

Mô hình Biến Công cụ và Hồi quy Gián đoạn (Instrumental Variables and Regression Discontinuity Design)

Kinh tế lượng ứng dụng

Lê Việt Phú

Chương trình Giảng dạy Kinh tế Fulbright

Ngày 20 tháng 5 năm 2015

Table of contents

Ôn tập lý thuyết

Lựa chọn không quan sát được - Selection on unobservables using IV

Thực hành phương pháp biến công cụ

Hồi quy gián đoạn - Regression Discontinuity Design

Ôn tập lý thuyết

Giả sử chúng ta muốn ước lượng tác động của chính sách tín dụng hộ gia đình lên thu nhập của hộ bằng một mô hình đơn giản sau:

$$Y_i = \alpha X_i + \beta * T_i + \varepsilon_i$$

- ▶ X_i là các đặc tính nhân khẩu học
- ▶ T là biến số chính sách, nhận giá trị 0 và 1 đối với gia đình không và có tham gia,

$$T = \begin{cases} 1 & \text{With participation} \\ 0 & \text{Without participation} \end{cases}$$

- ▶ ε_i là phần dư
- ▶ Nếu biến chính sách có tương quan với phần dư, $cov(T, \varepsilon) \neq 0$, ước lượng của β sẽ bị chệch.

Ôn tập lý thuyết (2)

Các cách xử lý vấn đề lựa chọn mẫu đã học:

- ▶ Ngẫu nhiên hoá việc chọn tham gia chương trình (random treatment assignment) $\Rightarrow cov(T, \varepsilon) = 0$ và β là tác động nhân quả của chính sách lên thu nhập.
- ▶ Nếu xác suất tham gia chính sách chỉ dựa vào các yếu tố có thể quan sát được $Y_i^0, Y_i^1 \perp T_i | X_i$ (conditional independence condition) thì có thể sử dụng phương pháp điểm xu hướng PSM.
- ▶ Nếu nhân tố không quan sát được có thể ảnh hưởng đến việc tham gia chương trình không thay đổi theo thời gian (time invariant unobserved heterogeneity) thì có thể dùng phương pháp khác biệt kép.

Lựa chọn không quan sát được - Selection on unobservables using IV

- ▶ Nếu việc tham gia chính sách là không ngẫu nhiên, hoặc tác động của nhân tố không quan sát được thay đổi theo thời gian, hoặc chỉ có dữ liệu chéo $\Rightarrow$ Tất cả các phương pháp đã học đều không sử dụng được $\Rightarrow$ sử dụng phương pháp biến công cụ (IV).
- ▶ IV cũng thường được sử dụng trong các trường hợp biến giải thích nội sinh (endogenous variables), hoặc có vấn đề sai số đo lường (measurement errors).

Khung phân tích sử dụng biến công cụ

Giả sử tồn tại một biến Z sao cho:

- ▶ $cov(T, Z) \neq 0$
- ▶ $cov(Z, \varepsilon) = 0$ (exclusion restriction)

$\Rightarrow$ biến Z ảnh hưởng đến việc lựa chọn tham gia chính sách nhưng không ảnh hưởng đến kết quả của chính sách. Khi đó có thể dùng biến Z để ước lượng tác động của chính sách bằng hồi quy 2 bước:

- ▶ Bước 1: Hồi quy $T_i = \gamma Z_i + \phi X_i + u_i$, ước lượng $\hat{T}_i$
- ▶ Bước 2: Hồi quy $Y_i = \alpha X_i + \beta \hat{T}_i + \varepsilon_i$

$$\beta_{IV} = \frac{cov(Y, Z)}{cov(T, Z)}$$

Khung lý thuyết phân tích

Hiểu thế nào về phương pháp biến công cụ?

- ▶ Bước 1: Tách biến động của việc tham gia chính sách (T) ra khỏi ảnh hưởng trực tiếp đến thu nhập (Y) thông qua tương quan giữa T và Z $\Rightarrow \hat{T}_i$ được coi là ngoại sinh đối với thu nhập.
- ▶ Bước 2: Dùng biến động $\hat{T}_i$ đã tách từ bước 1 để ước lượng tác động của chính sách lên thu nhập.

Khác biệt giữa IV với reduced-form regression (hồi quy rút gọn):

Tại sao không thay Z_i thay cho biến chính sách và ước lượng phương trình:

$$Y_i = \alpha X_i + \beta * Z_i + \varepsilon_i$$

mà phải dùng hồi quy 2SLS?

Lựa chọn biến công cụ như thế nào?

