Introducción a métodos econométricos en Stata

9 de septiembre de 2021

INSTITUTO TECNOLOGICO AUTONOMO DE MEXICO (ITAM)
Seminario de Investigación Económica
Instructor: Horacio Larreguy
Asistente: Manuel Quintero

1. Archivos en Stata

Hay distintas clases de archivos con los que Stata trabaja:

- .dta - archivo de datos.
- .do - archivo de comandos.
- .ado - programas o ficheros con Macros.
- .gph - archivos gráficos.
- .dct - archivos de diccionarios.
- .log - archivos de salida de Stata.

Stata mantiene los datos en la memoria por lo que para trabajar sobre una nueva base es necesario remover la base que esta cargada.

Algunos **comandos útiles** en Stata:

```
clear // Comando para limpiar la memoria
cd // Directorio en el que se esta trabajando
cd "Direccion" // Cambiar el directorio
mkdir "Direccion" // Crear un directorio
dir or ls // Muestra archivos existentes en el directorio
save // Guarda el archivo
use // Carga un archivo de datos en formato .dta
memory // Reporta la memoria que está siendo utilizada
compress // Comprime los datos, uso eficiente
list // Comando
```

1.1. Lectura de datos

Importación y exportación de archivos .dta

```
use "arhivo.dta" \\ Cargar archivo
save  archivo", replace \\ Guardar archivo y remplazarlo si ya existe
```

Importación y exportación de archivos .txt, y .csv

```
import delimited archivo.csv (archivo.txt) // Importar

outsheet using smauto1.csv , comma  // Exportar separado por coma todas
   las variables
outsheet var1 var2 using smauto1.csv , comma // Exportar separado por
   comas variables selectas
```

Importación y exportación de archivos .xlsx

```
import excel acentos.xlsx // Importar

export excel using "Nombre" // Exportar
```

Labels

Stata tiene la función de almacenar nombres o descripciones a las variables y base de datos que pueden ser utilizadas para un gran variedad de resultados, por ejemplo, para presentar los resultados en tablas con *labels* auto-explicativos.

```
label data \\ Etiquete para la base de datos que se está trabajando
label variable \\ Etiqueta para variable
label drop \\ Elimina las etiquetas
label dir \\ Lista todas las variables que tienen una etiqueta
```

1.2. Manejo de datos

una lista de comandos básicos de Stata la puedes encontrar aquí.

```
clear
set seed 32 // Semilla aleatoria
set obs 5 // Número de observaciones
gen ind = _n // Generamos variable ind de 1 a n = 5
expand 5 // Hacemos de cada ind 5 observaciones
gen x = rnormal() // Generamos x
gen y = rnormal() // Generamos x
gen z = 0 if x < 0.5 // Generamos z con condicional
replace z = 1 if z == . // Remplazamos valores NA de z con 1
keep if y > 0 // Filtramos aquellos renglones donde y > 0
collapse (sum) x y, by(ind) // Colapsamos los datos sumando las
   variables x e y, agrupando por ind
```

2. Regresión lineal

En Stata corremos una regresión lineal escribiendo el comando *regress/reg y x* . Para ver información más detallada de la sintaxis de regresión lineal véase aquí.

```
regress y1 x1 // Regresión lineal simple

regress y2 x1 x2 // Regresión lineal múltiple

regress y2 x1 x2, noconstant // Regresión lineal múltiple sin constante

regress y2 x1 x2 i.var // Regresión lineal múltiple con efectos fijos
   por la variable var

regress y x1 x2 i.var, vce(cluster var2)// Regresión lineal múltiple
   con efectos fijos por la variable var y errores estándar clustered
   por var2

regress y x1 x2 i.a [pweight = wvar] // Usando pesos de muestreo por
   wvar
```

