

Universidad Torcuato Di Tella
Estrategias y Técnicas Cuantitativas en Ciencia Política II



Profesor: Mauricio Grotz
Alumno: Germán Epelbaum
Trabajo Práctico N°3
Fecha de entrega: 22/9/2023

Datos de Panel

Algunos estados de EE.UU. han promulgado leyes que permiten a los ciudadanos llevar armas. Estas leyes son conocidas como leyes de «emisión obligatoria», debido a que obligan a las autoridades locales a emitir un permiso para llevar armas a todos los solicitantes que sean ciudadanos: sean mentalmente competentes y no hayan sido condenados por un delito grave (algunos estados imponen algunas restricciones adicionales). Sus defensores sostienen que, si más personas llevan armas, el crimen se reducirá debido a que los criminales serán disuadidos de atacar a otras personas. Sus opositores argumentan que el crimen aumentará debido al uso accidental o espontáneo de las armas. En este ejercicio, se analiza el efecto de las leyes sobre la tenencia de armas sobre los crímenes violentos. En el campus virtual se encuentra el archivo de datos Guns.dta que contiene un panel balanceado de datos sobre 50 estados de EE.UU., más el Distrito de Columbia para los años 1977 a 1999. Se ofrece una descripción detallada en el archivo Guns-Description, también disponible en el campus virtual.

