_i y decisiones pasadas.

$$P(d_i | \mu_i) = P(d_{i1} | \mu_i) * P(d_{i2} | \mu_i, d_{i1}) * \dots * P(d_{iT} | \mu_i, d_{i1}, d_{i2}, \dots, d_{i,T-1})$$

Caso D(1): $\mu = 0$ Considere $T=2$, $d_{i1} = d_{i2} = 1$. $P(d_{i1}=1, d_{i2}=1) = P(d_{i1}=1) * P(d_{i2}=1 | d_{i1}=1)$

$$P(d_{i1}=1) = P(u_{i1} > 0) = P(X_{i1}\beta + \varepsilon_{i1} > 0) = P(\varepsilon_{i1} > -X_{i1}\beta) = F(X_{i1}\beta)$$

$$P(d_{i2}=1 | d_{i1}=1) = P(u_{i2} > 0 | u_{i1} > 0) = P(X_{i2}\beta + \rho \varepsilon_{i1} + \eta_{i2} > 0 | X_{i1}\beta + \varepsilon_{i1})$$

$$= P(\eta_{i2} > -X_{i2}\beta - \rho \varepsilon_{i1} | \varepsilon_{i1} > -X_{i1}\beta)$$

$$= \int_{-X_{i1}\beta}^{\infty} P(\eta_{i2} > -X_{i2}\beta - \rho \varepsilon_{i1}) f(\varepsilon_{i1} > -X_{i1}\beta) d\varepsilon_{i1} \quad \left[\begin{array}{l} \text{Simular con} \\ \text{GHK} \end{array} \right]$$

$$\frac{f(\varepsilon_{i1})}{F(X_{i1}\beta)}$$

Caso D(2): $\mu \neq 0$ Requerimos que ε_{it} siga un proceso estacionario.

$$\text{Var}(\varepsilon_{it}) = \text{Var}(\rho \varepsilon_{it-1} + \eta_{it}) = \rho^2 \text{Var}(\varepsilon_{it-1}) + \text{Var}(\eta_{it})$$

$$\sigma_\varepsilon^2 = \rho^2 \sigma_\varepsilon^2 + \sigma_\eta^2$$

$$(1 - \rho^2) \sigma_\varepsilon^2 = \sigma_\eta^2$$

$$\sigma_\varepsilon^2 = \frac{\sigma_\eta^2}{1 - \rho^2}$$

$$\text{Cov}(w_{i1}, w_{i2}) = \text{Cov}(\mu_i + \varepsilon_{i1}, \mu_i + \varepsilon_{i2})$$

$$= \text{Var}(\mu_i) + \text{Cov}(\varepsilon_{i1}, \varepsilon_{i2})$$

$$= \sigma_\mu^2 + \text{Cov}(\varepsilon_{i1}, \rho \varepsilon_{i1} + \eta_{i2})$$

$$\text{Cov}(w_{i1}, w_{i2}) = \sigma_\mu^2 + \frac{\rho \sigma_\eta^2}{1 - \rho^2}$$

$$\begin{aligned} \text{Cov}(w_{i1}, w_{i3}) &= \text{Var}(\mu_i) + \text{Cov}(\varepsilon_{i1}, \varepsilon_{i3}) = \sigma_\mu^2 + \text{Cov}(\varepsilon_{i1}, \rho^2 \varepsilon_{i1} + \rho \eta_{i2} + \eta_{i3}) = \sigma_\mu^2 + \text{Cov}(\varepsilon_{i1}, \rho^2 \varepsilon_{i1} + \rho \eta_{i2} + \eta_{i3}) \\ &= \sigma_\mu^2 + \rho^2 \sigma_\varepsilon^2 = \sigma_\mu^2 + \frac{\rho^2 \sigma_\eta^2}{1 - \rho^2} \end{aligned}$$

$$(w_{it}) = \text{Cov}(\mu_i + \varepsilon_{it}, \mu_i + \varepsilon_{it}) = \sigma_\mu^2 + \sigma_\varepsilon^2$$

$$\Sigma = \begin{bmatrix} \sigma_\mu^2 + \frac{\sigma_\eta^2}{1 - \rho^2} & \sigma_\mu^2 + \frac{\rho \sigma_\eta^2}{1 - \rho^2} & \dots & \sigma_\mu^2 + \frac{\rho^{T-1} \sigma_\eta^2}{1 - \rho^2} \\ \sigma_\mu^2 + \frac{\rho \sigma_\eta^2}{1 - \rho^2} & \sigma_\mu^2 + \frac{\sigma_\eta^2}{1 - \rho^2} & \dots & \sigma_\mu^2 + \frac{\rho^{T-2} \sigma_\eta^2}{1 - \rho^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 + \frac{\rho^{T-1} \sigma_\eta^2}{1 - \rho^2} & \sigma_\mu^2 + \frac{\rho^{T-2} \sigma_\eta^2}{1 - \rho^2} & \dots & \sigma_\mu^2 + \frac{\sigma_\eta^2}{1 - \rho^2} \end{bmatrix} = T T'$$

Descomposición de Cholesky

SESGO DE SELECCIÓN

Truncamiento Incidental

↳ No se obs. $\forall$ característ.

Truncamiento incidental: No se observan todas las características para un subconjunto de la muestra.

↳ p.e: No observo el salario de las mujeres, pero sí su localización geográfica.

Indicador de selección

$$z_i^* = w_i \delta + u_i$$

Suponga un modelo $y_i = x_i \beta + \varepsilon_i$, donde la selección está determinada por:

$$z_i^* = w_i \delta + u_i \quad \text{t.q.} \quad z_i = \begin{cases} 1 & \text{si } z_i^* > 0 \rightarrow \text{Obs. } y_i \\ 0 & \text{e.o.c.} \rightarrow \text{No obs. } y_i \end{cases}$$

Donde $x_i \subset w_i \rightarrow$ Muestra no aleatoria.

Supuestos

Supuestos

① $x_i, w_i, z_i \forall i$

1. Se observan w_i, x_i, z_i para todos los individuos de la muestra, y x_i y w_i son exógenas.

② $(\varepsilon_i, u_i) \sim NB$

2. ε_i y u_i siguen distribución normal bivariada:

$$\begin{pmatrix} u_i \\ \varepsilon_i \end{pmatrix} \sim NB \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \sigma_\varepsilon \\ \rho \sigma_\varepsilon & \sigma_\varepsilon^2 \end{pmatrix} \right\} \quad \rho = \text{Corr}(u_i, \varepsilon_i)$$

Note que $\text{Cov}(\varepsilon_i, u_i) \neq 0 \rightarrow$ Choques exógenos de la decisión de trabajar afectan el salario.

V.E. de una var. truncada

Calculamos el valor esperado de una variable truncada.

$$\begin{aligned} E[y_i | y_i \text{ se observa}] &= E[y_i | z_i = 1] \\ &= E[y_i | z_i^* > 0] \\ &= E[y_i | w_i \delta + u_i > 0] \\ &= E[y_i | u_i > -w_i \delta] \\ &= E[x_i \beta + \varepsilon_i | u_i > -w_i \delta] \\ E[y_i | y_i \text{ se observa}] &= x_i \beta + E[\varepsilon_i | u_i > -w_i \delta] \end{aligned}$$

Truncamiento de y_i se traduce en truncamiento de ε_i .

Truncamiento de $y_i \Rightarrow$ trunc. de ε_i

Momentos de NB incid.

Tma: Momentos de la distribución normal bivariada incidental

Considere dos variables $\varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$ $u \sim N(\mu_u, \sigma_u^2)$, $\text{Corr}(\varepsilon, u) = \rho$.

$$\begin{aligned} E[\varepsilon | u > a] &= \mu_\varepsilon + \rho \sigma_\varepsilon \lambda(\alpha_u) \\ \text{Var}[\varepsilon | u > a] &= \sigma_\varepsilon^2 (1 - \rho^2 \lambda(\alpha_u)) \end{aligned}$$

$$\text{Donde } \alpha_u = \frac{a - \mu_u}{\sigma_u}, \quad \lambda(\alpha_u) = \frac{\phi(\alpha_u)}{1 - \Phi(\alpha_u)}, \quad \lambda(\alpha_u) = \lambda(\alpha_u) [\lambda(\alpha_u) - \alpha_u]$$

Por tanto:

$$\begin{aligned} E[y_i | y_i \text{ se observa}] &= x_i \beta + E[\varepsilon_i | u_i > -w_i \delta] \\ &= x_i \beta + \rho \sigma_\varepsilon \lambda(\alpha_u) \\ &= x_i \beta + \rho \sigma_\varepsilon \lambda(w_i \delta) \end{aligned}$$

Casos

Selección
exógena

↳ MCO

↳ Lotería

Selección
endógena

↳ Sesgo de
especificación
por v. omitida

Soluciones

① Controles

② Heckit

Est $\hat{\gamma}$ → Probit

$W_i \neq X_i$

Casos del modelo de selección

(A) Selección de muestra exógena

Si $p = 0 \Rightarrow E[y_i | z_i = 1] = X_i \beta \Rightarrow$ Estimamos por MCO.

Consideremos un ejemplo en el que la selección de la muestra se lleva a cabo por medio de una lotería; por tanto, la selección de la muestra es independiente de factores tanto observables como no observables.

(B) Selección de muestra endógena

Si $p \neq 0 \Rightarrow y_i = X_i \beta + p \sigma_\varepsilon \lambda(W_i \hat{\gamma}) + v_i \rightarrow y_i$ truncada

$\lambda(W_i \hat{\gamma})$ no es observable directamente $\Rightarrow$ Esta variable relevante se omite si implementamos MCO.

↳ Sesgo de selección $\subset$ Sesgo de especificación por v. omitida.

Soluciones

1. Inclusión de variables de control

Las variables que generan la selección se incluyen en la regresión, en caso de ser observables.

2. Heckit: Método de Heckman

Dado que se conoce la forma funcional de λ , pero no se conoce $\hat{\gamma}$, la sugerencia de Heckman es estimar $\hat{\gamma}$ usando un probit.

(a) Usar todas las observaciones para estimar el modelo:
 $z_i = W_i \hat{\gamma} + u_i$. Obtener $\hat{\gamma}$, y computar:
(PROBIT)

$$\lambda(W_i \hat{\gamma}) = \frac{\phi(W_i \hat{\gamma})}{\Phi(W_i \hat{\gamma})}$$

(b) Usar las observaciones de la muestra de selección y estimar $y_i = X_i \beta + p \sigma_\varepsilon \lambda(W_i \hat{\gamma}) + v_i$

Importante: Para la primera etapa, debe haber una o más variables que expliquen z_i , pero no a y_i directamente.

↳ Si $X_i = W_i$ hay multicolinealidad. Incluso, si la primera etapa se estima por un MPL $\Rightarrow$ Habrà multicolinealidad perfecta.

Más aún, es necesario calcular correctamente los errores estándar para incorporar la incertidumbre de la primera etapa. Los paquetes estadísticos ya hacen esta corrección.

Sea $\beta_1 = p \sigma_\varepsilon$. Se puede realizar una prueba de hipótesis para verificar si es necesario corregir por sesgo de selección.

$H_0: \beta_1 = 0 \rightarrow$ No se requiere corrección

$H_1: \beta_1 \neq 0 \rightarrow$ Se requiere corrección

EVALUACIÓN DE IMPACTO

Objetivo: Medir el impacto del tratamiento (p.e: un programa) sobre unas variables de resultado en un conjunto de individuos.

Problema: Efecto del tratamiento $\rightarrow \Delta$ entre la variable de resultado de los participantes con programa y la de los participantes en ausencia del programa (Contrafactual).

Notación

$$D_i = \begin{cases} 1 & \text{si el individuo recibe tratamiento} \\ 0 & \text{e.o.c} \end{cases}$$

Grupo de tratamiento

Grupo de individuos participantes en el programa.

Grupo de control

Grupo de individuos que no participan en el programa.

$Y_i(D_i)$: Variable de resultado para cada individuo i dado su estado D_i .
 $\rightarrow Y_i(1)$: Variable de resultado para individuo que fue tratado.
 $\rightarrow Y_i(0)$: Variable de resultado para individuo que no fue tratado.

El efecto del tratamiento sobre el individuo i estaría dado por:

$$\tau_i = Y_i(1) - Y_i(0)$$

Pero sólo se puede observar $Y_i(1)$ o $Y_i(0)$ para el individuo i . El resultado efectivamente observado es:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

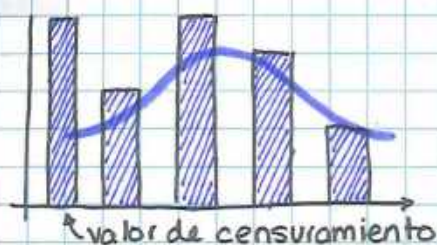
Se debe encontrar el impacto promedio del programa en la población.

$$TATE = E(\tau_i) = E[Y_i(1) - Y_i(0)] \rightarrow \Delta \text{ de } Y_i \text{ cuando un individuo pasa aleatoriamente de participar a no participar.}$$

Average Treatment Effect

RESUMEN: CENSURAMIENTO

Problema: Los valores de y_i por encima o por debajo de un valor de censura a , no se registran como los verdaderos valores de una variable, sino como a .