Sin embargo, es mejor utilizar el comando **reghdfe** para cualquiera de las regresiones anteriores. Podemos encontrar información de este paquete en la página de Correia (2016), aquí.

```
reghdfe y1 x1, noabsorb // Regresión lineal simple

reghdfe y2 x1 x2, noabsorb // Regresión lineal múltiple

reghdfe y2 x1 x2, absorb(var) // Regresión lineal múltiple con efectos
   fijos por la variable var

reghdfe y x1 x2, absorb(var) vce(cluster var2)// Regresión lineal
   múltiple con efectos fijos por la variable var y errores estándar
   clustered por var2

reghdfe y x1 x2  [pweight = wvar] , absorb(var)// Usando pesos de
   muestreo por wvar
```

3. De Stata a LaTeX

Cuando queremos reportar los resultados de una manera elegante y formal tenemos que exportar los resultados estadísticos a un formato más presentable. Una manera de mostrar tus resultados a un público es por medio de LaTeX. Para exportar nuestros resultados de Stata a LaTeX, por lo general, utilizamos el paquete **estout**.

Una guía muy completa sobre el uso de este paquete la pueden encontrar en Naqvi (2021). La documentación del paquete y todos los atributos que se pueden utilizar para personalizar las tablas los pueden encontrar aquí y diversos ejemplos aquí.

Un primer ejemplo, exportamos tablas de estadísticas descriptivas.

```
* Simulamos datos
clear
set seed 32 // Semilla aleatoria
set obs 100
generate x1 = rnormal()
generate x2 = rnormal(2)
generate y1  = 2*x1 + 1 + rnormal()
generate y2 = 2*x1 + 3*x2 + 1 + rnormal()

* Agregamos etiquetas a cada variable
label variable x1   "Var. Independiente 1"
label variable x2   "Var. Independiente 2"
label variable y1   "Var. Dependiente 1"
label variable y2   "Var. Dependiente 2"
```

```
* Guardamos tabla de estadísticas descriptivas para x1,x2,y1,y2
estpost tabstat x1 x2 y1 y2, c(stat) stat(mean sd min max n)

* Guardamos el codigo TeX
esttab using "./Tables/table1.tex", replace ///
 cells("mean(fmt(%13.3fc)) sd(fmt(%13.3fc)) min max count") ///
 title("Tabla de estadísticas descriptivas") label ///
 collabels("Sum" "Mean" "SD" "Min" "Max" "N") noobs nonumbers
```

Tabla 1: Tabla de estadísticas descriptivas

	Sum	Mean	SD	Min	Max
Var. Independiente 1	-0.000	0.924	-2.768	1.925	100.000
Var. Independiente 2	1.997	1.021	-0.576	3.887	100.000
Var. Dependiente 1	0.962	1.981	-3.931	5.985	100.000
Var. Dependiente 2	6.905	3.688	-2.631	14.072	100.000

Alguno de los atributos que hemos utilizado son:

- title: agregamos el título a la tabla.
- label: cambia el nombre de las variables por las etiquetas.
- noobs: no observaciones.
- nonumbers: no enumeración de columnas

Nota: Como podemos observar para imprimir una tabla con el formato que hemos especificado utilizamos el comando **esttab** y para guardar los resultados de una tabla hicimos uso de **estpost:**. De ahora en adelante para guardar los resultados de las regresiones utilizaremos **eststo:** antes del comando *regress* o similar.

Múltiples regresiones en una sola tabla

```
regress y1 x1  // corremos
eststo: reghdfe y1 x1, noabsorb // corremos y guardamos regresión
qui testparm* // prueba de hipótesis
estadd scalar p_value = round(r(p),3) // guardamos el valor p de la
   prueba F

eststo: reghdfe y2 x1 x2 noabsorb   // corremos y guardamos regresión
qui testparm*
estadd scalar p_value = round(r(p),3)

* Creamos tabla
esttab using "./Tables/table2.tex", replace  ///
title(Regresión Lineal) ///
addnote("This table includes ...") ///
se star(* 0.10 ** 0.05 *** 0.01) label ///
stats(N r2 p_value, labels("Observations" "R-squared" "F-stat
   (p-value)"))

eststo clear  // borramos todas las estimaciones
```

Tabla 2: Regresión Lineal

	(1)	(2)
	y1	y2
Var. Independiente 1	1.936***	1.958***
	(0.121)	(0.0853)
x2		2.806***
		(0.0846)
Constant	0.917***	1.408***
	(0.117)	(0.194)
Observations	100	100
R-squared	0.724	0.947
F-stat (p-value)	0	0

Standard errors in parentheses

This table includes ...

