

# Modelos para Datos de Panel o Longitudinales

Cesar Jimenez y Nestor Viñals

## Introduccion

En este proyecto se estima un modelo de la demanda de cigarrillos usando datos de panel con información de 46 estados de los Estados Unidos de America, abarcando el periodo 1963-92. Este modelo dinámico de demanda parte de, y expande, el trabajo realizado por Baltagi y Levin (1992) y cuyo objetivo principal era mostrar las trampas de depender modelos de series temporales para un estado dado o, alternativamente, estudios de corte transversal para periodo (para mas información sobre el análisis original, ver Baltagi y Levin, 1992).

El análisis original expande un trabajo anterior y lo expande para incluir, en total, el periodo 1963-88. En este trabajo, lo ampliamos para abarcar el periodo 1963-92.

## Importando y transformando la base de datos

### Descripcion

La base de datos utilizada en este análisis es el dataframe `Cigar` de la librería `plm`. Esta contiene información acerca del consumo de cigarrillos y ingreso disponible per capita para 46 estados de los Estados Unidos comprendiendo el periodo 1963-92 (para ver todas la variables de la base de datos, ver la viñeta de `Cigar` usando `?plm::Cigar`).

Para la estimación de los modelos, se realizaron las siguientes transformaciones:

```
data(Cigar) # Importada de la librería `plm`
Cigar <- Cigar %>%
  mutate(sales_lag1 = lag(sales, 1), lnC = log(sales), lnPm = log(pimin),
         lnC_lag1 = log(sales_lag1), lnP = log(price), lnY = log(ndi)) %>%
  select(state, year, lnC, lnC_lag1, lnP, lnY, lnPm) %>%
  tibble()
datos_panel <- pdata.frame(Cigar, index = c("state", "year"))
```

## Modelo y resultados

El modelo a estimar es el siguiente:

$$\ln C_{i,t} = \beta_0 + \beta_1 \ln C_{i,t-1} + \beta_2 \ln P_{i,t} + \beta_3 \ln Y_{i,t} + \beta_4 \ln Pn_{i,t} + \epsilon_{i,t}$$

Donde el subíndice  $i$  denota el *i*vo estado y el subíndice  $t$  indica el *t*vo periodo.  $C_{i,t}$  representa las ventas de cigarrillos en paquetes per capita (considerando solo a las personas mayores de 14 años).  $P_{i,t}$  es el precio promedio de venta en términos reales.  $Pn_{i,t}$  es el precio de venta mínimo por paquete de cigarrillos en los estados contiguos.  $Y_{i,t}$  representa el ingreso disponible per capita en términos reales.

En la siguiente tabla se muestran los distintos modelos a comparar: Pooling (OLS), Between, First Differences (primeras diferencias), Within y Random Effects (efectos aleatorios):

Table 1:

	<i>Dependent variable:</i>				
	Pooled	Between	lnC First Differences	Within	Random Effects
	(1)	(2)	(3)	(4)	(5)
lnC_lag1	0.861*** (0.011)	1.036*** (0.009)	-0.030** (0.013)	0.495*** (0.016)	0.861*** (0.011)
lnP	-0.235*** (0.026)	0.037 (0.026)	-0.380*** (0.024)	-0.551*** (0.030)	-0.235*** (0.026)
lnY	0.069*** (0.012)	-0.033*** (0.011)	0.259*** (0.035)	0.247*** (0.014)	0.069*** (0.012)
lnPm	0.106*** (0.022)	-0.032 (0.020)	0.031 (0.027)	0.200*** (0.029)	0.106*** (0.022)
Constant	0.597*** (0.052)	0.091 (0.107)	-0.004 (0.003)		0.597*** (0.052)
Observations	1,379	46	1,333	1,379	1,379
R <sup>2</sup>	0.888	0.998	0.228	0.783	0.888
Adjusted R <sup>2</sup>	0.888	0.998	0.226	0.775	0.888
F Statistic	2,729.869*** (df = 4; 1374)	5,127.576*** (df = 4; 41)	97.955*** (df = 4; 1328)	1,196.212*** (df = 4; 1329)	10,919.480***

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Las elasticidades estimadas de largo plazo pueden ser obtenidas a partir de las de corto plazo multiplicando este por  $1/(1 - \hat{\beta}_1)$ , donde  $\hat{\beta}_1$  es el estimador del consumo rezagado. Así, los estimadores de largo plazo para los modelos *Pooling*, *Between*, *First Differences*, *Within* y *Random Effects* son 7.207, -28.098, 0.971, 1.979 y 1.979, respectivamente.

## Diagnostico

Iniciamos el diagnostico aplicando la prueba de los multiplicadores de Langrange del modelo *pooled*. Aquí estamos contrastando el modelo de *random effects* (efectos aleatorios) con el modelo *pooled*. A partir de este, rechazamos la hipótesis nula de que el modelo correcto es el *pooled*.

```
##
## Lagrange Multiplier Test - (Honda) for unbalanced panels
##
## data: lnC ~ lnC_lag1 + lnP + lnY + lnPm
## normal = 14.164, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Luego procedemos a ejecutar la prueba F de contraste de *fixed effects* (efectos fijos) versus modelo *pooled*. Aquí, de nuevo, rechazamos la hipótesis nula de que el modelo *pooled* es el correcto (los efectos individuales son significativos) y concluimos que debemos utilizar un modelos de efectos fijos.

```
##
## F test for individual effects
##
## data: lnC ~ lnC_lag1 + lnP + lnY + lnPm
## F = 17.427, df1 = 45, df2 = 1329, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Finalmente, Luego procedemos a ejecutar la prueba Hausman de contraste de *fixed effects* (efectos fijos) versus *random effects* (efectos aleatorios), para descartar . A partir de este, rechazamos la hipótesis nula de que ambos modelos son iguales y concluimos que existe endogeneidad y el modelo correcto es a utilizar es el de *fixed effects*.

```
##
## Hausman Test
##
## data: lnC ~ lnC_lag1 + lnP + lnY + lnPm
## chisq = 1098.9, df = 4, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

## Conclusion

A partir del modelo seleccionado, el de efectos fijos (*within*), podemos interpretar lo siguiente, usando los estimadores obtenidos:

- La persistencia del consumo de cigarrillos se aprecia en el coeficiente, donde un aumento de un 1% en el consumo de cigarrillos en el periodo anterior se traduce en un aumento de 0.495% en el periodo evaluado.
- Un aumento de un 1% en el precio del paquete de cigarrillos esta asociado a una disminución de 0.55% en el consumo de estos.
- Contrariamente, un aumento de un 1% en el ingreso percibido se correlaciona con un aumento de un 0.25% en el consumo de cigarrillos.
- A su vez, un aumento en el precio minimo de un paquete de cigarrillos en los estando colindantes, se traduce en un aumento en el consumo de cigarrillos en el estado evaluado.