



Universidad de SanAndrés

ECONOMETRÍA AVANZADA

WALTER SOSA ESCUDERO
GASTÓN GARCÍA ZAVALETA

Trabajo Práctico 2

GARCÍA VASSALLO, HEDEMAN, IOLSTER, SURY

2022

Ejercicio 1

En la primera columna del cuadro 1 del apéndice se encuentran los resultados de la estimación del modelo between.

Dado que se adopta una especificación log-log, los coeficientes son interpretables como elasticidades. Procedemos a analizar las 4 variables de justicia criminal, cuyas estimaciones tienden a corroborar el trabajo empírico previo que se concentro en la estimación cross-section del modelo económico del crimen. Esto es, excepto la probabilidad de que te encarcelen dado que te condenaron, los coeficientes de las variables presentan el signo negativo esperado, es decir, hay presentan una correlación negativa con la tasa de criminalidad. Específicamente, si la probabilidad de que te arresten (P_a) aumenta en un 1 %, la tasa de criminalidad baja en 0.648 %. Luego, si la probabilidad de que te condenen dado que te arrestaron (P_c) aumenta en un 1 %, la tasa de criminalidad baja en 0.528 %. Por su parte, si la severidad del castigo (S) aumenta en un 1 % la tasa de criminalidad cae en 0.236 %. En cambio, en contra de lo esperado, si la probabilidad de que te encarcelen dado que te condenaron (P_p) aumenta en un 1 %, la tasa de criminalidad aumenta en 0.297 %. Sin embargo, solo los coeficientes estimados asociados a (P_a) y (P_c) son estadísticamente significativas. Por otro lado, el resto de las variables que correlacionan positivamente con la tasa de crímenes son la cantidad de policía per cápita; la densidad poblacional; porcentaje de hombres jóvenes; salario semanal de la construcción; salario semanal en el comercio mayorista y minorista; salario semanal en finanzas, seguros e inmobiliarias; salario semanal en el gobierno federal; salario semanal en el gobierno local. Por otro lado, las variables que correlacionan negativamente con la tasa de crimen son el salario semanal en transporte, servicios públicos y comunicación; el salario semanal en transporte, servicios públicos y comunicación; el salario semanal en el sector servicios; salario semanal en la industria manufacturera; salario semanal en el gobierno estatal; el hecho de ser un condado del oeste o del centro en comparación a ser de otra zona y el hecho de ser un condado urbano.

La discusión acerca de la validez del estimador between y potenciales sesgos se encuentra en la siguiente sección.

Ejercicio 2

La estimación between consiste en calcular los promedios para los condados y estimar el modelo a partir de los promedios, esto es, eliminar la varianza que al interior de cada condado y usar sólo la variabilidad que existe entre los condados para identificar el parámetro. De esta forma, se transforman la estructura de datos de panel en una estructura cross-section usando los promedios de cada condado. Por lo tanto, este modelo es muy sensible a presentar problemas de endogeneidad ya que este estimador es solo consistente solo si (X'_{it}, P'_{it}) no correlacionan con el término de error, lo cual en este caso, un supuesto demasiado fuerte. Esto es así ya que hay heterogeneidad entre los condados que es no observable y afecta a la tasa de criminalidad y entonces es capturada por el término de error. Luego, si estas variables inobservables específicas de cada condado correlacionan con las variables independientes cuyo efecto nos interesa identificar (lo cual es altamente probable), este procedimiento va a generar estimadores sesgados. En otras palabras, la variabilidad between es potencialmente endógena.