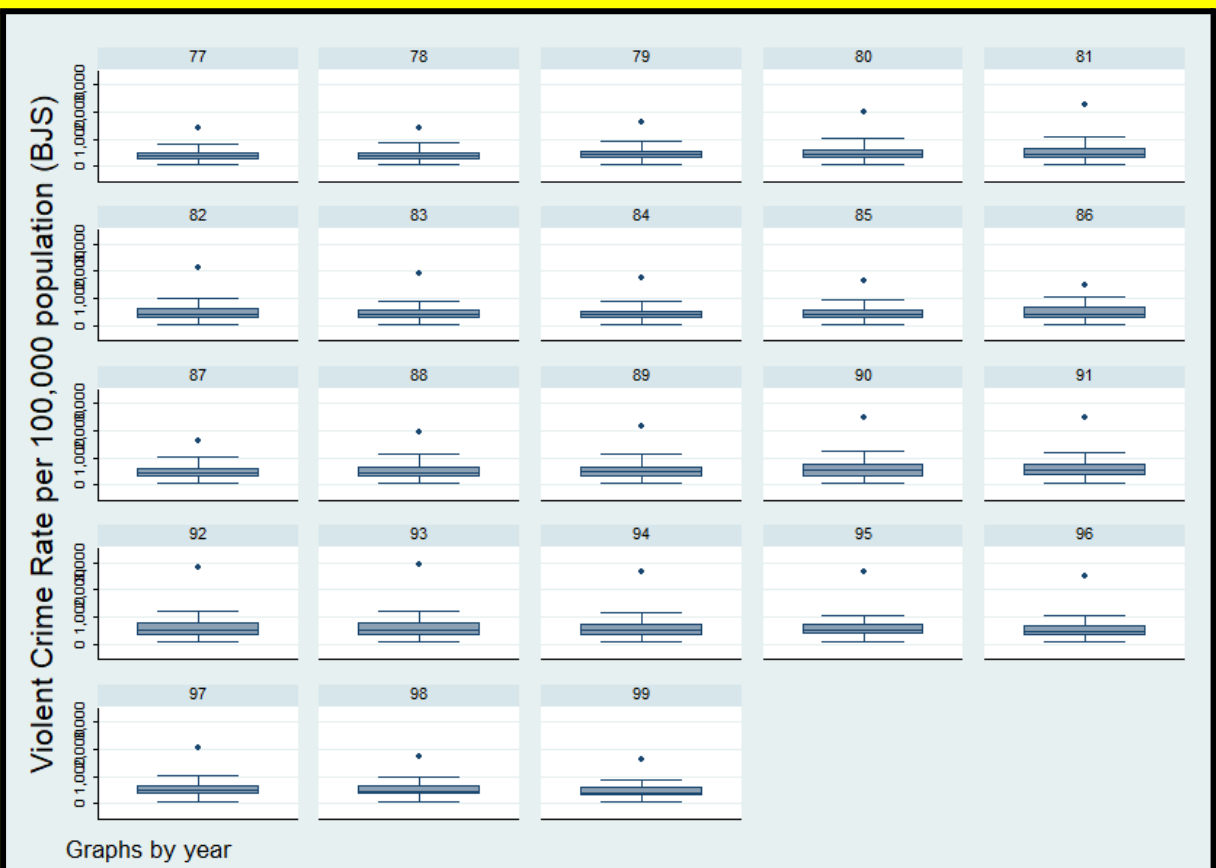
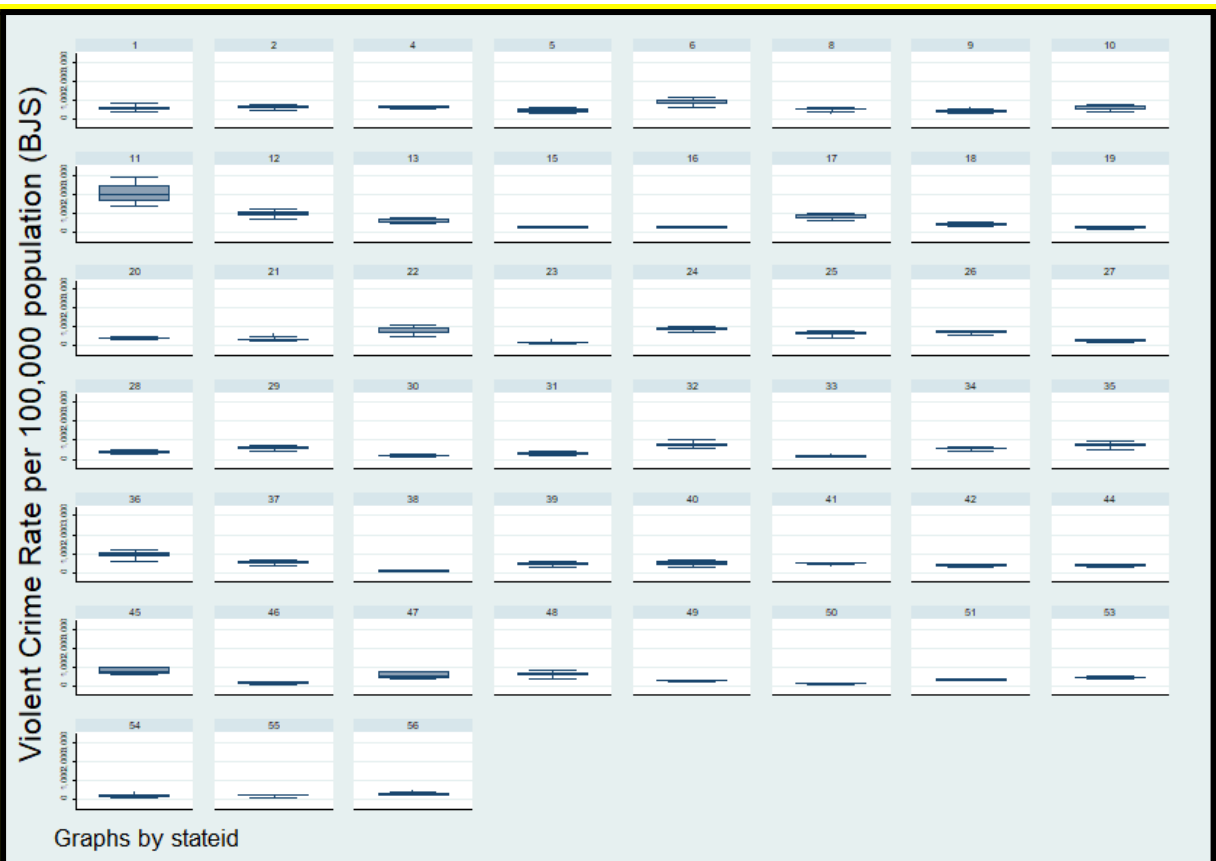
1) Presente las estadísticas descriptivas de las variables vio, rob, mur y shall. Interprete los resultados, analizando la variabilidad general, temporal y entre unidades de análisis.

```
. sum vio rob mur y shall
```

