Diferencias en Diferencias

Datos Panel

Lo primero que debemos hacer es abrir la base de datos del modelo de diferencias:

Esta base de datos cuenta con información para 4.000 individuos. Para cada individuo, tenemos información de dos periodos. El subíndice de cada variable corresponde al periodo de observación. Por ejemplo "ha\_nchs2" indica el ingreso mensual, en decenas de miles de pesos del jefe del hogar en el periodo 2.

. \*-------------------------------------------\*

. \*1. Veamos algunas estadísticas descriptivas\*

. \*-------------------------------------------\*

Talla para la edad de los individuos del grupo de tratamiento en el primer periodo

. sum ha\_nchs1 if D==1

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

ha\_nchs1 | 2000 -.6786743 1.265438 -5.919884 5.874913

Talla para la edad de los individuos del grupo de control en el primer periodo

sum ha\_nchs1 if D==0

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

ha\_nchs1 | 2000 -.7790041 1.264368 -5.585266 6.061383

.

Talla para la edad de los individuos del grupo de tratamiento en el segundo periodo

. sum ha\_nchs2 if D==1

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

ha\_nchs2 | 2000 -.2787444 1.293322 -5.45469 5.736451

.

Talla para la edad de los individuos del grupo de control en el segundo periodo

. sum ha\_nchs2 if D==0

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

ha\_nchs2 | 2000 -.5812385 1.29934 -5.159187 6.333417

Aparentemente, existe una diferencia entre los individuos de tratamiento y de control en el primer periodo en lo referente a la variable talla para la edad "ha\_nchs. Además, esta diferencia se incrementa en el tiempo.

Haciendo una prueba de diferencia de medias, podemos ver que en los dos periodos hay una diferencia estadísticamente significativa que es mayor en el segundo periodo:

. ttest ha\_nchs1, by(D)

Two-sample t test with equal variances

------------------------------------------------------------------------------

Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]

---------+--------------------------------------------------------------------

0 | 2000 -.7790041 .0282721 1.264368 -.83445 -.7235582

1 | 2000 -.6786743 .0282961 1.265438 -.7341672 -.6231815

---------+--------------------------------------------------------------------

combined | 4000 -.7288392 .0200131 1.26574 -.768076 -.6896024

---------+--------------------------------------------------------------------

diff | -.1003298 .0399998 -.1787516 -.0219079

------------------------------------------------------------------------------

diff = mean(0) - mean(1) t = -2.5083

Ho: diff = 0 degrees of freedom = 3998

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.0061 Pr(|T| > |t|) = 0.0122 Pr(T > t) = 0.9939

.

. ttest ha\_nchs2, by(D)

Two-sample t test with equal variances

------------------------------------------------------------------------------

Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]

---------+--------------------------------------------------------------------

0 | 2000 -.5812385 .0290541 1.29934 -.638218 -.5242589

1 | 2000 -.2787444 .0289196 1.293322 -.33546 -.2220287

---------+--------------------------------------------------------------------

combined | 4000 -.4299914 .0206334 1.30497 -.4704443 -.3895385

---------+--------------------------------------------------------------------

diff | -.3024941 .0409937 -.3828646 -.2221236

------------------------------------------------------------------------------

diff = mean(0) - mean(1) t = -7.3790

Ho: diff = 0 degrees of freedom = 3998

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

.

Mientras que en el primer periodo la diferencia es de 0.1, para el segundo se incrementa hasta 0.3.

. \*--------------------------------------------------------\*

. \*2. Modelo de diferencias en diferencias con datos panel \*

. \*--------------------------------------------------------\*

Al ver las estadísticas descriptivas de la variable talla para la edad y desnutrición crónica vemos que en realidad hay una diferencia preexistente entre los individuos pertenecientes al grupo de tratamiento y al grupo de control. En el periodo después de aplicar el tratamiento es posible que la diferencia entre los dos grupos se deba a la diferencia preexistente o a la aplicación del tratamiento. El modelo de diferencias en diferencias nos permite controlar por diferencias existentes antes de la aplicación del programa.

Para correr el modelo básico de diferencias en diferencias, utilizando una base de datos panel, utilizamos como variable dependiente el cambio en la talla para la edad en función del tratamiento.

Primero debemos generar la variable "delta\_ha\_nchs" y después si podemos correr el programa:

.

. gen delta\_ha\_nchs=ha\_nchs2-ha\_nchs1

.

.

. reg delta\_ha\_nchs D

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 1, 3998) = 450.77

Model | 40.8704296 1 40.8704296 Prob > F = 0.0000

Residual | 362.492281 3998 .090668404 R-squared = 0.1013

-------------+------------------------------ Adj R-squared = 0.1011

Total | 403.362711 3999 .100865894 Root MSE = .30111

------------------------------------------------------------------------------

delta\_ha\_n~s | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2021644 .009522 21.23 0.000 .1834959 .2208328

\_cons | .1977656 .0067331 29.37 0.000 .184565 .2109662

El coeficiente asociado a la variable "D" tiene una magnitud de 0.2 y es significativo. Esto nos indica que la diferencia de talla para la edad entre los individuos de tratamiento y control se incrementa en 0.2 por la aplicación del tratamiento.

.

. \*3. Modelo de diferencias en diferencias con datos panel utilizando regresores adicionales \*

.

.

Podemos suponer que el crecimiento en la variable talla para la edad, además de deberse al tratamiento y al crecimiento natural en el tiempo, puede tener su origen en características de los individuos. Por ejemplo, resulta razonable suponer que el crecimiento en la talla para la edad de un individuo también está asociado al ingreso en el primer periodo del jefe del hogar. En este punto verificamos esta hipótesis con los ingresos del jefe de hogar y la educación del jefe de hogar.

.

.

.

. reg delta\_ha\_nchs D ingresos\_hogar\_jefe1 educa\_jefe1 ingresos\_hogar\_jefe2 educa\_jefe2

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 5, 3994) = 1335.34

Model | 252.385721 5 50.4771442 Prob > F = 0.0000

Residual | 150.97699 3994 .037800949 R-squared = 0.6257

-------------+------------------------------ Adj R-squared = 0.6252

Total | 403.362711 3999 .100865894 Root MSE = .19442

------------------------------------------------------------------------------

delta\_ha\_n~s | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .0771181 .006429 12.00 0.000 .0645136 .0897225

ingresos\_h~1 | .078656 .0010516 74.79 0.000 .0765942 .0807178

educa\_jefe1 | .0136727 .018503 0.74 0.460 -.0226035 .0499489

ingresos\_h~2 | -.0786525 .0010527 -74.72 0.000 -.0807163 -.0765887

educa\_jefe2 | -.0145574 .01849 -0.79 0.431 -.0508081 .0216933

\_cons | .0871868 .0071875 12.13 0.000 .0730953 .1012783

------------------------------------------------------------------------------

.

Vemos que en realidad la única variable significativa en esta regresión es el coeficiente asociado a "D". No podemos decir entonces que un incremento en los ingresos del jefe de hogar ni en su nivel educativo tienen un efecto sobre el crecimiento de la talla para la edad (ha\_nchs).