

### Trabajo 3: Tópicos de Microeconometría – Regresión Discontinua

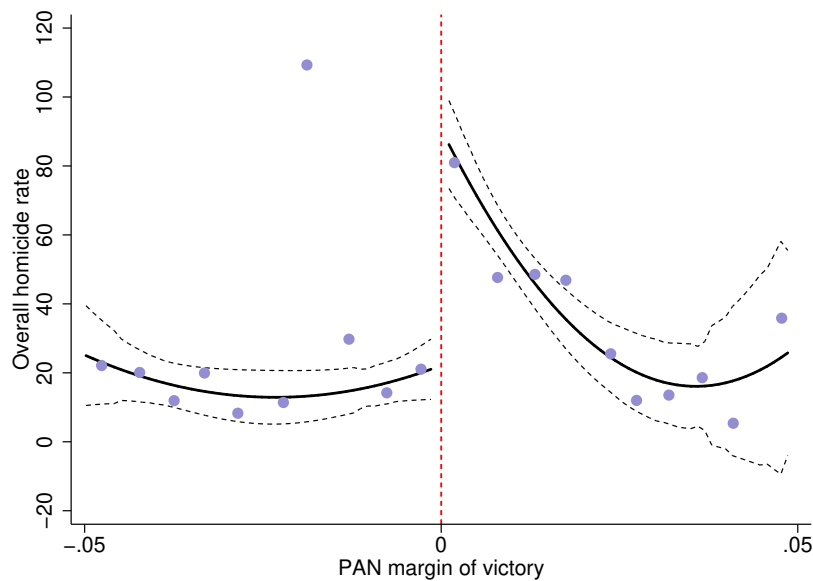
Universidad de Santiago de Chile  
Magíster en Economía  
Semestre 1, 2019

Este ejercicio tiene dos componentes: una sección aplicada, y una sección que vamos a simular.

El primer componente (pregunta A) se enfoca en el paper “Trafficking Networks and the Mexican Drug War”, escrito por Dell (2015). Su paper examina el impacto de la política Mexicana anti-droga sobre violencia, y específicamente homicidios. Explora la variación en el partido del/la alcalde/sa basada en elecciones cercanas para estimar un diseño de regresión discontinua. El partido PAN implementó una serie de medidas anti-drogas de grande escala, y ella estudia si estas medidas tuvieron un impacto sobre violencia. Más detalles están disponibles en el paper publicado (disponible en el sitio web del curso). Para la pregunta 1, se dispone de la base de datos de Dell (*DrugTrafficking.dta*), que contiene las variables que miden el porcentaje de votos en cada elección (solo se incluyen elecciones cercanas), la cantidad de homicidios, y la tasa de homicidios en cada municipalidad, y por último, un indicador de si una elección fue ganada por el partido PAN. Un resultado gráfico del paper (que replicarías en la pregunta A) se presenta a continuación.

Para la segunda parte (pregunta B), simularemos nuestros propios datos para examinar como la regresión discontinua funciona cuando conocemos exactamente el subyacente proceso generador de datos. La simulación es un muy herramienta muy útil cuando queremos examinar el rendimiento de un estimador, y su habilidad de recuperar un parametro conocido: algo que solamente podemos hacer si tenemos control sobre los no-observables.

Figure 1: Replicación de la figura 4B de Dell (2015) “Trafficking Networks and the Mexican Drug War”, *American Economic Review* 105(6):1738-1779.



La entrega consiste en un documento de trabajo corto (entre aproximadamente 1000 y 2000 palabras), y el código que utilizas para hacer el análisis detallada a continuación. La entrega se hace por correo elec-

trónico (a [damian.clarke@usach.cl](mailto:damian.clarke@usach.cl) con copia a [kathya.tapia@usach.cl](mailto:kathya.tapia@usach.cl)) antes de las 23:59 del jueves 6 de junio de 2019. **Todos los archivos entregados deben ser enviados en un solo archivo .zip llamado “NombreApellido.zip” (con tu nombre y apellido en vez de “Nombre” y “Apellido”).** Si no tienen un programa de compresión en tu computador para crear archivos .zip, hay sitios que permite hacerlo, por ejemplo: <https://archivo.online-convert.com/es/convertir-a-zip>.