Considere el modelo $y_i^* = x_i \beta + w_i$ donde y_i^* es la variable real. Tenemos y_i , la variable observada censurada en a . Asumimos $y_i^* \sim N(\mu, \sigma^2)$, donde $\mu = x_i \beta$

Ejemplo: Censuramiento por debajo

$$y_i = \begin{cases} a & \text{si } y_i^* \leq a \\ y_i^* & \text{si } y_i^* > a \end{cases}$$

$$P(y_i = a | x_i) = P(y_i^* \leq a | x_i) = P(x_i \beta + w_i \leq a | x_i) = P(w_i \leq a - x_i \beta | x_i)$$

$$P(y_i = a | x_i) = P\left(\frac{w_i}{\sigma} \leq \frac{a - x_i \beta}{\sigma} | x_i\right) = \Phi\left(\frac{a - x_i \beta}{\sigma}\right)$$

$$P(y_i = y_i^* | x_i) = P(y_i^* > a | x_i) = P(x_i \beta + w_i > a | x_i) = P(w_i > a - x_i \beta | x_i)$$

$$P(y_i = y_i^* | x_i) = P\left(\frac{w_i}{\sigma} > \frac{a - x_i \beta}{\sigma} | x_i\right) \Rightarrow f(y_i^* | x_i)$$

$$\text{Por lo tanto: } f(y_i | x_i) = \begin{cases} \Phi\left(\frac{a - x_i \beta}{\sigma}\right) & \text{si } y_i^* \leq a \\ f(y_i^* | x_i) & \text{si } y_i^* > a \end{cases}$$

Ejemplo: Censurando por arriba

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* < b \\ b & \text{si } y_i^* \geq b \end{cases} \Rightarrow P(y_i = b | x_i) = P\left(\frac{w_i}{\sigma} > \frac{b - x_i \beta}{\sigma} | x_i\right) = 1 - \Phi\left(\frac{b - x_i \beta}{\sigma}\right)$$

$$\text{Por lo tanto: } f(y_i | x_i) = \begin{cases} f(y_i^* | x_i) & \text{si } y_i^* < b \\ 1 - \Phi\left(\frac{b - x_i \beta}{\sigma}\right) & \text{si } y_i^* \geq b \end{cases}$$

Solución: Modelo Tobit

Este método considera la función de densidad censurada para plantear una función de verosimilitud. Considere el caso del censuramiento por arriba.

$$\ln \mathcal{L} = \sum_{y_i < b} \ln [f(y_i^* | x_i)] + \sum_{y_i = b} \ln \left[1 - \Phi\left(\frac{b - x_i \beta}{\sigma}\right) \right]$$

Se utilizan métodos numéricos para estimar β . La interpretación de los coeficientes

Media de una variable censurada

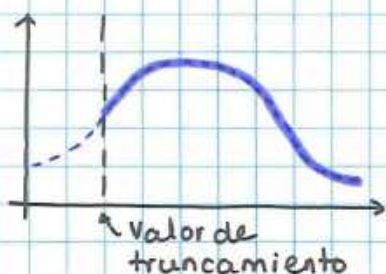
Considere el caso del censuramiento por arriba.

$$E[y_i] = P(y_i = b) E[y_i | y_i = b] + P(y_i < b | x_i) E[y_i | y_i < b]$$

$$E[y_i] = (1 - \Phi\left(\frac{b - x_i \beta}{\sigma}\right)) b + f(y_i^* | x_i) E[y_i^* | y_i^* < b]$$

RESUMEN: TRUNCAMIENTO

Problema A partir de un valor de y_i , no se registran todas o ninguna de las características del individuo.



Considere el modelo $y_i = X_i\beta + \varepsilon_i$. La variable y_i tiene una distribución truncada.

Tma 1: Si una variable aleatoria continua y tiene una función de densidad de probabilidad $f(y)$ y ' α ' es una constante, la función de densidad de la variable truncada es:

$$f(y|y > a) = \frac{f(y)}{P(y > a)} = \frac{\frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right)}{1 - \Phi(\alpha)}; \quad y_z = \frac{y-\mu}{\sigma}, \quad \alpha = \frac{a-\mu}{\sigma}$$

Tma 2: Los dos primeros momentos de la variable truncada son:

$$E[y|y > a] = \int_a^{\infty} y f(y|y > a) dy = \mu + \sigma \lambda(\alpha); \quad \lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \left. \begin{array}{l} \text{Inverse} \\ \text{Hills} \\ \text{Ratio} \end{array} \right\}$$

$$\text{Var}(y|y > a) = \sigma^2(1 - \lambda(\alpha)) \quad \text{donde} \quad \lambda(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$$

Considere el modelo $y_i = X_i\beta + \varepsilon_i$; $y_i|x_i \sim N(X_i\beta, \sigma^2)$ y $\varepsilon \sim N(0, \sigma^2)$

$$E[y_i | y_i > a] = E[X_i\beta | X_i\beta + \varepsilon_i > a] = E[X_i\beta | \varepsilon_i > a - X_i\beta]$$

Normalizando: $E[y_i | y_i > a] = X_i\beta + \frac{\sigma}{\sigma} E[\varepsilon_i | \varepsilon_i > a - X_i\beta]$

$$= X_i\beta + \sigma E[\varepsilon_i/\sigma | \varepsilon_i/\sigma > \frac{a - X_i\beta}{\sigma}]$$

Aplicando el tma 1: $E[y_i | y_i > a] = X_i\beta + \frac{\sigma \phi(\alpha)}{1 - \Phi(\alpha)} = X_i\beta + \sigma \lambda(\alpha)$

Sea v_i el error de predicción: $v_i = [y_i | y_i > a] - E[y_i | y_i > a]$

$$[y_i | y_i > a] = v_i + E[y_i | y_i > a]$$

$$y_i | y_i > a = v_i + X_i\beta + \sigma \lambda(\alpha)$$

$$= X_i\beta + v_i + \sigma \lambda(\alpha) \rightarrow \text{Si se omite, genera endogeneidad.}$$

Solución: Se plantea la siguiente función de verosimilitud:

$$\ln \mathcal{L} = \sum_{i=1}^N \ln \lambda_i(y_z, \alpha) = \sum_{i=1}^N \left\{ \ln \left[\frac{1}{\sigma} \phi\left(\frac{y - X_i\beta}{\sigma}\right) \right] - \ln [1 - \Phi(\alpha)] \right\}$$

Al maximizar $\mathcal{L}$, se obtienen $\hat{\beta}$ consistentes.

RESUMEN: SESGO DE SELECCIÓN

Truncamiento incidental No se observan todas las características para un subconjunto de la población.

Suponga un modelo $y_i = X_i\beta + \varepsilon_i$, donde la selección se determina según el modelo:

$$z_i^* = W_i\delta + u_i \xrightarrow{\quad} z_i = \begin{cases} 1 & \text{si } z_i^* > 0 \\ 0 & \text{e.o.c.} \end{cases}$$

Donde $X_i \subset W_i$.

Suponemos que: $\rightarrow X_i, z_i$ y W_i se observan $\forall i$.

$$\rightarrow \begin{pmatrix} u_i \\ \varepsilon_i \end{pmatrix} \sim NB \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{pmatrix} \right\}; \rho = \text{Corr}(u_i, \varepsilon_i)$$

Calculamos el valor esperado de y_i (una variable truncada):

$$E[y_i | y_i \text{ se observa}] = E[y_i | z_i^* > 0] = E[y_i | W_i\delta + u_i > 0]$$

$$= E[y_i | u_i > -W_i\delta]$$

$$\text{Reemplazando } y_i = E[X_i\beta + \varepsilon_i | u_i > -W_i\delta]$$

$$= E[X_i\beta | u_i > -W_i\delta] + E[\varepsilon_i | u_i > -W_i\delta]$$

$$E[y_i | y_i \text{ se observa}] = X_i\beta + \underbrace{E[\varepsilon_i | u_i > -W_i\delta]}$$

Truncamiento de ε_i

Tma: Momentos de la distribución normal bivariada incidental

Considere $\varepsilon_i \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$ y $u_i \sim N(\mu_u, \sigma_u^2)$; $\text{Corr}(\varepsilon_i, u_i) = \rho$.

$$E[\varepsilon | u > a] = \mu_\varepsilon + \rho \sigma_\varepsilon \lambda(\alpha_u)$$

$$\text{Var}[\varepsilon | u > a] = \sigma_\varepsilon^2 (1 - \rho^2 \lambda(\alpha_u))$$

$$\text{Donde } \alpha_u = \frac{a - \mu_u}{\sigma_u}; \lambda(\alpha_u) = \frac{\phi(\alpha_u)}{1 - \Phi(\alpha_u)}; \lambda(\alpha_u) = \lambda(\alpha_u)[\lambda(\alpha_u) - \alpha_u]$$

$$\Rightarrow E[y_i | y_i \text{ se observa}] = X_i\beta + \rho \sigma_\varepsilon \lambda\left(\frac{-W_i\delta}{\sigma_u}\right) = X_i\beta + \rho \sigma_\varepsilon \lambda(-W_i\delta)$$

Note que si $\rho = 0$: $E[y_i | y_i \text{ se observa}] = X_i\beta \rightarrow$ Estimamos MCO

Pero si $\rho \neq 0$: $E[y_i | y_i \text{ se observa}] = X_i\beta + \rho \sigma_\varepsilon \lambda(W_i\delta) \rightarrow$ Estimar MCO omite $\rho \sigma_\varepsilon \lambda(W_i\delta)$ (v. relevante) porque no es observable.

Soluciones

- ① Incluir en la regresión las variables que generan la selección.
- ② Heckit: Estimar δ usando todas las observaciones en el modelo de $z_i = W_i\delta + u_i \rightarrow$ Probit. Usar $\hat{\delta}$ para computar $\lambda(W_i\hat{\delta})$. Incluirlo en el modelo: $y_i = X_i\beta + \beta\lambda(W_i\hat{\delta}) + \varepsilon_i$
 - $\rightarrow$ Requiere al menos un W que no determine y , t.q $W_i \neq X_i$. De lo contrario, habrá multicolinealidad
 - $\rightarrow$ Los e.e. requieren calcularse diferente para incorporar la incertidumbre de la primera etapa.
 - $\rightarrow$ Si $\beta\lambda$ es no significativo $\Rightarrow$ Corrección era innecesaria.

CAPÍTULO I: INTRODUCCIÓN

Problema central de la E.I:

Encontrar el contrafactual

Pregunta central: ¿qué habría pasado con los beneficiarios en ausencia del programa?

CAPÍTULO II: DEF. PARÁMETROS DE IMPACTO DEL TRATAMIENTO

Obj: Encontrar valor de Y con y sin tratamiento para el mismo individuo

↳ Contrafactual: Resultado hipotético del participante

Variable de resultado: Variables sobre las cuales se espera que el programa tenga un impacto en los beneficiarios.

↳ Obj: Establecer Δ entre Y_i en presencia del programa, Y_i en ausencia del programa para el mismo i → Efecto del tratamiento

↳ Requiere contrafactual: Resultado hipotético del i participante si el programa no existiera.

Efecto del tratamiento

$$T_i = Y_i(1) - Y_i(0)$$

Average treatment effect

$$TATE = E[T_i] = E[Y_i(1) - Y_i(0)]$$

Se puede representar en un modelo de regresión lineal:

$$Y_i = \beta_0 + T_i D_i + u_i$$

$$\Rightarrow TATE = E[Y_i(1)] - E[Y_i(0)] = \beta_0 + T_i + u_i - (\beta_0 + u_i) = T_i$$

↳ Interpretación: Δ promedio en Y_i cuando un individuo escogido al azar pasa aleatoriamente de participar a no participar.

$$TATE = E[Y_i(1) - Y_i(0)]$$

↳ β_i en MCO

↳ Δ promedio en Y_i cuando i aleatorio pasa de participar a no participar

ATT: Parámetro más importante

↳ Requiere $E[Y_i(0) | D_i = 1]$

Average Treatment on Treated

$$T_{ATT} = E[T_i | D_i = 1] = E[Y_i(1) | D_i = 1] - \underbrace{E[Y_i(0) | D_i = 1]}_{\text{contrafactual}}$$

↳ Parámetro de mayor interés en eval. de impacto

↳ Efecto promedio del programa sobre los tratados.

Average Treatment on Untreated

$$T_{ATU} = E[T_i | D_i = 0] = \underbrace{E[Y_i(1) | D_i = 0]}_{\text{contrafactual}} - E[Y_i(0) | D_i = 0]$$

$$E[Y_i(0) | D_i = 1]$$

se puede reemplazar con

$$E[Y_i(0) | D_i = 0]$$

⇒ Son iguales

¿Cómo estimar el contrafactual?

Se puede usar $E[Y_i(0) | D_i = 0]$ para aproximar $E[Y_i(0) | D_i = 1]$. Considere el T_{ATT} :

$$T_{ATT} = E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 1]$$

$$E[Y_i(1) | D_i = 1] = T_{ATT} + E[Y_i(0) | D_i = 1]$$

$$E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 0] = T_{ATT} + E[Y_i(0) | D_i = 1] - E[Y_i(0) | D_i = 0]$$

$$T_{ATT} = E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 0]$$

$$\Leftrightarrow E[Y_i(0) | D_i = 1] - E[Y_i(0) | D_i = 0] = 0$$

Si hay igualdad
 $\Rightarrow$ Características
 de i no determinan
 participación en
 el tratamiento

$E[Y_i(0)|D_i=1] = E[Y_i(0)|D_i=0]$ es un supuesto fuerte. Implica que no hay sesgo de autoselección $\Leftrightarrow$ Las características de los individuos no determinan la participación en el programa.

$\rightarrow$ Si se incumple, $Y_i = \beta_0 + \tau_i D_i + u_i$ tiene endogeneidad porque u_i contiene las características de los individuos.

$\rightarrow$ Si se cumple, $\hat{\tau}_i$ estimado por MCO es insesgado y consistente.

$$\hat{\tau}_i = [\bar{Y}|D=1] - [\bar{Y}|D=0]$$

Importante: Elección de la variable de resultado no es trivial.

CAPÍTULO 3: SESGO DE SELECCIÓN

Sesgo de selección
 Participación es
 elección de i

Ppal reto de la evaluación de impacto: Determinar cuándo se cumple que $E[Y_i(0)|D_i=1] = E[Y_i(0)|D_i=0]$

$\rightarrow$ Se viola si la participación en el programa es elección de los individuos.

Sesgo de selección

Si se incumple: $\tau_{ATT} + E[Y_i(0)|D_i=1] - E[Y_i(0)|D_i=0] = E[Y_i(1)|D_i=1] - E[Y_i(0)|D_i=0]$

Si características
 que determinan
 selección son
 observables, se
 incluyen en la
 regresión

Si los determinantes de la selección son observables, se deben incluir en la regresión.

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \epsilon_i$$

Sin embargo, si no son observables, $\hat{\beta}_1$ será sesgado de la siguiente manera.

	$Cov(D_i, X_i) > 0$	$Cov(D_i, X_i) < 0$
$\beta_2 > 0$	$E[\hat{\beta}_1] > \beta_1$	$E[\hat{\beta}_1] < \beta_1$
$\beta_2 < 0$	$E[\hat{\beta}_1] < \beta_1$	$E[\hat{\beta}_1] > \beta_1$

CAPÍTULO 4: EXPERIMENTOS ALEATORIOS (SOCIALES)

Experimento
 social controlado
 $\Rightarrow$ Asignación
 aleatoria.

Cumple igualdad
 $E[Y_i(0)|D_i=1] =$
 $E[Y_i(0)|D_i=0]$

Experimento social controlado: Asignación aleatoria del tratamiento, de tal manera que los participantes no puedan elegir.

$\rightarrow$ De esta forma, se cumple que $E[Y_i(0)|D_i=0] = E[Y_i(0)|D_i=1]$.

Es decir, los grupos de tratamiento y control son idénticos ex ante.