En este contexto, una estimación con efectos fijos por condado resolvería el problema de sesgo. Esta estrategia opera al incluir una variable dummy por condado en el modelo regresión. De forma intuitiva, al controlar por las variables específicas de cada condado, efectos fijos incluye una ordenada al origen para cada condado. De esta forma, las características que varían por condado no van al termino de error ya que son incluidas en la regresión por lo que el supuesto de que (X'_{it}, P'_{it}) no correlacionan con el término de error ya no es tan fuerte como en el caso anterior. Así, se logra eliminar el sesgo causado por la correlación entre el término de error y las variable independientes, siempre y cuando estemos dispuesto a suponer que toda la heterogeneidad inobservable relevante es constante en el tiempo. Dicho de otro modo toda, la heterogeneidad inobservable que es constante en el tiempo esta capturada por los efectos fijos por lo que si la heterogeneidad inobservable que te generaba el sesgo o la inconsistencia es constante en el tiempo, esta estrategia soluciona el problema de sesgo, sino no.

En el cuadro 1 del apéndice se presentan los resultado de estimar el modelo con la estrategia de efectos fijos en la segunda columna.

Se puede ver claramente que al centrarse en los efectos estimados de las variables de justicia criminal la diferencia entre los estimadores within y between es impactante. En particular, al incluir los efectos fijos, en valor absoluto los efectos de estas variables sobre la tasa de crímenes se reducen drásticamente lo que da cuenta de

que las estimaciones between eran sesgadas, esto es, que las variables de justicia criminal estaban captando el efecto de la heterogeneidad inobservable sobre la tasa de crimen. Específicamente, la elasticidad asociada a la probabilidad de que te arresten (P_a) pasa de -0.648 a -0.385, la elasticidad asociada a la probabilidad de que te condenen dado que te arrestaron (P_c) pasa de -0.648 a -0.301. Por su parte, los efectos de las dos variables restantes de justicia criminal cambiaron su signo. En particular, la elasticidad asociada a la probabilidad de que te encarcelen dado que te condenaron (P_p) paso de 0.297 a -0.191 y la elasticidad asociada a la severidad del castigo paso de -0.236 a 0.026. En cuanto a la significatividad, el único cambio es que la probabilidad de que te encarcelen dado que te condenaron pasa a ser significativa al 10 %.

A continuación presentamos los resultados del test F que compara el modelo de efectos fijos con el de MCO.

| | |
|---|--------|
| F | 17.293 |
| df1 | 85 |
| df2 | 524524 |
| p-value | 0.0000 |
| alternative hypothesis: significant effects | |

Dado que el p-valor es 0.0000 rechazamos la hipótesis nula. Por lo tanto, hay evidencia suficiente para decir que hay efectos fijos.

Ejercicio 3

El test de Hausman evalúa el modelo de efectos aleatorios y el modelo de efectos fijos. A partir del test nos permite conjeturar sobre qué modelo es consistente. Para ello se establece la hipótesis de que no existe correlación entre las variables explicativas y los efectos individuales. Cuando no rechazamos H_0 , el test de Hausman debería ser aproximado a cero dado que el modelo de efectos fijos y efectos aleatorios estarían estimando lo mismo porque ambos modelos serían insesgados.

| | |
|---------|--------|
| chisq | 66.189 |
| df | 16 |
| p-value | 0.0000 |

Dado que el p-valor es 0.0000 rechazamos la hipótesis nula donde decimos que ambos modelos son consistentes. Entonces, como rechazamos H_0 podemos decir que el modelo de efectos fijos es consistente mientras que el modelo random effects no.

En cuanto a las estimaciones del modelo de efectos aleatorios, los resultados se aproximan a los del modelo de efectos fijos. En la tabla de regresiones se puede apreciar un sesgo de sobre estimación en términos absolutos en la elasticidad asociada a la probabilidad de que te arresten (P_a), la probabilidad de que te condenen dado que te arrestaron (P_c) y la probabilidad de que te encarcelen dado que te condenaron (P_p). También es relevante nombrar la diferencia en el coeficiente de la densidad por metro cuadrado de 0.469 y significatividad del 1 % en el caso de modelo con efectos aleatorios en comparación con un 0.409 y significatividad que no alcanza a ser significativa ni al 10 % en el modelo de EF. Por último, observamos sesgo de subestimación en términos absolutos en el coeficiente que acompaña a la severidad del castigo (S), la cantidad de policías per cápita (POLICE), y del gasto semanal (WFED, WSTA, WLOC, WMFG), en comparación con el modelo (2). Además, se modifica la significatividad en las variables de gasto semanal WTUC y WSTA.

