

Datos de Panel: Efectos Fijos

Controlando por Heterogeneidad No Observada

EC3003B - Economía Laboral Aplicada

Tecnologico de Monterrey

Jueves 19 de febrero, 2025 | 3-5pm

Contenido

- 1 Introduccion a Datos de Panel
- 2 El Modelo de Efectos Fijos
- 3 Implementacion en Stata
- 4 Consideraciones Practicas
- 5 Aplicacion al Proyecto
- 6 Resumen

¿Que son Datos de Panel?

Estructura de datos:

- Observamos las **mismas unidades** en **multiples periodos**
- $i = 1, \dots, N$ individuos
- $t = 1, \dots, T$ periodos
- Total: $N \times T$ observaciones

¿Que son Datos de Panel?

Estructura de datos:

- Observamos las **mismas unidades** en **múltiples periodos**
- $i = 1, \dots, N$ individuos
- $t = 1, \dots, T$ periodos
- Total: $N \times T$ observaciones

Ejemplos:

- ENOE: mismas personas seguidas trimestralmente
- Empresas seguidas anualmente
- Países en múltiples años

Ventaja clave

Podemos controlar por características **no observadas** que no cambian en el tiempo.

El Problema de Variables Omitidas

En corte transversal:

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

Si hay una variable omitida A_i (ej. habilidad) correlacionada con X_i :

$$\varepsilon_i = A_i + u_i \quad \text{con } \text{Corr}(X_i, A_i) \neq 0$$

\Rightarrow OLS esta sesgado.

El Problema de Variables Omitidas

En corte transversal:

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

Si hay una variable omitida A_i (ej. habilidad) correlacionada con X_i :

$$\varepsilon_i = A_i + u_i \quad \text{con } \text{Corr}(X_i, A_i) \neq 0$$

\Rightarrow OLS esta sesgado.

Solucion con panel

Si A_i es **constante en el tiempo**, podemos eliminarla usando la variacion **within** (dentro del individuo).

Ejemplo: Retorno a Educacion

Problema clasico:

- Personas mas habiles obtienen mas educacion
- Personas mas habiles ganan mas (independiente de educacion)
- Habilidad no observada \rightarrow sesgo hacia arriba en $\hat{\beta}_{educ}$

Ejemplo: Retorno a Educacion

Problema clasico:

- Personas mas habiles obtienen mas educacion
- Personas mas habiles ganan mas (independiente de educacion)
- Habilidad no observada \rightarrow sesgo hacia arriba en $\hat{\beta}_{educ}$

Con panel:

- Seguimos a la misma persona en el tiempo
- Su “habilidad innata” es constante
- Comparamos su salario **antes y despues** de obtener mas educacion
- La habilidad se “cancela”

Modelo con efectos individuales:

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

Donde:

- Y_{it} : variable dependiente (salario)
- X_{it} : variables independientes que varían en el tiempo
- α_i : **efecto fijo** individual (constante en t)
- ε_{it} : error idiosincrático

Modelo con efectos individuales:

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

Donde:

- Y_{it} : variable dependiente (salario)
- X_{it} : variables independientes que varían en el tiempo
- α_i : **efecto fijo** individual (constante en t)
- ε_{it} : error idiosincrático

α_i captura

Todo lo que es específico del individuo y no cambia: habilidad, motivación, antecedentes familiares, personalidad, etc.

Estimacion: Within Transformation

Paso 1: Calcular medias por individuo

$$\bar{Y}_i = \beta \bar{X}_i + \alpha_i + \bar{\varepsilon}_i$$

Paso 2: Restar la media (“demeaning”)

$$Y_{it} - \bar{Y}_i = \beta(X_{it} - \bar{X}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Estimacion: Within Transformation

Paso 1: Calcular medias por individuo

$$\bar{Y}_i = \beta \bar{X}_i + \alpha_i + \bar{\varepsilon}_i$$

Paso 2: Restar la media (“demeaning”)

$$Y_{it} - \bar{Y}_i = \beta(X_{it} - \bar{X}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Resultado:

- α_i desaparece (es constante, su media es el mismo)
- Usamos solo variacion **within** (dentro del individuo)
- OLS en datos transformados da estimador de efectos fijos

Efectos Fijos vs Efectos Aleatorios

	Efectos Fijos (FE)	Efectos Aleatorios (RE)
Supuesto	$\text{Corr}(\alpha_i, X_{it}) \neq 0$	$\text{Corr}(\alpha_i, X_{it}) = 0$
Variacion usada	Solo within	Within + between
Variables invariantes	No se pueden estimar	Se pueden estimar
Eficiencia	Menor	Mayor (si supuesto vale)

Efectos Fijos vs Efectos Aleatorios

	Efectos Fijos (FE)	Efectos Aleatorios (RE)
Supuesto	$\text{Corr}(\alpha_i, X_{it}) \neq 0$	$\text{Corr}(\alpha_i, X_{it}) = 0$
Variacion usada	Solo within	Within + between
Variables invariantes	No se pueden estimar	Se pueden estimar
Eficiencia	Menor	Mayor (si supuesto vale)

¿Cual usar?