$\rightarrow$ En la práctica, se puede hacer:

- Evaluando un piloto elegido aleatoriamente
- Incentivando a un subgrupo aleatorio de la población a participar.
 - No todo el subgrupo participará, y agentes externos al subgrupo participarán.
 - Haber recibido el incentivo se usa como instrumento de la variable D_i .

Si se incentiva
 a un subgrupo
 aleatorio $\Rightarrow$
 Instrumentar
 D_i con haber
 recibido incentivo

Conglomerados
> Individuos
si: Spillovers,
todos quieren.

Pero $N \gg$ para
que haya signifi-
ficancia.

Estimador de
diferencias

$\tau_{\pi} = \Delta Y$ si
 $D=1$ o si $D=0$
⇒ Si hay una alea-
torización
 β_1 en MCO

Regresores adicio-
nales mejoran
la eficiencia

Efectos hetero-
géneos: Impacto
del programa va-
ría según carac-
terísticas de los
individuos

Se incorporan
interacciones

Si no hay grupo
de control:
 Δ en el tiempo
de tratamiento

4.1. Intervenciones a nivel individual vs. conglomerados

Se asigna aleatoriamente la participación en el programa a nivel de barrios o comunidades.

Ventajas

- Adecuada si la intervención tendrá efectos sobre toda la comunidad. (spillovers)
- Es buena alternativa si es difícil negar los beneficios de una intervención a un segmento de la comunidad.

Desventajas

- El tamaño de muestra debe ser mayor que en el caso en el que la unidad de observación es el individuo, para que los efectos estimados sean significativos.

4.2. El modelo de diferencias

Un buen experimento social garantiza $E[Y_i(1)|D=0] = E[Y_i(0)|D=0]$. Si D_i es binario, el efecto del tratamiento puede estimarse como una diferencia de medias entre los grupos de tratamiento y control después del tratamiento.

Estimador de diferencias

$$\hat{\tau} = \hat{\beta}_1 = [\bar{Y}|D_i=1] - [\bar{Y}|D_i=0]$$

- ↳ Genera estimaciones insesgadas ⇒ Validez interna
- ↳ No es necesariamente eficiente.

4.3. Estimador de diferencias con regresores adicionales

Se puede mejorar la eficiencia del estimador de diferencias incluyendo características individuales. Esto mantiene el insesgamiento y la consistencia.

4.4. Estimador de diferencias con efectos heterogéneos

Ayuda a capturar los efectos de la intervención si difieren según las características del individuo. Considere una variable binaria X_i y el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 D_i X_i + \varepsilon_i$$

$$\text{Si } X_i = 1: E[Y_i(1)|D_i=1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

$$- E[Y_i(0)|D_i=0] = \beta_0 + \beta_2$$

$$\hat{\tau}_{\pi} = \beta_1 + \beta_3 \rightarrow \text{¿Qué tanto mejor es el trata- miento si } X_i = 1.$$

$$\text{Si } X_i = 0: E[Y_i(1)|D_i=1] = \beta_0 + \beta_1$$

$$- E[Y_i(0)|D_i=0] = \beta_0$$

$$\hat{\tau}_{\pi} = \beta_1$$

4.5. Estimador de diferencias en el tiempo

Calcula el impacto de estar expuesto como la diferencia en la variable de interés antes y después de la implementación del programa.

- ↳ Para el grupo de tratamiento únicamente.

Definimos $T_i = \begin{cases} 1 & \text{en el período de tratamiento} \\ 0 & \text{antes del tratamiento.} \end{cases}$

Estimador de Δ en el tiempo

$$\hat{\tau} = (\bar{y}_1 | D_i = 1) - (\bar{y}_0 | D_i = 0)$$

$y_i = \beta_0 + \beta_1 T_i + u_i$ sobre el segmento de la muestra que tenga $D_i = 1$.

Permite evaluar programas universales

Incorporar una variable $t = 1, 2, \dots, T$ para que β_1 no capture evoluciones naturales de variables en el tiempo.

Ventajas

Desventajas

- Permite evaluar programas universales.
- No requiere un grupo de control
- Algunas variables de resultado tienen una tendencia en el tiempo que quedaría capturada en T_i .

Solución: Generar una variable $t = \{1, 2, \dots, T\}$.

Estimar $y_i = \beta_0 + \beta_1 T_i + \beta_2 t_i + \varepsilon_i$

↳ Asume que los aumentos en y_i a través del tiempo.

4.6. Problemas potenciales de la aleatorización

Problemas con aleatorización

Problemas de aleatorización → Inferencia no es válida para población bajo estudio.

Efecto Hawthorne / John Henry
 $N \rightarrow 0 \Rightarrow IC$
no se cumple
Autoselección
Incumplimiento del protocolo

- Comportamiento cambia al participar en un experimento (efectos Hawthorne o John Henry).
- Tamaño de muestra es pequeño $\Rightarrow$ No garantiza I.C.
- Los que ganan la lotería deciden no participar, y los que no ganan pelean hasta que los meten.
- Incumplimiento del protocolo de tratamiento: la intención era aleatoria, el tratamiento no lo es.

4.7. Verificación de la aleatorización

Verificar aleatorización
 $D_i = X_i \alpha + \varepsilon_i$
 α no debe ser significativo

Considere el modelo $y_i = \beta_0 + \beta_1 D_i + u_i$. Si la aleatorización fue exitosa $\text{Corr}(D_i, X_i) = 0$. Esto se prueba con el estadístico F de la regresión $D_i = X \alpha + \varepsilon$, donde ε son errores estándar robustos.

4.8. Desventajas de los métodos experimentales

Ventajas
Técnicas económ.
sencillas $\Rightarrow$ Transparencia
No hay elección estratégica

Ventajas

- Técnicas de estimación son sencillas $\Rightarrow$ Transparencia
- No hay cómo elegir estratégicamente la técnica de modelación para manipular los resultados.

Desventajas
Costoso
Antiético
Ext. y efectos de GE $\leftrightarrow$ Validez externa + G. control
Difícil determinar contrafactual

Desventajas

- Costoso monetario y políticamente
- Injusticia fundamental al negarle el tratamiento a otro.
- Las externalidades (efectos externos del tratamiento) y los efectos de equilibrio general (efectos del tratamiento sobre el entorno) pueden afectar al grupo de control.
- Es difícil determinar el contrafactual.
- Validez externa sólo se da si la muestra es suficientemente similar en relación a la población.
 - ↳ Replicabilidad al escalar
 - ↳ Especificidades regionales
 - ↳ Corta duración

Experimento natural

Evento natural que genera aleatorización en factor no relacionado con relación causal.

Asignación puede ser:

- Completamente exógena
- Parcialmente exógena

CAPÍTULO 5: EXPERIMENTOS NATURALES O CUASI-EXPERIMENTOS

Experimento natural Un evento fortuito ocasiona una asignación entre tratamiento y control similar a la que generaría un experimento aleatorio controlado.

↳ Aleatoriedad en un factor no relacionado con el efecto causal que se estudia.

↳ Dos tipos de experimento

(1) Asignación en el grupo de tratamiento es completamente aleatoria.

- Herramientas metodológicas idénticas a los experimentos aleatorios controlados.

- Dif-en-dif: Si evento fortuito genera asignación que parece aleatoria, pero hay diferencias preexistentes entre ambos grupos.

(2) Evento fortuito determina sólo una parte del grupo de tratamiento de forma aleatoria.

- El tratamiento es parcialmente endógeno.

Dif-en-dif

ΔY si $D=1$ -

ΔY si $D=0$

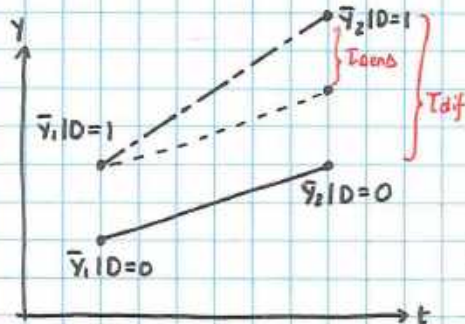
Si X no varían a través del tiempo $\Rightarrow$ Est. más eficiente

Req. tendencias paralelas

5.1. El modelo de diferencias - en - diferencias

Corrige diferencias sistemáticas entre el grupo de tratamiento y control antes de la aplicación del tratamiento.

↳ Requiere datos panel.



Sea $t=1$ el período antes del tratamiento (línea de base) y $t=2$ el período de seguimiento.

$$T_{dif\ endif} = [E(\bar{Y}_2 | D=1) - (E(\bar{Y}_1 | D=1))] - [E(\bar{Y}_2 | D=0) - (E(\bar{Y}_1 | D=0))]$$

$$\hat{T}_{dif\ endif} = (\Delta \bar{Y} | D=1) - (\Delta \bar{Y} | D=0)$$

Ventajas

- Es más eficiente que el estimador de diferencias, si algunos de los X no varían a través del tiempo.
- Elimina diferencias preexistentes $\Leftrightarrow$ se cumple el supuesto de tendencias paralelas.

↳ La tendencia temporal de Y entre $t=1$ y $t=2$ es la misma para ambos grupos.

MCO:

$$\Delta Y = X\beta + \delta D + u$$

$$\delta = T_{dif\ endif}$$

Por MCO: $Y_{2i} - Y_{1i} = \beta_0 + \beta_1 D_i + (u_{i2} - u_{i1})$

$$\Delta Y_i = \beta_0 + \beta_1 D_i + v_i$$

Note que $E[\Delta Y_i | D=1] = \beta_0 + \beta_1 + E[u_{i2} - u_{i1} | D=1]$

$$= \beta_0 + \beta_1 + E[u_{i2} | D=1] - E[u_{i1} | D=1]$$

$$E[\Delta Y_i | D=0] = \beta_0 + E[u_{i2} | D=0] - E[u_{i1} | D=0]$$

Asumimos $E[u_{i2} | D=1] = E[u_{i1} | D=1] = 0$.

$$\Rightarrow T_{dif\ en\ dif} = \beta_1$$

5.2. El estimador de diferencias-en-diferencias con regresores adicionales

Regresores adicionales: Características en línea de base

Si hay diferencias sistemáticas entre el grupo de tratamiento y el grupo de control, se deben incluir variables observables que midan características preexistentes antes del tratamiento, sobre todo de variables que varían en el tiempo.

→ No deben ser afectadas directamente por el tratamiento.

Mejora eficiencia

$$\Delta Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_{i1} + \dots + \beta_{K+1} X_{Ki} + v_i \quad \left. \vphantom{\Delta Y_i} \right\} \text{En el período inicial}$$

Ventajas

- Mejora la eficiencia del estimador.
- Evalúa la validez del supuesto de tendencias paralelas
- Ajustar el estimador si la aleatorización se hizo controlando por X.

5.3. El estimador de diferencias-en-diferencias para múltiples períodos

Dif-en-dif

múltiples períodos

↳ E.F. de individuo

↳ E.F. de tiempo

Considere una base de datos que incluye $i = 1, 2, \dots, N$ individuos seguidos en $t = 1, 2, \dots, T$ períodos de tiempo.

→ Permite incorporar efectos fijos de tiempo y de individuo.

En el período inicial $\left\{ \begin{array}{l} Y_{it} = \beta_0 + \beta_1 D_{it} + \delta_2 G_2 + \dots + \delta_N G_N + \beta_2 T_2 + \dots + \beta_T T_T + u_{it} \end{array} \right.$

Dif-en-dif

corte transversal repetido ↔

Muestras aleatorias de la misma población

5.4. Diferencias-en-diferencias utilizando datos de corte transversal repetido

El supuesto que permite utilizar datos de corte transversal repetidos en vez de datos panel es que si las muestras son aleatorias de la misma población. → Individuos de la base de datos de corte transversal inicial pueden ser utilizados como sustitutos de los individuos de los grupos de tratamiento y control del corte transversal posterior.

Requiere

- Que sea posible identificar qué individuos o unidades pertenecen al tratamiento antes y después de su aplicación.
- Que u_{it} sea igual para los individuos que pertenecen al grupo de tratamiento y control.

Requiere:

- ↳ Identificar tratamiento y control
- ↳ ser igual para ambos grupos

Por MCO: $Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 I[t = 2] + \beta_3 D_i I[t = 2] + u_{it}$

Dummy de período de seguimiento + Interacción

$$T_{\text{dif-en-dif}} = [E[Y|D=1, t=2] - E[Y|D=1, t=1]] - [E[Y|D=0, t=2] - E[Y|D=0, t=1]]$$

$$\begin{array}{l} E[Y|D=1, t=2] = \beta_0 + \beta_1 + \beta_2 + \beta_3 \\ E[Y|D=1, t=1] = \beta_0 + \beta_1 \\ E[Y|D=0, t=2] = \beta_0 + \beta_2 \\ E[Y|D=0, t=1] = \beta_0 \end{array} \quad \left. \vphantom{\begin{array}{l} \\ \\ \\ \end{array}} \right\} \begin{array}{l} \beta_2 + \beta_3 \\ \beta_2 \end{array}$$

$$T_{\text{dif-en-dif}} = \beta_3$$

Es inseguro si el experimento natural origina una asignación al tratamiento aleatoria.

CAPITULO 6: EL MÉTODO DE EMPAREJAMIENTO

Condición de IC: Todos los determinantes de asignación al tratamiento se incluyen en el modelo

Supuesto central Todas las variables que afectan simultáneamente la asignación al tratamiento y los resultados potenciales (Y_i) son observadas por el investigador e incluidas en el modelo que se estima.

Condición de independencia condicional

$$Y(0), Y(1) \perp D | X \quad \forall X$$

↳ Implica que $E[Y_i(0) | D_i = 1] = E[Y_i(0) | D_i = 0] \Rightarrow$ Elimina el sesgo de selección.

Obj: Emparejar individuos de tratamiento y control con probabilidad de participación similar

Obj: Encontrar un 'clon' de cada individuo tratado en el grupo de control y contrastar las variables de resultado de cada uno.

- Idénticos en términos de $X \rightarrow$ Pero si hay muchas X , es mejor que sean idénticos en $Pr(D=1|X)$

$$P(X) = P(D=1|X)$$

Propensity score

Las diferencias entre grupos de tratamiento y control se distinguen en un área donde todos tengan las mismas características observables (soporte común).

Soporte común: Área donde se traslapan características de tratamiento y control.

Condición de soporte común

Los individuos con el mismo vector de variables X tienen probabilidad positiva de ser tanto participantes como no participantes $\Rightarrow 0 < P(D=1|X) < 1$.