Variable	Obs	Mean	Std. Dev.	Min	Max
vio	1173	503.0747	334.2772	47	2921.8
rob	1173	161.8202	170.51	6.4	1635.1
mur	1173	7.665132	7.52271	.2	80.6
year	1173	88	6.636079	77	99
shall	1173	.2429668	.4290581	0	1

El panel presenta las estadísticas descriptivas de las variables, mostrando 1173 observaciones por variable. Las observaciones se repiten entre estados, de modo que para analizar la variabilidad temporal necesitamos otros recursos. Esto será realizado en el siguiente inciso.

2) Presente los gráficos que considere adecuados para dar cuenta de la heterogeneidad individual y temporal observada de la variable vio.



En el gráfico de caja (boxplot) se observa la existencia de heterogeneidad entre los estados. Se destaca la cantidad de hechos de violencia acaecidos en el estado N°11.

El gráfico de caja anualizado, permite observar la existencia de variabilidad temporal al interior de los propios estados, aunque no en la medida que se da entre los estados. Ante esta situación, se presenta como valiosa la posibilidad de realizar una prueba de efectos fijos, a fin de profundizar en esta hipótesis,

3) Estime una regresión lineal simple de la variable vio sobre la variable shall y una regresión lineal múltiple con las variables independientes shall, incarc_rate, density, avginc, pop, pb1064, pw1064 y pm1029.

```
. reg vio shall
```

Source	SS	df	MS	Number of obs = 1173		
Model	5605560.24	1	5605560.24	F(1, 1171) = 52.36		
Residual	125355176	1171	107049.681	Prob > F = 0.0000		
				R-squared = 0.0428		
				Adj R-squared = 0.0420		
Total	130960737	1172	111741.243	Root MSE = 327.18		

vio	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
shall	-161.1868	22.27475	-7.24	0.000	-204.8897	-117.484
_cons	542.2377	10.97959	49.39	0.000	520.6959	563.7796

```
. reg vio shall incarc_rate density avginc pop pb1064 pw1064 pm1029
```

Source	SS	df	MS	Number of obs = 1173		
Model	94776406.5	8	11847050.8	F(8, 1164) = 381.10		
Residual	36184330.1	1164	31086.1942	Prob > F = 0.0000		
				R-squared = 0.7237		
				Adj R-squared = 0.7218		
Total	130960737	1172	111741.243	Root MSE = 176.31		

vio	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
shall	-92.72125	13.4257	-6.91	0.000	-119.0625	-66.37997
incarc_rate	.8155948	.044174	18.46	0.000	.7289253	.9022642
density	94.66882	5.428439	17.44	0.000	84.0182	105.3194
avginc	1.234759	3.207344	0.38	0.700	-5.058062	7.527581
pop	18.50567	1.054859	17.54	0.000	16.43603	20.57531
pb1064	11.29196	6.864417	1.64	0.100	-2.176053	24.75998
pw1064	2.702383	3.453603	0.78	0.434	-4.0736	9.478366
pm1029	9.316543	4.441374	2.10	0.036	.602549	18.03054
_cons	-178.7774	224.0104	-0.80	0.425	-618.2867	260.7319

a. Interprete el coeficiente de la variable shall de la regresión lineal simple e indique si considera que el efecto de la legislación es sustantivo.

En la regresión lineal simple se observa que la modificación de la variable independiente tiene un efecto por cada unidad de -161.19 unidades sobre la variable dependiente (crímenes violentos cada cien mil habitantes). Siendo el coeficiente p significativo, podemos afirmar con un 95% de certeza que el verdadero parámetro se encuentra entre -204,89 y -117.48.

