# PUBLICACIONES DE 4er CURSO

**Curso: 4º**

**Grado: Economía**

**Asignatura: ECONOMETRÍA III**

**TRANSPARENCIAS PARTE 4: OTROS TEMAS**

**TEMA 9: MODELOS PARA DATOS PANEL**

**Profesores: Antonio Aznar**

**Departamento de ANÁLISIS ECONÓMICO**

**Curso Académico 2015/16**


Facultad de
Economía y Empresa
Universidad Zaragoza

# Tema 9. MODELOS PARA DATOS PANEL

Índice

1. Motivación
2. Datos de Panel con dos Periodos
3. Regresión de Efectos Fijos

# 1. Motivación

A banco de datos de panel contiene observaciones de múltiples entidades (individuos, estados, empresas,…), en donde cada entidad es observada en dos o más periodos temporales.

*Ejemplos:*

- Datos de 420 distritos escolares de California en 1999 y en 2000.
- Datos sobre los 50 estados americanos en tres años diferentes.
- Datos de 400 individuos en cuatro meses diferentes.

Un doble subíndice distingue entidades (estados) y periodos temporales (años).

$i$ = entidad (estado), $n$ = número de entidades

por lo que, $i = 1,\ldots,n$

$t$ = periodo temporal (año), $T$ = número de periodos temporales, por lo que $t = 1,\ldots,T$

Datos: suponiendo un regresor, los datos son:

$$(X_{it}, Y_{it}),\ i = 1,\ldots,n,\ t = 1,\ldots,T$$

Si hay k regresores:

$$(X_{1it}, X_{2it},\ldots,X_{kit}, Y_{it}),\ i = 1,\ldots,n,\ t = 1,\ldots,T$$

$n$ = número de entidades (estados)
$T$ = número de periodos temporales (años)

Otros términos…

- Los datos panel también se llaman **datos longitudinales**.

***Un panel es balanceado cuando no falta ninguna observación.***

## Regresión con Datos Panel ¿Por qué es util?

La idea clave es esta:

Si una variable omitida no cambia en el tiempo, entonces cualquier cambio de Y en el tiempo no puede ser causado por la variable omitida.