↳ Si se incumple $\Rightarrow$ Hay un conjunto de individuos para el cual no se puede decir nada acerca del efecto del programa.

Requisitos

- Igualdad en factores no observados que determinan la decisión de participar.
 - ↳ ¿cómo se prueba? Calcular diferencias de v. observables. Si son muy distintas $\Rightarrow$ No observables también lo serán.

Método no paramétrico
Alta validez externa, baja validez interna

Ventajas

- Método no paramétrico
- Soporte común garantiza comparar grupos comparables.
 - $\rightarrow$ Sacrifica validez interna
- Cuando se duda de la posibilidad de extrapolación de los resultados (grupos de tratamiento y control disímiles), el emparejamiento disminuye la extrapolación al reducir la comparación al soporte común.

PSM se estima usando logit o probit

↳ Sensible a v. incluidas

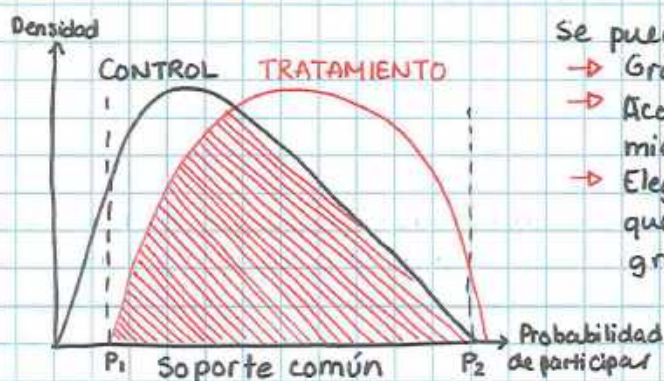
6.1. Probabilidad de participación

En primer lugar, se estima un modelo $P(D_i = 1 | X) = f(X)$, frecuentemente se usa $f(\cdot)$ de logit o probit.

- ↳ Los PSM son muy sensibles a la selección de variables usadas para estimar $P(X)$
- ↳ Incluir variables no relacionadas puede exacerbar el problema de soporte común y aumentar la varianza.

6.2. Soporte común

Soporte común
Se puede definir con algoritmo
↳ Entre más pequeño, + variación
↳ Entre más grande más insesgado



Se puede definir:

- Gráficamente
- Acotar con el mínimo de tratamiento y máximo de control.
- Elegir valores de $P(X)$ para los que hay densidad > 0 para ambos grupos (trimming).

A mayor % de ind. emparejados $\Rightarrow$ Mayor precisión.

6.3. Selección de un algoritmo de emparejamiento

Metodologías de emparejamiento

Obj: Emparejar los individuos más parecidos posibles.

Vecino más cercano

- (a) **Vecino más cercano** El vecino más cercano del individuo i está dado por: $C_i = \{j \in D = 0 \mid \arg \min \|P_i(X) - P_j(X)\|\}$ donde $i \in D=1$. El estimador está dado por $\tau_{ATT} = \left\{ \sum_{i=1}^n \tilde{w}_i \cdot \text{Promedio}[Y_i \mid D=1] - [Y_{C_i} \mid D=0] \right\}$
- ↳ Si $P_i(X) = P_j(X) \Rightarrow \hat{\tau}_{ATT}$ es insesgado, pero la varianza es pequeña
 - ↳ Reemplazo: Permitir que más de un individuo del grupo de tratamiento tenga como vecino al mismo j del grupo de control (estimaciones insesgadas, pero menos precisas).

Distancia máxima

- (b) **Emparejamiento de distancia máxima** Se impone una cota superior a la distancia de probabilidades, tal que el "vecino" emparejado es $C_i = \{j \in D = 0 \mid \|P_i(X) - P_j(X)\| \leq k\}$.

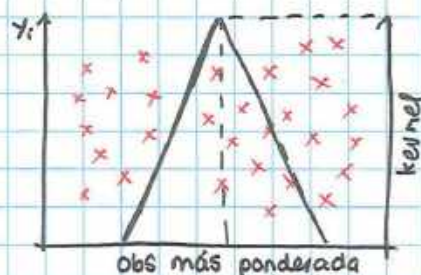
Estratificación

- (c) **Estimador de PSM por estratificación** Se divide la muestra según las probabilidades de participación en estratos. El estimador es una diferencia entre las medias ponderadas de las diferencias por estrato.
- $$\tau_{ATT} = \text{Promedio}(Y_i \mid D=1, j) - \text{Promedio}(Y_i \mid D=0, j)$$
- ↳ Requiere $E[X_i \mid D=1] = E[X_i \mid D=0] \rightarrow$ Que los grupos estén balanceados.
 - ↳ Menos sensible a la especificación de la probabilidad de participación.

Kernel

- (d) **Estimador por Kernel y regresión lineal local** Estimación no paramétrica que empareja a cada i del grupo de tratamiento con un promedio ponderado de todos los $j \in D=0$.

Dobles diferencias emparejadas



- (e) **Estimador de dobles diferencias emparejadas** Similar a dif-en-dif, hay una línea de base $t=1$ y un período de seguimiento $t=2$. La condición de IC se traduce en $u_2 - u_1 \perp D \mid X \forall X$.
- $$\tau = E\{E[Y_2 - Y_1 \mid D=1, P(X)] - E[Y_2 - Y_1 \mid D=0, P(X)]\}$$

¿Qué algoritmo elegir?

- Muestras grandes: Cualquiera es bueno.
- Muestras pequeñas: Si los resultados difieren, es necesario especificar por qué lo hacen.

Emparejamiento es bueno
 $\Leftrightarrow$ Grupos balanceados

Prueba:
 $D = \beta_0 + \beta_1 P(X) + u$
 β no debe ser significativo

Errores estándar deben incorporar incertidumbre de estimar PSM
 $\Rightarrow$ Bootstrapping

Pruebas de falsificación

Pseudoresultado

Pseudotrataimiento

6.4. Calidad del emparejamiento

Para que PSM funcione, condicional en la probabilidad de participación, los grupos de tratamiento y control deben estar balanceados.

$\rightarrow$ Después de controlar por $P(X)$, D no debe estar determinada por X .

Se estima $D_i = \beta_0 + \beta_1 P(X) + \beta_2 X_1 + \dots + \beta_{K+1} X_K + \varepsilon_i$

$\rightarrow H_0: \beta_2 = \beta_3 = \dots = \beta_{K+1} = 0 \rightarrow$ El emparejamiento fue bueno (Se cumple IC.)

Se prueba $E[X|D=1, P(X)] = E[X|D=0, P(X)]$ para demostrar el balance de la muestra.

6.5. Errores estándar

Para que sean correctos, los errores estándar deben incorporar la varianza debida a la estimación de la probabilidad de participación.

Bootstrapping: Estimar propiedades de un estimador midiendo dichas propiedades en la distribución empírica de los datos.

$\rightarrow$ No funciona para los estimadores de n vecinos cercanos. En ese caso, es mejor calcular los errores estándar de forma analítica.

6.6. Pruebas de falsificación

Intentan comprobar indirectamente si se cumple el supuesto central de IC.

① Estimar el efecto del programa sobre una variable que sabemos no puede ser alterada por éste. (Pseudoresultado)

② Estimar el efecto causal de una intervención ficticia que no debería afectar la variable de resultado (Pseudotrataimiento).

CAPÍTULO 7: MÉTODO DE VARIABLES INSTRUMENTALES

Objetivo: Estimar el efecto de un programa cuando no creemos válido utilizar $E[Y=0|D=0]$ para aproximar el contrafactual $E[Y=0|D=1]$.

$\rightarrow \text{Cov}(D, u) \neq 0$ a causa de variable omitida no observable.

7.1. Definición de variable instrumental

Sea Z_i una variable instrumental para D_i tal que cumple:

(I) **Relevancia:** $\text{Cov}(Z, D) \neq 0$ (Condición de rango completo)

(II) **Exogeneidad:** $\text{Cov}(Z, u) = 0$. (Restricción de exclusión)

+ **Monotonicidad:** D_i es una función monótona creciente de Z_i (necesaria en caso de efectos heterogéneos para garantizar la identificación local del efecto.)

$\rightarrow$ Implica que no existe algún i tal que $D_i=0|Z_i=1$ o $D_i=1|Z_i=0$.

Si hay efectos heterogéneos
 $\Rightarrow$ Requiere monotonicidad
 $D_i = f(Z_i)$

7.2. Estimación por el método de variables instrumentales

Considere el modelo $Y_i = \beta_0 + \beta_1 D_i + u_i$ y sea Z_i instrumento de D_i .
 $\text{Cov}(Y_i, Z_i) = \beta_1 \text{Cov}(D_i, Z_i) + \text{Cov}(Z_i, u_i)$

Asumiendo relevancia y exogeneidad: (*) $\beta_1 = \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} \rightarrow \frac{\sum (Y - \bar{Y})(Z - \bar{Z})}{\sum (D - \bar{D})(Z - \bar{Z})}$

Si hay más de una variable exógena en el modelo, se estima por MC2E.

Primera etapa: $D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_1 + \dots + \alpha_{K+1} X_K + \varepsilon_i$

↳ Se puede estimar linealmente o por medio de logit o probit, corrigiendo los errores estándar.

Segunda etapa: $Y_i = \beta_0 + \beta_1 \hat{D}_i + \beta_2 X_2 + \dots + \beta_{K+1} X_K + u_i$

Si hay efectos heterogéneos: Modelo estructural: $Y_i = D_i \gamma + X_i \beta + D_i X_i \lambda + u_i$

Primera etapa: $D_i = X_i \alpha + Z_i \beta + X_i Z_i \varphi + \varepsilon_i$

7.3. Elección de los instrumentos

Instrumentos
más usados
en las Ev. Im.

- Inst. de disponibilidad o facilidad de acceso al programa
- Instr. de diferencias de precios de acceso al programa
- Políticas de asignación que generan aleatoriedad en la participación.

7.4. Evaluación de la variable instrumental

(1) Comprobar relevancia En la primera etapa, $\hat{\alpha}_1$ deb ser significativo y el estadístico $F > 10$.

(2) Comprobar monotonicidad No es fácil de comprobar directamente, pero si hay monotonicidad hay mejor ajuste de la primera etapa.

(3) Comprobar exogeneidad Sólo se puede probar si hay más de un instrumento por variable exógena.

↳ Prueba de restricciones sobreidentificadas: Estimar el modelo por MC2E utilizando sólo uno de los instrumentos y verificar que el error del modelo no se correlaciona con el otro instrumento.

H_0 : Al menos un Z cumple $\text{Cov}(Z, u) = 0$

Estadístico de prueba: $\chi^2_q = NR^2_{aux}$

- Sólo vale si no hay efectos heterogéneos.

También se puede aplicar una prueba de pseudo resultado.

7.5. Problemas potenciales del estimador de variables instrumentales

- Es difícil encontrar variables instrumentales exógenas y relevantes.
- Si el instrumento es débil, la magnitud del sesgo $\text{Cov}(Z, u) / \text{Cov}(Z, D)$ aumenta.
- La monotonicidad no se puede comprobar.
- Los efectos son locales y no globales.

CAPÍTULO 8: MÉTODO DE REGRESIÓN DISCONTINUA

Objetivo: Aprovechar una discontinuidad en la probabilidad de participación para definir el grupo de tratamiento y control.

↳ El cambio se da con una variable continua observada Z , llamada variable de focalización.

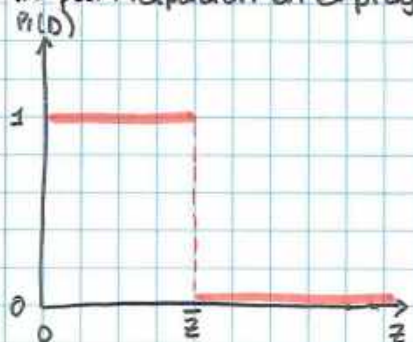
- La asignación al tratamiento depende completa o parcialmente de Z , y se determina según un umbral $\bar{Z}$.

- Si $\bar{Z}$ es arbitrario, los individuos a ambos lados del umbral son relativamente parecidos.

↳ El parámetro estimado es un efecto promedio local del tratamiento.

8.1. El diseño de regresión discontinua nítida (RDN)

Ocurre cuando Z determina completamente (de forma determinística) la participación en el programa. $\rightarrow D(Z) = I(Z \leq \bar{Z})$

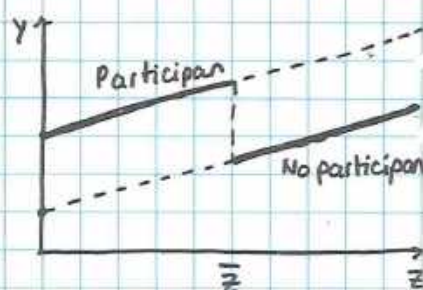


Supuesto: Z no es manipulable.

Condición de identificación

$$\lim_{z \uparrow \bar{Z}} D(Z=z) = 1 \neq 0 = \lim_{z \downarrow \bar{Z}} D(Z=z)$$

$$E[Y|Z=z] = \sum_{i=0,1} E[Y|D=i, z] Pr[D=i|Z=z]$$



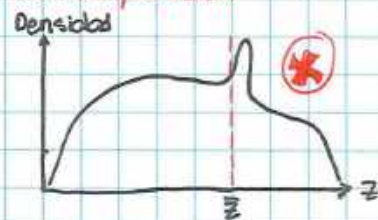
Se aplica la metodología de emparejamiento

↳ Se cumple CI: $Y_i(0), Y_i(1) \perp D_i | Z_i$

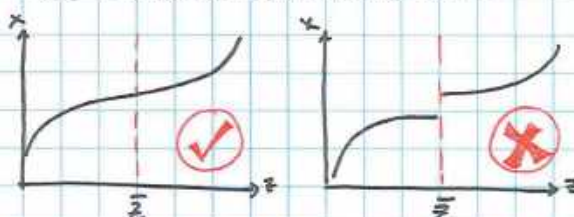
↳ No hay soporte común: Control y tratamiento no tienen valores de Z que se traslapen.

- Obliga a asumir que los i cerca del umbral son muy similares.

No manipulación



Continuidad local: $E[Y(0)|Z=\bar{Z}]$ y $E[Y(1)|Z=\bar{Z}]$ deben ser continuas en $\bar{Z}$.



↳ Discontinuidad de Y se debe únicamente a Z .

