

ANÁLISIS DE LA TASA DE HOMICIDIOS EN EUA: ¿EXISTE UNA DIFERENCIA A PARTIR DE 1991?

**INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO
FUNDAMENTOS DE ESTADÍSTICA**

Entrega Final - 16 diciembre 2015

Fernanda Alcalá
Mauricio González

Introducción

A partir de 1991 hubo una considerable reducción en ciertos índices de violencia de EUA, generalizada en todo el territorio, para todos los estados. Esto levantó varias cejas, pues las diferentes variables estereotípicas para medición de seguridad (policias, pobreza, etc.) eran sumamente dispares entre estados y no existía un buen modelo que pudiera capturar alguna relación latente que pudiera explicar el comportamiento a la baja de la violencia.

Se propuso que existía un factor externo que afectó a todos los niveles de violencia (estudiados) a partir de 1991: el hecho de que fue precisamente en 1973, 18 años antes, cuando fue legalizado el aborto en todo el territorio de los Estados Unidos. Un estudio bastante reconocido, publicado por Donohue y Levitt, defiende la hipótesis de que este acontecimiento es la causa de la baja en la violencia.

En el artículo se comenta ampliamente que existe evidencia de que el periodo de los 18-24 años de edad es precisamente el pico de violencia para la población en general, y que en 1991 se alcanza la primera oleada de niños nacidos bajo el nuevo esquema permissivo.

Se teoriza que la legalización del aborto es la causa, pues las mujeres más susceptibles de necesitar tal recurso pertenecen a sectores poco favorecidos de la población: mujeres adolescentes, o pertenecientes a grupos marginalizados (como ciertas razas). Al tener esta decisión, y poder elegir el momento de procreación, aseguran un mejor futuro para sus hijos, en lugar de estar forzados a vivir en condiciones poco idóneas tanto económicas como sociales y familiares, lo cual los habría orillado a tener comportamiento violento más adelante en su vida.

Entonces, el abrupto cambio en la tendencia podría ser explicado por el hecho de que el grupo de población susceptible de crecer para ser violento, efectivamente era más pequeño (comparativamente con la población), pues muchos de ellos habrían venido de familias poco privilegiadas que no habrían tenido la oportunidad de decidir si podrían darles una vida digna.

Contamos con datos proporcionados por el Profesor de la Materia. Se tiene información de las siguientes variables para cada uno de los 50 estados, del año 1966 al año 1999.

- xxprison.- Número de prisioneros (por cada mil habitantes).
- xxpolice.- Número de policías (por cada mil habitantes).
- xxunemp.- Tasa de desempleo (porcentaje).
- xxincome.- Variable de ingreso (tasa).
- xxpover.- Tasa de pobreza (porcentaje de hogares debajo de la línea de pobreza).

Buscamos explicar alguna de las variables relacionadas con el crimen, entre las cuales se encuentran crimen con violencia, crimen de propiedad y homicidio.

Análisis Exploratorio de los Datos

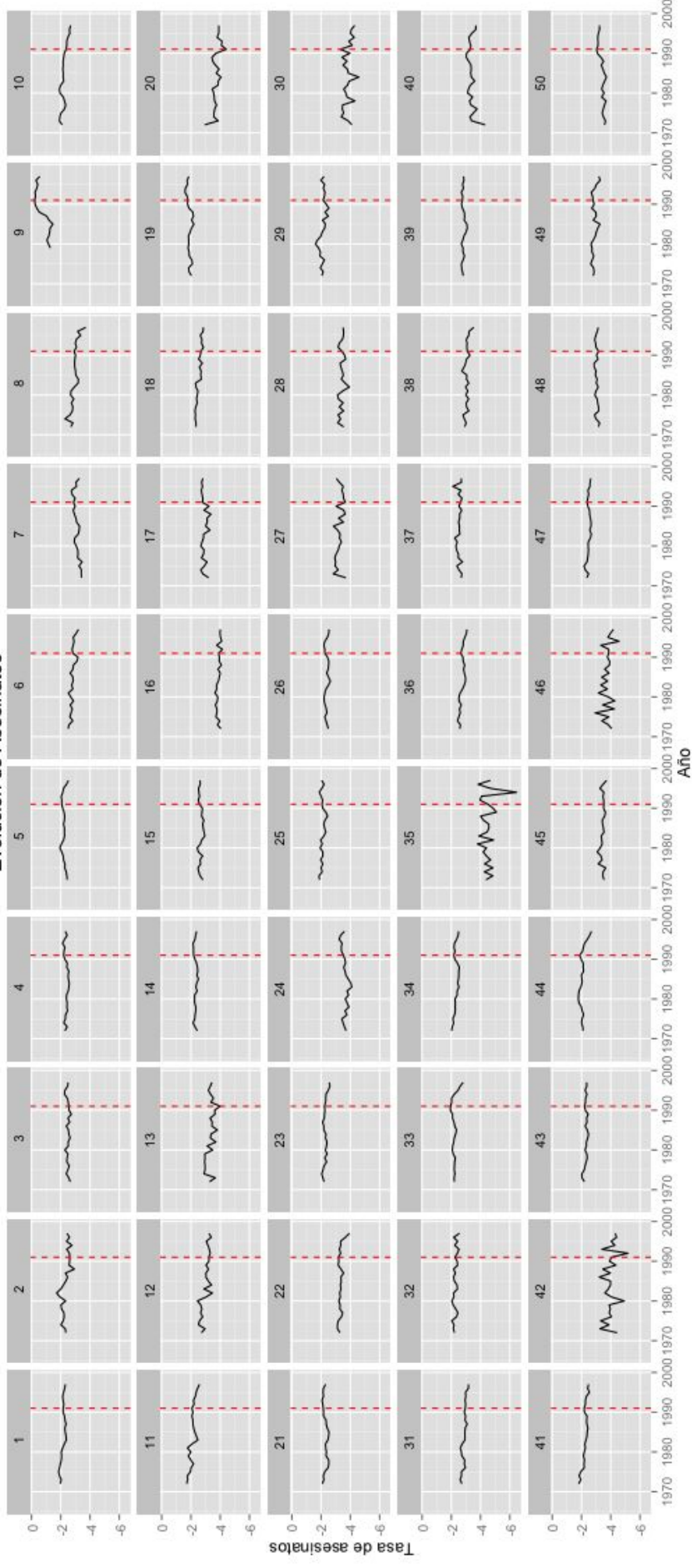
Un primer vistazo a los datos, para familiarizarnos con ellos:

statenum	year	popul	lpc_viol	lpc_prop	lpc_murd
Min. : 1.0	Min. :66.0	Min. : 304	Min. : -0.7774	Min. :2.561	Min. : -6.4583
1st Qu.:13.0	1st Qu.:74.0	1st Qu.: 1198	1st Qu.: 0.9923	1st Qu.:3.601	1st Qu.: -3.2543
Median :25.5	Median :82.5	Median : 3266	Median : 1.4402	Median :3.774	Median : -2.6813
Mean :25.5	Mean :82.5	Mean : 4742	Mean : 1.3594	Mean :3.756	Mean : -2.7859
3rd Qu.:38.0	3rd Qu.:91.0	3rd Qu.: 5513	3rd Qu.: 1.8140	3rd Qu.:3.967	3rd Qu.: -2.2703
Max. :50.0	Max. :99.0	Max. :33145	Max. : 3.3748	Max. :4.560	Max. : -0.2089
		NA's :200	NA's :406	NA's :406	NA's :406
efamurd	efaviol	efaprop	xxprison	xxpolice	xxunemp
Min. :0.0000	Min. :0.0006	Min. :0.0026	Min. : -1.5938	Min. :0.2686	Min. :0.0192
1st Qu.:0.0154	1st Qu.:0.0804	1st Qu.:0.2291	1st Qu.: -0.0853	1st Qu.:0.7611	1st Qu.:0.0490
Median :0.1349	Median :0.3091	Median :0.6985	Median : 0.4450	Median :0.8877	Median :0.0615
Mean :0.3760	Mean :0.5867	Mean :1.0379	Mean : 0.4219	Mean :0.9267	Mean :0.0639
3rd Qu.:0.5175	3rd Qu.:0.8223	3rd Qu.:1.4801	3rd Qu.: 0.9267	3rd Qu.:1.0483	3rd Qu.:0.0756
Max. :6.2191	Max. :7.5094	Max. :9.7653	Max. : 2.9554	Max. :2.2060	Max. :0.1803
NA's :1050	NA's :1050	NA's :1050	NA's :450	NA's :450	NA's :400
xxincome	xxpover	xxafdc15	xxgunlaw	xxbeer	
Min. : 9.397	Min. : 2.90	Min. : 1526	Min. :0.0000	Min. : 2.50	
1st Qu.: 9.770	1st Qu.:10.50	1st Qu.: 4856	1st Qu.:0.0000	1st Qu.:20.90	
Median : 9.900	Median :13.00	Median : 6515	Median :0.0000	Median :23.10	
Mean : 9.904	Mean :13.73	Mean : 6823	Mean :0.2509	Mean :23.51	
3rd Qu.:10.028	3rd Qu.:16.20	3rd Qu.: 8650	3rd Qu.:1.0000	3rd Qu.:25.60	
Max. :10.487	Max. :27.20	Max. :15436	Max. :1.0000	Max. :40.40	
NA's :400	NA's :800	NA's :1000	NA's :50	NA's :1000	

Para este trabajo, hemos decidido analizar el posible cambio en la tasa de asesinatos a través del tiempo. Contamos con datos temporales para cada estado (1-50). Observemos el comportamiento general de esta variable de interés:



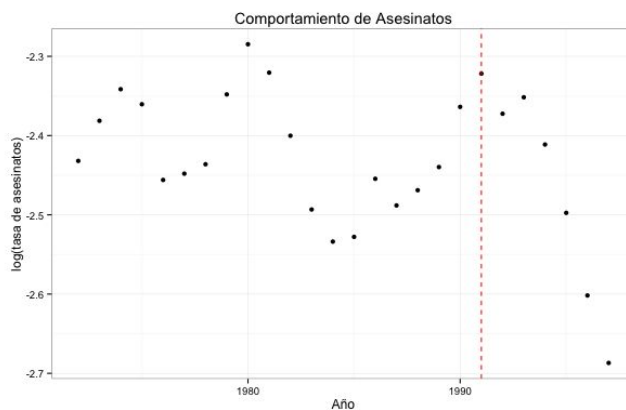
Evolución de Asesinatos



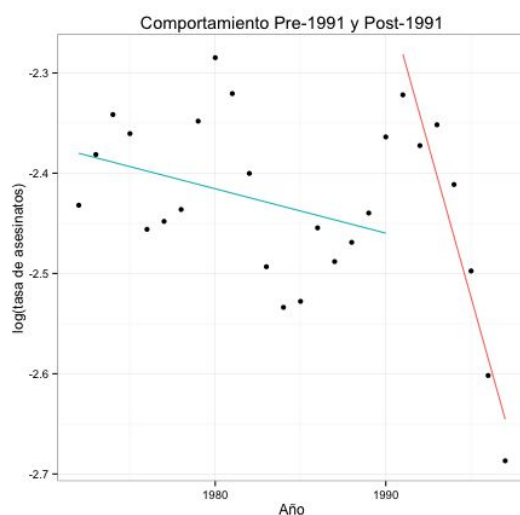
Tasas de Criminalidad antes y después: ¿hay un cambio significativo?