Ejercicio 4

En la tabla de regresiones se puede apreciar como las elasticidades de las variables de justicia criminal en el primer modelo son mayores en términos absolutos que en el modelo de efectos fijos. Esto es, P_a en el modelo between en términos absolutos es 0.648 mientras que en el modelo de efectos fijos es 0.385, P_c en el modelo between en términos absolutos es 0.528 mientras que en el modelo de efectos fijos es 0.301, P_p en el modelo between en términos absolutos es 0.297 mientras que en el modelo de efectos fijos es 0.191, S en el modelo between en términos absolutos es 0.236 mientras que en el modelo de efectos fijos es 0.026. Por otro lado, en cuanto a la significatividad de las variables podemos observar que en el primer modelo las variables de justicia criminal P_a y P_c , son significativas al 10 %. Luego, en el segundo modelo de efectos fijos, tanto P_a como P_c , y

P_p son significativas al 10 %. La severidad del castigo no es significativo en ninguno pero tiene un mayor valor en términos absolutos en el modelo between.

En cuanto a las variables del mercado de trabajo, en el primer modelo ninguna de ellas es significativa. Por otro lado, en el modelo de efectos fijos las variables salario semanal en transporte, servicios y comunicación (WTUC), salario semanal en la industria manufacturera (WMFG), salario semanal en el gobierno federal (WFED), salario semanal en el gobierno estatal (WSTA) y salario semanal en la administración local (WLOC) son significativas. Todas son significativas al 10 % mientras que la variable salario semanal en el gobierno estatal (WSTA) es significativa al 5 %.

El modelo between no solo no les da relevancia dada la significancia a las variables de mercado de trabajo sino que también presenta sesgos de subestimación y de signo en comparación con el modelo de efectos fijos. Esto es así ya que el coeficiente indicando la elasticidad es mayor en términos absolutos en el modelo between y en el caso de WTUC, WFED y WSTA el signo es distinto.

En conclusión, las afirmaciones de Cornwell y Trumbull son acertadas dado lo remarcado previamente. No es solo la justicia criminal la que afecta la criminalidad, hay que tener en cuenta el mercado laboral a la hora de pensar en crimen.

Ejercicio 5

Cuando tenemos autocorrelación lo que se trata de evaluar es si los residuos del modelo tienen asociación con sus mismos valores rezagados o retardados. Para ellos realizaremos una serie de tests.

Test conjuntos de correlación serial y efectos aleatorios

| | |
|---------|--------|
| chisq | 715.24 |
| df | 2 |
| p-value | 0.0000 |

Este test es una prueba de correlación serial residual (o efectos aleatorios individuales) localmente robusta frente a efectos aleatorios individuales (correlación serial) para modelos de panel y prueba conjunta de correlación serial y especificación de efectos aleatorios. Aparte, le indicamos que realice una prueba conjunta para la correlación serial de primer orden (".ar") y efectos aleatorios (re"). Dado que el p-valor es 0.0000 rechazamos la hipótesis nula y aceptamos la hipótesis alternativa donde los errores son AR(1) o hay random effects.

Test robusto local para correlación serial o Efectos Aleatorios

| | |
|---------|--------|
| chisq | 42.356 |
| df | 1 |
| p-value | 0.0000 |

Al igual que el test anterior, este test genera una prueba de correlación serial residual (o efectos aleatorios individuales) localmente robusta frente a efectos aleatorios individuales (correlación serial) para modelos de panel y prueba conjunta de correlación serial y especificación de efectos aleatorios pero, la diferencia con el test anterior es que no realiza la prueba conjunta. Dado que el p-valor es 0.0000 rechazamos la hipótesis nula y aceptamos la hipótesis alternativa donde los errores son sub AR(1) con random effects.

