Nota 8
Regresión Discontinua

1 Introducción

El método de regresión discontinua es un método cuasiexperimental utilizado para identificar efectos causales. Este método se basa en cortes que surgen por ley o por diseño y que implican una discontinuidad en alguna variable (X_1). Este método hará posible determinar la relación causal de cambios en esta variable sobre alguna otra (la variable dependiente). Un ejemplo tradicional que se ha utilizado son las votaciones por mayoría. En casos donde hay dos contrincantes en una elección, si se usa la regla de mayoría sabremos que un candidato gana si obtiene mas del 50% de las votaciones. Supongamos que nos interesa ver si los gobiernos de izquierda hacen una diferencia en la política fiscal o económica (Pettersson-Lidbom, 2007). Si el partido de izquierda recibe más del 50% votos gana la elección y si recibe menos la pierde. Llevar a cabo una estimación de MCO donde la variable dependiente es la tasa impositiva y tu variable de interés es que el gobierno de izquierda haya ganado muy probablemente te llevará a un sesgo por variable omitida (en particular, que tan liberal es el electorado puede llevar a un sesgo). Intuitivamente, el método de regresión discontinua compara situaciones en las cuales el gobierno de izquierda apenas gana (i.e. reciba poco mas del 50% de los votos) con situaciones donde apenas pierda (i.e. reciba poco menos del 50% de los votos). En este caso se puede argumentar que justo en la discontinuidad pasamos de una situación en la que el partido de izquierda pierde a una en la que gana. Sin embargo, en ambos casos el electorado debe ser similarmente liberal (de igual manera no debe haber un brinco abrupto en cualquier otra variable que provoque sesgo). Lo que haremos entonces es ver si en esa discontinuidad hay un brinco discontinuo en la variable dependiente. De haberlo, el método de regresión discontinua atribuirá dicho brinco al hecho de que el gobierno de izquierda haya ganado la elección.

2 Planteamiento

Este método nos permitirá investigar el efecto de la variable T_i sobre Y_i, donde T_i será una variable dummy que especifica si el individuo i forma parte del grupo de "tratamiento" ($T_i = 1$) o de "control" ($T_i = 0$). El planteamiento utiliza nuevamente el concepto de resultados potenciales que describimos en la *Nota 7*. Cada individuo tiene dos resultados potenciales, de los cuales únicamente observamos uno (el realizado). Y_i^T es el nivel de la variable dependiente que i tendría si forma parte del grupo de tratamiento, Y_i^C el que tendría si forma parte del grupo de control y Y_i el nivel observado.

A diferencia que en el caso de experimentos donde ambos grupos (tratamiento y control) eran determinados de forma aleatoria, en este caso son determinados por una regla objetiva de decisión. Dicha regla deberá especificar cortes especificos en los cuales la probabilidad de formar parte del grupo de tratamiento o control cambie de forma abrupta (i.e. discontinua). La regla deberá estar basada en una variable, que llamaremos la *variable definitoria* (G_i). Esta variable puede estar relacionada directamente con los resultados potenciales (y por tanto, con la variable dependiente), sin embargo, dicha relación (así como con el resto de las variables de control) se asume como continua. El cambio discontinuo en la probabilidad de formar parte del grupo de tratamiento puede ser de dos tipos:

2.1 Regresión Discontinua "Sharp"

Este tipo de discontinuidad se refiere al caso en el cual la probabilidad de formar parte del grupo de tratamiento pasa de 0 a 1 después del corte que determina la discontinuidad. Ejemplos de este tipo de cortes incluyen: (i) márgenes por los cuales se pierde una elección; (ii) cortes administrativos que definen diferencias en precio (e.g. adulto mayor de 65 años); (iii) política en México de apoyo a los 125 municipios más pobres.