## Ejemplo: Muertes de Tráfico e impuestos sobre el alcohol

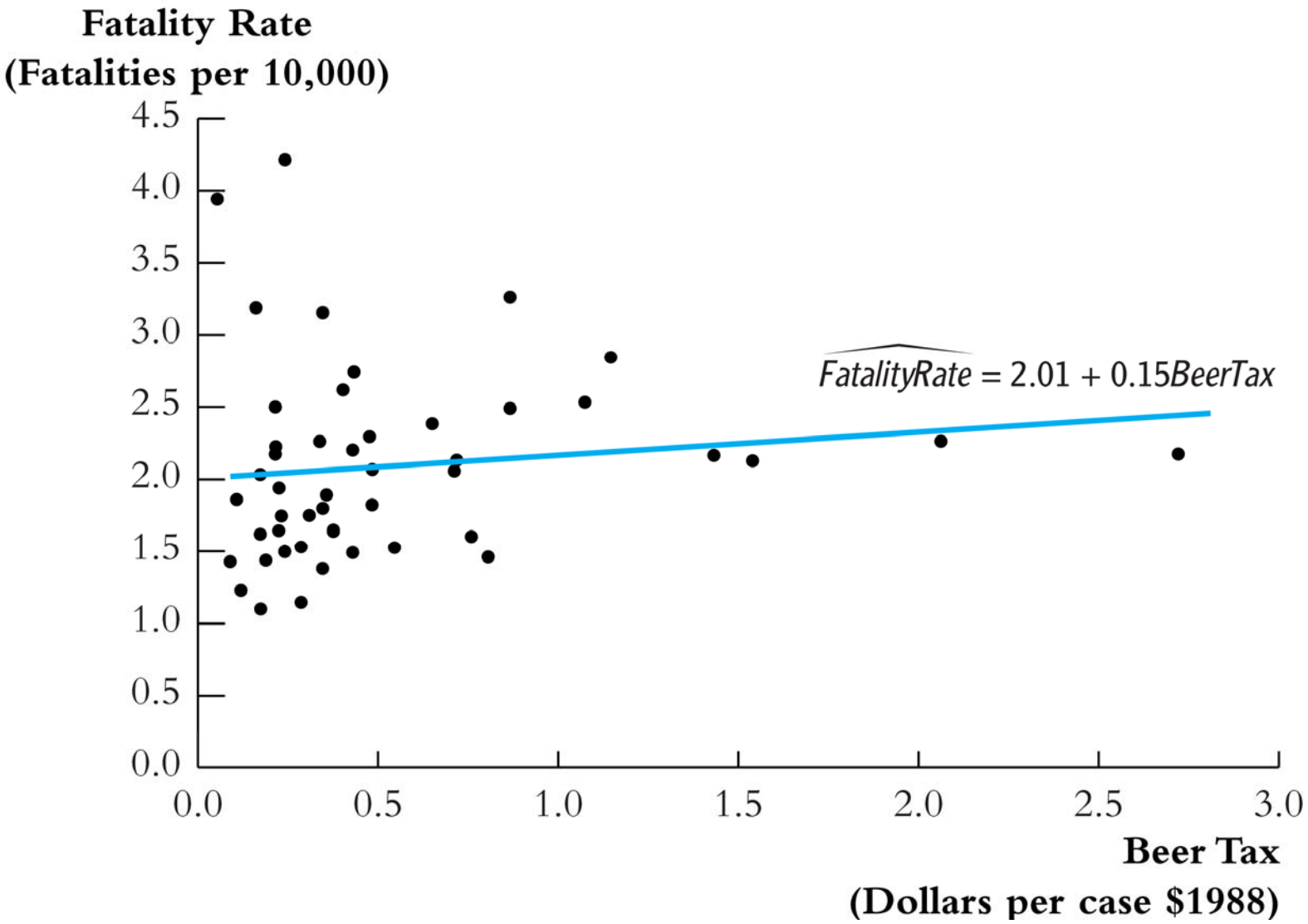
Unidad de observación: un año de un estado de USA

- 48 Estados, por lo que $n$ = # de entidades = 48
- 7 años (1982,…, 1988), por lo que $T$ = # de periodos temporales = 7
- panel balanceado, de forma que el # total de observaciones = $7 \times 48 = 336$

Variables:

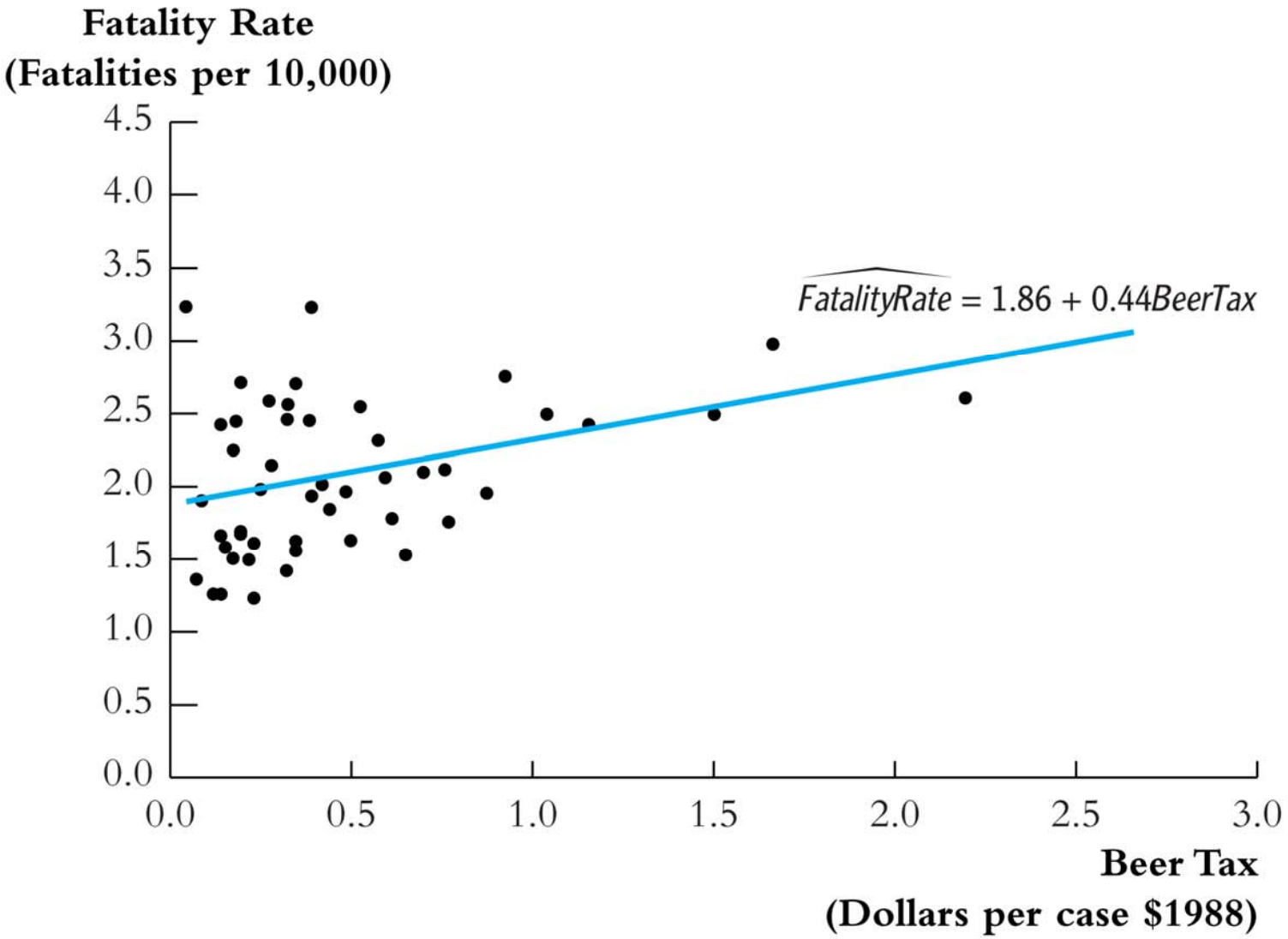
- Tasa de mortalidad de tráfico (# número de muertes de tráfico en un estado en cada año por cada 10.000 residentes)
- Impuesto sobre una caja de cerveza
- Otras (edad permitida, leyes sobre conducir bebido,…)

# Datos sobre muertos de tráfico en USA para 1982:




¿Impuestos más altos, más muertes?

Datos sobre muertos de tráfico en USA para 1988:




¿Impuestos más altos, más muertes?

¿Cómo puede explicarse esta asociación positiva entre las muertes de tráfico y los impuestos?
Otros factores que determinan la tasa de mortalidad por tráfico

- Calidad de las carreteras
- "Cultura" sobre el beber y el conducir
- Densidad de coches en la carretera

Estos factores omitidos pueden causar el sesgo de variable omitida.
*Ejemplo* #1: densidad de tráfico. Suponer:

(i) Cuanto mayor la densidad de tráfico más muertes
(ii) Los estados del oeste tienen menor densidad de tráfico y, también, impuestos sobre el alcohol más bajos.

- Entonces, se satisfacen las dos condiciones para la existencia del sesgo de variable omitida. Específicamente, “impuestos altos” pueden reflejar “densidad de tráfico alta” (de forma que el estimador MCO del coeficiente tenga un sesgo positivo. – impuestos altos, más muertes)
- Los datos panel nos permiten eliminar el sesgo de variable omitida si las variables omitidas son constantes en el tiempo dentro de cada estado.

## 2. Datos de Panel con dos periodos temporales

Considar el modelo de datos de panel,

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_{it}$$

$Z_i$ es un factor que no cambia en el tiempo (densidad), al menos en los años para los que tenemos datos.