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4. Diferencias en diferencias

Buscamos estimar el siguiente modelo:

$$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 \text{post}_t + \delta T_i \cdot \text{post}_t + \epsilon_{it},$$

donde, Y_{it} es la variable endógena, T_i y post_i son indicadoras de tratamiento y periodo, respectivamente. El parámetro δ es el verdadero efecto en el tratamiento o la diferencia en diferencia y ϵ_i es el error aleatorio.

4.1. Ejemplo 2 periodos

Para correr una regresión de diferencias en diferencias podemos utilizar el comando *regress* o el comando *diff* y obtener los mismos resultados.

```
clear
* Cargar los datos en la base dd_data a la memoria
use "dd_data.dta"
label variable post   "Post"
label variable wage   "Salario"
label variable Treatment   "Tratamiento"

* Calculamos la regresión DID
eststo: reghdfe wage post##Treatment, noabsorb

eststo: diff wage, t(Treatment) p(post)

esttab using "./Tables/table3.tex", replace noomitted nobaselevels
   label se r2 title("Regresión DID en 2 periodos") ///
mtitles("Modelo reg" "Modelo diff") ///
star(* 0.10 ** 0.05 *** 0.01) ///
nonumbers

eststo clear
```

Tabla 3: Regresión DID en 2 periodos

	Modelo reg	Modelo diff
Post=1	10***	
	(1.432)	
Post		10.00***
		(1.432)
Tratamiento=1	-1.400	
	(1.432)	
Post=1 × Tratamiento=1	-9.400***	
	(2.025)	
Tratamiento		-1.400
		(1.432)
Diff-in-diff		-9.400***
		(2.025)
Constant	16.60***	16.60***
	(1.012)	(1.012)
Observations	20	20
R^2	0.842	0.842

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2. Two-way fixed-effect regression

De acuerdo con Imai y Kim (2020), la regresión lineal bidireccional de efectos fijos (2FE) se ha convertido en un método predeterminado para estimar efectos causales a partir de datos panel. Muchos investigadores utilizan el modelo 2FE para ajustar doble efectos fijos: por unidad y por tiempo, que interactúan en un lapso determinado. La forma estándar del modelo es:

$$Y_{it} = \gamma_i + \lambda_t + \tau T_{it} + \epsilon_{it},$$

donde, γ_i es el efecto fijo por individuo, λ_t es el efecto fijo por periodo, T_{it} es una variable dicótoma igual a 1 si el individuo i es tratado para el tiempo $t \geq \sigma$, ϵ_{it} es el error aleatorio y τ es el parámetro de interés.

Lags and leads

Una estrategia para probar el supuesto de tendencias paralelas es correr una regresión de *Lags and leads*:

$$Y_{it} = \gamma_i + \lambda_t + \sum_{t' \in \mathbf{F}} \delta_{t'} T_{it+t'} + \epsilon_{it}, \qquad \text{donde } \mathbf{F} = \{-f, ..., f\},$$

Donde γ_i y λ_t son los efectos fijos por individuo y periodo, respectivamente. $T_{it+t'}$ $\forall t^{'} \in \mathbf{F}$ son las variables rezagadas y adelantadas, incluyendo el periodo 0, donde inicia el tratamiento, y verificar que las variables adelantadas son estadísticamente igual a 0.

Una aplicación ilustrativa de Lags and leads se puede encontrar en Autor (2003) y una explicación más detallada del método se puede encontrar en las notas de clase de Pischke (2005).