- ▶ Đặc tính về địa lý như khoảng cách, hay các thay đổi có yếu tố bất ngờ như các hiện tượng thời tiết cực đoan, chiến tranh, hay thay đổi chính sách vĩ mô chỉ ảnh hưởng đến một số cá nhân trong xã hội (RD).
- ▶ Một số ví dụ điển hình: tác động của chương trình đào tạo để giúp người thất nghiệp. Việc tham gia chương trình là không ngẫu nhiên. Cần biến công cụ tương quan với việc tham gia, nhưng không trực tiếp tương quan với xác suất xin việc. Dùng khoảng cách quan sát được giữa nhà với trung tâm đào tạo làm biến công cụ.
- ▶ Nghiên cứu về thu nhập và nội chiến (Miguel et al 2005, JPE)
- ▶ Nghiên cứu về tác động lâu dài của bom Mỹ đến tăng trưởng kinh tế ở VN.
- ▶ Nghiên cứu về tỷ suất thu nhập của số năm đi học.

IV-LATE (Local Average Treatment Effect): Biến công cụ ước lượng tác động tham gia trung bình nội tại

- ▶ Đây là trường hợp đặc biệt đối với biến nội sinh là nhị phân (binary) như có hay không tham gia chính sách. Thay đổi của biến công cụ Z ảnh hưởng đến việc tham gia chính sách, ví dụ Z tăng làm cho biến chính sách chuyển từ 0 (không tham gia) thành 1 (có tham gia).
- ▶ IV-LATE sẽ đánh giá tác động của việc tham gia chính sách đối với những người mà trạng thái tham gia chính sách thay đổi do thay đổi của biến công cụ (khái niệm "nội tại"). Những người không thay đổi trạng thái tham gia chính sách (những người luôn tham gia, hoặc không tham gia) không bị ảnh hưởng.

Thực hành phương pháp biến công cụ

- ▶ STATA file hh_98.dta và iv.do
- ▶ STATA code: ivregress, xtivreg, treatreg (sử dụng khi biến nội sinh là biến trạng thái chính sách)

Các kiểm định trong phương pháp biến công cụ:

- ▶ Kiểm định Wu-Hausman kiểm định tính nội sinh của biến chính sách T
- ▶ Biến công cụ yếu (weak instruments): 1st-stage F-stat > 10 (Stock and Yogo, 2005)
- ▶ Exclusion restriction untestable

Hồi quy gián đoạn - Regression Discontinuity Design

- ▶ Là trường hợp đặc biệt của IV-LATE: việc tham gia chính sách phụ thuộc vào một biến khác, nhưng với tiêu chí tham gia quan sát được (selection on observables).
- ▶ Tiêu chí lựa chọn được thiết lập không phụ thuộc vào ý muốn chủ quan của cá nhân hay hộ gia đình $\Rightarrow$ việc tham gia chương trình là một tình huống bán thực nghiệm. Ví dụ:
 - ▶ Sinh viên tốt nghiệp loại giỏi nếu điểm trung bình lớn hơn 75%, hay học sinh tiểu học được miễn học phí nếu thu nhập của bố mẹ thấp hơn 2 triệu đồng một tháng.
 - ▶ Những hộ gia đình ở hai bên cận biên của tiêu chí được kỳ vọng giống nhau về mọi mặt, ngoại trừ việc tham gia chính sách $\Rightarrow$ Có thể so sánh các hộ này để tính ra tác động của việc tham gia chính sách.

Khung lý thuyết phân tích

Giả sử việc cấp học bổng cho học sinh chỉ dựa vào điểm trung bình S . Học sinh đạt được điểm số cao hơn S^* sẽ được nhận học bổng.

Trường hợp thứ nhất: hồi quy cắt với tiêu chuẩn tham gia rõ ràng (Sharp RD)

Đối với trường hợp này miễn là học sinh đạt được điểm trung bình lớn hơn hoặc bằng S^* thì được nhận học bổng.

$$T_i = \begin{cases} 0 & \text{if } S_i < S^* \\ 1 & \text{if } S_i \geq S^* \end{cases} \quad (1)$$

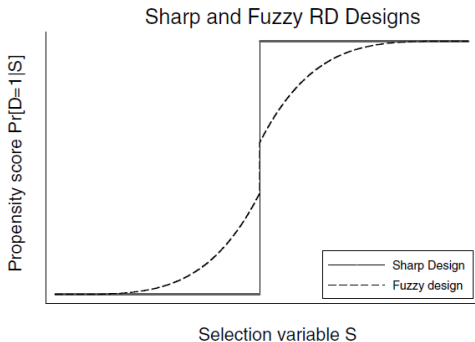
Xác xuất nhận học bổng tăng từ 0 đến 1 sau khi điểm trung bình chạm và vượt qua mốc S^* .