Después de observar las gráficas anteriores, en la cual se contrasta, por medio de un indicador rojo, el comportamiento de las tasas antes y después de 1991, sospechamos que la mayoría tienen una pendiente negativa, independientemente de su comportamiento previo.

Analicemos la tasa de criminalidad a lo largo de todo el país. Tomemos un promedio ponderado (con la población) de dichas tasas. En la siguiente gráfica, la línea punteada roja representa el punto de corte teórico, en 1991.

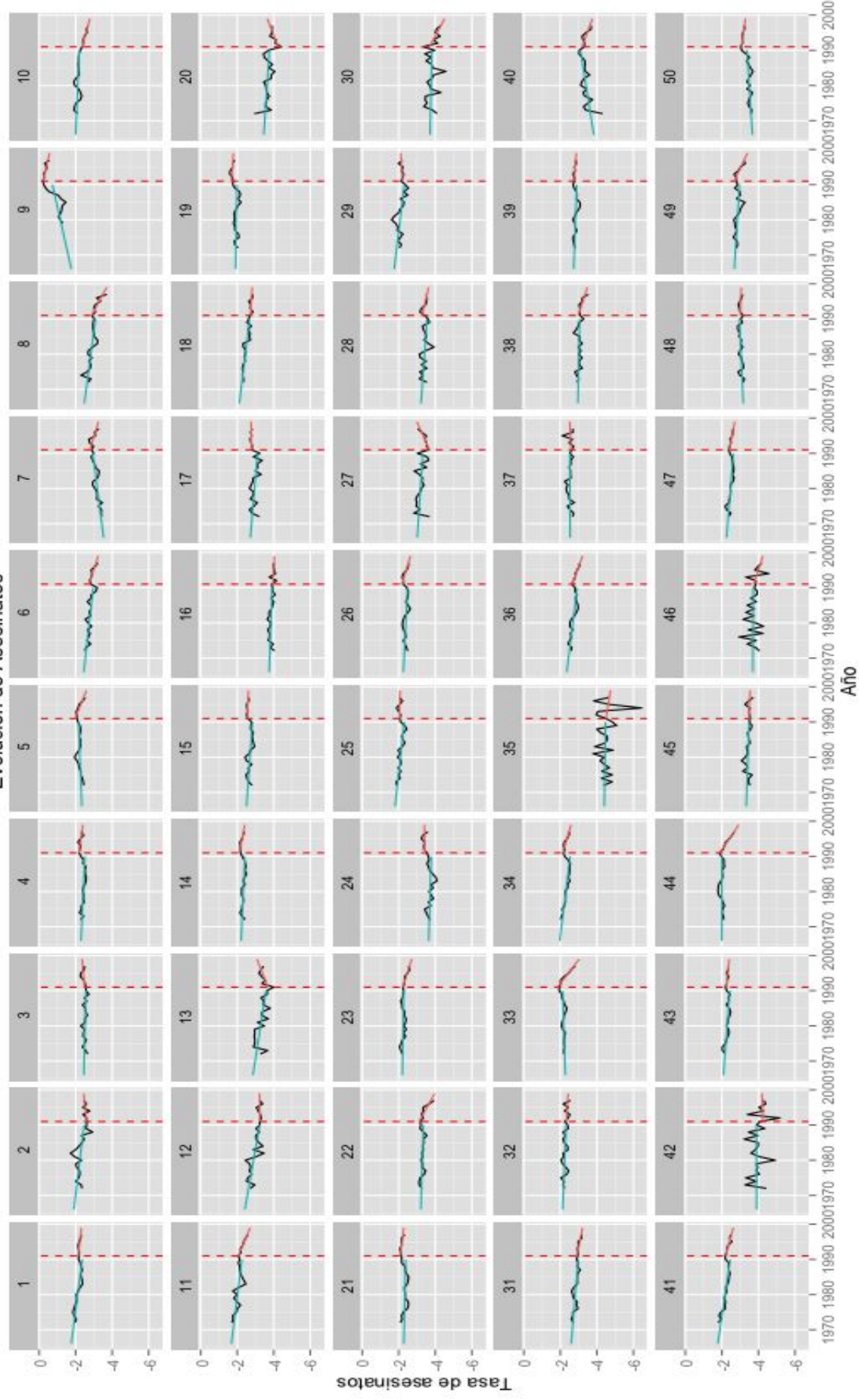


Un buen acercamiento a determinar si existe un cambio, es obtener regresiones lineales para cada uno de los subconjuntos de datos (antes/después). Como las pendientes son suficientemente diferentes, entonces tenemos evidencia de que ha habido un cambio.



A continuación se muestra el análisis desglosado por estado. Comparemos, para cada uno de ellos, una regresión lineal comprendida de una combinación de dos modelos: pre y post 1991.

Evolución de Asesinatos



Se sospecha una tendencia a la baja en general, independientemente de la diferencia del año. Hay algunos estados, como el 13 y el 20, que muestran una ligera tendencia a la alta. Es notable, sin embargo, el abrupto cambio de pendiente en algunos estados como el 9, 33, 34 y 44.