Did plot

Una segunda alternativa para detectar tendencias antes del tratamiento y efectos anticipados es utilizando *Did plot*:

$$Y_{it} = \gamma_i + \lambda_t + \sum_{t'=-T}^{0} \delta_{t'} T_i + \sum_{t'=1}^{T} \delta_{-t'} T_i + \epsilon_{it}.$$

Ejemplo 1: *No differential pre-trends*

Para este ejemplo simulamos datos panel para 20 individuos a lo largo de 10 años ($2010 - 2019$). El tratamiento se dio en cualquiera de los años $2013 - 2017$.

```
* Ejercio con multiples periodos
   ------------------------------------------
* Simulamos datos
clear all
set obs 20
gen ind = _n
expand 10
bysort ind: egen time = seq(), f(2010) t(2019) // dentro de cada ind
gen treat = 1 if ind <= 10
replace treat = 0 if treat == .

* Generamos la varible donde empieza el evento
set seed 1 // Semilla aleatoria
gen year_treated = runiformint(2013,2017)
bysort ind: replace year_treated = year_treated[1]
replace year_treated = 0 if treat == 0

* Dummy de treatment con 1 después del tratamiento
bysort ind: gen post_treatment = 1 if time >= year_treated
replace post_treatment = 0 if post_treatment == .
replace post_treatment = 0 if treat == 0

* Generamos Leads y lags
* Llenamos las observacioens de control con 0
bysort ind: gen lead3 = post_treatment[_n+3]
replace lead3 = 0 if treat == 0
bysort ind: gen lead2 = post_treatment[_n+2]
replace lead2 = 0 if treat == 0
bysort ind: gen lead1 = post_treatment[_n+1]
replace lead1 = 0 if treat == 0
bysort ind: gen lag1 = post_treatment[_n-1]
replace lag1 = 0 if treat == 0
bysort ind: gen lag2 = post_treatment[_n-2]
replace lag2 = 0 if treat == 0
bysort ind: gen lag3 = post_treatment[_n-3]
replace lag3 = 0 if treat == 0

* Simulamos la variable endógena
by ind: gen outcome =  runiform(0,1) if time < year_treated  // Pre
by ind: replace outcome = runiform(2,3) if time == year_treated //
   Evento
by ind: replace outcome = runiform(3.5,4.5) if time > year_treated //
   Post
```

```
* Generamos la variable endógena de control con pre tendencia paralela
replace outcome =  runiform(0,1) if treat == 0
```

Para estimar un modelo 2FE o de múltiples niveles de efectos fijos, podemos hacer uso del paquete **reghdfe**. Como mencionamos anteriormente, podemos encontrar información de este paquete en la página de Correia (2016), aquí.

Por ejemplo, supongamos que queremos correr un 2FE controlando por efectos fijos de unidad y tiempo:

```
* Two-way fixed effects (2FE) ----------------------------------------
reghdfe outcome post_treatment, a(ind time)
```

Podemos hacer uso el mismo paquete o del paquete **xtreg** para correr la especificación de *Lads and lags*.

```
* Especificación Leads and Lags ---------------------------------------
* Usando xtreg
xtset ind time
xtreg outcome lead3 lead2 lead1 post_treatment lag1 lag2 lag3 i.time, fe

* Usando reghdfe
reghdfe outcome lead3 lead2 lead1 post_treatment lag1 lag2 lag3, a(ind
   time)
sum outcome
return list
estadd local Cluster "Normal" // Var aux
estadd scalar Min = r(min) // valor min
estadd scalar Max = r(max) // valor max
estadd scalar Count = r(N) // obs
estadd scalar Mean = r(mean) // media
estadd scalar SD = r(sd) // std. dev.
est sto est1

label variable lead3   "Lead 3"
label variable lead2   "Lead 2"
label variable lead1   "Lead 1"
label variable post_treatment   "Periodo 0"
label variable outcome   "Var. Dependiente"

* Tabla de leads and lags, omitimos lags
esttab est1 using "./Tables/table5.tex", replace noomitted nobaselevels
   label se r2 title("Leads and lags") keep(lead3 lead2 lead1
   post_treatment) scalars("DF" "Cluster") stats(N r2 Mean SD Min Max
   Cluster, label(N \(R^{2}\) "DepVar: Mean" "DepVar: Std.Dev." "Scale:
   min" "Scale: max" "Std. Errors")) star(* 0.10 ** 0.05 *** 0.01)
   notes addnotes("Lag variables are included but not shown.") nonumbers
```