Estimador: $\tau_{RDN}(\bar{Z}) = \lim_{z \uparrow \bar{Z}} E[Y_i|Z_i=z] - \lim_{z \downarrow \bar{Z}} E[Y_i|Z_i=z]$

Note que $\tau_{RDN}(\bar{Z}) = \lim_{z \uparrow \bar{Z}} E[D_i|Z_i=z] - \lim_{z \downarrow \bar{Z}} E[D_i|Z_i=z] + \lim_{z \uparrow \bar{Z}} E[-u_i|Z_i=z] - \lim_{z \downarrow \bar{Z}} E[u_i|Z_i=z]$

Si hay CL: $\tau_{RDN}(\bar{Z}) = E[\tau_i|Z=\bar{Z}]$

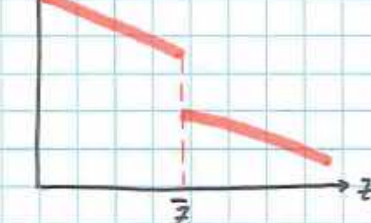
↳ Impacto local para un i con $z=\bar{Z}$. Validez externa limitada.

8.2. El diseño de regresión discontinua borrosa

Ocurre cuando la probabilidad condicional de participación en el tratamiento es discontinua en $\bar{z}$, en vez de la participación efectiva.

$Pr(D)$

↳ Participación efectiva sólo depende parcialmente de z .



Condición de identificación

$$\lim_{z \uparrow \bar{z}} Pr(D=1|Z=z) \neq \lim_{z \downarrow \bar{z}} Pr(D=0|Z=z)$$

$$E[Y|Z=\bar{z}] = \sum_{i=0,1} E[Y(i)|D=i, Z=\bar{z}] Pr(D=i|Z=\bar{z})$$



El salto de Y en $\bar{z}$ sólo puede usarse para evaluar el impacto de la elegibilidad del tratamiento sobre Y , no del tratamiento como tal. $\Rightarrow$ Estimador de intento de tratamiento

Supuesto: Quienes son elegibles para participar pero no participan, no reciben beneficios del programa en forma de externalidades.

↳ Si no se cumple, se subestima el efecto del programa.

$$\tau_{RD}(\bar{z}) = \frac{\lim_{z \uparrow \bar{z}} E[Y_i | Z_i = z] - \lim_{z \downarrow \bar{z}} E[Y_i | Z_i = z]}{\lim_{z \uparrow \bar{z}} Pr(D=1|Z=z) - \lim_{z \downarrow \bar{z}} Pr(D=1|Z=z)}$$

Coincide con estimador de Wald

Supuesto (Independencia condicional a nivel local): Para cualquier z en una vecindad arbitrariamente pequeña de $\bar{z}$, la participación es determinística condicional en z :

$$E(\tau_i | D, z) = E[\tau_i | z] \text{ para } z \in [\bar{z} - \varepsilon, \bar{z} + \varepsilon] \forall \varepsilon > 0.$$

↳ Supuesto fuerte.

8.3. Implementación empírica del estimador de RD

(A) Método paramétrico Considere la regresión $Y_i = m(Z_i) + \beta_i D_i + e_i$, donde $m(\cdot)$ es una función continua (RDN). Se estima en dos etapas, donde la primera es: $D_i = f(Z_i) + \gamma I(Z_i \leq \bar{z}) + v_i$. Si $f(\cdot)$ y $m(\cdot)$ están bien especificadas, β_i es consistente.

↳ RDB: $Y_i = m(Z_i) + \beta_i E[D_i | Z_i] + u_i$

(B) Método no paramétrico Se estima primero una regresión no paramétrica de Y sobre Z y una regresión no paramétrica de D sobre Z a cada lado de $\bar{z}$.

En ambos casos, el investigador debe definir el ancho de banda h dentro del cual se estimará el efecto del tratamiento:

↳ Si $h \rightarrow 0$: Estimadores más insesgados, menos precisos.

↳ Si $h \rightarrow \infty$: Estimadores precisos pero sesgados.

I. REPASO MCO y TEORÍA ASINTÓTICA

Identificación del estimador de MCO

Considere el modelo

$$Y = X\beta + u.$$

$$X'Y = X'X\beta + X'u \leftarrow \text{Asumo que } X \text{ es de rango completo.}$$

$$E(X'Y) = E(X'X\beta + X'u)$$

$$E(X'Y) = E(X'X)\beta + E(X'u) \rightarrow E(X'u) = 0 \text{ por supuesto de IC.}$$

$$\beta = E^{-1}(X'X)E(X'Y) \quad \text{QED.}$$

Consistencia del estimador de MCO

Considere

 $\beta = E^{-1}(X'X)E(X'Y)$. usando el método de momentos:

$$\hat{\beta} = \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N x_i' y_i\right)$$

$$\hat{\beta} = \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N x_i' [x_i\beta + u_i]\right)$$

$$\hat{\beta} = \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \left[\left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)\beta + \left(\frac{1}{N} \sum_{i=1}^N x_i' u_i\right)\right]$$

$$\hat{\beta} = \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \left[\left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)\beta + \frac{1}{N} \sum_{i=1}^N x_i' u_i\right]$$

$$\Rightarrow \hat{\beta} = \beta + \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N x_i' u_i\right)$$

 $x_i' x_i$ es i.i.d y $E(|x_i' x_i|) < \infty$. Por ley de los grandes números:

$$\frac{1}{N} \sum_{i=1}^N x_i' x_i \xrightarrow{P} E(X'X)$$

Por el corolario 3.1: $\left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \xrightarrow{P} E^{-1}(X'X)$

Similarmente, $x_i' u_i$ es i.i.d y $E(|x_i' u_i|) < \infty$. Por ley de los grandes números:

$$\frac{1}{N} \sum_{i=1}^N x_i' u_i \xrightarrow{P} E(X'u) = 0$$

Supuesto de IC.

$$\Rightarrow \text{plim } \hat{\beta} = \text{plim } \beta + \text{plim} \left[\left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N x_i' u_i\right) \right]$$

Por tma. Slutsky: $\text{plim } \hat{\beta} = \beta + \text{plim} \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} * \text{plim} \left(\frac{1}{N} \sum_{i=1}^N x_i' u_i\right)$

$$\text{plim } \hat{\beta} = \beta + 0 = \beta \quad \text{QED}$$

Distribución asintótica del estimador de MCOTenemos que $\hat{\beta} - \beta = \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N x_i' u_i\right) \Rightarrow X$ es de rango completo

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \frac{\sqrt{N}}{N} \sum_{i=1}^N x_i' u_i$$

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N} \sum_{i=1}^N x_i' x_i\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N x_i' u_i$$

Considere $x_i'x_i$: es i.i.d y $E[|x_i'x_i|] < \infty$. Por ley de los grandes números:

$$\frac{1}{N} \sum x_i'x_i \xrightarrow{p} E(x'x)$$

Por corolario 3.1: $\left(\frac{1}{N} \sum x_i'x_i\right)^{-1} \xrightarrow{p} E^{-1}(x'x)$

$$\Rightarrow \left(\frac{1}{N} \sum x_i'x_i\right)^{-1} E^{-1}(x'x) \xrightarrow{p} 0 \Rightarrow \left(\frac{1}{N} \sum x_i'x_i\right)^{-1} E^{-1}(x'x) \text{ es } o_p(1).$$

Ahora, $x_i'u_i$ es i.i.d y $E[|x_i'u_i|^2] < \infty$ por supuesto de IC. Por el teorema del límite central:

$$\frac{1}{\sqrt{N}} \sum x_i'u_i \xrightarrow{d} N\left(\overset{IC}{0}, v\right)$$

Donde $v = \text{Var}(x_i'u_i) = \overset{IC}{E[(x_i'u_i)(x_i'u_i)']} = E(x_i'x_i)\sigma^2$
 Asumo homoscedasticidad.

Como $\frac{1}{\sqrt{N}} \sum x_i'u_i$ converge en distribución, sabemos que está acotado $\Rightarrow$ es $O_p(1)$.

$$\begin{aligned} o_p(1) * O_p(1) &= o_p(1) = \left[\left(\frac{1}{N} \sum x_i'x_i \right)^{-1} - E^{-1}(x'x) \right] \left[\frac{1}{\sqrt{N}} \sum x_i'u_i \right] \\ &= \underbrace{\left(\frac{1}{N} \sum x_i'x_i \right)^{-1}}_{\text{Asintóticamente equivalentes}} \left(\frac{1}{\sqrt{N}} \sum x_i'u_i \right) - E^{-1}(x'x) \left(\frac{1}{\sqrt{N}} \sum x_i'u_i \right) \end{aligned}$$

$$o_p(1) = \underbrace{\sqrt{N}(\hat{\beta} - \beta) - [E(x'x)]^{-1} \left(\frac{1}{\sqrt{N}} \sum x_i'u_i \right)}_{\text{Asintóticamente equivalentes}}$$

Por el teorema del mapa continuo:

$$E^{-1}(x'x) \frac{1}{\sqrt{N}} \sum x_i'u_i \xrightarrow{d} N(0, B) \text{ donde } E(x'x)^{-1} V E(x'x)^{-1} = B.$$

$$B = E(x_i'x_i)^{-1} E(x_i'x_i) \sigma^2 E(x_i'x_i)^{-1} = \sigma^2 E(x_i'x_i)^{-1}$$

$$\Rightarrow \sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 E(x_i'x_i)^{-1}) \text{ QED.}$$

$$\hat{\beta} - \beta \xrightarrow{d} N\left(0, \frac{\sigma^2 E(x'x)^{-1}}{N}\right)$$

$$\hat{\beta} \xrightarrow{d} N\left(\beta, \frac{\sigma^2 E(x'x)^{-1}}{N}\right)$$

Usando método de momentos sobre $A\text{Var}(\hat{\beta}) = \frac{\sigma^2 \left(\frac{1}{N} \sum x_i'x_i \right)^{-1}}{N} = \sigma^2 (x'x)^{-1}$

Estimación de σ^2 de MCO

$\sigma^2 = \text{Var}(u) = E(u'u)$. Utilizando el método de momentos:

$$\hat{\sigma}^2 = \frac{\sum u_i' u_i}{N} \quad \text{Pero dado que los errores no son observables, reemplazamos por los residuales.}$$

$$\hat{\sigma}^2 = \frac{\hat{u}' \hat{u}}{N} = \frac{(Y - \hat{Y})'(Y - \hat{Y})}{N} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{N} = \frac{(Y' - \beta'X')(Y - X\hat{\beta})}{N}$$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ (X\beta + u)' - \hat{\beta}'X' \right\} (X\beta + u - X\hat{\beta}) \}$$

Recuerde que $\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u$
 $X\hat{\beta} = X(X'X)^{-1}X'u + X\beta$
 $\hat{\beta}'X' = u'X(X'X)^{-1}X' + \beta'X'$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ (\beta'X' + u' - \beta'X' - u'X(X'X)^{-1}X') (X\beta + u - X\beta - X(X'X)^{-1}X'u) \right\}$$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ (u' - u'X(X'X)^{-1}X') (u - X(X'X)^{-1}X'u) \right\}$$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ (u - X(X'X)^{-1}X'u)' \underbrace{(I_N - X(X'X)^{-1}X')}_M u \right\}$$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ (Mu)' Mu \right\} = \frac{1}{N} u' M' M u$$

$$M'M = (I_N - X(X'X)^{-1}X')(I_N - X(X'X)^{-1}X') = -X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' + I_N$$

$$M'M = I_N - 2X(X'X)^{-1}X' + X(X'X)^{-1}X' = I_N - X(X'X)^{-1}X' = M \Rightarrow \text{Idempotente.}$$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ \underbrace{u'}_{1 \times N} \underbrace{Mu}_{N \times 1} \right\} = \text{Escalar} = E[u'Mu]$$

Propiedades de la traza

$$\begin{aligned} \hat{\sigma}^2 &= \text{tr}(E(u'Mu)) = E(\text{tr}(u'Mu)) = E(\text{tr}(Mu u')) = \text{tr}(E[Mu u']) \\ &= \text{tr}(\sigma^2 E[M]) = \sigma^2 \text{tr}(E[M]) = \sigma^2 E[\text{tr}(M)] = \sigma^2 E[\text{tr}(I_N - X(X'X)^{-1}X')] \\ &= \sigma^2 E[\text{tr}(I_N) - \text{tr}(X(X'X)^{-1}X')] = \sigma^2 E[N - \text{tr}(X'X(X'X)^{-1})] \\ &= \sigma^2 E[N - \text{tr}(I_{K+1})] = \sigma^2 E[N - (K+1)] \end{aligned}$$

$$\Rightarrow \hat{\sigma}^2 = \sigma^2(N - K - 1) \rightarrow \text{Estimador sesgado.}$$

Insesgado: $\hat{\sigma}^2 = \frac{\frac{1}{N} \sum u_i' u_i}{N - K - 1} \quad \text{QED.}$

Demostración Teorema de Gauss-Markov

Considere cualquier estimador lineal $\tilde{\beta} = A'Y$ donde A es una función de X . Definimos $D = A' - (X'X)^{-1}X'$ como la diferencia entre $\tilde{\beta}$ y $\hat{\beta}_{MCO}$.

$$\begin{aligned} D'X &= A'X - (X'X)^{-1}X'X \\ D'X &= \underbrace{A'X}_{\downarrow} - I_{k+1} \end{aligned}$$

Asumiendo insesgamiento:

$$\begin{aligned} E[\tilde{\beta}] &= \beta \\ E[A'Y] &= \beta \\ E[A'(X\beta + u)] &= \beta \\ E[A'X\beta] + \underbrace{E[A'u]}_{=0 \text{ por IC}} &= \beta \\ \beta E[A'X] &= \beta \\ A'X &= I_{k+1} \end{aligned}$$

$$\Rightarrow D'X = I_{k+1} - I_{k+1} = 0.$$

Note que si $D' = A' - (X'X)^{-1}X' \Rightarrow A' = D' + (X'X)^{-1}X'$.

$$\begin{aligned} \tilde{\beta} &= (D' + (X'X)^{-1}X')(X\beta + u) \\ \tilde{\beta} &= \underbrace{D'X\beta}_{=0} + (X'X)^{-1}X'X\beta + D'u + (X'X)^{-1}X'u \\ \tilde{\beta} &= \beta + (D' + (X'X)^{-1}X')u \\ \tilde{\beta} - \beta &= (D' + (X'X)^{-1}X')u \end{aligned}$$

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] = E[(D' + (X'X)^{-1}X')u u'(D + X(X'X)^{-1})] \\ \text{Var}(\tilde{\beta}) &= (D' + (X'X)^{-1}X')E[uu'](D + X(X'X)^{-1}) \\ &= \sigma^2(D' + (X'X)^{-1}X')(D + X(X'X)^{-1}) \\ &= \sigma^2(\underbrace{D'D}_{=0} + \underbrace{D'X(X'X)^{-1}}_{=0} + \underbrace{(X'X)^{-1}X'D}_{=0} + (X'X)^{-1}X'X(X'X)^{-1}) \\ \text{Var}(\tilde{\beta}) &= \sigma^2(D'D + (X'X)^{-1}) \end{aligned}$$

Note que $\text{Var}(\tilde{\beta}) = \text{Var}(\hat{\beta}_{MCO}) \Leftrightarrow D'D = 0 \Rightarrow D' = 0 \Rightarrow$ No hay diferencia entre $\tilde{\beta}$ y $\hat{\beta}_{MCO}$.