### **Evaluación:**

Este trabajo forma 15% de la nota final del curso. La evaluación tomará en cuenta la parte escrita, y también el código de análisis que demuestra los resultados producidos. La nota para cada ítem está indicada en la pregunta, y la nota final se calcula como:  $nota = 1 + \frac{total}{10}$  (de 1,0 a 7,0). Si el código *no* se replica los resultados entregados, se descontará hasta 5 décimas de la nota final. Los puntos para cada pregunta dependen tanto de los resultados entregados (tablas o gráficos), cómo también las respuestas a cada pregunta.

### **Preguntas:**

**(A) Estimando una Regresión Discontinua con Dell (2015)** Abra la base de datos DrugTrafficking.dta, y estime la siguiente regresión, como descrito en la ecuación de Dell:

$$HomicideRate_m = \beta_0 + \beta_1 PANwin_m + \beta_2 PANwin_m \times f(VoteDif_m) + \beta_3(1 - PANwin_m) \times f(VoteDif_m) + \varepsilon_m \quad (1)$$

$PANwin_m$  es un término binario indicando si PAN ganó en una elección cercana, y los términos de interacción son funciones del margen de victoria en cada lado del umbral de victoria (50% de votos), permitiendo que esta “variable corredora” (*running variable*) comporta de forma distinta en cada lado de la discontinuidad. En cada caso utilizaremos la variable  $HomicideRate_m$ , que es la tasa de homicidio al nivel de la municipalidad, como nuestra variable dependiente de interés.

1. (3 puntos) Estime la regresión utilizando una función lineal para  $f(VoteDif_m)$  en cada lado de la discontinuidad.
2. (2 puntos) Estime la regresión utilizando una función cuadrático para  $f(VoteDif_m)$  en cada lado de la discontinuidad. Esto requiere dos términos (lineal y al cuadrado) en cada lado de la discontinuidad.
3. (10 puntos) Replique la figura en la página anterior (figura 4B del paper de Dell). No es necesario preocuparse del formato final, ni tampoco los intervalos de confianza, que son las líneas punteadas. Note que cada punto es la tasa promedio de homicidio en “bins” (o grupos) de votos de 0.005. Se puede recrear la línea en cada lado de la discontinuidad utilizando una función cuadrática (por ejemplo `qfit`).
4. (5 puntos) Por qué solamente nos enfocamos en elecciones cuyo margen de victoria es entre -0.05 to +0.05?
5. (10 puntos) En ests caso, ¿en qué consiste el test de McCrary (2008)? Realice este test, e incluya el gráfico producido en la entrega final. ¿Cuál es el resultado de este test?

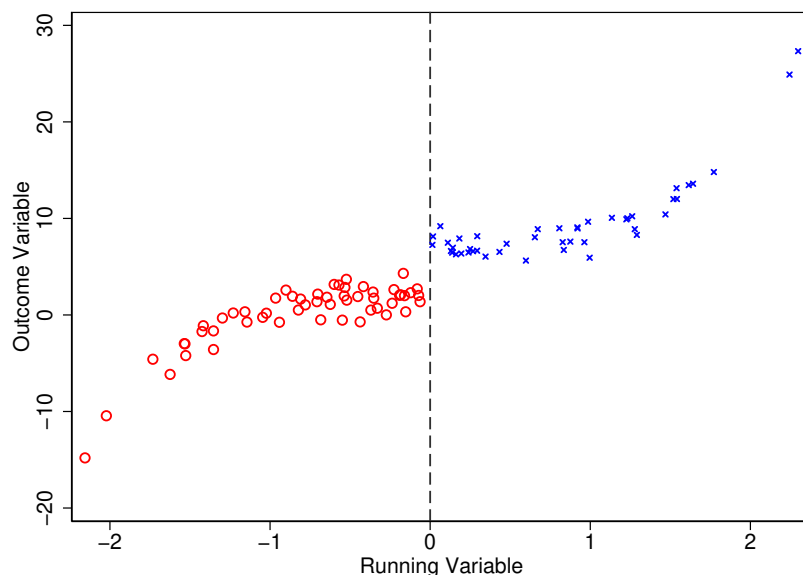
**(B) Simulando una Regresión Discontinua** En esta pregunta, simularemos una relación discontinua, y examinaremos si, mediante una regresión local-lineal, podemos recuperar el parametro de interés cuando