En este caso T_i es definida por la *variable definitoria* (G_i) y el punto de discontinuidad es k:

$$T_i = 1\{G_i \geq k\} \tag{1}$$

Todos los individuos con $G_i \geq k$ estarán en el grupo de tratamiento (i.e. su participación es obligatoria) y todos con $G_i < k$ estarán en el grupo de control (i.e. su participación en el tratamiento está prohibida).

Para identificar el efecto del tratamiento sobre la variable dependiente tendremos que asumir:

1. Independencia (condicional en la variable definitoria). $Y_i^T, Y_i^C \perp T_i | G_i$
2. Continuidad. $E(Y_i^T | G_i = g)$ y $E(Y_i^C | G_i = g)$ son continuas en $G_i = k$.

Además de esos supuestos utilizaremos la *ley de esperanzas iteradas* que indica que el valor esperado de Y_i puede ser calculado como:

$$\begin{aligned} E(Y_i | G_i = g) = E(Y_i^T | T_i = 1, G_i = g) * Pr(T_i = 1 | G_i = g) + \\ + E(Y_i^C | T_i = 0, G_i = g) * Pr(T_i = 0 | G_i = g) \end{aligned} \tag{2}$$

Utilizando los supuestos establecidos y [2] tenemos que:

$$\begin{aligned} E[Y_i^C | G_i = k] &= \lim_{g \to k^-} E[Y_i^C | G_i = g] \\ &= \lim_{g \to k^-} E[Y_i^C | T_i = 0, G_i = g] \\ &= \lim_{g \to k^-} E[Y_i | G_i = g] \end{aligned} \tag{3}$$

$$E[Y_i^T | G_i = k] = \lim_{g \to k^+} E[Y_i | G_i = g] \tag{4}$$

Por lo tanto, utilizando estos supuestos podemos obtener el efecto de tratamiento en el punto $G_i = k$:

$$\tau_s = E[Y_i^T - Y_i^C | G_i = k] \tag{5}$$

En este caso será necesario utilizar extrapolación, ya que si G_i es continua, la probabilidad de observar alguna unidad con $G_i = k$ será cero. Por lo tanto, no tendríamos observaciones para llevar a cabo la estimación. En este caso utilizaremos observaciones con G_i arbitrariamente cerca del valor de corte k.

Para estimar el valor de τ_s podemos estimar el siguiente modelo mediante MCO utilizando únicamente las observaciones con $|G_i - k| < h$ (es decir, observaciones con G_i a menos de h de distancia del valor de corte k):[1]

$$Y_i = \alpha_0 + \alpha_1(G_i - k) + \tau_s T_i + \alpha_2(G_i - k)T_i \tag{6}$$

2.2 Regresión Discontinua "Fuzzy"

En este tipo de discontinuidad la probabilidad de formar parte del grupo de tratamiento tiene un cambio discontinuo en el punto de corte $G_i = k$. Sin embargo, la probabilidad no cambia de 0 a 1. Este tipo de discontinuidad suele darse en casos en los cuales el punto de corte determina la elegibilidad al tratamiento, pero que la participación en el grupo de tratamiento no es obligatoria. Ejemplos: (i) PMT, como Progresa/Oportunidades; (ii) becas basadas en la calificación de algún examen de admisión estandarizado.

Sea $T_i(g)$ el estatus potencial de tratamiento para un valor dado de la *variable definitoria*, donde g es un valor cercano a k (el punto de corte). En este caso, $T_i(g) = 1$ si i tomaría el tratamiento en caso de que el punto de corte fuera igual a g. En el caso de regresión discontinua tipo "fuzzy" necesitaremos agregar el siguiente supuesto:

3. Monotonicidad. $T_i(g)$ es no-creciente en x si $x = k$. Este supuesto corresponde al supuesto de variables instrumentales que indica que no deben existir los "defiers" (contreras).