Método Delta

Se usa para encontrar la varianza de transformaciones no lineales de parámetros que ya estimamos.

Suponga $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ y tenemos una función continua $c(\theta)$.
 $\Rightarrow \sqrt{N}(c(\hat{\theta}) - c(\theta)) \xrightarrow{d} N(0, J(\theta)' V J(\theta))$ donde $J(\theta)$: Jacobiano de $c(\theta)$.

II. VIOLACIÓN DE SUPUESTOS: HETEROSCEDASTICIDAD, AUTOCORRELACIÓN Y MCG

A) HETEROSCEDASTICIDAD

Características: σ^2 varía según los individuos de la muestra.

Efectos: $\hat{\beta}_{MCO}$ insesgado, pero ineficiente.

Dem: Considere $\hat{\beta} - \beta = (X'X)^{-1}X'u$

$$\text{Var-Cov}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1}X'u u'X(X'X)^{-1}]$$

$$\text{Por no aleatoriedad de } X: = (X'X)^{-1}X'E[UU']X(X'X)^{-1}$$

Si no hay heteroscedasticidad: $E[UU'] = \sigma^2 I_N$, y $\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

Si hay heteroscedasticidad: $E[UU'] = \Sigma$, donde:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix} \quad \text{Note que } \sigma_{12} = 0 \Rightarrow \text{Asumimos no autocorrelación.}$$

$$\text{Var}(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1} \neq \sigma^2 (X'X)^{-1} \Rightarrow \text{Ineficiente.}$$

Detección: Test de White

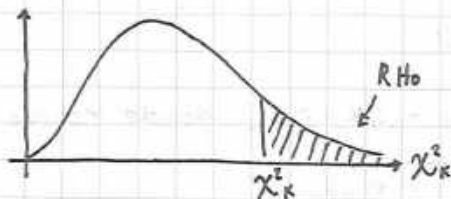
Obj: Saber si $\hat{u}$ varía con las características de los individuos $\rightarrow \text{Var}(\hat{u}) = f(x)$.

① Estimar el modelo $\hat{u}_i^2 = \alpha_0 + \alpha_1 x_i^2 + \alpha_2 x_i^2 + \dots + \alpha_{12} x_1 x_2 \dots + \epsilon$.

② H_0 : Homoscedasticidad,
 $\alpha_0 = \alpha_1 = \dots = 0$

H_1 : Heteroscedasticidad
Al menos un $\alpha_i \neq 0$.

Estadístico de prueba: $\chi^2_K = NR_{\hat{u}_i^2}^2$, $K = \#$ de alfas, sin contar intercepto.



Solución: Mínimos cuadrados generalizados factibles.

Para $Y = X\beta + u$, $u \sim N(0, \sigma^2)$ $\Rightarrow \text{Var}(u) = \sigma^2 \Omega = \sigma^2 \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_N \end{bmatrix}$, $w_i = f(x)$.

Definimos una matriz $P = \begin{bmatrix} \frac{1}{\sqrt{w_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{w_N}} \end{bmatrix} \Rightarrow P'P = \Omega^{-1}$

Transformamos el modelo usando la matriz P : $PY = PX\beta + Pu \rightarrow Y^* = X^*\beta + u^*$

Note que $\text{Var}(u^*) = E[u^* u^{*'}] = E[P V U U' P']$. Por no aleatoriedad de P :
 $\text{Var}(u^*) = P E[V U U'] P'$, y $E[V U U'] = \sigma^2 \Omega$
 $\text{Var}(u^*) = P \sigma^2 \Omega P'$

Como Ω y P son matrices triangulares superior e inferior, el producto de matrices es conmutable.

$$\text{Var}(u^*) = \sigma^2 \Omega P' P = \sigma^2 \Omega \Omega^{-1} = \sigma^2 \quad \forall \in D.$$

Errores estándar robustos

Tenemos que $\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}$. Lo problemático de esta expresión es Σ . Los errores estándar robustos intentan estimarlo.

Considere el término $X' \Sigma X$. Se puede escribir como $\sum a_i^2 x_i' x_i$.

↳ Estimador de White: $S_0 = \widehat{X' \Sigma X} = \sum \hat{u}_i^2 x_i' x_i$

⇒ $\text{Var}(\hat{\beta}) = (X'X)^{-1} S_0 (X'X)^{-1}$. Sin embargo, en general se cumple que:
 $\text{Var}(\hat{\beta})_{HT} = (X'X)^{-1} S_0 (X'X)^{-1} > \sigma^2 (X'X)^{-1} = \text{Var}(\hat{\beta})_{HM}$

- Intervalos de confianza serán más pequeños. ⇒ $\hat{\beta}$ serán menos significativos.

⑧ AUTOCORRELACIÓN

Características: $\text{Cov}(u_i, u_j) \neq 0 \quad \forall i \neq j$. Es más frecuente en las series de tiempo.

Efectos: $\hat{\beta}_{MCO}$ insesgado, pero $\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1} \Rightarrow$ Ineficiente.

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \dots & \sigma_{n1} \\ \sigma_{12} & \sigma^2 & \dots & \sigma_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma^2 \end{bmatrix}$$

Detección: Cochran - Orlutt

Considere el modelo $y_t = \beta_0 + \beta_1 x_t + u_t$, donde $\text{cov}(u_t, u_s) \neq 0 \quad \forall t \neq s$.

① Estimar $y_t = \beta_0 + \beta_1 x_t + u_t$.

② Obtener $\hat{u}_t$.

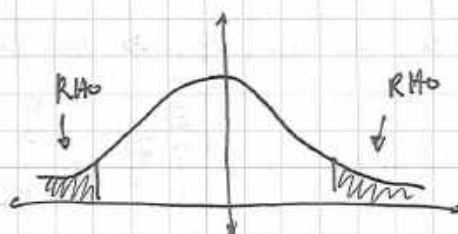
③ Correr el modelo $\hat{u}_t = \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \dots + \rho_j \hat{u}_{t-j} + \epsilon_t$

H_0 : No autocorrelación

$$\rho_i = 0$$

Estadístico de prueba

$$t = \frac{\hat{\rho}}{\text{se}(\hat{\rho})} \sim t_{n-j}$$



Soluciones: Mínimos cuadrados generalizados factibles

Suponga que la autocorrelación es de la forma AR(2): $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$

$$\begin{aligned}
 Y_t &= \beta_0 + \beta_1 X_t + u_t \\
 - \rho_1 Y_{t-1} &= \rho_1 \beta_0 + \rho_1 \beta_1 X_{t-1} + \rho_1 u_{t-1} \\
 - \rho_2 Y_{t-2} &= \rho_2 \beta_0 + \rho_2 \beta_1 X_{t-2} + \rho_2 u_{t-2}
 \end{aligned}$$

$$\begin{aligned}
 Y_t - \rho_1 Y_{t-1} - \rho_2 Y_{t-2} &= \beta_0(1 - \rho_1 - \rho_2) + \beta_1(X_t - \rho_1 X_{t-1} - \rho_2 X_{t-2}) + (u_t - \rho_1 u_{t-1} - \rho_2 u_{t-2}) \\
 \tilde{Y}_t &= \tilde{\beta}_0 + \beta_1 \tilde{X}_t + \tilde{u}_t
 \end{aligned}$$

I. Cochrane - Orcutt

- ① Estimar el modelo $Y_t = X_t \beta_1 + \beta_0 + u_t$
- ② Obtener $\hat{u}_t$.
- ③ Estimar $\hat{u}_t = \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \varepsilon_t$
- ④ Obtener $\hat{\rho}_1$ y $\hat{\rho}_2$.
- ⑤ Construir $\tilde{Y}_t = \tilde{\beta}_0 + \beta_1 \tilde{X}_t + \tilde{u}_t$ con $\hat{\rho}_1$ y $\hat{\rho}_2$.
- ⑥ Estimar $\tilde{Y}_t$.

II. Praise - Winsten: El hecho de rezagar el modelo hace perder una observación. Este método previene el problema.

Considere la primera observación: $Y_1 = \beta_0 + \beta_1 X_1 + u_1$

↳ Para esta observación: (I) $\text{Cov}(u_1, \varepsilon_t) = 0$.

$$(II) \text{Var}(u_1) = \sigma_1^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}, \quad \text{var}(\varepsilon_t) = \sigma_\varepsilon^2.$$

Para homogeneizar la varianza: $\sqrt{1 - \rho^2} Y_1 = \sqrt{1 - \rho^2} \beta_0 + \sqrt{1 - \rho^2} \beta_1 X_1 + \sqrt{1 - \rho^2} u_1$.

Se corre el modelo con $\tilde{Y}_t$ en todas las observaciones y $\sqrt{1 - \rho^2} Y_1$ en la primera.

Errores estándar robustos

$$\text{Var}(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}.$$

$$\text{Para un AR(1): } \Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \dots & \rho^{T-1}\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \dots & \rho^{T-2}\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \dots & \rho^{T-3}\sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1}\sigma^2 & \rho^{T-2}\sigma^2 & \rho^{T-3}\sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

El objetivo de los errores estándar robustos es estimar Σ .

$$\sigma^2 \rightarrow \hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{T - k - 1}$$

$\rho \rightarrow$ Obtenemos de $u_t = \rho u_{t-1} + \varepsilon_t$

Materia:

Hora:

0 M A

Pruebe que si $\sqrt{N}(\hat{\theta}_N - \theta) = O_p(1) \Rightarrow \hat{\theta}_N - \theta = o_p(N^{-c}) \forall c \in [0, \frac{1}{2})$

Se que $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{P} a \Rightarrow \hat{\theta}_N - \theta \xrightarrow{P} \frac{a}{\sqrt{N}} = N^{-1/2}a$ $\frac{\hat{\theta}_N - \theta}{N^{-c}} \xrightarrow{P} 0$ es $o_p(1)$

$\frac{\hat{\theta}_N - \theta}{N^{-c}} \xrightarrow{P} N^{-\frac{1}{2}+c}a$ $P\left(\left|\frac{\hat{\theta}_N - \theta}{N^{-c}}\right| \geq c\right) \xrightarrow{N \rightarrow \infty} 0$

$\lim_{N \rightarrow \infty} (N^{-\frac{1}{2}+c})a = 0$ $\frac{\hat{\theta}_N - \theta}{N^{-c}} \xrightarrow{P} 0$

$c \in [0, \frac{1}{2})$

Sea $\hat{\theta}$ un estimador $\sqrt{N}$ -asintóticamente normal para el escalar $\theta > 0$. Sea $\hat{\gamma} = \log(\hat{\theta})$ un estimador de $\gamma = \log(\theta)$. $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$

(a) ¿Por qué $\hat{\gamma}$ es un estimador consistente de γ ?

Considere $\text{plim } \hat{\gamma} = \text{plim } \log(\hat{\theta})$. Como el logaritmo es una función continua, por el teorema de Slutsky: $\text{plim } \hat{\gamma} = \log(\text{plim } \hat{\theta}) = \log(\theta) = \gamma$. QED

(b) Encuentre la varianza asintótica de $\sqrt{N}(\hat{\gamma} - \gamma)$ en términos de la varianza asintótica de $\sqrt{N}(\hat{\theta} - \theta)$.

Tenemos que $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ donde $\text{Avar}(\sqrt{N}[\hat{\theta} - \theta]) = V$
 $\sqrt{N}(\log(\hat{\theta}) - \log(\theta)) \xrightarrow{d} N(0, B)$ $B = f(V)$

$$\hat{\theta} - \theta \xrightarrow{d} N(0, \frac{V}{N})$$

$$\hat{\theta} \xrightarrow{d} N(\theta, \frac{V}{N}) \quad \text{¿Tiene mapeo continuo?}$$

Método delta: $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ y $c(\cdot)$ que es continua
 $\sqrt{N}(c(\hat{\theta}) - c(\theta)) \xrightarrow{d} N(0, J(\theta)' V J(\theta))$

$$\frac{d\hat{\gamma}}{d\hat{\theta}} = \frac{1}{\hat{\theta}} \Rightarrow \sqrt{N}(\log(\hat{\theta}) - \log(\theta)) \xrightarrow{d} N(0, \frac{V}{\theta^2})$$

(c) Suponga que $\hat{\theta} = 4$ y $ee(\hat{\theta}) = 2$. ¿Qué es $\hat{\gamma}$ y su ee asintótico?

$$\hat{\gamma} = \log 4 \quad \hat{\gamma} - \gamma \xrightarrow{d} N(0, \frac{V}{N\theta^2})$$

$$\hat{\gamma} \xrightarrow{d} N(\gamma, \frac{V}{N\theta^2}) \Rightarrow \text{Aee}(\hat{\gamma}) = \sqrt{\frac{V}{N\theta^2}} = \frac{\text{Aee}(\hat{\theta})}{\sqrt{N}\theta}$$

$$\text{Aee}(\hat{\gamma}) = \frac{2}{\sqrt{N}\theta} \quad \hat{\text{Aee}}(\hat{\gamma}) = \frac{2}{\sqrt{N}\hat{\theta}} = \frac{1}{2\sqrt{N}}$$

(d) Considere $H_0: \theta = 1$. ¿Cuál es el estadístico t asintótico para probar H_0 , dados los números en (c)?

$$t = \frac{\hat{\theta} - 1}{ee(\hat{\theta})} = \frac{3}{2}$$

(e) Ahora plantee la prueba anterior en términos de $\hat{\gamma}$.

$$H_0: \log(\theta) = \log(1) \Rightarrow \gamma = 0$$

$$t = \frac{\hat{\gamma}}{\text{Aee}(\hat{\gamma})} = \frac{\log 4 \sqrt{N}\theta}{2}$$

Sea $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$ un estimador $\sqrt{N}$ -asintóticamente normal de $\theta = (\theta_1, \theta_2)'$ con $\theta_2 \neq 0$.
 Sea $\hat{\gamma}$ un estimador de $\gamma = \theta_1/\theta_2$, $\hat{\gamma} = \hat{\theta}_1/\hat{\theta}_2$.