Test robusto local para correlación serial o Efectos Aleatorios

| | |
|---------|--------|
| z | 18.445 |
| p-value | 0.0000 |

Al igual que los tests anteriores, este test realiza una prueba de correlación serial residual (o efectos aleatorios individuales) localmente robusta frente a efectos aleatorios individuales (correlación serial) para modelos

de panel y prueba conjunta de correlación serial y especificación de efectos aleatorios. Pero, para este test le indicamos que la prueba a realizar es la de efectivos aleatorios (re"). Dado que el p-valor es 0.0000 rechazamos la hipótesis nula por lo que aceptamos la hipótesis alternativa donde los errores con random effects son sub AR(1).

Test condicional LM para errores AR(1) o MA(1) bajo Efectos Aleatorios

| | |
|---------|--------|
| z | 8.4341 |
| p-value | 0.0000 |

El test que realizamos es una prueba del multiplicador de Lagrange para el nulo de no correlación serial, contra la alternativa de un proceso AR(1) o MA(1), en el componente idiosincrático del término de error en un modelo de panel de efectos aleatorios. Dado que el p-valor es 0.0000 rechazamos la hipótesis nula por lo que aceptamos la hipótesis alternativa donde los errores son AR(1)/MA(1) en un modelo de panel con random effects.

Test de correlación serial general

| | |
|---------|--------|
| chisq | 83.517 |
| df | 7 |
| p-value | 0.0000 |

El test que realizamos es una prueba del multiplicador de Lagrange que utiliza el modelo auxiliar en datos (cuasi) degradados tomados de un modelo de clase plm que puede ser una combinación (predeterminado para la interfaz de fórmula), aleatorio o dentro del modelo. A diferencia de la mayoría de las otras pruebas de correlación serial en paneles, esta permite elegir el orden de correlación para probar. Dado que el p-valor es 0.0000 rechazamos la hipótesis nula y aceptamos la hipótesis alternativa donde hay correlación serial en errores idiosincráticos.

Cuadro 1: Tabla conjunta

| | <i>Dependent variable:</i> | | |
|-------------------------|----------------------------|------------------------|---------------------|
| | Crime rate | | |
| | (1) Between | (2) Within | (3) Random Effects |
| Pa | −0.648* (0.088) | −0.385* (0.033) | −0.415* (0.030) |
| Pc | −0.528* (0.067) | −0.301* (0.021) | −0.325* (0.020) |
| Pp | 0.297 (0.231) | −0.191* (0.033) | −0.196* (0.033) |
| S | −0.236 (0.174) | 0.026 (0.025) | 0.019 (0.026) |
| POLICE | 0.364* (0.060) | 0.424* (0.027) | 0.415* (0.025) |
| DENSITY | 0.168* (0.077) | 0.409 (0.279) | 0.469** (0.051) |
| Percent Young Male | −0.095 (0.158) | 0.381 (0.325) | −0.132 (0.123) |
| WCON | 0.195 (0.210) | −0.034 (0.039) | −0.020 (0.039) |
| WTUC | −0.196 (0.170) | 0.029* (0.018) | 0.025 (0.018) |
| WTRD | 0.129 (0.278) | −0.039 (0.041) | −0.035 (0.042) |
| WFIR | 0.113 (0.220) | −0.013 (0.029) | −0.017 (0.029) |
| WSER | −0.106 (0.163) | 0.004 (0.019) | −0.004 (0.020) |
| WMFG | −0.025 (0.134) | −0.388* (0.102) | −0.264* (0.078) |
| WFED | 0.156 (0.287) | −0.553* (0.165) | −0.394* (0.144) |
| WSTA | −0.284 (0.256) | 0.216** (0.101) | 0.009 (0.086) |
| WLOC | 0.010 (0.463) | 0.341* (0.107) | 0.209** (0.099) |
| WEST | −0.230* (0.108) | | −0.207* (0.105) |
| CENTRAL | −0.164* (0.064) | | −0.180** (0.062) |
| URBAN | −0.035 (0.132) | | −0.236** (0.114) |
| Percent Minority | 0.148* (0.049) | | 0.199* (0.043) |
| CONSTANT | −2.097 (2.822) | | 0.252 (0.561) |
| Observations | 90 | 630 | 630 |
| R ² | 0.880 | 0.425 | 0.566 |
| Adjusted R ² | 0.846 | 0.310 | 0.551 |
| F Statistic | 25.412* (df = 20; 69) | 24.220* (df = 16; 524) | 793.199* |