Tabla 4: Leads and lags

	Var. Dependiente
Lead 3	0.361
	(0.362)
Lead 2	-0.101
	(0.223)
Lead 1	-0.0792
	(0.176)
Periodo 0	2.111***
	(0.152)
N	140
R^2	0.953
DepVar: Mean	1.377
DepVar: Std.Dev.	1.454
Scale: min	0.0169
Scale: max	4.493
Std. Errors	Normal

Standard errors in parentheses

Lag variables are included but not shown.

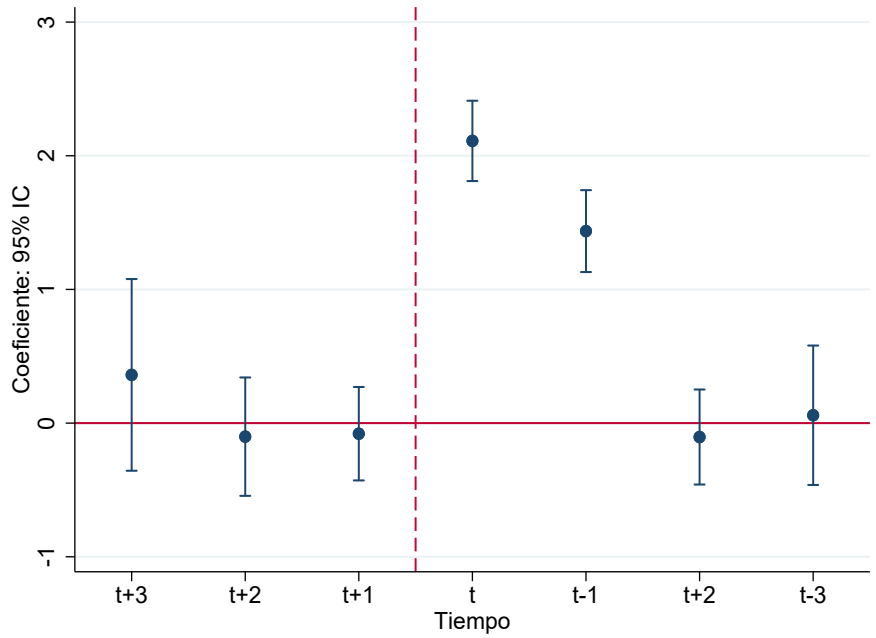
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Las variables en que se muestran son las variables de interés (las variables adelantadas) y observamos que no hay evidencia para rechazar la hipótesis nula $H0 : Lead_i = 0$, para $i = 1, 2, 3$.

Gráficamente:

```
* Gráfica de Leads and lags
ssc install coefplot, replace
coefplot, vertical drop(_cons) yline(0) coeflabels(lead3 = "t+3" lead2
   = "t+2" lead1 = "t+1" post_treatment = "t" lag1 = "t-1" lag2 = "t+2"
   lag3 = "t-3") xline(3.5, lp(dash)) ciopts(recast(rcap))
   xtitle(Tiempo) ytitle(Coeficiente: 95% IC) graphregion(color(white))

* Guardamos la gráfica en formato pdf
graph export "LeadsLags.pdf", replace
```