- Suponer que $Z_i$ no es observada, por lo que su omisión puede provocar el sesgo de variable omitida.
- El efecto de $Z_i$ puede eliminarse utilizando $T = 2$ años.

La idea clave:

Cualquier cambio en la tasa de mortalidad (fatality rate) de1982 a 1988 no puede ser explicado por $Z_i$, poque $Z_i$ (por hipótesis) no cambia entre 1982 y 1988.

$$FatalityRate_{i1988} = \beta_0 + \beta_1 BeerTax_{i1988} + \beta_2 Z_i + u_{i1988}$$
$$FatalityRate_{i1982} = \beta_0 + \beta_1 BeerTax_{i1982} + \beta_2 Z_i + u_{i1982}$$

Suponer $E(u_{it}|BeerTax_{it}, Z_i) = 0$.

Haciendo la resta 1988 – 1982 (esto es, calculando el cambio), elimina el efecto de $Z_i$…

*Restando, se tiene que:*

$$FatalityRate_{i1988} - FatalityRate_{i1982} = \beta_1(BeerTax_{i1988} - BeerTax_{i1982}) + (u_{i1988} - u_{i1982})$$

- El nuevo término de error , $(u_{i1988} - u_{i1982})$, no está correlacionado ni con $BeerTax_{i1988}$ ni con $BeerTax_{i1982}$.
- Esta ecuación en diferencias puede ser estimada con MCO , incluso aunque $Z_i$ no es observada.
- La variable omitida $Z_i$ no cambia por lo que no puede explicar el cambio de Y.
- Esta regresión en diferencias no tiene constante- fue eliminada con la resta.

*Ejemplo: Muertes de tráfico e impuestos sobre la cerveza*

1982 data:

$$\widehat{FatalityRate} = 2.01 + 0.15BeerTax \qquad (n = 48)$$
$$(.15)\quad(.13)$$

1988 data:

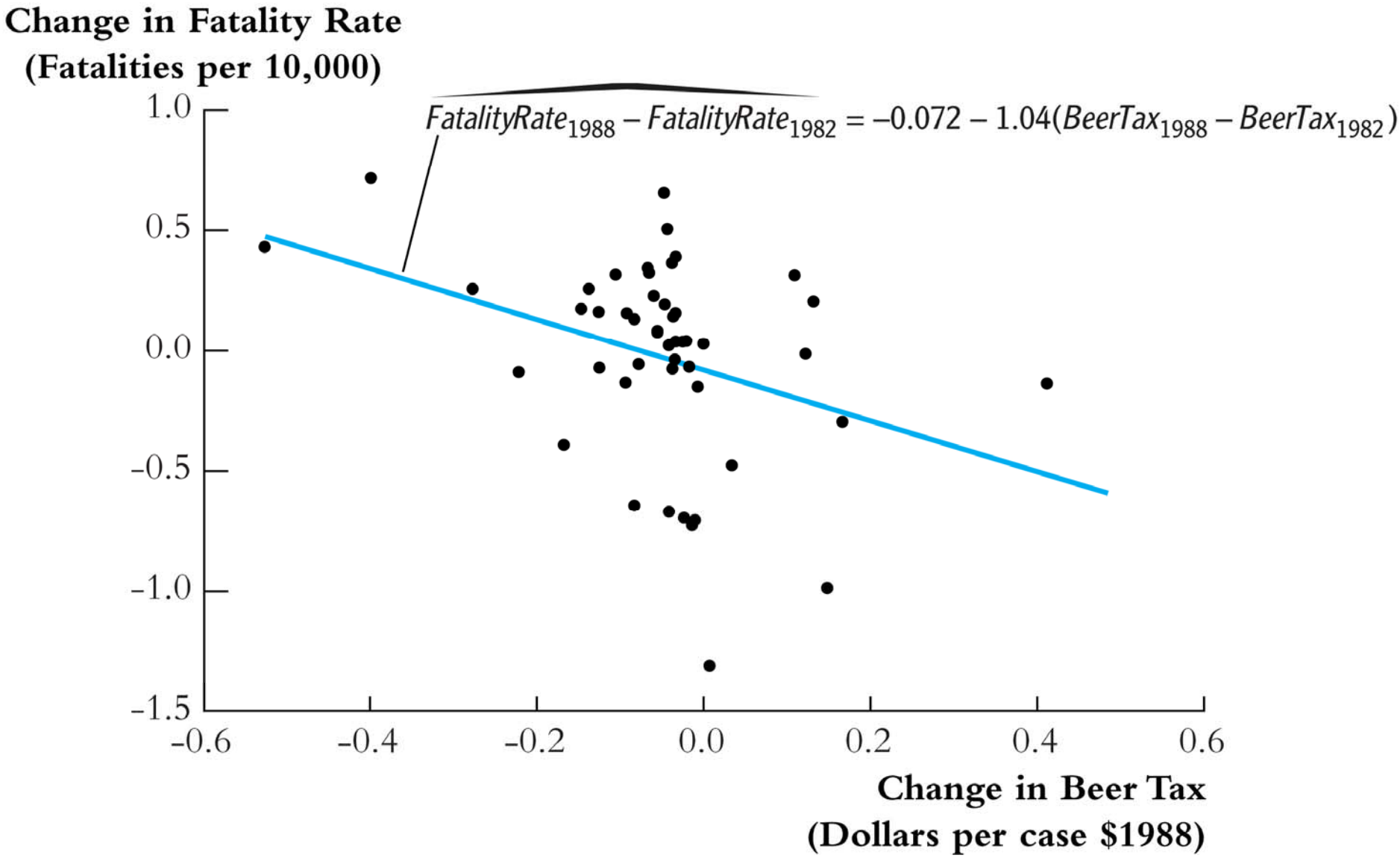
$$\widehat{FatalityRate} = 1.86 + 0.44BeerTax \qquad (n = 48)$$
$$(.11)\quad(.13)$$