Note:

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

```

#install.packages("stargazer")
#install.packages("plm")
#install.packages("Formula")
#install.packages("plyr")
#install.packages("tidyverse")
#install.packages("dplyr")

library(tidyverse)
library(dplyr)
library(plyr)
library(stargazer)
library(plm)
library(Formula)

data = read_csv("cornwell.csv")

#Renombramos las columnas para que coincidan con las del paper
data2 = plyr::rename(data, c("prbarr" = "Pa",
                             "prbconv" = "Pc",
                             "prbpris" = "Pp",
                             "avgsen" = "S",
                             "polpc" = "police",
                             "pctymle" = "pym",
                             "pctmin80" = "pm"))

#Establecemos el index en el año y county
data2 = pdata.frame(data2, index = c("county", "year"))

#Definimos el modelo
modelo = Formula(log(crmrte) ~ log(Pa) + log(Pc) + log(Pp) + log(S)
+ log(police) + log(density) + log(pym) + log(wcon)
+ log(wtuc) + log(wtrd) + log(wfir) + log(wser) +
log(wmfg) + log(wfed) + log(wsta) + log(wloc) +
west + central + urban + log(pm))

##### 1

#Corremos la regresión between
between = plm(modelo, data = data2, model = "between")

#Exportamos la tabla de regresion
stargazer(between,
          type="latex",
          dep.var.labels=c("Crime rate"),
          out="between")

##### 2

#Corremos la regresión within
within = plm(modelo, data = data2, model = "within")

```



```

#Exportamos la tabla de regresion
stargazer(within,
           type="latex",
           dep.var.labels=c("Sales"),
           out="within")

#Corremos la regresión OLS
ols = lm(modelo, data = data2)

# F test for individual effects
pFtest(within, ols)

##### 3

#Corremos la regresión random effects
random = plm(modelo, data = data2, model = "random")

#Exportamos la tabla de regresion
stargazer(random,
           type="latex",
           dep.var.labels=c("Sales"),
           out="random")

#Test de Hausman
phptest(within,random)

#Exportamos todas las regresiones juntas
stargazer(between, within, random,
           type="latex",
           dep.var.labels=c("Crime rate"),
           out="regresiones")

##### 5

# Test conjuntos de correlación serial y efectos aleatorios
pbsytest(modelo, data = data2, test = "J")

# Test robusto local para correlación serial o Efectos Aleatorios
pbsytest(modelo, data = data2)

# Test robusto local para correlación serial o Efectos Aleatorios
pbsytest(modelo, data = data2, test = "RE")

# Test condicional LM para errores AR(1) o MA(1) bajo Efectos
Aleatorios
pbltest(log(crmrte) ~ log(Pa) + log(Pc) + log(Pp) + log(S) +
log(police) + log(density) + log(pym) + log(wcon)
        + log(wtuc) + log(wtrd) + log(wfir) + log(wser) + log(wmfg)
        + log(wfed) + log(wsta) + log(wloc) +

```

```
      west + central + urban + log(pm), data = data2,  
alternative = "onesided")  
  
# Test de correlación serial general  
pbgtest(random)
```