(a) Muestre que $\text{plim } \hat{\gamma} = \gamma$.

$$\text{plim } \hat{\gamma} = \text{plim } \frac{\hat{\theta}_1}{\hat{\theta}_2} \underset{\text{Slutsky}}{=} \frac{\text{plim } \hat{\theta}_1}{\text{plim } \hat{\theta}_2} = \frac{\theta_1}{\theta_2} = \gamma. \quad \text{QED}$$

(b) Halle $\text{Avar}(\hat{\gamma})$ en términos de θ y $\text{Avar}(\hat{\theta})$ usando el método delta.

Sean $\hat{\theta}$ y $\tilde{\theta}$ dos estimadores $\sqrt{N}$ -asintóticamente normales del vector $\theta \in \mathbb{R}^p$, con $\text{Avar}(\sqrt{N}[\hat{\theta} - \theta]) = V_1$ y $\text{Avar}(\sqrt{N}[\tilde{\theta} - \theta]) = V_2$.

Defina un vector de parámetros $\gamma \in \mathbb{R}^q = g(\theta)$ donde $g(\cdot)$ es continuamente diferenciable.

Muestre que: Si $\hat{\theta}$ es más asintóticamente eficiente que $\tilde{\theta} \Rightarrow \hat{\gamma} = g(\hat{\theta})$ es asintóticamente eficiente en relación a $\tilde{\gamma} = g(\tilde{\theta})$.

$$\hat{V}_1 - \hat{V}_2 \text{ tiene que ser p.s.d} \Rightarrow \underbrace{\text{Avar}(\sqrt{N}[\hat{\gamma} - \gamma])}_{W_1} - \underbrace{\text{Avar}(\sqrt{N}[\tilde{\gamma} - \gamma])}_{W_2} \text{ p.s.d.}$$

$$\text{Avar } W_1 = J(\theta) V_1 J(\theta)'$$

$$W_2 = J(\theta) V_2 J(\theta)'$$

$$\begin{aligned} W_1 - W_2 &= J(\theta) V_1 J(\theta)' - J(\theta) V_2 J(\theta)' \\ &= J(\theta) [V_1 - V_2] J(\theta)' \end{aligned}$$

$$W_1 = \underbrace{J(\theta)}_{Q \times P} \underbrace{V_1}_{P \times P} \underbrace{J'(\theta)}_{P \times Q}$$

$$J(\theta) = \begin{bmatrix} dg/d\theta_1 \\ dg/d\theta_2 \\ \vdots \\ dg/d\theta_p \end{bmatrix}_{P \times 1}$$

$$g: \mathbb{R}^p \rightarrow \mathbb{R}^q$$

$$g = [g_1(\theta), g_2(\theta), \dots, g_q(\theta)]$$

$$J = \begin{bmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_1}{\partial \theta_2} & \dots & \frac{\partial g_1}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_q}{\partial \theta_1} & \frac{\partial g_q}{\partial \theta_2} & \dots & \frac{\partial g_q}{\partial \theta_p} \end{bmatrix}_{Q \times P}$$

$$A \text{ es p.s.d} \rightarrow BAB' \text{ es p.s.d.}$$

Consider $\ln(w) = \beta_0 + \beta_1 \text{married} + \beta_2 \text{educ} + \varepsilon + u$, $E(u | \text{married}, \text{educ}, \varepsilon) = 0$

(a) Muestre que $E[u | x] = 0 \Rightarrow \theta_1 = 100 \cdot [\exp(\beta_1) - 1]$

$$E(\text{wage} | \text{married} = 1, \text{educ}, \varepsilon) - E(\text{wage} | \text{married} = 0, \text{educ}, \varepsilon) \\ = \beta_0 + \beta_1 + \beta_2 \text{educ} + \varepsilon - (\beta_0 + \beta_2 \text{educ} + \varepsilon) = \beta_1$$

$$\frac{E(\text{wage} | \text{married} = 1) - E(\text{wage} | \text{married} = 0)}{E(\text{wage} | \text{married} = 0)} \times 100 = \frac{\beta_1}{\beta_0 + \beta_2 \text{educ} + \varepsilon} \times 100$$

$$E \rightarrow \text{wage} = e^{\ln w} = e^{\beta_0} * e^{\beta_1 \text{married}} * e^{\beta_2 \text{educ}} * e^{\varepsilon} * e^u$$

$$E(\text{wage} | m = 1) = e^{\beta_0 + \beta_1 + \beta_2 e + \varepsilon + u} \\ - E(\text{wage} | m = 0) = e^{\beta_0 + \beta_1(0) + \beta_2 e + \varepsilon + u}$$

$$\theta_1 = \frac{e^{\beta_0 + \beta_2 e + \varepsilon} (e^{\beta_1} - 1)}{e^{\beta_0 + \beta_2 e + \varepsilon}} \times 100$$

$$\theta_1 = 100(e^{\beta_1} - 1) \Rightarrow \hat{\theta}_1 = 100(e^{\hat{\beta}_1} - 1)$$

(b) Use el método delta para encontrar $ee(\hat{\theta}_1)$

$$\text{Avar}(\hat{\theta}_1) = J(\beta_1) * \text{Avar}(\hat{\beta}_1) * J(\beta_1)' = \frac{d\hat{\theta}_1}{d\beta_1} \text{Avar}(\hat{\beta}_1)$$

$$\frac{d\hat{\theta}_1}{d\beta_1} = 100e^{\beta_1} \Rightarrow \text{Avar}(\hat{\theta}_1) = (100e^{\beta_1})^2 \text{Avar}(\hat{\beta}_1) \\ \Rightarrow ee(\hat{\theta}_1) = 100e^{\beta_1} * ee(\hat{\beta}_1)$$

4.4. Show that $\hat{B} = \frac{1}{N} \sum \hat{u}_i^2 x_i' x_i$ es consistente para $B = E(u^2 x' x)$ by showing that $\frac{1}{N} \sum \hat{u}_i^2 x_i' x_i = N^{-1} \sum u_i^2 x_i' x_i + o_p(1)$

Hint: write $\hat{u}_i^2 = u_i^2 - 2x_i u_i(\hat{\beta} - \beta) + [x_i(\hat{\beta} - \beta)]^2$ & use the facts that sample averages are $O_p(1)$ when expectations exist & that $\hat{\beta} - \beta = o_p(1)$

$$\hat{B} = \frac{1}{N} \sum \{ u_i^2 x_i' x_i - 2x_i u_i(\hat{\beta} - \beta) x_i' x_i + [x_i(\hat{\beta} - \beta)]^2 x_i' x_i \}$$

$$\hat{B} = \frac{1}{N} \sum u_i^2 x_i' x_i - \frac{1}{N} \sum 2x_i u_i(\hat{\beta} - \beta) x_i' x_i + \frac{1}{N} \sum x_i' x_i (\hat{\beta} - \beta)^2 x_i' x_i \\ = \frac{1}{N} \sum u_i^2 x_i' x_i - \frac{1}{N} 2(\hat{\beta} - \beta) \sum x_i u_i x_i' x_i + \frac{1}{N} (\hat{\beta} - \beta)^2 \sum x_i' x_i x_i' x_i$$

$$\hat{u}_i^2 = u_i^2 - 2x_i u_i(\hat{\beta} - \beta) + x_i'(\hat{\beta} - \beta)^2 x_i$$

$$4.5 \quad E(y|\vec{x}, z) = \vec{x}'\beta_{k+1} + \gamma z \quad \text{Var}(y|\vec{x}, z) = 0 \quad \gamma \neq 0$$

$$\hat{\beta} \Rightarrow E(y|\vec{w}) = \vec{w}'\hat{\beta} \quad \text{donde } \vec{w} = [x_1, x_2, \dots, x_k, z]_{1 \times (k+1)} \quad y \quad (k+1) \times 1$$

$$\tilde{\beta} \Rightarrow E(y|\vec{x}) = \vec{x}'\tilde{\beta} \quad \tilde{\beta} = (\vec{x}'\vec{x})^{-1}\vec{x}'y$$

Ambos consistentes y $E(x'z) = 0$

(a) Mostrar que $\tilde{\beta}$ es menos asintóticamente eficiente que $\hat{\beta}$

Como $\hat{\beta}$ y $\tilde{\beta}$ son consistentes: $\hat{\beta} - \beta \xrightarrow{P} 0, \tilde{\beta} - \beta \xrightarrow{P} 0$

$$\hat{\beta} = (w'w)^{-1}w'y = (w'w)^{-1}w'(x\beta + \gamma z + u) = (w'w)^{-1}w'x\beta + (w'w)^{-1}w'\gamma z + (w'w)^{-1}w'u$$

$$\Rightarrow (w'w)^{-1}w'x:$$

$$w'x = \begin{bmatrix} x' \\ z \end{bmatrix}_{k+1 \times 1} x_{1 \times k} = \begin{bmatrix} x'x \\ z'x \end{bmatrix}_{k+1 \times k}$$

$$\begin{bmatrix} x' \\ z \end{bmatrix}_{k+1 \times 1} \begin{bmatrix} x & z \end{bmatrix}_{1 \times (k+1)} = \begin{bmatrix} x'x & x'z \\ z'x & z^2 \end{bmatrix}_{k+1 \times k+1} \Rightarrow (w'w)^{-1}w'x = \begin{bmatrix} x'x x'x + x'x z'x \\ z'x x'x + z^3 x \end{bmatrix}_{k+1 \times k}$$

$$(w'w)^{-1}w'x\beta = \begin{bmatrix} x'x x'x + x'x z'x \\ z'x x'x + z^3 x \end{bmatrix}_{k+1 \times k} \beta_{k \times 1}$$

Suponga $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ $i = 1, \dots, N$ donde y_i, x_i son continuas y $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$

(a) Escriba la función de log-verosimilitud de ε_i en una muestra de tamaño N .

$$f(\varepsilon_i | \theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{\left(\frac{\varepsilon_i}{\sigma_\varepsilon} - 0\right)^2}{2}\right\} = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma_\varepsilon}\right)^2\right\}$$

$$\ln f(\varepsilon_i | \theta) = -\frac{1}{2} [\ln 2\pi + \ln \sigma_\varepsilon^2] + \frac{\varepsilon_i^2}{\sigma_\varepsilon^2} \left(-\frac{1}{2}\right)$$

$$\sum_{i=1}^N \ln f(\varepsilon_i | \theta) = -\frac{1}{2} \left[\ln 2\pi(N) + N \ln \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^N \varepsilon_i^2 \right] = \mathcal{L}(\theta | \varepsilon_i)$$

(b) Obtenga los estimadores $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_\varepsilon^2$ por el método de máxima verosimilitud. Sea detallado en cada uno de los pasos realizados.

Para encontrar $\hat{\sigma}_\varepsilon^2$:

$$\max_{\sigma_\varepsilon^2} \mathcal{L}(\theta | \varepsilon_i) = -\frac{1}{2} \left[N \ln 2\pi + N \ln \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2} \sum \varepsilon_i \right]$$

$$\text{CPO: } \frac{d\mathcal{L}}{d\sigma_\varepsilon^2} = -\frac{1}{2} \left[\frac{N}{\sigma_\varepsilon^2} - \frac{2}{(\sigma_\varepsilon^2)^2} \sum \varepsilon_i \right] = 0$$

$$\frac{\sum \varepsilon_i}{(\sigma_\varepsilon^2)^2} = \frac{N}{\sigma_\varepsilon^2}$$

$$\frac{\sum \varepsilon_i}{N} = \frac{(\sigma_\varepsilon^2)^2}{\sigma_\varepsilon^2}$$

$$\Rightarrow \hat{\sigma}_\varepsilon^2 = \frac{\sum \hat{\varepsilon}_i}{N}$$

Para encontrar $\hat{\beta}_0$ y $\hat{\beta}_1$:

$$\max_{\beta_0, \beta_1} \mathcal{L}(\theta | \varepsilon_i) = -\frac{1}{2} \left[N \ln 2\pi + N \ln \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2} \sum (y_i - \beta_0 - \beta_1 x_i) \right]$$

$$\text{CPO: } \frac{d\mathcal{L}}{d\beta_0} = -\frac{1}{2} \left[\frac{1}{\sigma_\varepsilon^2} \sum (y_i - \beta_0 - \beta_1 x_i)(-1) \right] = 0$$

$$\frac{1}{\sigma_\varepsilon^2} \sum (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum y_i - N\beta_0 - \beta_1 \sum x_i = 0$$

$$N\beta_0 = \sum (y_i - \beta_1 x_i)$$

$$\hat{\beta}_0 = \frac{\sum y_i}{N} - \beta_1 \frac{\sum x_i}{N}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\frac{d\mathcal{L}}{d\beta_1} = -\frac{1}{2} \left[\frac{1}{\sigma_\varepsilon^2} \sum (y_i - \beta_0 - \beta_1 x_i)(-x_i) \right] = 0$$

$$\beta_1 \sum x_i^2 = \sum [y_i x_i - \beta_0 x_i]$$

$$\Rightarrow \sum (y_i x_i) - \beta_0 \sum x_i$$

$$\begin{aligned}
 \max \mathcal{L}(\theta, \epsilon) &= -\frac{1}{2} [N \ln 2\pi + N \ln \sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} \sum (y_i - \beta_0 - \beta_1 x_i)] \\
 &= -\frac{1}{2} [N \ln 2\pi + N \ln \sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} (\sum y_i - N\beta_0 - \beta_1 \sum x_i)] \\
 &= -\frac{1}{2} [N \ln 2\pi + N \ln \sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} \sum y_i - \frac{N\beta_0}{\sigma_\epsilon^2} - \frac{\beta_1}{\sigma_\epsilon^2} \sum x_i]
 \end{aligned}$$

$$\text{cpo: } \frac{d\mathcal{L}}{d\beta_1} = -\frac{1}{2} \left[-\frac{1}{\sigma_\epsilon^2} \sum x_i \right]$$

$$\frac{d\mathcal{L}}{d\beta_1} = -\frac{1}{2} \left[\frac{1}{\sigma_\epsilon^2} \sum x_i \right]$$

Monitoría - Máxima Verosimilitud

Contenido: ① Definición ② Problemas de la estimación por MV
③ Ejemplo: Regresión lineal.