Difference regression ($n = 48$)

$$\widehat{FR_{1988} - FR_{1982}} = -.072 - 1.04(BeerTax_{1988} - BeerTax_{1982})$$
$$(.065)\quad(.36)$$

*Se incluye una constante para permitir una media de la tasa de mortalidad (FR ) diferente de cero –mas sobre esto más tarde…*

# ΔTasa de mortalidad v. ΔImpuesto cerveza:



## *Regresión de Efectos Fijos*

¿Qué ocurre si hay más de dos periodos ($T > 2$)?

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it},\ i = 1,\ldots,n,\ T = 1,\ldots,T$$

Podemos reescribir este modelo en dos formas útiles:

1. Modelo de regresión con n-1 variables ficticias
2. Modelo de regresión de “Efectos Fijos” .

Primero, escribimos la forma de “efectos fijos” . Suponer que tenemos $n = 3$ estados: California, Texas, and Massachusetts.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it},\ i = 1,\ldots,n,\ T = 1,\ldots,T$$

La regresión para California (esto es, $i$ = CA):

$$\begin{aligned} Y_{CA,t} &= \beta_0 + \beta_1 X_{CA,t} + \beta_2 Z_{CA} + u_{CA,t} \\ &= (\beta_0 + \beta_2 Z_{CA}) + \beta_1 X_{CA,t} + u_{CA,t} \end{aligned}$$

or

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

- $\alpha_{CA} = \beta_0 + \beta_2 Z_{CA}$ no cambia en el tiempo
- $\alpha_{CA}$ es la constante para CA, y $\beta_1$ es la pendiente.
- La constante es única para CA, pero la pendiente es la misma para todos los estados: líneas paralelas.

