

Análisis Econométrico con Datos de Panel

Practica 3

José Burgos 25-0140

2025-08-15

Preguntas conceptuales

1. ¿Qué se entiende por datos de panel y qué ventajas ofrece respecto a datos de corte transversal?

Los **datos de panel** observan un conjunto de individuos a lo largo de múltiples periodos de tiempo, combinando dimensiones de corte transversal y series de tiempo. Pueden ser balanceados o no balanceados.

Ventajas principales: Proporcionan información sobre el comportamiento individual a través del tiempo. Permiten controlar la heterogeneidad no observada entre individuos que podría estar correlacionada con los regresores (e.g., modelos de efectos fijos). Ofrecen un mayor número de observaciones ($N * T$), lo que puede mejorar la precisión de las estimaciones.

2. Explique la diferencia entre los modelos de efectos fijos y efectos aleatorios. ¿Qué implicaciones tiene sobre la correlación entre efectos individuales no observados y las covariables?

La diferencia clave radica en la correlación entre los efectos individuales no observados (α_i) y los regresores (covariables): Efectos Fijos (EF): Permite que α_i esté correlacionado con los regresores. Si existe esta correlación, el estimador EF es consistente. Efectos Aleatorios (EA): Asume que α_i es independiente de los regresores. Si esta independencia no se cumple, el estimador EA es inconsistente. Si se cumple, EA es más eficiente que EF.

3. ¿En qué consiste la prueba de Hausman y qué criterio se utiliza para decidir entre efectos fijos y aleatorios?

La **prueba de Hausman** es un contraste estadístico que evalúa si existe una diferencia significativa entre los estimadores de efectos fijos y efectos aleatorios. Solo se aplica a regresores que varían en el tiempo.

Criterio de decisión: Si el contraste es no significativo (p-valor alto): No se rechaza la hipótesis nula, lo que sugiere que no hay correlación entre los efectos individuales y los regresores. Se elige el modelo de Efectos Aleatorios (EA) por ser más eficiente. Si el contraste es significativo (p-valor bajo): Se rechaza la hipótesis nula, lo que indica *correlación* entre los efectos individuales y los regresores. Se elige el modelo de Efectos Fijos (FE) por ser consistente en esta situación.

4. ¿Qué es el “within transformation”? ¿Cómo permite eliminar el efecto no observado en el modelo de efectos fijos?

El “**within transformation**” es la técnica utilizada por el estimador de efectos fijos que emplea solo la variación dentro de cada individuo a lo largo del tiempo. Consiste en descontar la media temporal de cada variable para cada individuo.

Esta transformación elimina el efecto individual no observado (α_i) porque, al ser constante para cada individuo en el tiempo, se cancela al restarle su propia media temporal ($\alpha_i - \alpha_i = 0$). Esto permite estimar los coeficientes sin sesgo debido a α_i . Las variables que no varían en el tiempo son eliminadas.

5. ¿Qué problemas pueden surgir si hay autocorrelación o heterocedasticidad en paneles no balanceados? ¿Cómo pueden corregirse?

- **Problemas:** Tanto la autocorrelación (correlación serial de errores) como la heterocedasticidad (varianza no constante de los errores) hacen que los errores estándar estimados sean incorrectos. Esto lleva a inferencias estadísticas (pruebas de hipótesis, intervalos de confianza) inválidas, aunque los estimadores de los coeficientes puedan ser consistentes.
- **Corrección:** Suelen corregirse usar errores estándar robustos adecuados para paneles no balanceados, como: Robustos a heterocedasticidad: `vcovHC()` con tipo `arellano` o `HC1`. Robustos a autocorrelación y heterocedasticidad: errores estándar agrupados (`clustered`) por individuo o por dos vías (individuo y tiempo).