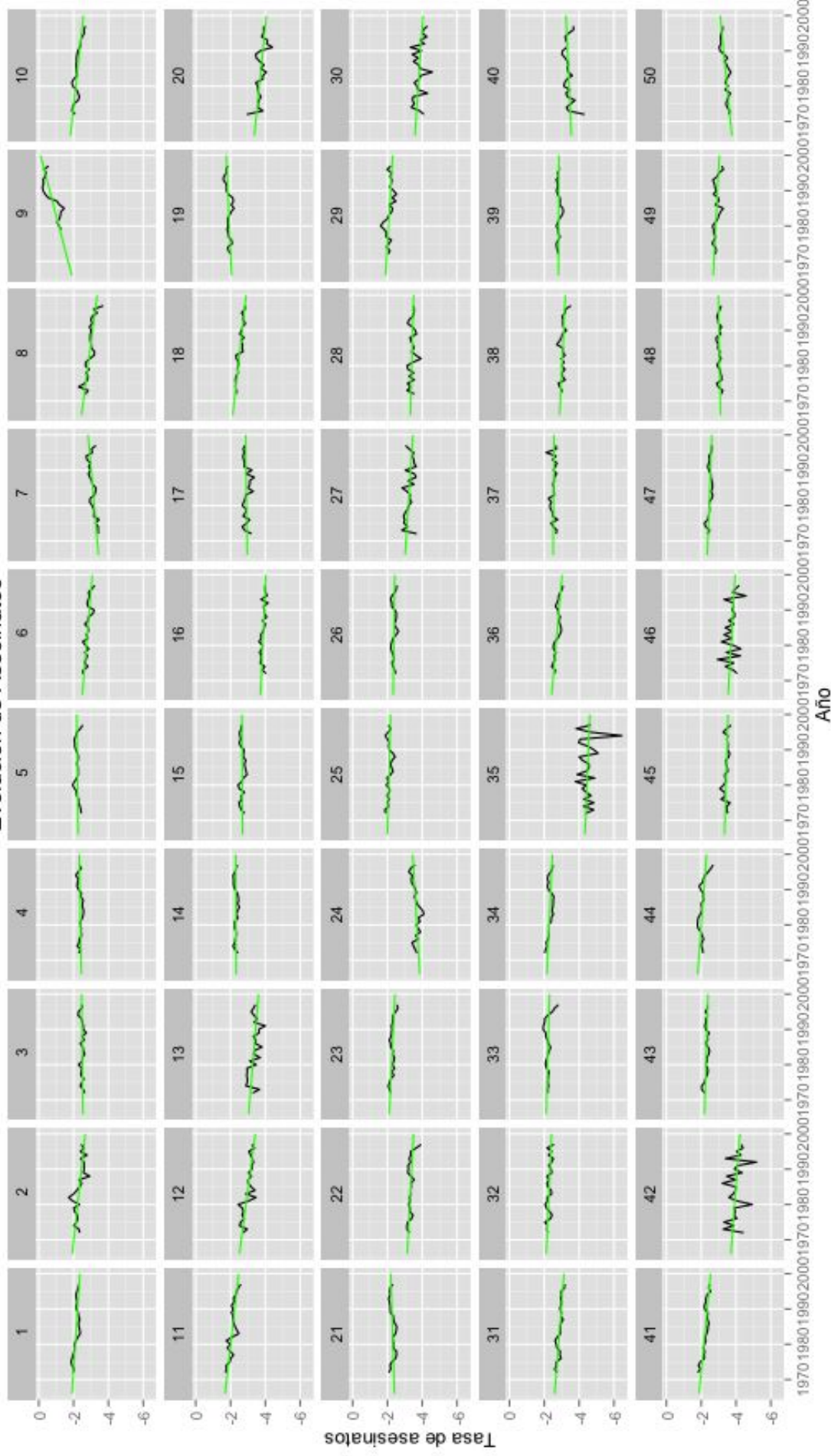
Veamos ahora predicciones para el año 2000 para cada estado, con ambos modelos. Los colores indican, para cada estado, cuál de las predicciones es mayor o menor: una disminución en la predicción implicaría menor tasa de asesinatos, lo cual conlleva la connotación positiva que le hemos dado con el color verde.

Es bastante claro que existe evidencia empírica que sustenta la hipótesis de que los datos se comportan sustancialmente diferente a partir de 1991: para más de dos tercios de los estados, la tasa de asesinatos viene claramente a la baja desde 1991.

Estado	Prediccion	Predicción datos>1991	Estado	Prediccion	Predicción datos>1991
1	● -2.36087	● -2.33652	26	● -2.42215	● -2.69978
2	● -2.66793	● -2.42261	27	● -3.46412	● -2.90316
3	● -2.45552	● -2.33903	28	● -3.50376	● -3.68150
4	● -2.32527	● -2.42275	29	● -2.30308	● -2.09691
5	● -2.17642	● -2.67024	30	● -4.04488	● -4.61491
6	● -3.07059	● -3.29014	31	● -3.12754	● -3.25315
7	● -2.83271	● -3.30741	32	● -2.41215	● -2.43493
8	● -3.35753	● -3.80301	33	● -2.29612	● -3.18713
9	● -0.09639	● -0.61899	34	● -2.46404	● -2.64547
10	● -2.56379	● -2.82803	35	● -4.61497	● -4.77286
11	● -2.46737	● -2.79081	36	● -3.02138	● -3.29000
12	● -3.42838	● -3.21949	37	● -2.55534	● -2.50273
13	● -3.60860	● -2.99870	38	● -3.20679	● -3.54113
14	● -2.29828	● -2.44897	39	● -2.83800	● -2.89467
15	● -2.65581	● -2.62809	40	● -3.23928	● -3.82717
16	● -4.02673	● -4.04387	41	● -2.53853	● -2.67492
17	● -2.87555	● -2.76010	42	● -4.22064	● -4.18512
18	● -2.89873	● -2.83360	43	● -2.38884	● -2.42211
19	● -1.75682	● -1.81261	44	● -2.30618	● -3.06012
20	● -4.06240	● -3.57042	45	● -3.53688	● -3.55703
21	● -2.19149	● -2.29684	46	● -3.95298	● -4.29976
22	● -3.49561	● -4.07712	47	● -2.60977	● -2.74970
23	● -2.43639	● -2.77352	48	● -2.98288	● -3.05773
24	● -3.45120	● -3.39753	49	● -3.03999	● -3.49692
25	● -2.17531	● -2.07998	50	● -3.08646	● -3.32177

En la figura la estimación para el año 2000, para cada uno de los estados. Podemos observar que, si no suponemos que existe un punto de corte en 1991, la mayor parte de los modelos tienen una pendiente negativa pero cercana a cero, lo cual es consistente e intuitivo respecto a los resultados vistos anteriormente.