① DEFINICIÓN

Función de verosimilitud

Sea $X_1, \dots, X_N$ una muestra i.i.d con densidad $f(X_i, \theta)$.
 $\theta_{K \times 1}$ caracteriza $f(X_i, \theta)$

P.e: Distribución normal.

$$\theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix} \rightarrow f(X_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right\}$$

Función de densidad conjunta

$$f(X_1, \dots, X_N, \theta) = \prod_{i=1}^N f(X_i | \theta) \text{ por i.i.d}$$

P.e (no hay i.i.d): AR(1) $\rightarrow y_t = \rho y_{t-1} + \epsilon_t$

Estimador de máxima verosimilitud

Considere $y = \{y_1, \dots, y_N\}$ con densidad conjunta $\ell(y, \theta)$.
Una vez se observa y ; $\hat{\theta}_{MV} = \operatorname{argmax}_{\theta} \ell(y, \theta)$ donde se tiene que $\theta \in \Theta$.

$\hookrightarrow$ Dado que $\ln(\cdot)$ es estrictamente creciente:
 $\hat{\theta}_{MV} = \operatorname{argmax}_{\theta} \ln \ell(y, \theta)$.

P.e: $X_i \sim \text{Bernoulli}(\theta)$ donde $\theta = \Pr(X_i=1)$, $1-\theta = \Pr(X_i=0)$.

$f(X_i, \theta) = \theta^{x_i} (1-\theta)^{1-x_i}$, $X_i = \{0, 1\}$ si la muestra es i.i.d de tamaño N :

$$\mathcal{L}(\theta|x) = \ln \ell(y, \theta) = \sum_{i=1}^N \{X_i \ln \theta + (1-X_i) \ln (1-\theta)\}$$

$$\frac{d\mathcal{L}}{d\theta} = \sum_{i=1}^N \frac{X_i}{\theta} - \frac{1-X_i}{1-\theta} = 0 \rightarrow \frac{X_i}{\theta} = \frac{1-X_i}{1-\theta}$$

$$\frac{1-\theta}{\theta} = \frac{1}{X_i} - 1$$

$$\frac{1}{\theta} - 1 = \frac{1}{X_i} - 1$$

$$\frac{1}{X_i} = \frac{1}{\theta}$$

$$\hat{\theta}_{MV} = X_i$$

$$\frac{1}{\theta} \sum X_i = \frac{1}{1-\theta} \sum (1-X_i)$$

$$\frac{1-\theta}{\theta} = \frac{\sum (1-X_i)}{\sum X_i} = \frac{\sum 1 - \sum X_i}{\sum X_i} = \frac{N}{\sum X_i} - 1 = \frac{1-\theta}{\theta} = \frac{1}{\theta} - 1$$

$$\frac{1}{\theta} = \frac{N}{\sum X_i}$$

$$\hat{\theta}_{MV} = \frac{\sum X_i}{N} = \bar{x}$$

② PROBLEMAS DE LA ESTIMACIÓN POR MV

- I. No unicidad de la función de máxima verosimilitud si la función es continua.
- II. No existencia de la solución del problema de maximización. Puede ocurrir para una muestra en particular.
 ↳ ¿Por qué? Si Θ es abierto o si la función de log-verosimilitud es discontinua.
- IV. Múltiples soluciones al problema de maximización. Sin embargo, este problema puede subsanarse acotando el problema de maximización.

P.e: Considere $Y_1, \dots, Y_N$ donde Y es i.i.d $f(y_i, \theta) = \frac{1}{\pi} \left(\frac{1}{1+[y-\theta]^2} \right)$
 $N=2$.

$$f(y_1, y_2, \theta) = \frac{1}{\pi^2} \left(\frac{1}{1+[y_1-\theta]^2} \right) \left(\frac{1}{1+[y_2-\theta]^2} \right) = \frac{1}{\pi^2} [1+(y_1-\theta)^2]^{-1} [1+(y_2-\theta)^2]^{-1}$$

$$\ln f(y_1, y_2, \theta) = \left\{ \frac{1}{\pi^2} \right\} - \ln(1+[y_1-\theta]^2) - \ln(1+[y_2-\theta]^2)$$

$$\frac{d \ln f}{d \theta} = - \frac{2(y_1-\theta)(-1)}{1+(y_1-\theta)^2} - \frac{2(y_2-\theta)(-1)}{1+(y_2-\theta)^2} = 0$$

$$\frac{(y_1-\theta) + (y_1-\theta)(y_2-\theta)^2 + (y_2-\theta) + (y_2-\theta)(y_1-\theta)^2}{[1+(y_1-\theta)^2][1+(y_2-\theta)^2]} = 0$$

$$0 = \frac{1}{\pi} (y_1 + y_2) + (y_1-\theta)(y_2-2\theta y_2 + \theta^2) + (y_2-\theta)(y_1^2 - 2\theta y_1 + \theta^2)$$

$$0 = \frac{1}{\pi} [(y_1 + y_2) + y_1 y_2 - \theta y_2 - 2\theta y_1 y_2 + 2\theta^2 y_2 + \theta^2 y_1 - \theta^3 + y_1^2 y_2 - \theta y_1^2 - 2\theta y_1 y_2 + 2\theta^2 y_1 + \theta^2 y_2 - \theta^3]$$

$$0 = \frac{1}{\pi} [y_1(1+\theta^2 - \theta y_1 + 2\theta^2) + y_2(1-\theta)]$$

③ PROBLEMA DE REGRESIÓN LINEAL

Sea $Y_1, \dots, Y_N$ i.i.d donde $Y_i = X_i \beta + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$
 ↳ Asumimos $X'X$ no singular $\Rightarrow Y|X \sim N(X\beta, \sigma^2)$

$$f(y_i|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{y_i - x_i \beta}{\sigma} \right)^2 \right\}$$

$$f(y_1, y_2, \dots, y_N | X) \Rightarrow \mathcal{L}(\beta, \sigma | X) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(Y - X\beta)'(Y - X\beta)}{\sigma^2} \right\}$$

$$\ln \mathcal{L}(\beta, \sigma | X) = - \left[\frac{N}{2} \ln 2\pi + N \ln \sigma + \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \right]$$

$$\begin{aligned} \frac{d \ln \mathcal{L}}{d \beta} &= - \frac{d}{d \beta} \left\{ \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \right\} = - \frac{d}{d \beta} \left\{ \frac{(Y' - \beta' X')(Y - X\beta)}{2\sigma^2} \right\} \\ &= - \frac{d}{d \beta} \left\{ \frac{Y'Y - \beta' X'Y - Y'X\beta + \beta' X'X\beta}{2\sigma^2} \right\} \\ &= - \left[\frac{-2X'Y + 2(X'X)\beta}{2\sigma^2} \right] = 0 \Rightarrow \hat{\beta}_{MV} = (X'X)^{-1} X'Y \end{aligned}$$

$$\frac{d \ln \mathcal{L}}{d\sigma} = \left[\frac{N}{\sigma} + \frac{-2}{2\sigma^3} (Y - X\beta)'(Y - X\beta) \right] = 0$$

$$\frac{N}{\sigma} = \frac{1}{\sigma^3} \hat{\varepsilon}' \hat{\varepsilon}$$

$$\frac{\sigma^3}{\sigma} = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{N}$$

$$\hat{\sigma}_{MV}^2 = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{N}$$

Para encontrar la varianza de $\hat{\theta}_{MV} = [\hat{\beta}_{MV}, \hat{\sigma}_{MV}^2]'$:

$$\frac{d \ln \mathcal{L}}{d\beta} = - \left[\frac{-2X'Y + 2(X'X)\beta}{2\sigma^2} \right] = - \left[\frac{-X'Y + (X'X)\beta}{\sigma^2} \right]$$

$$\frac{d^2 \ln \mathcal{L}}{d\beta d\beta} = - \frac{1}{\sigma^2} (X'X) \quad \frac{d^2 \ln \mathcal{L}}{d\beta d\sigma^2} = - \frac{(-2)(-X'X[X'X]\beta)}{\sigma^3} = - \frac{2X'(Y - X\beta)}{\sigma^3}$$

$$\frac{d^2 \ln \mathcal{L}}{d\sigma^2} = - \left[- \frac{N}{\sigma^2} + \frac{3}{\sigma^4} (\varepsilon' \varepsilon) \right] = \frac{N}{\sigma^2} - \frac{3}{\sigma^4} [\varepsilon' \varepsilon]$$

$$\begin{bmatrix} - \frac{X'X}{\sigma^2} & - \frac{2X'(Y - X\beta)}{\sigma^3} \\ - \frac{2X'(Y - X\beta)}{\sigma^3} & \frac{N}{\sigma^2} - \frac{3}{\sigma^4} [\varepsilon' \varepsilon] \end{bmatrix} = A \Rightarrow E[A]$$

$$E \left[- \frac{X'X}{\sigma^2} \right] = - \frac{X'X}{\sigma^2} \quad E \left[- \frac{2X'(Y - X\beta)}{\sigma^3} \right] = - \frac{2}{\sigma^3} E[X(Y - X\beta)] = - \frac{2}{\sigma^3} E[X \varepsilon] = 0$$

$$E \left[\frac{N}{\sigma^2} - \frac{3}{\sigma^4} \varepsilon' \varepsilon \right] = \frac{N}{\sigma^2} - \frac{3}{\sigma^4} E[\varepsilon' \varepsilon] = \frac{N}{\sigma^2} - \frac{3}{\sigma^4} \sigma^2 N = \frac{1}{\sigma^2} (N - 3) = - \frac{2N}{\sigma^2}$$

$$E[A] = \begin{bmatrix} - \frac{X'X}{\sigma^2} & 0 \\ 0 & - \frac{2N}{\sigma^2} \end{bmatrix} \Rightarrow E[A]^{-1} = \text{Var-cov}(\hat{\theta}_{MV}) = \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \frac{\sigma^2}{2N} \end{bmatrix}$$

María Medellín Esquerma (201125036)

Pregunta 3

(a) Tenemos que $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Por ende, la función de densidad acumulada de ε_i (como es i.i.d) es:

$$f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N | \theta) = \frac{1}{\sqrt{2\sigma_\varepsilon^2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\varepsilon_1}{\sigma_\varepsilon}\right)^2\right\} \cdot \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2}\left(\frac{\varepsilon_2}{\sigma_\varepsilon}\right)^2\right\} \dots$$

$$= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma_\varepsilon}\right)^2\right\} = \mathcal{L}(\theta | \varepsilon)$$

$$\ln \mathcal{L}(\theta | \varepsilon) = \sum_{i=1}^N -\frac{1}{2} \left[\ln 2\pi + \ln \sigma_\varepsilon^2 + \frac{(\varepsilon_i)^2}{\sigma_\varepsilon^2} \right]$$

$$= -\frac{1}{2} \left[N \ln 2\pi + N \ln \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2} \sum (\varepsilon_i)^2 \right]$$

I EVALUACIÓN DE IMPACTO

$$Y = X\alpha + D\beta + u; \quad E[u] = 0; \quad D = \begin{cases} 1 & \text{si es tratado} \\ 0 & \text{e.o.c} \end{cases}$$

El tratamiento se asigna aleatoriamente: $\varepsilon = \begin{cases} 1 & \Rightarrow \text{se asigna} \\ 0 & \text{e.o.c} \end{cases}$

$D \rightarrow$ Participación en un mundo sin experimento (yo lo quiero)

$A = 1$ si participa efectivamente

(a) Conformidad perfecta con el experimento

$A = 1$ si $\varepsilon = 1 \rightarrow$ ¡quería el tratamiento, y efectivamente lo recibió

$A = 0$ si $\varepsilon = 0$

(b) Conformidad imperfecta con el experimento

$A = D\varepsilon \rightarrow$ Sólo recibe el tratamiento si quiere y "puede".

ATE $E(Y_i - Y_0 | X = x) = E[Y_i | X = x] - E[Y_0 | X = x]$
 Average treatment effect
 $ATE = X\alpha + \beta + E[u | X = x] - X\alpha - E[u | X = x]$
 $ATE = \beta$

TT $E[Y_i - Y_0 | X = x, A = 1] = \beta$
 Treatment on treated

$\tilde{TT}$ $E[Y_i - Y_0 | X = x, D = 1] = \beta$

TUT $E[Y_i - Y_0 | X = x, D = 0]$

Conformidad perfecta: $E[Y | A = 1, X = x] = X\alpha + \beta + E[u | A = 1, X = x]$
 $= X\alpha + \beta + E[u | X = x]$
 $E[Y | A = 0, X = x] = X\alpha + E[u | A = 0, X = x]$
 $= X\alpha + E[u | X = x]$

$E[Y | A = 1, X = x] - E[Y | A = 0, X = x] = \beta$

Conformidad imperfecta:

$E[Y | A = 1, X = x] = E[Y | D = 1, \varepsilon = 1, X = x]$
 $= X\alpha + \beta + E[u | D = 1, \varepsilon = 1, X = x]$
 $= X\alpha + \beta + E[u | D = 1, X = x]$

$E[Y | A = 0, X = x] = X\alpha + E[u | A = 0, X = x]$
 $= X\alpha + \left\{ E[u | D = 0, \varepsilon = 0, X = x] \frac{P(D = 0, \varepsilon = 0 | X = x)}{1 - P(A = 1 | X = x)} \right.$
 $+ E[u | D = 0, \varepsilon = 1, X = x] \frac{P(D = 0, \varepsilon = 1 | X = x)}{1 - P(A = 1 | X = x)}$
 $\left. + E[u | D = 1, \varepsilon = 0, X = x] \frac{P(D = 1, \varepsilon = 0 | X = x)}{1 - P(A = 1 | X = x)} \right\}$

II TRUNCAMIENTO

Considere $y_i = d_i (x_i' \beta_0 + \varepsilon_i) + (1 - d_i) k_i \quad i = 1, 2, \dots, n$

Donde $d_i = I(x_i' \beta_0 + \varepsilon_i \leq k_i)$

Se observa $\{y_i, x_i, k_i\}$, asumimos $\{\varepsilon_i, x_i, k_i\} \sim \text{i.i.d } N(0, \sigma^2)$

Considere $y_i^* = x_i' \beta_0 + \varepsilon_i \rightarrow y_i = \begin{cases} 1 & \text{si } y_i^* \leq k \rightarrow d_i = 1 \\ 0 & \text{e.o.c} \rightarrow d_i = 0 \end{cases}$

(a) Derivar la log-verosimilitud.

$$\begin{aligned} P(y_i = k | X, k) &= P(y_i^* \leq k | X, k) = P(x_i' \beta_0 + \varepsilon_i \leq k | X, k) \\ &= P(\varepsilon_i \leq k_i - x_i' \beta_0 | X, k) \\ &= \Phi\left(\frac{\varepsilon_i}{\sigma} \leq \frac{k_i - x_i' \beta_0}{\sigma} | X, k\right) \end{aligned}$$