Parte II: Ejercicios aplicados con Datos de Wooldridge

Ejercicio 1: Efectos fijos y aleatorios (base wagepan.csv)

Variable dependiente: `lwage` (log del salario horario)

Variables explicativas: `educ`, `exper`, `black`, `married`, `union`

```
library(plm)
library(stargazer)
library(wooldridge) # #ID T
wagepan <- pdata.frame(wagepan, index = c("nr", "year"))
```

(a) Estime un modelo de efectos fijos.

```
modelo1 <- plm(lwage ~ educ + exper + black + married + union,  
  data = wagepan, model= "between")  
  
stargazer(modelo1, type = "text")
```

```
##  
## =====  
##                Dependent variable:  
##                -----  
##                lwage  
## -----  
## educ                0.091***  
##                   (0.011)  
##  
## exper                0.028**  
##                   (0.011)  
##  
## black               -0.143***  
##                   (0.048)  
##  
## married             0.141***  
##                   (0.041)  
##  
## union               0.259***  
##                   (0.046)  
##  
## Constant            0.290  
##                   (0.176)  
## -----  
## Observations        545  
## R2                  0.215  
## Adjusted R2         0.208  
## F Statistic         29.601*** (df = 5; 539)  
## =====  
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

(b) Estime un modelo de efectos aleatorios.

```
# Modelo de efectos aleatorios -----  
modelo12 <- plm(lwage ~ educ + exper + black + married + union,  
  data = wagepan, model = "random")  
  
stargazer(modelo12, type = "text")
```

```
##  
## =====  
##                Dependent variable:  
##                -----  
##                lwage  
## -----  
## educ                0.107***  
##                   (0.009)  
##  
## exper                0.058***  
##                   (0.003)  
##  
## black               -0.144***  
##                   (0.047)  
##  
## married             0.076***  
##                   (0.017)  
##  
## union               0.110***  
##                   (0.018)  
##  
## Constant            -0.037  
##                   (0.107)  
##  
## -----  
## Observations        4,360  
## R2                  0.169  
## Adjusted R2         0.168  
## F Statistic         887.465***  
## =====  
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

(c) Realice la prueba de Hausman y determine qu´e modelo es m´as apropiado.

```
## Test de Hausman:  
phtest(modelo1, modelo2)
```

```
##  
## Hausman Test  
##  
## data: lwage ~ educ + exper + black + married + union  
## chisq = 21.801, df = 5, p-value = 0.0005711  
## alternative hypothesis: one model is inconsistent
```

La prueba de Hausman contrasta la hipótesis nula de que el estimador de efectos aleatorios (EA) es consistente y eficiente contra la alternativa de que solo el estimador de efectos fijos (EF) es consistente. En tu caso, el estadístico de Hausman para comparar EF y EA es: Con modelo “between”: $\chi^2(5) = 21.801$, $p = 0.000571$. Debido a que el $p < 0.01$, se rechaza la hipótesis nula. Esto indica que el modelo de efectos aleatorios es inconsistente y, por tanto, el modelo de efectos fijos es el más apropiado para este conjunto de datos.

(d) Interprete el efecto de pertenecer a un sindicato (union).

El coeficiente de union es 0.259, lo que indica que, en promedio, los trabajadores que pertenecen a un sindicato (generando un efecto fijo) ganan un 25.9% más que aquellos que no pertenecen a un sindicato, manteniendo constantes las demás variables del modelo. Este efecto es estadísticamente significativo al nivel del 1% ($p < 0.01$), lo que sugiere que la pertenencia a un sindicato tiene un impacto positivo y significativo en los salarios de los trabajadores.

Ejercicio 2: Cambios en el crimen

- Variable dependiente: crmrte (tasa de criminalidad)
- Variables explicativas: prbarr (probabilidad de arresto), prbconv (probabilidad de condena), avgsen (sentencia promedio), polpc (policía por habitante), density (densidad poblacional), income (ingreso per cápita).