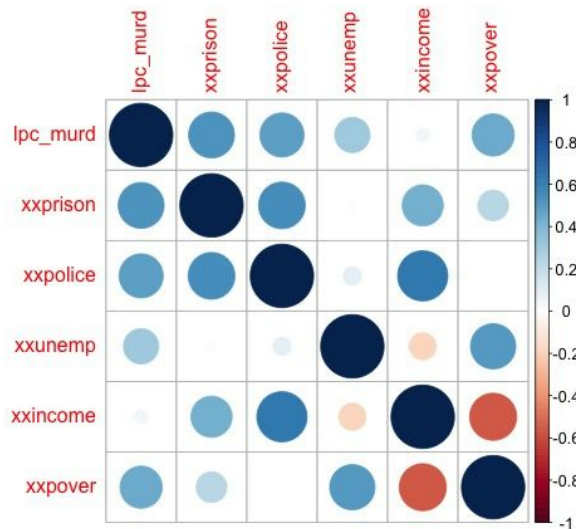
Evolución de Asesinatos



Relación entre la tasa de criminalidad y otras covariables

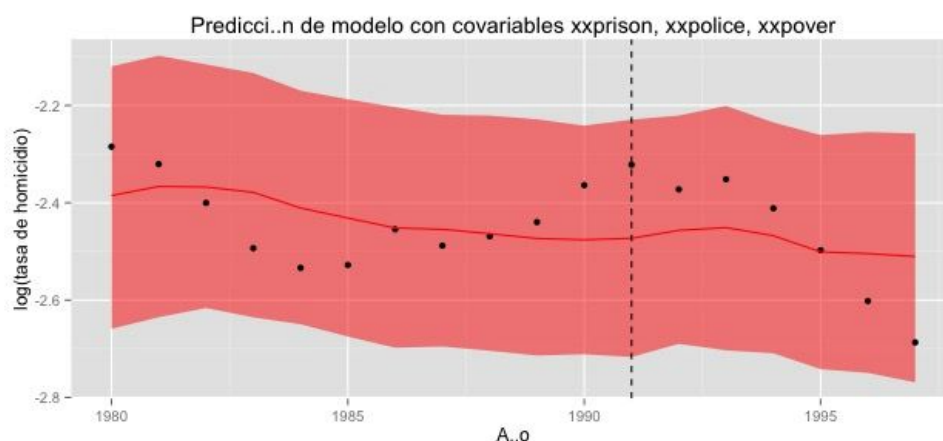
Primero buscaremos un modelo de regresión lineal que relacione la incidencia criminal, a nivel nacional, con algún conjunto de covariables.

Por incidencia criminal, estamos utilizando la variable `lpc_murd` (relativa a homicidios), entonces revisaremos la correlación entre esta y las demás variables:



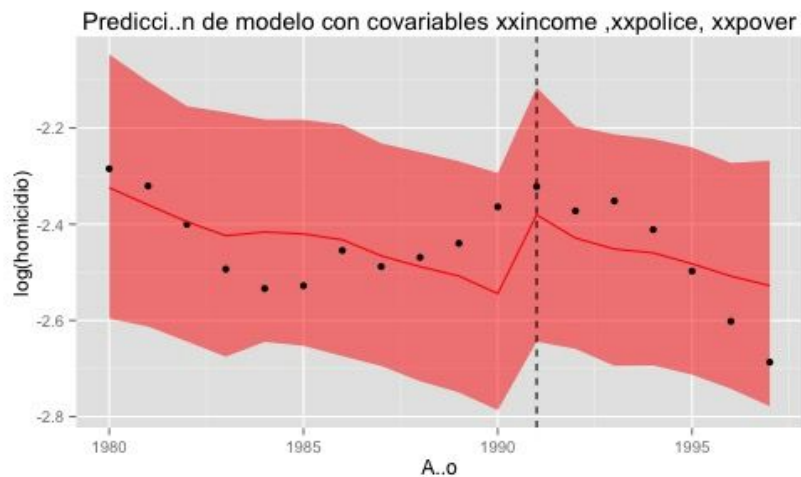
Vemos que las variables `xxprison`, `xpxpolice` y `xpxpover` tienen una alta correlación (de hecho, positiva) con nuestra variable de interés.

La implementación de un modelo lineal con los datos separados se puede visualizar en la siguiente figura:



Vemos que el ajuste propuesto no describe de manera adecuada los datos.

Ahora, lo que haremos será considerar una variable auxiliar que indique los 2 periodos de tiempo, antes y después de 1991. Obtenemos:



Podemos ver que esto da un mejor ajuste, ya que incorpora la información del periodo.

Modelo Jerárquico: ¡Combinemos los Datos!

Los modelos de efectos fijos nos permiten diferenciar entre grupos contenidos en los datos, por medio del primer parámetro. Intuitivamente, en vez de asignarle una sola regresión lineal a los datos, suponemos que están diferenciados por una de sus variables (categórica) y separamos tal regresión lineal en dos: estas regresiones compartirían los parámetros asociados a variables, pero tendrían diferente “intercepto” por grupo.