Para Texas(TX):

$$Y_{TX,t} = \beta_0 + \beta_1 X_{TX,t} + \beta_2 Z_{TX} + u_{TX,t}$$
$$= (\beta_0 + \beta_2 Z_{TX}) + \beta_1 X_{TX,t} + u_{TX,t}$$

o

$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}, \text{ where } \alpha_{TX} = \beta_0 + \beta_2 Z_{TX}$$

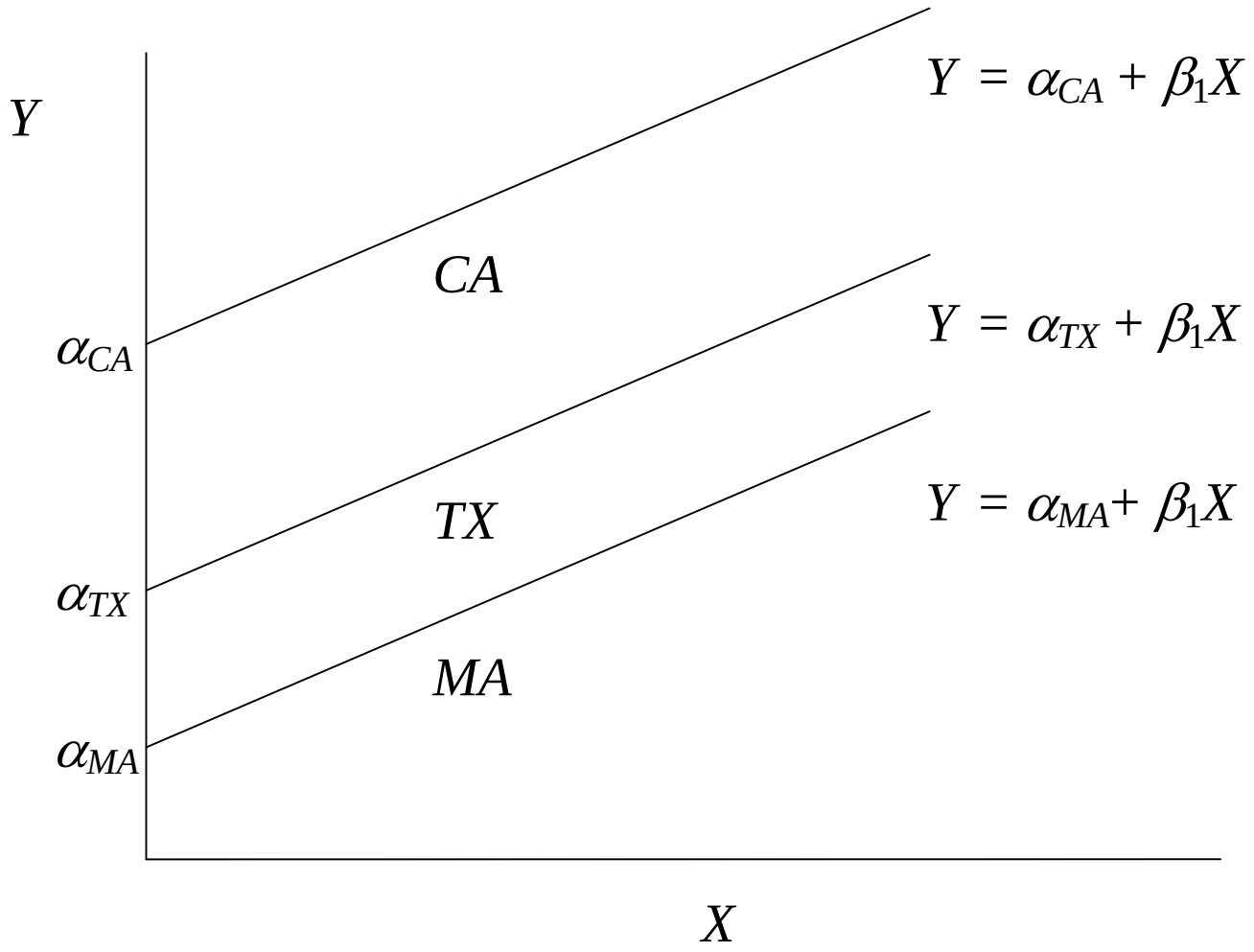
Las regresiones para los tres estados son:

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$
$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}$$
$$Y_{MA,t} = \alpha_{MA} + \beta_1 X_{MA,t} + u_{MA,t}$$

