



MAESTRÍA EN ECONOMÍA

**Economía Aplicada**

PROF. MARTIN A. ROSSI

TUTORES: PAOLA LLAMAS Y TOMÁS  
PACHECO

## **Problem Set 7: Cluster Robust Inference**

Garcia Ojeda, Juan  
Hausvirth, Martina  
Hayduk, Gaspar  
Salvatierra, Elias Lucas D.

Fecha de entrega: 1 de Noviembre de 2024

## PROBLEM SET 7: CLUSTER ROBUST INFERENCE

GARCÍA OJEDA - HAUSVIRTH - HAYDUK - SALVATIERRA

Este breve problem set está basado en “The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial” de Angrist y Levy (American Economic Review, 2009). El objetivo no es replicar los resultados del paper, si no simplemente comparar los resultados que se obtienen al emplear distintas estrategias de inferencia. Deberán estimar regresiones de la siguiente forma:

$$Bagrut_{i,g} = \beta Treated_{t,g} + \delta X + u_{i,g}$$

donde  $Bagrut_{i,g}$  es una variable binaria que toma el valor 1 si el individuo  $i$ , perteneciente a cluster  $g$ , cuenta con un high school matriculation certificate (Bagrut).  $Treated_{i,g}$  es la variable de interés y  $X$  es un vector de controles que incluye las variables *semarab* y *semrel*, además de efectos fijos por grupo (en caso de ser posible); también decidimos incluir como control al sexo del individuo.

### 1. RESOLUCIÓN DE EJERCICIO.

Para este ejercicio lo que haremos es tomando como referencia el paper de Angrist y Levy, comparar las cuatro estrategias posibles para hacer inferencia.

El estudio analiza cómo los incentivos financieros influyen en la obtención del certificado de Bagrut entre estudiantes de secundaria en escuelas con las tasas de certificación más bajas. Este certificado es condición necesaria para acceder a la universidad, y por lo tanto representa una barrera significativa en la vida profesional de los individuos en Israel, lo que hace que su obtención sea especialmente relevante.

Con la base de datos disponible, consideramos las estrategias Robust Standard Errors, Cluster Robust SE, Wild Bootstrap, y ARTs para hacer inferencia y determinaremos, dado este caso particular, cuál es la mas adecuada.

En primer lugar, el uso de la estrategia de errores estándar robustos (Robust Standard Errors) supone que las observaciones entre individuos son independientes, y se emplea comúnmente cuando se sospecha que los errores presentan heterocedasticidad. En el paper de Angrist y Levy, dado que en las observaciones se presentan grupos de individuos que asisten a distintas escuela, es altamente probable que aquellos que compartan escuela también compartan características inobservables como la inteligencia y motivación de los estudiantes o profesores. Así, los errores pueden no ser independientes y presentar correlación. Que los errores no sean independientes significa que todo lo que

afecta al outcome que no está en mi modelo (inteligencia, motivación) para una observación está correlacionado con todo lo que afecta al outcome que no está en mi modelo (inteligencia, motivación) para otra observación. Cuando los errores están correlacionados dentro de un grupo, como en este caso, esta estrategia no captura la estructura de dependencia entre observaciones, lo que provoca una mala estimación de la varianza del estimador y lleva a errores estándar incorrectos, afectando la precisión de las pruebas de significancia. En función de esta consideración, concluimos que esta estrategia no es óptima para este caso.

En una segunda instancia, si asumimos como mencionamos anteriormente que los errores dentro de cada escuela están correlacionados, podríamos optar por dividir las observaciones en grupos o muestras (clusters). De esta forma, utilizando Cluster Robust Standard Errors, permitimos que las observaciones dentro de cada cluster estén correlacionadas internamente (intra-cluster correlation), pero no entre clusters. En un caso como el analizado en el paper, *a priori* sería razonable implementar una estrategia como tal. Sin embargo, existen dos problemas subyacentes a esta estrategia: por un lado, el desempeño cuando la cantidad de clusters ( $n$ ) es chico no es bueno. En la literatura se toma como referencia un mínimo de 50 clusters para que esta estrategia funcione, y los datos obtenemos 11 grupos emparejados (clusters). Por otro lado, el desempeño empeora cuanto mayor sea la heterogeneidad entre clusters. En el paper se observan 40 escuelas, de las cuales 10 son musulmanas, 10 judías, y 20 laicas. Con dichas divergencias religiosas, podríamos presuponer que hay heterogeneidad entre clusters, lo cual nos lleva a sospechar de esta estrategia.