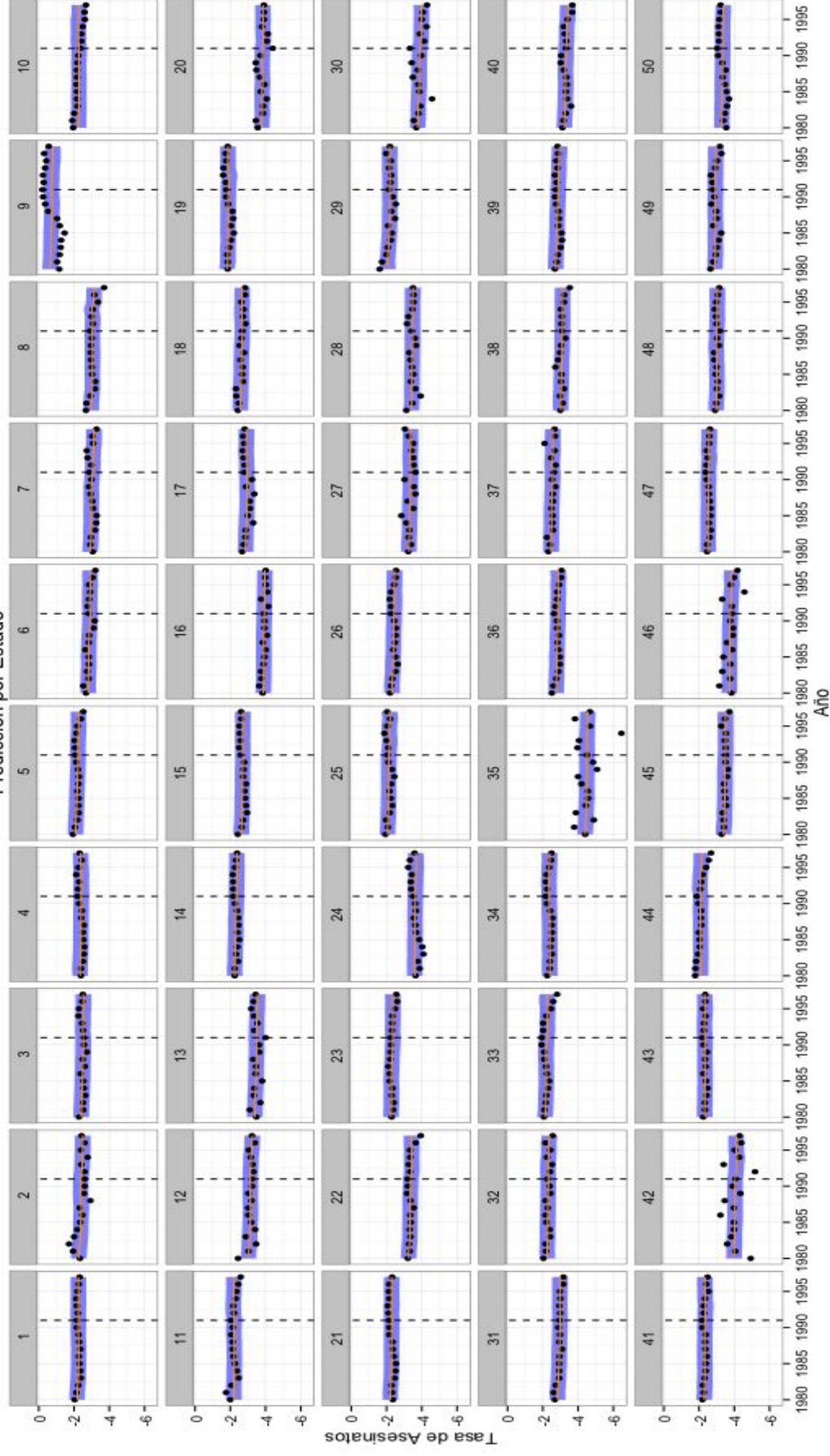
Los resultados obtenidos fueron:

Parámetros Compartidos	
Variable	Coficiente
Prison	-0.1090988
Police	0.23099467
Poverty	0.0060145

Estado	β_0
1	0.79339
2	0.80345
3	0.96309
4	1.08287
5	0.34386
6	0.31709
7	0.26219
8	2.61180
9	0.89945
10	1.05001
11	0.08458
12	-0.21371
13	0.91385
14	0.65728
15	-0.67992
16	0.35220
17	0.60170
18	1.39302
19	-0.59659
20	1.07905
21	-0.15856
22	1.02517
23	-0.29443
24	1.17464