Khung lý thuyết phân tích

Trường hợp thứ hai: hồi quy cắt với tiêu chuẩn tham gia không rõ ràng (Fuzzy RD)

- ▶ Trường hợp này tiêu chí không được áp dụng một cách nghiêm túc, do đó có thể xảy ra trường hợp học sinh điểm thấp hơn S^* một chút có thể được nhận học bổng, hoặc học sinh điểm cao hơn S^* một chút không được nhận học bổng.
- ▶ Ví dụ: Nếu ngưỡng thu nhập để nhận hỗ trợ là 1 triệu đồng/tháng, thì hộ gia đình thu nhập trên 1 triệu đồng một chút sẽ có xu hướng khai thấp để được nhận hỗ trợ.

Xác suất tham gia chương trình và ước lượng tác động đối với Sharp-RD và Fuzzy-RD

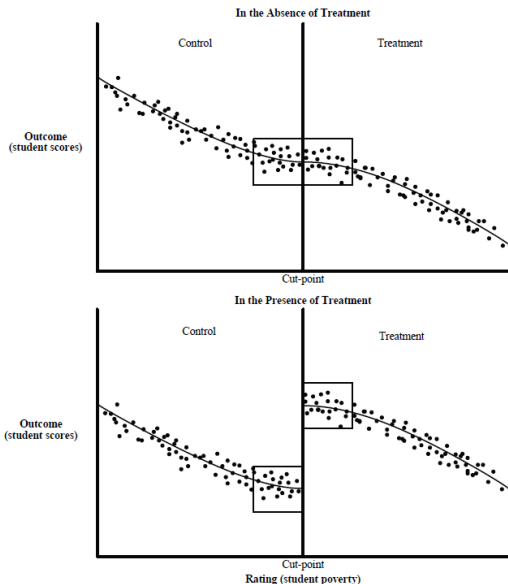


Đối với Sharp-RD, ATT được ước lượng phi tham số:

$$ATT_{SRD} = \lim_{x \rightarrow C^+} E[Y^T | X] - \lim_{x \rightarrow C^-} E[Y^C | X]$$

$$ATT_{FRD} = \frac{\lim_{x \rightarrow C^+} E[Y^T | X] - \lim_{x \rightarrow C^-} E[Y^C | X]}{\lim_{x \rightarrow C^+} P[T = 1 | X] - \lim_{x \rightarrow C^-} P[T = 0 | X]}$$

Ví dụ hồi quy gián đoạn

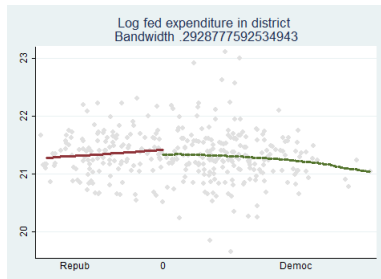


NOTE: Dots represent individual schools. The vertical line in the center of each graph designates a cut-point, above which candidates are assigned to the treatment and below which they are not assigned to the treatment. The boxes represent the proportion of the distribution proximal enough to the cut-point to be used in regression discontinuity analysis when the relationship is viewed as local randomization.

Các bước phân tích hồi quy gián đoạn

Hồi quy gián đoạn dựa vào đồ thị thay vì ước lượng ra các tham số (đặc tính của phương pháp phi tham số - nonparametric):

- ▶ Thay đổi về trạng thái tham gia chương trình tại ngưỡng chính sách, đồ thị có bị đứt gãy hay không?
- ▶ Thay đổi về kết quả tại ngưỡng thực hiện chính sách, có khác biệt rõ rệt không?
- ▶ Kiểm tra điều kiện cân bằng: đồ thị các biến giải thích có liên tục hay bị đứt gãy?



Thực hành hồi quy gián đoạn

- ▶ Tải chương trình lpoly (findit lpoly) cho STATA (bản cập nhật cho chương trình locpoly dùng để ước lượng hồi quy nội tại - local polynomial regression)
- ▶ File dữ liệu hh_98.dta
- ▶ STATA dofile rd.do, chỉnh sửa "locpoly" thành "lpoly"
- ▶ Học viên có thể dùng chương trình rd của STATA ("findit rd") để thực hiện tất cả các bước tự động

So sánh IV và RD

	IV	IV-LATE	RD
Giả định			
Dữ liệu			
Tác động			
Ưu điểm			
Nhược điểm			