la relación entre la variable corredora ( $x$ ) y el outcome ( $y$ ) no es lineal. Utilizaremos la siguiente forma funcional para simular nuestros datos:

$$y = 2 - 0.5x + 0.1x^2 + 1.7x^3 + 5w + \varepsilon \quad (2)$$

donde aquí  $y$  es la variable dependiente de interés,  $x$  es la variable corredora, y  $w$  es la variable de tratamiento. Solamente los individuos que tienen  $x \geq 0$  reciben tratamiento, y por lo tanto  $w$  estará igual a 1 si  $x \geq 0$  y 0 si  $x < 0$ .

1. (5 puntos) Simule 100 observaciones siguiendo a esta especificación, donde tanto  $x$  como  $w$  son extraídos de de la distribución normal, con medio 0 y desviación estándar 1. El comando `set obs` se puede utilizar para definir la cantidad de observaciones para simular.
2. (5 puntos) Replique la siguiente figura que grafica los datos simulados.



No es necesario replicar el mismo estilo. Si quiere basarse en los mismos número pseudo-aleatorios utilizado en esta figura, antes de simular, utilice el comando `set seed 110`.

3. (10 puntos) Estime el coeficiente en el tratamiento  $w$  utilizando un control local para la variable  $x$ , concentrando solamente en las observaciones en el rango  $x \in (-2, 2), x \in (-1.9, 1.9), \dots, x \in (-0.1, 0.1)$ . La forma funcional para  $x$  es la misma en ambos lados del corte, y por lo tanto puede estimar el modelo simplemente incluyendo un solo control lineal para  $x$ , sin el término de interacción. Esto resultará en 20 distintas estimaciones (una para cada uno de los rangos de  $x$  considerado). Presente estos resultados como quiere – un gráfico podría ser útil para visualizar todas a la vez. **Pista:** En vez de hacer todo esto “a mano”, podría ser útil utilizar un bucle. Como ejemplo, considere la estimación de una regresión de  $y$  sobre  $x$  solamente para las observaciones que tienen una realización de  $x$  superior a una serie de valores, en cada caso guardando el coeficiente estimado sobre  $x$  en la variable `coefficients`, y el punto de corte de  $x$  en cada regresión en la variable `cutoff`:

```
gen coefficients = .
gen cutoff      = .
```

```
local i = 1
foreach num of numlist 0.1(0.1)2 {
    reg y x if x > `num'
    replace coefficients = _b[x] in `i'
    replace cutoff      = `num' in `i'
    local i = `i'+1
}
```

Necesitaría aplicar este código el ejemplo específico en la pregunta, que requeriría algunas modificaciones!

4. (10 puntos) ¿Qué nos dicen los resultados anteriores acerca del comportamiento de una regresión discontinua utilizando regresiones lineales locales? ¿Cuál es el punto de corte óptimo según Imbens y Kalyanaraman, (2012) y Calonico, Cattaneo and Titiunik, (2014b) (ambas referencias en los apuntes del curso)? Exlique sus resultados, y si logran rescatar el valor “poblacional” en cada caso.