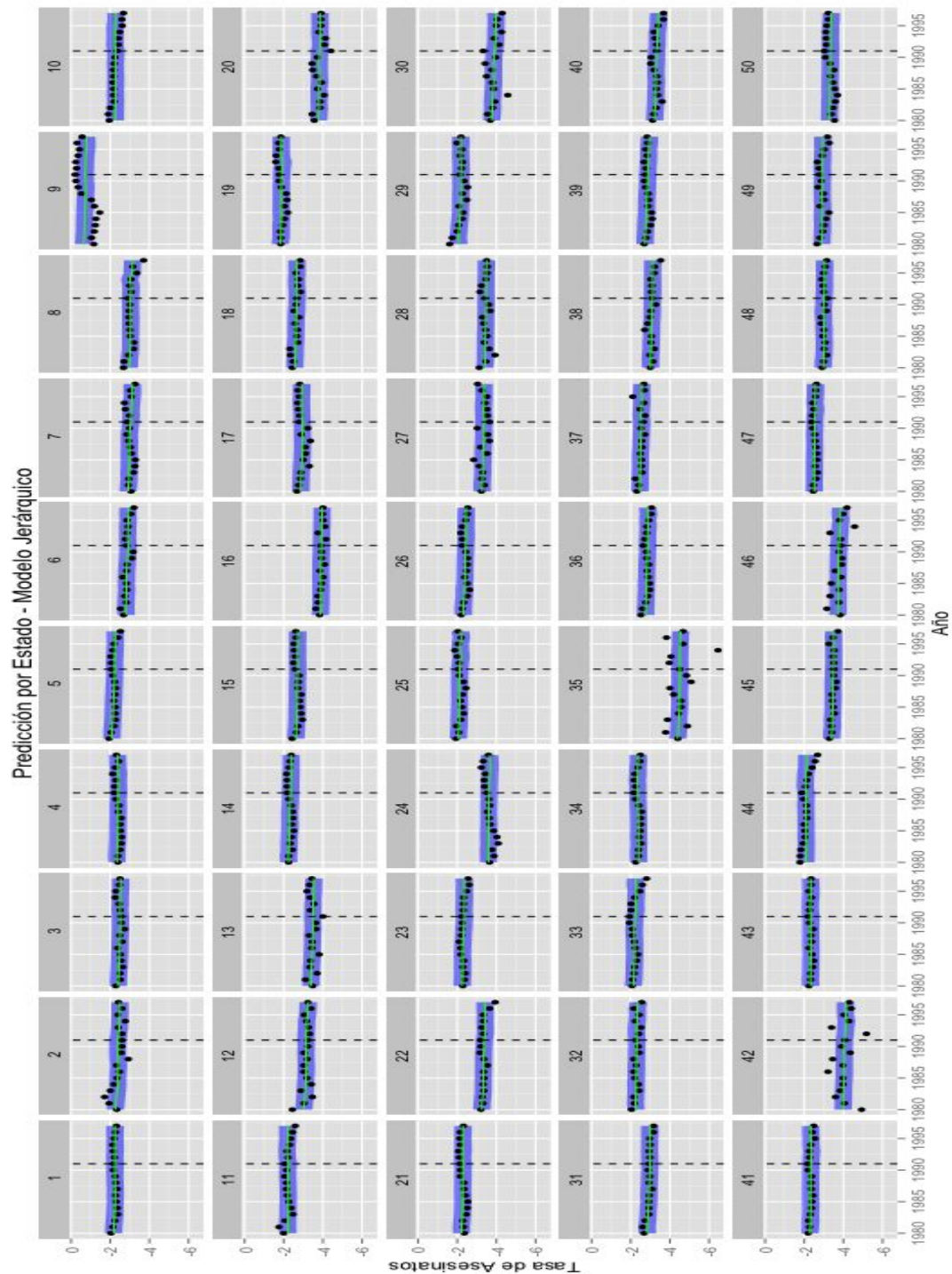
Estado	β_0
25	0.89231
26	-0.09843
27	-0.17012
28	1.07753
29	-0.59596
30	0.23225
31	0.87245
32	0.98776
33	0.93254
34	-1.35344
35	0.48601
36	0.75351
37	0.24503
38	0.48576
39	-0.06374
40	0.98962
41	-0.73466
42	0.94768
43	1.10474
44	-0.21734
45	-0.55632
46	0.79911
47	0.30587
48	0.29371

Predicción por Estado



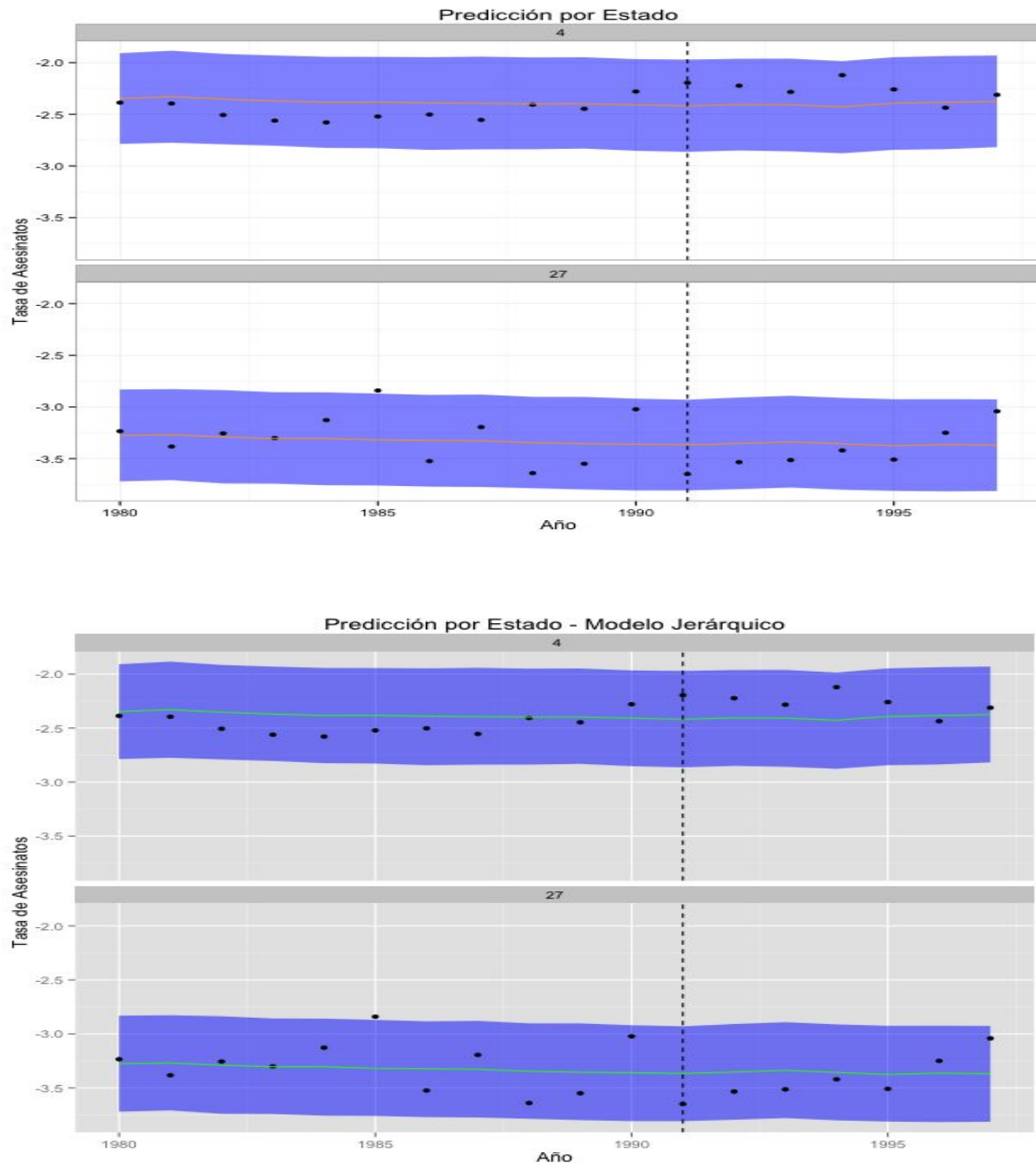
Los modelos jerárquicos nos permiten modelar datos en distintos grupos, pero que están regidos por una relación latente de un nivel más “alto”. Hemos visto ya los detalles teóricos de ellos durante el curso. Intuitivamente, estos modelos tienen la facultad de diferenciar cada grupo con su propio vector de parámetros, pero apoyándose en información de los demás grupos para comprender mejor el comportamiento de ese mismo vector, como parte de un modelo más general.