- En economia laboral, casi siempre hay correlacion → **FE**
- Test de Hausman: compara FE vs RE
- Si rechazamos H_0 , usar FE

```
* Declarar datos como panel
```

```
xtset id_persona año
```

```
* Efectos fijos
```

```
xtreg ln_salario experiencia experiencia2 formal, fe
```

```
* Efectos aleatorios
```

```
xtreg ln_salario experiencia experiencia2 formal, re
```

```
* Test de Hausman
```

```
hausman fe re
```

Ejemplo Completo

```
* Cargar ENOE panel
use "datos/enoe_panel.dta", clear

* Declarar panel
xtset id trimestre

* Variables
gen ln_salario = ln(salario_hora)
gen experiencia2 = experiencia^2

* Efectos fijos con errores cluster
xtreg ln_salario experiencia experiencia2 formai i.sector, fe ///
      vce(cluster id)

* Comparar con pooled OLS
reg ln_salario experiencia experiencia2 formai i.sector, cluster(id)
```


Test de Hausman

* Guardar estimaciones

```
quietly xtreg ln_salario experiencia experiencia2 formal, fe  
estimates store fe
```

```
quietly xtreg ln_salario experiencia experiencia2 formal, re  
estimates store re
```

* Test de Hausman

```
hausman fe re
```

* Interpretacion:

* H_0 : Diferencia en coeficientes no es sistematica (usar RE)

* Si $p < 0.05$: Rechazar $H_0 \rightarrow$ usar FE

Variable	Pooled OLS	FE	RE
Experiencia	0.042*** (0.002)	0.028*** (0.005)	0.035*** (0.003)
Formal	0.285*** (0.010)	0.120*** (0.025)	0.180*** (0.015)
R^2 within	–	0.12	0.10
R^2 between	–	0.35	0.38
N	150,000	150,000	150,000

Nota: FE da coeficientes menores → parte del efecto en OLS era heterogeneidad no observada.

❶ No estima variables invariantes en el tiempo

- Genero, raza, lugar de nacimiento
- Se “absorben” en α_i

① No estima variables invariantes en el tiempo

- Genero, raza, lugar de nacimiento
- Se “absorben” en α_i

② Requiere variacion within

- Si X casi no cambia para nadie, poca precision
- Ej: educacion cambia poco en adultos

① No estima variables invariantes en el tiempo

- Genero, raza, lugar de nacimiento
- Se “absorben” en α_i

② Requiere variacion within

- Si X casi no cambia para nadie, poca precision
- Ej: educacion cambia poco en adultos

③ Solo controla por invariantes en el tiempo

- No controla por shocks que varian en el tiempo
- Ej: cambios en motivacion, salud

Problemas potenciales:

- Correlacion serial: $\text{Corr}(\varepsilon_{it}, \varepsilon_{is}) \neq 0$
- Heterocedasticidad entre individuos

Problemas potenciales:

- Correlacion serial: $\text{Corr}(\varepsilon_{it}, \varepsilon_{is}) \neq 0$
- Heterocedasticidad entre individuos

Solucion: Errores cluster por individuo

- Permiten correlacion arbitraria dentro del cluster
- Robustos a heterocedasticidad

Recomendacion

Siempre usar `vce(cluster id)` en panel.

Para la empresa cliente:

- Si tuvieramos datos de empleados en multiples anos...
- Podriamos estimar efecto de promociones controlando por α_i
- Ver como cambia el salario cuando alguien sube de nivel

Para la empresa cliente:

- Si tuviéramos datos de empleados en múltiples años...
- Podríamos estimar efecto de promociones controlando por α_i
- Ver cómo cambia el salario cuando alguien sube de nivel

Con datos de ENOE (panel rotativo):

- Seguimos personas 5 trimestres
- Estimamos efecto de cambio de empleo formal/informal
- Controlamos por habilidad no observada

Efectos Fijos:

- Controla por α_i invariante
- Usa variación within
- Robusto a correlación α_i - X

Comandos Stata:

- `xtset id t`
- `xtreg y x, fe`
- `vce(cluster id)`
- `hausman`

Mensaje clave

Panel permite controlar por heterogeneidad no observada constante, acercandonos mas a efectos causales.

¿Preguntas?

Proxima Sesión:

M09: Panel - Antigüedad y Trayectorias

Lunes 23 de febrero, 3-5pm