(a) Estime un modelo de efectos fijos para evaluar el impacto de la severidad del castigo sobre la criminalidad.

```
crimen <- pdata.frame(crime4, index = c("county", "year"))

modelo21 <- plm(crmrte ~ prbarr + prbconv + avgsen + polpc + density,
  data = crimen, model= "between")

stargazer(modelo21, type = "text",
  title = "Modelo de Efectos Fijos para el Crimen")
```

```
##
## Modelo de Efectos Fijos para el Crimen
## =====
##               Dependent variable:
##               -----
##               crmrte
## -----
## prbarr          -0.046***
##                 (0.008)
##
## prbconv          -0.005***
##                 (0.001)
##
## avgsen           -0.001
##                 (0.001)
##
## polpc            3.685***
##                 (0.604)
##
## density          0.007***
##                 (0.001)
##
## Constant         0.039***
##                 (0.006)
##
## -----
## Observations           90
## R2                    0.729
## Adjusted R2           0.713
## F Statistic    45.305*** (df = 5; 84)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

(b) ¿Cómo cambia el coeficiente de prbarr si se omite polpc? Interprete.

```
modelo22 <- plm(crmrte ~ prbarr + prbconv + avgsen + density,
  data = crimen, model= "between")
```

```
stargazer(modelo21, modelo22,
  type = "text", title = "Comparación de Modelos con y sin polpc")
```

```
##
## Comparación de Modelos con y sin polpc
## =====
##                               Dependent variable:
##                               -----
##                               crmrte
##                               (1)           (2)
## -----
## prbarr          -0.046***          -0.031***
##                  (0.008)           (0.010)
##
## prbconv          -0.005***           0.0001
##                  (0.001)           (0.001)
##
## avgsgen          -0.001             -0.0001
##                  (0.001)           (0.001)
##
## polpc            3.685***
##                  (0.604)
##
## density          0.007***           0.008***
##                  (0.001)           (0.001)
##
## Constant         0.039***           0.032***
##                  (0.006)           (0.007)
## -----
## Observations      90                90
## R2                 0.729             0.609
## Adjusted R2       0.713             0.591
## F Statistic  45.305*** (df = 5; 84) 33.157*** (df = 4; 85)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

Al omitir polpc, el coeficiente de prbarr disminuye en magnitud, lo que indica que parte del efecto de la probabilidad de arresto sobre la criminalidad estaba siendo capturado por la variable polpc. Esto sugiere que la presencia de más policías por habitante puede estar relacionada con una menor tasa de criminalidad, y al no incluirla, se subestima el efecto de prbarr.

(c) Incluya efectos temporales fijos (años) y analice si la tendencia de criminalidad cambia.

```
modelo23 <- plm(
  crmrte ~ prbarr + prbconv + avgsgen + density,
  data = crimen,
  model= "within",
  effect = "twoways"
)
```

```
stargazer(
  modelo21, modelo23,
  type = "text",
  float = TRUE,
  title = "Comparación de Modelos con y sin Efectos Temporales"
)
```

```
##
## Comparación de Modelos con y sin Efectos Temporales
## =====
##                               Dependent variable:
##                               -----
##                               crmrte
##                               (1)          (2)
## -----
## prbarr          -0.046***          -0.002
##                  (0.008)          (0.003)
##
## prbconv          -0.005***          -0.00003
##                  (0.001)          (0.0002)
##
## avgsen           -0.001             0.0001
##                  (0.001)          (0.0001)
##
## polpc            3.685***
##                  (0.604)
##
## density          0.007***             0.003
##                  (0.001)          (0.004)
##
## Constant         0.039***
##                  (0.006)
## -----
## Observations           90             630
## R2                     0.729           0.003
## Adjusted R2            0.713          -0.170
## F Statistic  45.305*** (df = 5; 84) 0.373 (df = 4; 536)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Al incluir **efectos temporales fijos** (años) en el modelo, los coeficientes de las variables clave pierden significancia y el R^2 cae de 0.729 a 0.003, lo que indica que gran parte de la variación en la criminalidad se explica por factores comunes a todos los individuos en cada año. Esto sugiere que la tendencia de criminalidad está fuertemente influida por shocks o condiciones agregadas más que por variaciones específicas de cada unidad.

Ejercicio 3: Productividad agrícola

Variable dependiente: `crop output`

Variables explicativas: `fertilizer, labor, land`

(a) Estime un modelo de efectos fijos para evaluar el impacto del fertilizante.