Figura 1: Gráfica de *Leads and lags* con pre-tendencias paralelas

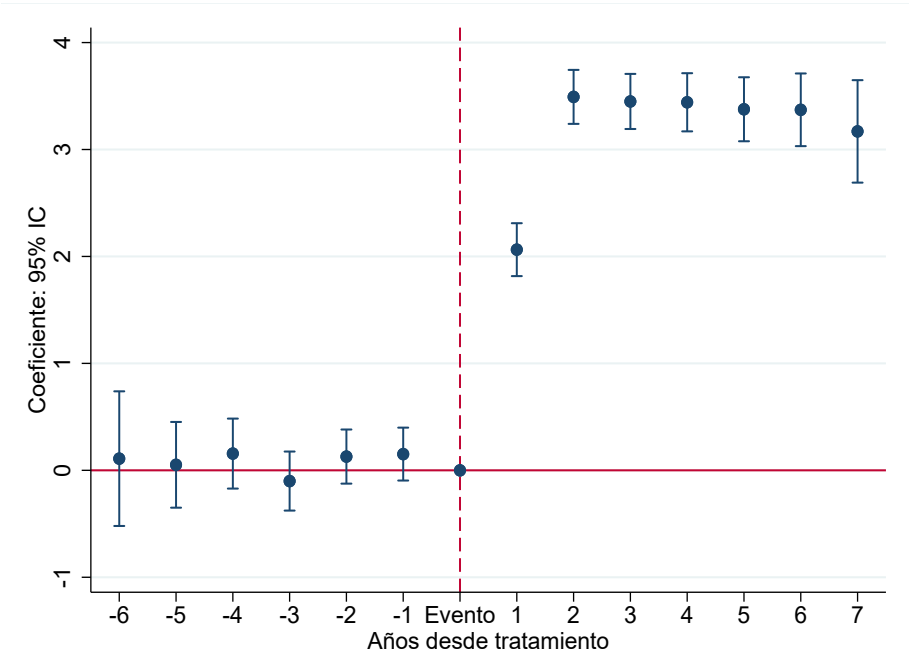


Análisis utilizando Did plot:
Un ejemplo de una gráfica de un *event study* o para nuestro caso, la pueden encontrar en DID plot.

```
* DID plot (event study) ------------------------------------------------
gen time_to_treat = time - year_treated // Creamos variable tiempo al
   tratamiento

replace time_to_treat = 0 if treat == 0 // Cambiamos valores NA a 0 del
   grupo control
```

```
summ time_to_treat
g shifted_ttt = time_to_treat - r(min) // Stata toma solo factores de
    valores positivos, hacemos un mapeo de time to treat (ttt) a
    shifted_ttt para valores positivos. (Sugerencia, vean los datos como
    se van creando las variables)

summ shifted_ttt if time_to_treat == -1
local true_neg1 = r(mean) // Variable local que nos da la categoria
    base, ttt = 0

* Corremos la regresión especificando la categoria base y con 2FE
reghdfe outcome ib‘true_neg1’.shifted_ttt, a(ind time)

* Gráficamos el event study
coefplot, keep(*.shifted_ttt) vertical  base ///
rename(0.shifted_ttt="-6" 1.shifted_ttt="-5" 2.shifted_ttt="-4"
    3.shifted_ttt="-3" 4.shifted_ttt="-2" 5.shifted_ttt="-1"
    6.shifted_ttt="Evento" 7.shifted_ttt = "1" 8.shifted_ttt = "2"
    9.shifted_ttt = "3" 10.shifted_ttt = "4" 11.shifted_ttt = "5"
    12.shifted_ttt = "6" 13.shifted_ttt = "7") ///
yline(0) xline(7, lp(dash)) ciopts(recast(rcap)) xtitle(Años desde
    tratamiento) ytitle(Coeficiente: 95% IC) graphregion(color(white))

* Guardamos la gráfica en formato pdf
graph export "DIDplot.pdf", replace
```

Figura 2: Gráfica DID con pre-tendencias paralelas



Note que no se viola el supuesto de tendencias paralelas, ya que las estimaciones para los periodos anteriores al tratamiento ($t < 0$) no son significativamente distintas de 0.

5. Variables instrumentales

Conceptualmente, la estimación por variables instrumentales se puede interpretar como dos etapas independientes de mínimos cuadrados ordinarios:

1. Primer etapa
$$X \sim Z\delta + \epsilon,$$
donde X son los predictores endógenos, Z los potenciales instrumentos y ϵ el vector de errores. Dado la estimación de $\hat{\delta}$, podemos obtener $\hat{X} = Z\hat{\delta}$.

