

### Conceptos

$\alpha (i)$

recoge todos aquellos factores individuales (diferentes entre individuos, pero que no son observables (no somos capaces de explicar)

$\gamma (t)$

recoge todos aquellos factores temporales pero que no son observables (no somos capaces de explicar)

$\epsilon (it)$

recoge aquello que no somos capaces de explicar con nuestras variables independientes.

En los Modelos de Regresión Lineal Simple, Múltiple y Polinómico los efectos sobre la variable dependiente  $y$  que varían en  $i$  se consideraban en el término de error  $\epsilon_i$ . Ahora NO. En los Modelos de Regresión con Datos de Panel nos interesarán tres tipos de efectos sobre la variable dependiente  $y$  :

Efectos sobre  $y$  que varían en  $i$  pero no en  $t$  (effect='individual'). Es decir, la variación entre individuos en un mismo instante de tiempo.

Efectos sobre  $y$  que varían en  $t$  pero no en  $i$  (effect='time'). Es decir, la variación en el tiempo para un individuo dado.

Efectos sobre  $y$  que varían en  $i$  y  $t$  se considerarán en el término de error  $\epsilon_{it}$  (effect='twoways').

### Interpretación de la salida de los modelos

1. output de la regresión

```
output <- summary(fixed)
```

2. coeficientes de regresión estimados

```
output$coef
```

3. residuos estimados

```
uhat <- resid(fixed)
```

4. valores ajustados de la variable dependiente

```
yajustado <- wdata$fare-uhat
```

Comprobamos  $y = y_{\text{estimado}} + \text{residuos}_{\text{estimados}}$

```
head(cbind(wdata$fare,yajustado,uhat))
```

5. grados de libertad

```
gdl = head(output$df)[2]
```

6. suma de cuadrados de los residuos

```
ssr = sqrt(deviance(fixed)/gdl)
```

7. errores de los parámetros estimados

```
se = sqrt(diag(output$vcov))
```

8. coeficientes de determinación

```
R2 = output$r.squared
```

9. intervalos de confianza para los coeficientes de regresión

```
ci.bhat <- confint(fixed)
```

10. estadístico F y su p-valor

```
f_estad = output$fstatistic
```

11. obtener predicciones de nuevos datos

```
yhat1 = predict(fixed)
```

### Criterios de selección del mejor modelo

A diferencia de los casos lineal simple y múltiple no se usan las métricas AIC, BIC y  $R^2_{adj}$ .

#### Test de agrupamiento

Se contrasta la hipótesis de que existen efectos individuales inobservables distintos entre individuos, países, etc. Se usa para saber si tenemos que usar los primeros métodos visto o within/random.  $H_0$  es que existen efectos individuales. Estimador pooling vs estimador de efectos fijos.

```
require(plm)
```

```
pFtest(fixed, pooling)
```

#### Test de Hausmann

necesitamos saber si se rechaza o no que  $Cov(a_i, x_{it})=0$  para saber si optamos por el estimador de efectos fijos o efectos aleatorios. bajo el supuesto de que la  $Cov(a_i, x_{it})=0$  obtenemos que el estimador de efectos aleatorios es consistente y eficiente. Sin embargo, bajo la hipótesis alternativa  $Cov(a_i, x_{it}) \neq 0$  el estimador de efectos fijos es consistente, mientras que el estimador de efectos aleatorios es inconsistente. si rechazamos la hipótesis nula  $Cov(a_i, x_{it}) \neq 0$  optaremos por el estimador de efectos fijos y si no, por el estimador de efectos aleatorios.

```
require(plm)
```

```
phtest(random, fixed)
```



### Exploración inicial (preparación)

```
require(wooldridge)
data(airfare)
wdata = airfare[,c("fare", "dist")]

1. Análisis gráfico de la existencia de
relación lineal

scatter.smooth(x = wdata$dist, y =
wdata$fare,
main = "fare ~ dist")
```

```
2. Evidencia numérica de la existencia de
relación lineal

cor(wdata$fare, wdata$dist)
```

```
Evidencia contrastada

cor.test(wdata$fare, wdata$dist)
```

```
Análisis de ajuste a una distribución normal

require(e1071)
par(mfrow=c(1, 2))
hist(wdata$fare, freq = FALSE)
lines(density(wdata$fare))
hist(wdata$dist, freq = FALSE)
lines(density(wdata$dist))
par(mfrow=c(1, 1))
```

```
Prueba Shapiro-Wilk con la función shapiro.test

shapiro.test(wdata$fare)
shapiro.test(wdata$dist)
```

```
Prueba Anderson-Darling con la función
ad.test de la librería nortest

require(nortest)
ad.test(wdata$fare)
ad.test(wdata$dist)
```

```
Prueba Cramer-von Mises con la función
cvm.test de la librería nortest

cvm.test(wdata$fare)
cvm.test(wdata$dist)
```

### Exploración inicial (preparación) (cont)

```
Prueba Kolmogorov-Smirnov con la función
lillie.test del paquete nortest

lillie.test(wdata$fare)
lillie.test(wdata$dist)
```