```
fertil <- read.csv("./fertilizer.csv")
# Convertir a datos de panel, controlando por country y year
fertil <- pdata.frame(fertil, index = c("country", "year"))
modelo3 <- plm(crop_output ~ fertilizer + labor + land,
  data = fertil, model = "within")
stargazer(modelo3, type = "text")
```

```
##
## =====
##                Dependent variable:
##                -----
##                crop_output
## -----
## fertilizer      -367,381.400
##                  (666,513.200)
##
## labor           3,036.418***
##                  (356.234)
##
## land            -169.203
##                  (211.276)
##
## -----
## Observations      174
## R2                 0.320
## Adjusted R2       0.288
## F Statistic      25.937*** (df = 3; 165)
## =====
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

(b) Incluya el término cuadrático de fertilizer y comente si hay rendimientos decrecientes.

```
modelo31 <- plm(crop_output ~ fertilizer + I(fertilizer^2) + labor + land,
  data = fertil, model = "within")
stargazer(modelo31, type = "text")
```

```
##
## =====
##                Dependent variable:
##                -----
##                crop_output
## -----
## fertilizer      2,548,735.000
##                  (1,974,002.000)
##
## I(fertilizer2)   -7,645.618
##                  (4,874.343)
##
```

```
## labor                2,853.892***
##                      (373.270)
##
## land                 -172.474
##                      (210.358)
##
## -----
## Observations          174
## R2                    0.331
## Adjusted R2           0.294
## F Statistic          20.240*** (df = 4; 164)
## =====
## Note:                 *p<0.1; **p<0.05; ***p<0.01
```

El coeficiente de `fertilizer` es positivo, mientras que el coeficiente de `I(fertilizer^2)` es negativo. Esto indica que hay rendimientos decrecientes a medida que se aumenta el uso de fertilizante, ya que el efecto marginal del fertilizante disminuye con niveles más altos de uso.

(c) Estime un modelo con efectos aleatorios y compare con el modelo anterior.

```
modelo32 <- plm(crop_output ~ fertilizer + I(fertilizer^2) + labor + land,
  data = fertil, model = "random")
stargazer(modelo31, modelo32, type = "text",
  title = "Comparación de Modelos de Efectos Fijos y Aleatorios")
```

```
##
## Comparación de Modelos de Efectos Fijos y Aleatorios
## =====
##                               Dependent variable:
##                               -----
##                               crop_output
##                               (1)           (2)
## -----
## fertilizer                2,548,735.000      2,788,752.000
##                          (1,974,002.000)      (1,959,437.000)
##
## I(fertilizer2)            -7,645.618         -7,957.465
##                          (4,874.343)         (4,847.792)
##
## labor                     2,853.892***       2,788.455***
##                          (373.270)          (368.691)
##
## land                     -172.474           -181.917
##                          (210.358)          (187.868)
##
## Constant                  946,987,107.000
##                          (717,772,961.000)
##
## -----
## Observations              174                174
## R2                        0.331                0.323
## Adjusted R2               0.294                0.306
```

```
## F Statistic    20.240*** (df = 4; 164)    80.455***
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

La comparación entre el modelo de efectos fijos y el modelo de efectos aleatorios muestra que los coeficientes son similares, pero el modelo de efectos aleatorios es más eficiente (menor error estándar).

```
phptest(modelo31, modelo32)
```

```
##
## Hausman Test
##
## data:  crop_output ~ fertilizer + I(fertilizer^2) + labor + land
## chisq = 2.3975, df = 4, p-value = 0.6631
## alternative hypothesis: one model is inconsistent
```

En el test de hausman el p-value es 0.6631 lo que indica que no se rechaza la hipótesis nula de que el modelo de efectos aleatorios es consistente. Sin embargo, dado que el modelo de efectos fijos es más robusto ante la correlación entre los efectos individuales y las variables explicativas, se prefiere este último para inferencias más confiables.