En este contexto, un *complier* es aquel que $\lim_{g \to G_i^+} T_i(g) = 0$ y $\lim_{g \to G_i^-} T_i(g) = 1$. Es decir, si el punto de corte es menor que el valor de su *variable definitoria* (G_i), i tomaría el tratamiento; y si es mayor no lo tomaría.

Siguiendo con el procedimiento de variables instrumentales, en este caso podremos estimar el efecto del tratamiento para los *compliers* que además tienen $G_i = k$. Este estimador corresponderá a:

$$\tau_f = \frac{\lim_{g \to k^+} E[Y_i | G_i = g] - \lim_{g \to k^-} E[Y_i | G_i = g]}{\lim_{g \to k^+} E[T_i | G_i = g] - \lim_{g \to k^-} E[T_i | G_i = g]} \tag{7}$$

[1] En el argot econométrico esto se conoce como utilizar *mínimos cuadrados ordinarios locales*. El término *locales* se refiere a que no se utilizan todas las observaciones de la muestra, sino sólamente un subconjunto. En este caso, solo las observaciones alrededor del punto de corte.

Para estimar el valor de τ_f generamos un indicador de la intención de tratamiento ($Z_i = 1\{G_i \geq k\}$) y estimamos los siguientes modelos mediante MCO utilizando únicamente las observaciones con $|G_i - k| < h$ (que corresponden a la primera etapa y la forma reducida):

$$T_i = \eta_0 + \eta_1(G_i - k) + \phi Z_i + \eta_2(G_i - k) * Z_i \tag{8}$$

$$Y_i = \psi_0 + \psi_1(G_i - k) + \theta Z_i + \psi_2(G_i - k) * Z_i \tag{9}$$

Utilizando los resultados de [8] y [9]:

$$\tau_f = \theta/\phi \tag{10}$$

3 Análisis gráfico

La representación gráfica permite visualizar de manera muy clara la estrategia del método de regresión discontinua para identificar los efectos causales. Existen cuatros tipos de gráficas que se recomienda incluir como parte de este método:

1. **Media condicional de la variable dependiente.** Como podemos observar, tanto el modelo [6] como el [9] representan estimaciones del valor esperado de Y_i como función de G_i. Por lo tanto, una gráfica que comúnmente se incluye como parte de este método es simplemente mostrar la media de Y_i condicional en distintos valores de G_i. En caso de que el método de regresión discontinua identifique un efecto significativo del tratamiento, es de esperarse un cambio discontinuo de la media condicional alrededor del valor $G_i = k$.

2. **Media condicional del tratamiento.** En el caso de regresión discontinua tipo "fuzzy" es común agregar una gráfica de la media de la probabilidad de recibir tratamiento condicional en G_i. Al igual que en el caso anterior, es de esperarse observar un cambio discontinuo de esta probabilidad condicional alrededor del valor $G_i = k$. En el caso de la regresión tipo "sharp" no es necesario incluir esta gráfica si se sabe que el cumplimiento de la regla es obligatorio.

3. **Media condicional de controles.** Para asegurarnos que el cambio discontinuo de Y_i en $G_i = k$ se da solo como resultado del tratamiento, es común analizar de manera gráfica si dicha discontinuidad no implica cambios en otras variables. Si estas variables cambian de manera discontinua y están asociadas con la variable dependiente Y_i, no será posible saber de dónde proviene la causalidad.

4. **Densidad de G_i.** Es importante que los individuos no puedan manipular su valor de G_i para hacerse elegibles al tratamiento. Por ello, suele graficarse la densidad de G_i, esperando que no haya cambios en la densidad abruptos alrededor del valor $G_i = k$. Dar ejemplo del caso de corrupción en Colombia.