A su vez, el R ajustado nos indica que la modificación de esta variable se asocia en un 4,2% en el cambio de la variable dependiente. Este es un valor sustantivo, de modo que podemos asociar la mencionada ley con una reducción de los incidentes armados.

Estos resultados aportan información bastante relativa, puesto que puede verse afectado por el sesgo de variable omitida.

b. Al añadir las variables de control en la regresión múltiple ¿cambia el efecto estimado de la ley de emisión obligatoria? Justifique.

Al añadir nuevas variables de control, la variación de la variable independiente en cuestión se asocia, *ceteris paribus*, con una modificación de la variable dependiente de -92,72 unidades por cada modificación de la variable independiente, una curva menos pronunciada que en el modelo de regresión simple.

Aplicando las variables de control a la regresión, sigue siendo muy significativo el peso de la variable *shall*, aunque con menos fuerza que en el modelo con sólo dos variables. Su efecto es negativo también, de manera que la aplicación de esta ley se asocia con una reducción de estos incidentes. Esto puede afirmarse al ser significativos los resultados del test P, con un 95% de certeza. Podemos sostener que el verdadero parámetro se encuentra entre menos 119.06 y -66.38.

Cuando se incorporan las variables de control se evidencia que la influencia de la variable *shale* sobre la dependiente está sobredimensionada. A su vez, disminuye el error estándar, que pasa de 22,27 a 13,42, haciendo más acotado el intervalo de confianza.

Destacamos que los resultados de la regresión múltiple pueden aportar indicios sobre nuestras hipótesis, pero deben ser analizados con reparos: se trata de observaciones repetidas en estados. Es relevante la aplicación de una técnica de panel, que permita estudiar en qué medida los cambios se deben a las características/proiedades no observadas de cada estado o a modificaciones dadas en el tiempo. Esto se da mediante la contemplación de la correlación de los errores, tomando como error del modelo a la variación de la que no se puede dar cuenta.

c. Sugiera una variable que cambie entre los estados, pero que probablemente cambie poco, o nada en absoluto, en el tiempo y que pudiera causar sesgo de variable omitida en la regresión.

Un factor que hipotetizamos tiene que ver con los incidentes y la violencia armada radica en características de las sociedades en las cuales se producen estos hechos. Centralmente, me interesaría detectar grados de conflictividad social como una variable explicativa de estos incidentes.

La hostilidad o conflictividad entre las personas no necesariamente se expresa a través del uso de agresiones mediante armamento. Sin embargo, entendemos que este puede ser un factor determinante.

Al respecto, construiría un índice que pondere el nivel de resolución violenta de conflictos en un estado. Utilizaría, por ejemplo, los registros de inspectores escolares en escuelas sobre hechos de violencia y agresión en centros educativos, consideraría peleas y reyertas en espectáculos deportivos, musicales, como también a nivel familiar. De este modo, a partir de la operacionalización de esta dimensión de la unidad de análisis podríamos dar cuenta del grado de tensión o violencia en un estado. Suponemos que este nivel de violencia puede ser una de las variables explicativas de incidentes armados, siendo idiosincrática de cada uno de estos estados y bastante durable en el tiempo.

d. ¿Resulta apropiado aplicar un modelo POLS? Justifique.

El modelo POLS se presenta en la regresión lineal múltiple -con variables de control- más arriba en este trabajo. Estima individualmente los coeficientes, tomando unidades como diferentes y asumiendo que la estructura de panel no agrega información, cuando en realidad esto no es así, puesto que hay observaciones que se repiten en los estados.

A partir del gráfico de cajas, habíamos encontrado indicios de que no se puede ignorar la heterogeneidad entre las distintas unidades de análisis. Por ejemplo, la unidad de análisis “estado N1995” comparte características con la unidad de análisis “estado N1196”, que impactan sobre la dependiente pero que este modelo no puede ver.

4) ¿Cambian los resultados de la regresión múltiple cuando se agregan los efectos fijos individuales de cada estado? Si es así, ¿qué modelo considera más adecuado y por qué?