En tercer lugar, probamos la estrategia de ARTs (Analysis of Resources of Trials). Esta última si bien no requiere de homogeneidad entre los clusters y funciona bien con pocos clusters, presupone que no todas las observaciones pueden ser tratadas dentro de un mismo cluster, es decir, dentro de una misma escuela deberían haber tanto grupos de tratamiento como de control. Esto, además de representar un problema moral por la aleatoriedad del experimento, y ser difícil de aplicar en la práctica, podría llegar a comprometer el experimento. En el estudio de Angrist y Levy, todos los alumnos en la escuela seleccionada para tratamiento eran elegidos para el pago. Esto ocurrió así dado que es una medida más popular socialmente, ya que reduce el efecto de "injusticia" asociado a la randomización, y además, de aplicar tratamiento solamente a algunos alumnos de la escuela, se puede presuponer que no se cumple SUTVA, ya que probablemente los alumnos que no reciben el tratamiento dentro de esa escuela, sean afectados por aquellos que sí lo reciben. La estrategia de ARTs tiene los siguientes pasos. En primer lugar, se estima la regresión  $Bagrut_{i,g} = \beta Treated_{i,g} + \delta X + u_{i,g}$  para cada uno de los 11 clusters por separado, obteniendo un efecto de la intervención para cada cluster. En segundo lugar, se calcula el estadístico t para cada cluster. En tercer lugar, se calcula un estadístico T que es el promedio de todos los estadísticos del paso anterior. En cuarto lugar, se computan todos los posibles cambios de signo de los coeficientes estimados del paso 1, vamos a tener  $2^{11}=2048$  vectores posibles; para cada uno de estos vectores repetimos los pasos 2 y 3. De este paso nos quedamos 2048 estadísticos  $T^*$  y se usan a estos 2048 estadísticos como la distribución del estadístico T. Finalmente, se rechaza la hipótesis nula al 5% si T es mayor al 95% de los  $T^*$ .

Finalmente, consideramos la estrategia Wild-Bootstrap. Esta estrategia solucionaría el problema del número pequeño de clusters, que para la estrategia RSE significaba un problema; y además no requiere que las observaciones dentro de un mismo cluster sean diferentes entre sí. No obstante, el potencial problema que tiene esta estrategia es que, tal y como sucedía con Robust Standard Errors, requiere que haya poca heterogeneidad en la muestra y que el número de clusters vaya a infinito, por lo que teniendo únicamente 11 clusters no parece la mejor estrategia en este ejemplo. El procedimiento de Wild-Bootstrap es el siguiente. En primer lugar, una vez hecha la estimación, se generan  $B$  pseudo-muestras. En segundo lugar, para cada pseudo-muestra se vuelve a regresar el modelo y se obtiene un estadístico  $T_b^*$  para cada pseudo-muestra, obteniendo un total de  $B$  estadísticos  $T_b^*$ . Finalmente, usamos estos  $B$   $T_b^*$  para inferir la distribución del estadístico original.

En nuestra consideración, la mejor estrategia para estimar el efecto del tratamiento es ARTs pues esta estrategia funciona bien cuando el número de clusters es chico, que es lo que ocurre en esta estimación donde hay 11 clusters. La desventaja es que se requiere que el número de observaciones dentro de cada cluster tienda a infinito. Wild-Bootstrap no nos parece la mejor estrategia dado que la validez teórica de este método requiere que el número de clusters tienda a infinito y requiere que haya poca heterogeneidad entre clusters.

La Tabla 1 contiene los resultados de las estimaciones para cada una de estas cuatro estrategias. En primer lugar, podemos observar que el coeficiente estimado por las estrategias Robust Standard Errors, Cluster Robust SE y Wild-Bootstrap son idénticos, dado que se estima la misma regresión y estas estrategias difieren en la corrección de los errores estándar y en el cómputo del estadístico y el p-valor; la estrategia ARTs estima un coeficiente diferente ya que se estima una regresión para cada cluster. En segundo lugar, solo la estrategia Robust Standard Errors (que no utiliza clusters) obtiene un efecto del tratamiento estadísticamente significativo, mientras que el efecto del tratamiento estimado para las estrategias que utilizan clusters es no significativo. Esto ocurre pues las estrategias que utilizan clusters internalizan el hecho de que las observaciones dentro de un mismo cluster están correlacionadas, cambiando la forma en la que se estima la varianza del estimador y la distribución del estadístico, generando p-valores más altos y menores chances de obtener coeficientes significativos.