2. Segunda etapa
$$Y \sim \hat{X}\beta_{IV} + \mu$$
donde Y es la variable dependiente de interés y μ es el error.

```
clear
* Generamos datos
set seed 32 // Semilla aleatoria
set obs 100
generate x1 = rnormal()
generate x2 = rnormal(2)
generate y1  = 2*x1 + 1 + rnormal()
generate y2 = 2*x1 + 3*x2 + 1 + rnormal()
set seed 32 // Semilla aleatoria
gen z1 = 0.85*x1 + 0.2*x2^3 + rnormal(0,1)
gen z2 = 0.5*x2 + 0.1*x1^2 + rnormal(0,1)
```

```
// IV paso a paso
// First stage
reghdfe x1 z1, noabsorb vce(robust) // fs
estimates store IVfs // Guardamos la regresión
qui testparm* // Prueba de hipotesis
estadd scalar p_value = round(r(p),3) // Guardamos el p value del
   partial f test

* Corremos la estimación en 2 pasos con un solo comando
ivreghdfe y1 (x1 = z1), robust
estimates store IV2sls // Guardamos la regresión

* Generamos la tabla en .tex
esttab IVfs IV2sls using "./Tables/table5.tex", ///
title(IV: First stage and 2SLS) ///
replace se label ///
star(* 0.10 ** 0.05 *** 0.01) ///
mtitles("First Stage" "2SLS") ///
stats(N r2 p_value, labels("Observations" "R-squared" "F-stat
   (p-value)"))
```

```
eststo clear  // Borramos todas las estimaciones

estimates clear // Borramos las regresiones guardadas
```

Tabla 5: IV: First stage and 2SLS

	(1) First Stage	(2) 2SLS
z1	0.187***	
	(0.0346)	
x1		1.912***
		(0.171)
Constant	-0.466***	0.840***
	(0.0835)	(0.101)
Observations	100	100
R-squared	0.366	0.811
F-stat (p-value)	0	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

La siguiente gráfica incluye muchos periodos, pero algo importante es como agregamos las notas y los comandos adicionales que tenemos que escribir en el compilador de LaTeX.

```
// Multiple regresiones IV
eststo: ivreghdfe y1 (x1 = z1)
eststo: ivreghdfe y1 x1 (x2 = z2)
eststo: ivreghdfe y2 (x1 x2 = z1 z2)
eststo: ivreghdfe y2 (x1 x2 = z1 z2), robust

esttab using "./Tables/table6.tex", replace noomitted nobaselevels
   label se r2 star(* 0.10 ** 0.05 *** 0.01) ///
nonotes postfoot("\hline  \hline \\[-1.8ex]  \multicolumn{5}{l}
   {\parbox[t]{11cm}{ \textit{Nota:} De esta manera podemos modificar
   las notas. * denota p$<$0.1, ** denota p$<$0.05, y *** denota
   p$<$0.01.}}  \\  \end{tabular} ")

* En LaTeX queremos escribirlo de la sigueinte manera
* \begin{table}[H]
* \centering
* \caption{Multiple IV regressions}
* \input{table6.tex}
* \end{table}

eststo clear  // borramos todas las estimaciones
```

Tabla 6: Múltiples regresiones de VI

	(1)	(2)	(3)	(4)
	y1	y1	y2	y2
x1	1.912***	2.048***	2.013***	2.013***
	(0.162)	(0.0973)	(0.126)	(0.120)
x2		-0.120	3.098***	3.098***
		(0.0979)	(0.101)	(0.0944)
Constant	0.840***	1.068***	0.781***	0.781***
	(0.0990)	(0.206)	(0.213)	(0.196)
Observations	100	100	100	100
R^2	0.811	0.816	0.937	0.937

Nota: De esta manera podemos modificar las notas. * denota p<0.1, ** denota p<0.05, y *** denota p<0.01.