Fixed-effects (within) regression		Number of obs	=	1173
Group variable: stateid		Number of groups	=	51
R-sq: within	= 0.2038	Obs per group: min	=	23
between	= 0.3446	avg	=	23.0
overall	= 0.2658	max	=	23
corr(u_i, Xb) = -0.8388		F(8,1114)	=	35.64
		Prob > F	=	0.0000

vio	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
shall	-18.57508	11.56324	-1.61	0.108	-41.26327	4.113097
incarc_rate	.0967879	.0573672	1.69	0.092	-.015772	.2093479
density	-155.9533	52.11761	-2.99	0.003	-258.213	-53.69351
avginc	-4.066645	3.621108	-1.12	0.262	-11.17161	3.038316
pop	12.24406	5.346751	2.29	0.022	1.753224	22.7349
pbl064	12.80803	10.88266	1.18	0.239	-8.54479	34.16085
pwl064	10.32325	3.110094	3.32	0.001	4.220951	16.42556
pml029	-23.86385	3.924761	-6.08	0.000	-31.56461	-16.16309
_cons	203.0109	235.8217	0.86	0.389	-259.6938	665.7157
sigma_u	500.37955					
sigma_e	98.505026					
rho	.96269175	(fraction of variance due to u_i)				

F test that all u_i=0:	F(50, 1114) =	52.30	Prob > F = 0.0000
------------------------	---------------	-------	-------------------

El modelo de regresión incorporando los factores fijos, permite dar cuenta de aquellos aspectos constantes en el tiempo pero no observados entre los estados. Permite dar un mayor grado de certeza sobre ciertos aspectos de la variable que previamente aparecen como errores de medición, y que permanecen en el tiempo.

Esto resulta interesante considerando que el gráfico de cajas nos había dado indicios de que la variabilidad depende más de las características de cada estado que del paso del tiempo. Ofrece un mayor grado de acercamiento sobre el papel de las variables independientes en la modificación de la dependiente, por lo que los valores se expresan con mayor seguridad. El valor de la variable shall es (coeficiente de -18,58) es menor que en los modelos anteriores,

aunque con un valor p tan elevado que impide descartar la hipótesis nula, por lo que deja de ser significativa. Esto también sucede con las variable avginc, con Pbl064 y con la constante. Los estadísticos p son menores en el primer modelo de regresión para la variable shall, incarc_rate, density, pop, Pbl064. Pbl064 y Pml 029 tienen un estadístico p menor en el segundo modelo.

A su vez, el menor valor del R2 en el segundo modelo, supone que este explica menos la variación de la variable que el anterior. Sin embargo, lo hace con mayor grado de confianza.

Dados estos resultados, considerando los aspectos que el modelo permite revelar y las observaciones que se vienen realizando sobre nuestro objeto de estudio, resultaría relevante incorporar efectos fijos al estudio de la problemática en cuestión.

5) ¿Incorporaría efectos fijos temporales? Justifique su respuesta.

```
. xtset stateid year
      panel variable:  stateid (strongly balanced)
      time variable:  year, 77 to 99
              delta:  1 unit
```

```
. xtreg vio shall incarc_rate density avginc pop pbl064 pw1064 pml029 i.year, fe

Fixed-effects (within) regression              Number of obs   =       1173
Group variable: stateid                      Number of groups =        51

R-sq:  within = 0.4202                      Obs per group:  min =        23
        between = 0.4545                      avg           =       23.0
        overall = 0.2462                      max           =        23

                                           F(30,1092)       =       26.38
corr(u_i, Xb)  = -0.7698                     Prob > F         =       0.0000
```