4 Consideraciones finales

4.1 Diferentes especificaciones: polinomios

Dado que en muchos casos las especificaciones de MCO locales que se utilizan para estimar el efecto del tratamiento utilizan datos cerca de la discontinuidad, utilizar una especificación lineal para describir la relación entre Y y la variable definitoria (G) podría no ser adecuado. Como mostraremos en clase, esto podría generar sesgo en la estimación de τ_s o τ_f. En estos casos, se sugiere utilizar un polinomio de mayor grado para caracterizar la relación entre Y_i y G_i. Por ejemplo, si en vez de utilizar la especificación lineal descrita en [6], utilizaramos un polinomio de grado 3, la estimación de τ_s se llevaría a cabo con la siguiente especificación:

$$\begin{aligned} Y_i = &\alpha_0 + \alpha_1(G_i - k) + \tau_s T_i + \alpha_2(G_i - k)T_i + \alpha_3(G_i - k)^2 + \alpha_4(G_i - k)^2 T_i + \dots \\ &\dots + \alpha_5(G_i - k)^3 + alpha_6(G_i - k)^3 T_i \end{aligned}$$

Para generalizar este caso a un polinomio de cualquier orden, la especificación del modelo de regresión discontinua típicamente se establece como:

$$Y_i = \tau_s T_i + f(G_i - k) \tag{11}$$

Lo mismo aplica para el caso de la regresión tipo *fuzzy*. En dicho caso, las ecuaciones [8] y [9] se transformarían al mismo polinomio y la estimación de τ_f se llevaría a cabo con las siguientes especificaciones:

$$T_i = \phi Z_i + f(G_i - k) \tag{12}$$

$$Y_i = \theta Z_i + f(G_i - k) \tag{13}$$

4.2 Selección de h

Para la estimación de los modelos [6], [8] y [9] es relevante determinar cuántas observaciones se tomarán en cuenta para la estimación. Dado que restringimos las observaciones a individuos cuyo valor de G_i se encuentra a menos de h de diferencia respecto al punto de corte k, es importante determinar el valor de h.

Una manera óptima de determinar h consiste en utilizar el método de *cross-validation*. Este método consiste en encontrar el valor de h que minimice:

$$CV_Y(h) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu}(G_i))^2 \tag{14}$$

donde $\hat{\mu}(G_i) = \alpha_0$ si $k-h < G_i < k$ y $\hat{\mu}(G_i) = \alpha_0 + \tau_s$ si $k < G_i < k+h$. Dado que $CV_Y(h)$ es una función de h, una manera de resolver esto será utilizando métodos númericos.

Otra alternativa recientemente desarrollada consiste en utilizar el procedimiento sugerido por Imbens y Kalyanaraman (2008). Este procedimiento se enfoca en encontrar la h que minimice:

$$E[((\hat{\mu}_r(k) - \hat{\mu}_l(k)) - (\mu_r(k) - \mu_l(k)))^2] \tag{15}$$

donde $\hat{\mu}_r(k) - \hat{\mu}_l(k) = \hat{\tau}_s$ y $\mu_r(k) - \mu_l(k) = \tau_s$.

Independientemente de la selcción de h, al utilizar el método de regresión discontinua se sugiere mostrar que el estimador de τ no ds muy sensible a cambios en h. Por ello se sugiere llevar a cabo las estimaciones cambiando h al doble o la mitad y mostrar los resultados omo pruebas de robustez.

4.3 Validez externa

Un problema del uso del método de regresión discontinua es que los estimadores son válidos para un grupo muy reducido de la población ($G_i = k$). Por lo tanto, en caso de que el efecto de tratamiento sea muy heterogéneo para distintos individuos de la población, se puede concluir poco acerca del efecto del tratamiento para la población en general. Hay casos en los cuales el valor del tratamiento en el corte es muy relevante. Por ejemplo, en el caso en que el corte está dado por alguna política pública, es interesante determinar el efecto para aquellas personas en el margen. Mas aun, en el caso de regresión discontinua “fuzzy,” el efecto estimado es un $LATE$, es decir, es el efecto para los *compliers* con $G_i = k$. A comaración de otros métodos, la gran ventaja de regresión discontinua surge de su alta validez interna.