6. Regresión discontinua

Consideramos el modelo

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \delta X_i * Z_i + \mu_i,$$

donde Y_i denota la variable endógena, X_i una variable explicativa continua y Z_i una variable por partes, discontinua, y función de X_i.

$$Z_i = \begin{cases} 0 & X_i \leq \bar{x} \\ 1 & X_i > \bar{x}, \end{cases}$$

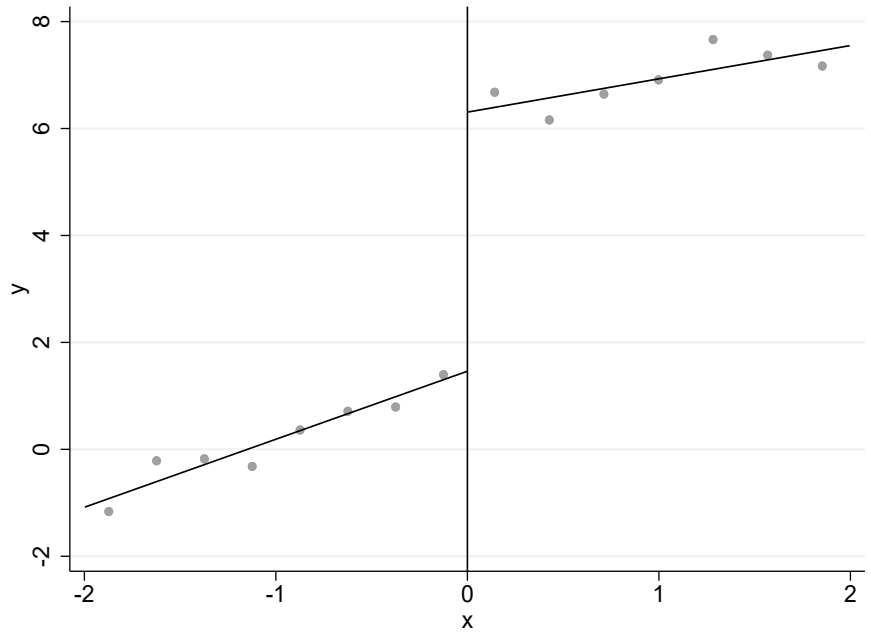
donde $\bar{x}$ es una constante que se conoce como el punto de quiebre. En este caso, todos los individuos tales que $X_i > \bar{x}$ reciben el tratamiento y aquellos que tienen un valor por debajo o igual al punto de quiebre, $X_i \leq \bar{x}$, no reciben el tratamiento. El estimador asociado a Z_i, β_2, es el *Average treatment effect* (ATE) de los individuos con $X_i = \bar{x}$.

```
clear all
set seed 1 // Semilla aleatoria
set obs 100
generate x = runiform(-2, 2)
generate y = 1 + x + 5*(x >= 0) + rnormal()
gen z = 0
replace z = 1 if x > 0
```

Test de manipulación de datos

```
* McCrary Test (bajad el archivo .ado)
DCdensity x, breakpoint(0) generate(Xj Yj r0 fhat se_fhat)
 drop Xj Yj r0 fhat se_fhat

* Uando rddensity
rddensity x // Test de manipulation, H0: cutoff manipulado

* Paquete rd
ssc install rd, replace

* Gráfica automática del diseño RD
rdplot y x
```

Figura 3: Gráfica de regresión discontinua




```
* Bandwidth
rdbwselect y x, bwselect(IK) // bw optimo

* Estimación
rdrobust y x rdbwselect(ik)
```

Tabla 7: RDD

	(1)
	y
RD_Estimate	4.722***
	(0.751)
Observations	100
R^2	

Standard errors in parentheses

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

Referencias

Autor, David H. (2003). "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing". En: *Journal of Labor Economics* 21.

Correia, Sergio (2016). *Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator*. Inf. téc. Working Paper.

Imai, Kosuke e In Song Kim (2020). "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data". En: *Cambridge University Press*.

Naqvi, Asjad (2021). *The Stata-to-LaTeX guide*. URL: https://medium.com/the-stata-guide/the-stata-to-latex-guide-6e7ed5622856 (visitado 03-09-2021).

Pischke, Steve (oct. de 2005). *Empirical Methods in Applied Economics*.