vio	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
shall	-.8390586	10.40307	-0.08	0.936	-21.25132	19.5732
incarc_rate	.1915715	.0547407	3.50	0.000	.0841626	.2989803
density	-95.20334	46.25107	-2.06	0.040	-185.9544	-4.452321
avginc	6.887878	3.901603	1.77	0.078	-.7676083	14.54337
pop	2.057769	4.770199	0.43	0.666	-7.302025	11.41756
pbl064	-36.09287	13.75854	-2.62	0.009	-63.08904	-9.0967
pw1064	-10.93647	4.766677	-2.29	0.022	-20.28935	-1.583584
pml029	47.17105	9.466965	4.98	0.000	28.59555	65.74655

```

F test that all u_i=0:      F(50, 1092) =      66.05      Prob > F = 0.0000

. testparm i.year

( 1)  78.year = 0
( 2)  79.year = 0
( 3)  80.year = 0
( 4)  81.year = 0
( 5)  82.year = 0
( 6)  83.year = 0
( 7)  84.year = 0
( 8)  85.year = 0
( 9)  86.year = 0
(10)  87.year = 0
(11)  88.year = 0
(12)  89.year = 0
(13)  90.year = 0
(14)  91.year = 0
(15)  92.year = 0
(16)  93.year = 0
(17)  94.year = 0
(18)  95.year = 0
(19)  96.year = 0
(20)  97.year = 0
(21)  98.year = 0
(22)  99.year = 0

      F( 22, 1092) =      18.53
      Prob > F =      0.0000

```

El modelo de efectos temporales es útil cuando tenemos la sospecha de que en cierto año pasó determinado evento que puede incidir en los resultados. Por el contrario, la debilidad emerge ante sus dificultades para estimar variables constantes.

En este modelo la variable shall tiene un coeficiente de -8390586 , con un parámetro que pasa va desde el -21.25132 al 19.5732, por lo cual pasa por el valor 0 (no se rechaza hipótesis nula). El error estándar es de 10.40307 con un estadístico p de 0.936, que tampoco permite negar la hipótesis nula.

El hecho de que aparezca el valor 0.0 junto a cada año supone que el modelo necesita efectos temporales. Cuando dichos efectos se agregan, la variable shale pierde incidencia, a la vez que los valores p son altísimos.

Se devela según este modelo que las leyes sobre portación de armas no tienen incidencia en la reducción de los crímenes violentos, dentro de cada estado, a lo largo del tiempo.

6) Discuta si le parece adecuado o no utilizar un modelo con efectos aleatorios. Realice el test correspondiente. Justifique brevemente.


```
estimates store re
```

```
.
```

```
. hausman fe re
```

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fe	(B) re		
shall	-18.57508	-37.19344	18.61836	.
incarc_rate	.0967879	.3931534	-.2963654	.0413254
density	-155.9533	104.7166	-260.6699	49.60782
avginc	-4.066645	-6.025157	1.958512	.8710399
pop	12.24406	13.15851	-.9144534	4.374196
pbl064	12.80803	20.61077	-7.802738	8.163496
pw1064	10.32325	8.026554	2.2967	.3248615
pml029	-23.86385	-9.640604	-14.22325	1.508542

b = consistent under Ho and Ha; obtained from xtreg

B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

```
chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
          = 64.49
Prob>chi2 = 0.0000
(V b-V B is not positive definite)
```

El test de Hausman nos permite averiguar si cambia sustantivamente la estimación de coeficientes agregando efectos fijos o aleatorios.

En este caso el resultado del test es cercano a cero, sugiriendo el uso del modelo de efectos fijos.

La probabilidad del chi2 del test es:

Prob > chi2 = 0.0000

7) Presente una tabla que compare las distintas regresiones múltiples: POLS, Efectos Aleatorios, Efectos Fijos Individuales, Efectos Fijos Individuales y Temporales. (Nota: incluya en la tabla solo las variables independientes: shall, incarcerate, density, avginc, pop, pbl064, pw1064 y pm1029). ¿Qué modelo le parece el más adecuado? Justifique brevemente utilizando los tests correspondientes.