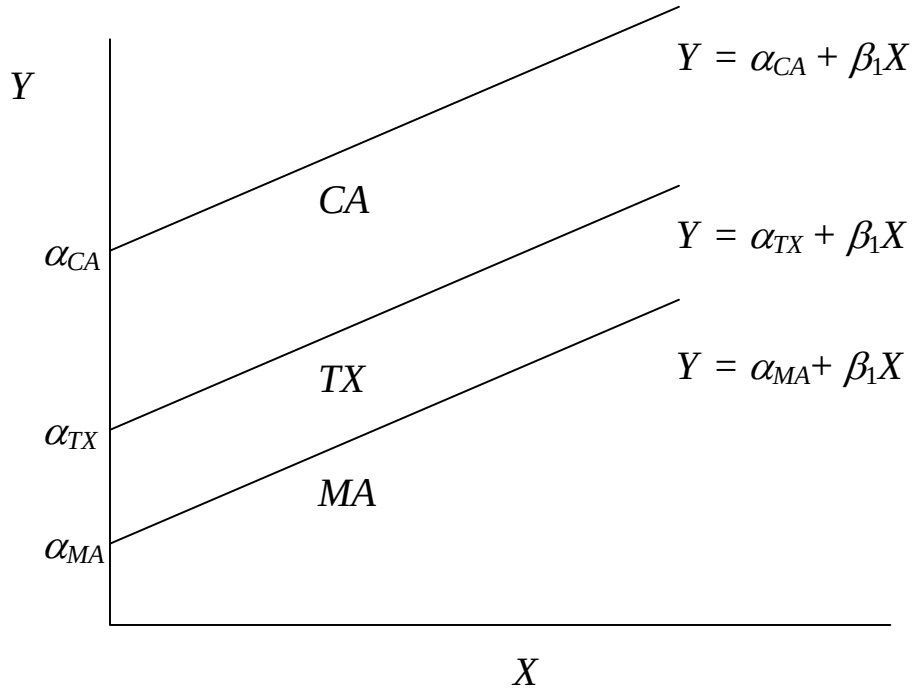
o

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}, \; i = \text{CA, TX, MA}, \; T = 1,\ldots,T$$

# La linea de regression para cada estado en un gráfico




Recordar que los cambios en la constante pueden ser representados utilizando regresores binarios.




Con variables binarias:

$$Y_{it} = \beta_0 + \gamma_{CA} DCA_i + \gamma_{TX} DTX_i + \beta_1 X_{it} + u_{it}$$

- $DCA_i$ = 1 si el estado es *CA*, = 0 en otro caso
- $DTX_t$ = 1 si el estado es *TX*, = 0 en otro caso
- Dejar fuera (leave out) $DMA_i$ (*¿Por qué*?)

## Resumen: dos formas de escribir el modelo de efectos fijos

1. **La forma con variables ficticias**

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \ldots + \gamma_n Dn_i + u_{it}$$

donde $D2_i = \begin{cases} 1 \text{ para } i=2 \text{ (estado \#2)} \\ 0 \text{ en otro caso} \end{cases}$, etc.

2. **La forma de efectos fijos:**

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

- $\alpha_i$ se llama un "efecto fijo de estado" o "efecto estado" – es el efecto constante (fijo) de estar en el estado i.

## Regresión de Efectos Fijos: Estimación

Tres métodos de Estimación:

a) Regresión MCO con variables ficticias
b) Regresión MCO con variables en desviación con respecto a la media temporal.
c) Especificación en incrementos (solamente funciona para $T = 2$)

- Los tres métodos producen idénticas estimaciones de los coeficientes e idénticos errores estándar.
- Ya comentado el tercer método que solo funciona para T=2.
- Los métodos #1 y #2 sirven para cualquier T.
- El método #1 poco práctico si n es demasiado grande.

## a. Regresión MCO con *n*-1 variables ficticias

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \ldots + \gamma_n Dn_i + u_{it} \qquad (1)$$

donde $D2_i = \begin{cases} 1 \text{ for } i=2 \text{ (state \#2)} \\ 0 \text{ otherwise} \end{cases}$ etc.