Los resultados obtenidos fueron:



Comparemos ahora ambos modelos. Hemos tomado a los estados 4 (porque tiene un comportamiento parecido a muchos otros estados) y 27 (porque parece tender a la alta).

Antes de realizar comparaciones numéricas, veamos las gráficas de efectos fijos y jerárquico para ambos estados:



Parecen ser sumamente similares. Esto indicaría que, de hecho, el modelo de efectos fijos es suficiente para describir los datos categorizados. No es inverosímil pensar que la tendencia fuera igual para ambos estados, a pesar de que sus datos se vean diferentes: es posible concluir que, de hecho, existe una tendencia ligerísima a la baja generalizada y esta es la relación latente que existe entre los Estados.

Sin embargo, es muy importante notar la importancia de los coeficientes asignados a cada una de las variables tomadas para explicar la tasa de homicidios. Una rápida inspección a los resultados nos muestra lo siguiente.

Para el modelo con efectos fijos:

	mean	sd	2.5%	25%	50%	75%	97.5%
beta[1]	-2.319241284	0.14166008	-2.5812302	-2.41054572	-2.319557589	-2.22919332	-2.061093607
beta[2]	-0.202552840	0.08779414	-0.3644988	-0.25699796	-0.202354373	-0.14757753	-0.039502423
beta[3]	-0.297885142	0.08008608	-0.4490976	-0.35053622	-0.296880436	-0.24506912	-0.145626557
beta[4]	-0.139883219	0.08087957	-0.2889830	-0.18994551	-0.139825095	-0.08831956	0.007142061
beta[5]	-0.005689248	0.08015583	-0.1547629	-0.05703163	-0.005353824	0.04631976	0.143799013
beta[6]	-0.731326998	0.08844080	-0.8989090	-0.78715712	-0.730900415	-0.67493560	-0.565515380

Para el modelo jerárquico, los primeros parámetros:

	mean	sd	2.5%	25%	50%	75%	97.5%
a.t	1.25380035	0.282034878	0.81991510	1.06092058	1.215275232	1.40179126	1.926326841
b.t	0.02704064	0.006992125	0.01500238	0.02221456	0.026472351	0.03120001	0.042227115
beta[1,1]	12.16337574	29.425766055	-3.78201623	-2.87070577	-2.336184501	-1.52928132	71.021653718
beta[2,1]	0.19685142	0.402198416	-0.05161393	-0.01570703	0.003704336	0.03558322	1.001252969
beta[1,2]	-1.97678402	0.049667854	-2.07624472	-2.00848931	-1.976380617	-1.94420745	-1.879945654
beta[2,2]	-0.01314613	0.003220492	-0.01950123	-0.01525384	-0.013187546	-0.01106565	-0.006791082

Conclusiones

Hemos observado el comportamiento de la tasa de interés de Homicidio, a través de los años, para todos los estados de Estados Unidos. Se observó un gran cambio en la tendencia al suponer que en 1991 existe un punto de corte: este punto es teorizado, relacionado con el momento en el cual llegaría a la edad pico de crimen la primera generación de bebés cuyas madres tuvieron derecho legal al aborto.

Esta tendencia, generalizada para todos los estados, muestra que el decrecimiento ha sido más rápido a partir de 1991. Esto confirmaría la teoría de Donohue y Levitt. Sin embargo, encontramos que las variables altamente correlacionadas sí explican de manera adecuada el declive de la tasa de interés.

Bibliografía

J. Donohue and S. Levitt, *The Impact of Legalized Abortion on Crime*, The Quarterly Journal of Economics Vol. CXVI, May 2001

J. Martínez-Ovando, *Modelos de Regresión - Enfoque Bayesiano*, publicado en la página del autor, consultado en <http://jcmartinezovando.github.io/cursos/EST46111/> 16/12/2015