VARIABLES	POLS	EFFECTOS ALEATORIO S	EFFECTOS FIJOS	EFFECTOS FIJOS Y TEMPORAL ES
shall	-92.72*** (13.43)	-37.19*** (11.60)	-18.58 (11.56)	-839 (10.40)
incarc_rate	0.816*** -442	0.393*** -398	0.0968* -574	0.192*** -547
density	94.67*** -5.428	104.7*** (15.98)	-156.0*** (52.12)	-95.20** (46.25)
avginc	1.235 -3.207	-6.025* -3.515	-4.067 -3.621	6.888* -3.902
pop	18.51*** -1.055	13.16*** -3.075	12.24** -5.347	2.058 -4.770
pb1064	11.29 -6.864	20.61*** -7.197	12.81 (10.88)	-36.09*** (13.76)
pw1064	2.702 -3.454	8.027*** -3.093	10.32*** -3.110	-10.94** -4.767
pm1029	9.317** -4.441	-9.641*** -3.623	-23.86*** -3.925	47.17*** -9.467
Constant	-178.8 (224.0)	-54.70 (231.1)	203.0 (235.8)	315.3 (285.9)
Observations	1,173	1,173	1,173	1,173
R-squared	0.724		0.204	0.420
Number of stateid		51	51	51

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Según lo que puede verse en las tablas, en el modelo POLS el valor de R2 es el mayor de todos los comparativos. A su vez, en esta variable shall -de interés para este trabajo- resulta significativa según los estadísticos p.

No obstante, hay errores de especificación que no deben pasarse por alto, dando la pauta de que se está trabajando con modelos inadecuados para el tipo de datos disponibles.

Más adelante se verá como ninguna de estas variables es significativa, por lo que estos modelos no son útiles para el trabajo emprendido.

8) A partir de su análisis previo, ¿qué conclusiones sacaría sobre los efectos de las leyes de tenencia de armas sobre los crímenes violentos?

Los primeros modelos parecían mostrarnos que las leyes de portación de armas tienen incidencia sobre los crímenes violentos.

No obstante, al ir aplicando los modelos subsiguientes se pudo observar que el pretendido efecto de esta variable sobre la dependiente desaparece. **Podemos afirmar que no hay indicios estadísticos que vinculen a las leyes de tenencia de armas con la reducción de crímenes violentos.**

9) Para el modelo seleccionado, analice si presenta problemas de heteroscedasticidad, correlación contemporánea y/o correlación serial. A partir de los resultados obtenidos, ejecute una regresión utilizando Panel-Corrected Standard Errors (Beck y Katz, 1995) y analice si se ven afectadas las conclusiones a las que arribó en el punto anterior.

```
. xtreg vio shall incarc_rate density avginc pop pb1064 pw1064 pml029 i.year, fe robust
```

```
Fixed-effects (within) regression                Number of obs   =       1173
Group variable: stateid                        Number of groups =        51

R-sq:  within = 0.4202                          Obs per group: min =        23
        between = 0.4545                             avg =       23.0
        overall = 0.2462                             max =        23

                                                F(30,50)        =       183.67
corr(u_i, Xb)  = -0.7698                        Prob > F         =       0.0000
```

(Std. Err. adjusted for 51 clusters in stateid)

vio	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
shall	-.8390586	19.83465	-0.04	0.966	-40.67812	39
incarc_rate	.1915715	.1188741	1.61	0.113	-.0471942	.4303371
density	-95.20334	85.25139	-1.12	0.269	-266.4358	76.02911
avginc	6.887878	8.380836	0.82	0.415	-9.945525	23.72128
pop	2.057769	10.21112	0.20	0.841	-18.45188	22.56741
pb1064	-36.09287	39.27422	-0.92	0.363	-114.9775	42.79172
pw1064	-10.93647	15.52461	-0.70	0.484	-42.11857	20.24564
pml029	47.17105	32.3086	1.46	0.151	-17.72268	112.0648

Esta regresión corrige errores de panel, tornando más robustos a los errores. No cambia sustancialmente, permaneciendo iguales los coeficientes. Es decir que el efecto promedio sobre la variable dependiente es el mismo,

El coeficiente de la variable shall es el mismo, aunque se incrementa el error estándar. Esto nos da una menor noción del intervalo de confianza y vuelve menos significativa a las variables.

Los errores estándar corregidos por panel hicieron que ninguna variable sea significativa, pasando todas por el cero.

Como se viene afirmando, los crímenes violentos dependen más de características propias del estado que de la evolución de las propias variables.