- Primero crear las variables binarias $D2_i,\ldots,Dn_i$
- Después, estimar (1) con MCO(OLS)
- La inferencia (contraste de hipótesis, intervalos de confianza) es la ya comentada (utilizando errores estandar heterocedásticos robustos).
- Este método es poco práctico cuando n es muy grande. (por ejemplo si trabajamos con $n$ = 1000 entidades)

## b. Regresión MCO con variables en desviaciones respecto a la media temporal

El modelo de efectos fijos es:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

El promedio temporal para cada entidad,

$$\frac{1}{T}\sum_{t=1}^{T} Y_{it} = \alpha_i + \beta_1 \frac{1}{T}\sum_{t=1}^{T} X_{it} + \frac{1}{T}\sum_{t=1}^{T} u_{it}$$

La desviación respecto a los promedios

$$Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \beta_1 \left( X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it} \right) + \left( u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it} \right)$$

o

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \qquad (2)$$

donde $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ y $\tilde{X}_{it} = X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}$

- $\tilde{X}_{it}$ y $\tilde{Y}_{it}$ son datos en desviaciones
- Para $i$=1 y $t$ = 1982, $\tilde{Y}_{it}$ es la diferencia entre la tasa de mortalidad en Alabama en 1982, y el promedio para ese estado calculado con los 7 años.

- Primero, construir las desviaciones $\tilde{Y}_{it}$ y $\tilde{X}_{it}$
- Después, estimar (2) regresando $\tilde{Y}_{it}$ sobre $\tilde{X}_{it}$ utilizando MCO
- Esto es como el enfoque de “cambios” pero en lugar de desviaciones con respecto a $Y_{i1}$ se toman con respecto a los promedios temporales dentro de cada estado.

# Ejemplo: Muertes de tráfico e impuesto cerveza (Gretl)

```
. xModelo 2: Efectos fijos, utilizando 336 observaciones
Se han incluido 48 unidades de sección cruzada
Largura de la serie temporal = 7
Variable dependiente: fatalityrate
Desviaciones típicas robustas (HAC)

             Coeficiente   Desv. Típica   Estadístico t    Valor p
  ------------------------------------------------------------------
  const        2.37707       0.148007        16.06        7.47e-042 ***
  beertax     -0.655874      0.288368        -2.274       0.0237    **

Media de la vble. dep.  2.040444   D.T. de la vble. dep.   0.570194
Suma de cuad. residuos  10.34537   D.T. de la regresión    0.189859
R-cuadrado MCVF (LSDV)  0.905015   R-cuadrado 'intra'      0.040745
F(48, 287) MCVF         56.96916   Valor p (de F)          2.0e-120
Log-verosimilitud       107.9727   Criterio de Akaike     -117.9454
Criterio de Schwarz     69.09305   Crit. de Hannan-Quinn  -43.38662
rho                     0.240535   Durbin-Watson           1.106864
```

```
Contraste conjunto de los regresores nombrados -
  Estadístico de contraste: F(1, 287) = 12.1904
  con valor p  = P(F(1, 287) > 12.1904) = 0.000555971

Contraste de diferentes interceptos por grupos -
  Hipótesis nula: Los grupos tienen un intercepto común
  Estadístico de contraste: F(47, 287) = 52.1792
  con valor p  = P(F(47, 287) > 52.1792) = 7.74337e-115tset state year;
```

Ejemplo**.** **Para** $n = 48$, $T = 7$:

$$\widehat{FatalityRate} = -.66 BeerTax + \text{Efectos Fijos estatales}$$
$$(.29)$$

- ¿Debe informarse de la constante?
- ¿Cuántas variables ficticias se incluirán para estimar los coeficientes del modelo?

- Comparar la pendiente y el error estándar con los obtenidos con la especificación en variaciones entre 1988 v. 1982 ($T = 2$, $n = 48$) (notar que esta incluye una constante):

$$\widehat{FR_{1988} - FR_{1982}} = -.072 - 1.04(BeerTax_{1988} - BeerTax_{1982})$$
$$(.065) \quad (.36)$$