Tabla 1: Comparación de las cuatro estrategias

	(1)	(2)	(3)	(4)
	Robust Standard Errors	Cluster Robust SE	Wild-Bootstrap	ARTs
Treated	0.0782*** (0.0139)	0.0782 (0.0486)	0.0782 (0.0486)	0.0666 (0.0139)
P-Valor	0.000	0.139	0.161	0.475
Observaciones	3821	3821	3821	3821

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

```

1  /*****
   *****/
2          Problem Set 7: CLUSTER ROBUST INFERENCE
3          Universidad de San Andrés
4          Economía Aplicada
5  *****/
6  * Gaspar Hayduk; Juan Gabriel García Ojeda; Elías Lucas
   Salvatierra; Martina Hausvirth
7
8  /*****
   *****/
9
10 * 0) Set up environment
11 *=====
   =====*
12 global main "/Users/gasparhayduk/Desktop/Economía Aplicada/PS7"
13 global input "$main/input"
14 global output "$main/output"
15
16
17 * Abrimos la base de datos:
18 use "$input/base01.dta", clear
19
20
21 * La ecuacion a estimar es:
22 *      bagrut_i,g = B*Treated_i,g + alpha*X_i,g + u_i,g
23
24 *Lo primero que hacemos para asegurarnos el balance entre los
   grupos tratamiento y control, es utilizar esta técnica de
   emparejamiento planteada en el paper.
25 * En este caso, emparejar significa que cada escuela en el grupo
   de tratamiento fue emparejada con una escuela en el grupo de
   control que tenía una tasa de Bagrut similar en el pasado; y una
   vez hecho este emparejamiento, se asigna de manera aleatoria a
   uno de los dos. Esto ayuda a aislar el efecto del tratamiento al
   reducir el sesgo debido a las diferencias iniciales entre las
   escuelas.
26
27 gen group = 1
28 replace group = 2 if pair == 2 | pair == 4
29 replace group = 3 if pair == 5 | pair == 8
30 replace group = 4 if pair == 7
31 replace group = 5 if pair == 9 | pair == 10
32 replace group = 6 if pair == 11
33 replace group = 7 if pair == 12 | pair == 13
34 replace group = 8 if pair == 14 | pair == 15
35 replace group = 9 if pair == 16 | pair == 17
36 replace group = 10 if pair == 18 | pair == 20
37 replace group = 11 if pair == 19
38
39 *ahora utilizaremos una por una las cuatro estrategias

```

```

39  *ahora utilizaremos una por una las cuatro estrategias
    propuestas para hacer inferencia.
40
41
42  eststo clear
43
44
45
46  *----- ESTRATEGIA 1: ROBUST
    STANDARD ERRORS -----
47  reg zakaibag treated semarab semrel boy i.group, robust
48  eststo: test treated
49  estadd scalar p_value=r(p)
50
51
52
53
54  *----- ESTRATEGIA 2: CLUSTER
    ROBUST SE -----
55  reg zakaibag treated semarab semrel boy i.group, cluster(group)
56  eststo: test treated
57  estadd scalar p_value=r(p)
58
59
60  *----- ESTRATEGIA 3:
    WILD-BOOTSTRAP -----
61  reg zakaibag treated semarab semrel boy i.group, cluster(group)
62  eststo: boottest {treated} , boottype(wild) cluster(group) robust
    seed(123)
63  mat p_wb = r(p)
64  estadd scalar p_value=r(p)
65
66
67  *----- ESTRATEGIA 4:
    ARTs -----
68  do "$main/programs/art.ado"
69
70  *Corremos:
71  eststo: art zakaibag treated semarab semrel boy, cluster(group)
    m(regress) report(treated) //aca no podemos meter FE por cluster.
72  estadd scalar p_value=r(pvalue_joint)
73
74
75
76  *----- Junto todo en una tabla. Reporto solo el coeficiente
    asociado a treated, el SE y el p-valor. Una columna por
    estrategia:
77  esttab using "$output/table_results.tex", p se replace label ///
78  keep(treated, relax) ///
79  cells(b(fmt(4) star) se(fmt(4) par)) ///
80  stats(p_value N, fmt(3 0) labels("P-Valor" "Observaciones"))
81

```

82  
83  
84  
85  
86  
87