3. Definir la estructura de una base de datos de panel

```
Obtenemos el número de dimensiones

table(airfare$year) # Qué años hay
nrow(table(airfare$id)) #Vuelos cada
año
```

```
Construimos el panel de datos

library(plm)
pdata = pdata.frame(airfare, index=c("i-
d", "year"))
```

```
Eliminamos las columnas year e id

pdata = pdata.frame(airfare, index=c("i-
d", "year"),
drop.index=TRUE, row.names =
TRUE)
```

```
porcentajes de la variabilidad

summary(pdata$fare)
las tarifas de billetes sólo ida (fare) se
ven afectadas en gran medida (94,41 %
) por las diferencias existentes entre las
agencias ( i ) y residualmente por el
paso del tiempo ( t ).
```

```
calcula la tarifa media de cada vuelo
considerando todo el periodo

between(pdata$fare)
```

```
calcula la diferencia de cada vuelo y año

Within(pdata$fare)
```

### Análisis de regresión con datos de panel

#### Modelos de regresión con datos de panel

Modelos de regresión con datos de panel cumpliendo que  $Cov(\alpha_i, \epsilon_{it}) \neq 0$

se supone que los efectos individuales no observables están correlacionados con el resto de variables explicativas del modelo.

1.1. Estimador común de MCO

ooling

Problema: las estimaciones que obtenemos NO son consistentes (no se aproxima al valor del parámetro cuanto mayor es el tamaño de la muestra)

```
require(plm)
pooling <- plm(lfare ~ concen + ldist +
l(ldist^2),
data=pdata, model= "pooling")
# Es lo mismo que
pooling2 <- lm(lfare ~ concen + ldist +
l(ldist^2),
data=pdata)
```

1.2. Estimador transformación

etween

primera de las transformaciones. estima el modelo con las medias de cada individuo, país, etc.. el estimador es consistente aunque no considera el comportamiento dinámico de las variables. NO recoge bien el efecto de las variables que varían en el tiempo aunque SI el efecto entre individuos en el término de error compuesto. Luego, no podemos extraer los efectos individuales. No tiene en cuenta el efecto temporal.

```
between <- plm(lfare ~ concen + ldist +
l(ldist^2),
data=pdata, model= "between")
summary(between)
```

### Análisis de regresión con datos de panel (cont)

1.3. Estimador transformación  $\square$   
 $\square$  `first-differences` transformar las variables en primeras diferencias entre periodos consecutivos y aplicar MCO al modelo así transformado. no permite identificar los efectos individuales, además de perderse una observación temporal en todos los individuos. suele usarse en modelos donde la dimensión temporal es grande y el interés se centra en la estimación de las pendientes y no en la estimación de los efectos individuales. el estimador es consistente y SI considera el comportamiento dinámico de las variables pero no permite obtener los efectos individuales.

```
firstdiff <-plm(lfare ~ concen + ldist +
l(ldist^2),
data=pdata, model= "fd")
summary(firstdiff)
```

### Análisis de regresión con datos de panel (cont)

1.4 Estimador de efectos fijos `Within` tercera de las transformaciones. En este caso, el estimador es consistente, SI considera el comportamiento dinámico de las variables y SI permite obtener los efectos individuales pero si la  $\square$   
 $\square$  `cov(ai,xit)=0` no es eficiente. Consiste en transformar las variables en diferencias respecto a las medias dentro de cada submuestra y aplicar MCO al modelo así transformado

En este caso, el estimador es consistente, SI considera el comportamiento dinámico de las variables y SI permite obtener los efectos individuales pero si la  $\square$   
 $\square$  `cov(ai,xit)=0` no es eficiente.

```
fixed <- plm(lfare ~ concen + ldist +
l(ldist^2), data=pdata, model= "within",
effect = "individual")
summary(fixed)
fixed2 <- plm(lfare ~ concen + ldist +
l(ldist^2), data=pdata, model= "within",
effect = "time")
summary(fixed2)
fixed3 <- plm(lfare ~ concen + ldist +
l(ldist^2), data=pdata, model= "within",
effect = "twoways")
summary(fixed3)
```

```
para extraer los efectos individuales:
fixef(fixed, type="level")
Para obtener una estimación local de las
betas de cada uno de los individuos
summary(fixef(fixed, type="dmean"))
fixef(fixed2, type="level")
summary(fixef(fixed2, type="dmean"))
```

### Análisis de regresión con datos de panel (cont)

1.5. Estimador de efectos aleatorios  $\square$   
 $\square$  `random` el estimador es consistente, SI considera el comportamiento dinámico de las variables, SI permite obtener los efectos individuales y si la  $\text{Cov}(ai,xit)=0$  el estimador de efectos aleatorios es eficiente (mínima varianza). El inconveniente es que si  $\text{Cov}(ai,xit)\neq 0$  el estimador de efectos aleatorios es inconsistente.

```
random <- plm(lfare ~ concen + ldist +
l(ldist^2), data=pdata, model= "random",effect = "-
individual")
summary(random)
random2 <- plm(lfare ~ concen + ldist + l(ldist^2),
data=pdata, model= "random",effect = "-
time")
summary(random2)
random3 <- plm(lfare ~ concen + ldist +
l(ldist^2), data=pdata, model= "ran-
dom",effect = "twoways")
summary(random3)
```

1.6 Comparar en una misma tabla los resultados de todas las transformaciones

```
# Comparamos los resultados obtenidos
en los modelos MCO vs efectos fijos
library(stargazer)
stargazer(pooling)
stargazer(pooling, between, firstdiff,
fixed,
type="text")
# Comparamos los resultados obtenidos
en los modelos de efectos fijos
stargazer(fixed, fixed2, fixed3, type="t-
ext")
# Comparamos los resultados obtenidos
en los modelos de efectos aleatorios
stargazer(random, random2, random3,
type="text")
```

