

Chương trình Giảng dạy Kinh tế Fulbright**Học kỳ Xuân, 2013****KINH TẾ LƯỢNG ỨNG DỤNG****PHÂN TÍCH TÁC ĐỘNG CỦA CHÍNH SÁCH CÔNG:
CÁCH TIẾP CẬN KHÁC BIỆT TRONG KHÁC BIỆT****1. Giới thiệu phương pháp khác biệt trong khác biệt**

Bài trước, chúng ta đã nghiên cứu phương pháp phân tích tác động của chính sách công dựa trên *thí nghiệm ngẫu nhiên*. Mặc dù về mặt học thuật thì đây là cách làm lý tưởng, nhưng lại không khả thi trong đa số các trường hợp ước lượng tác động của chính sách hay chương trình đầu tư. Một phương pháp thay thế là sử dụng *thí nghiệm tự nhiên*, trong đó vận dụng các tình huống đặc biệt để tạo tính ngẫu nhiên trong việc phân bổ các đối tượng điều tra vào nhóm xử lý và nhóm kiểm soát. Tình huống đặc biệt thường được khai thác là sự thay đổi của luật pháp hay chính sách trong đó có một nhóm đối tượng hay vùng địa lý bị tác động nhưng nhóm đối tượng hay vùng địa lý khác lại không bị chi phối.

Ước lượng khác biệt trong khác biệt (difference in difference hay viết tắt là DID) là một phương pháp thông dụng trong thí nghiệm tự nhiên. Để áp dụng phương pháp này ta cần số liệu bảng trong đó vừa chứa thông tin chéo về các đối tượng khác nhau, vừa có thông tin theo thời gian.

Sử dụng các ký hiệu tượng tự như trong bài thí nghiệm ngẫu nhiên, ta có Y là kết quả chịu tác động của một chính sách công (ví dụ như việc làm, thu nhập, điểm thi trắc nghiệm của học sinh,...); D là biến giả xác định một đối tượng có thuộc diện chi phối của chính sách công đó hay không ($D = 1$ nếu đối tượng chịu chi phối của chính sách công; $D = 0$ nếu đối tượng không bị chi phối bởi chính sách công).

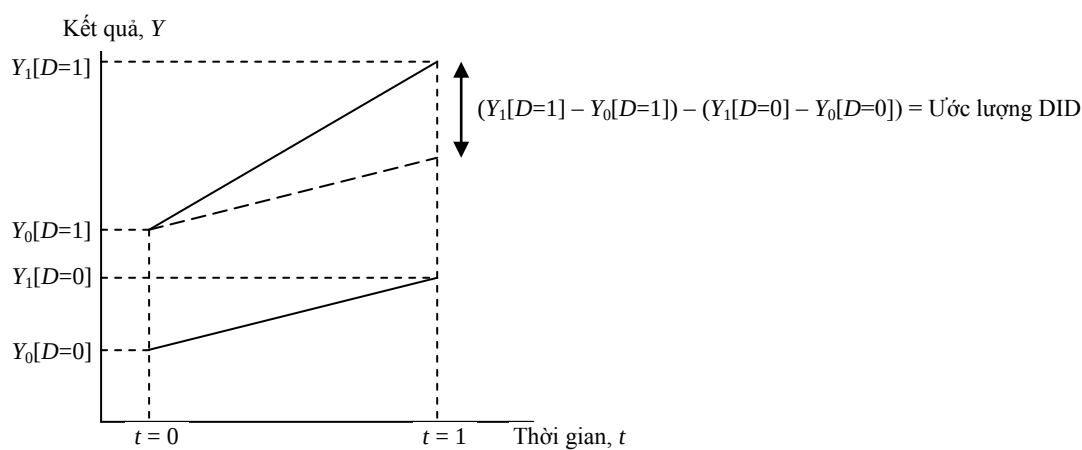
Về mặt thời gian, ta có Y_0 là kết quả tại thời điểm chưa thi hành chính sách công và Y_1 là kết quả tại thời điểm đã thi hành chính sách công. Vậy, đối với nhóm bị chi phối bởi chính sách công, kết quả thay đổi từ $Y_0[D = 1]$ thành $Y_1[D = 1]$; đối với nhóm không bị chi phối bởi chính sách công, kết quả thay đổi từ $Y_0[D = 0]$ thành $Y_1[D = 0]$.

Ta không thể coi tác động của chính sách công là khác biệt giữa kết quả sau và trước khi thi hành chính sách của nhóm bị chi phối (tức là, $Y_1[D = 1] - Y_0[D = 1]$). Lý do là một sự so sánh như vậy sẽ bị tác động bởi những biến động theo thời gian. Ví dụ như mặc dù trên thực tế chính sách không hề có tác động gì tới thu nhập nhưng theo thời gian thu nhập của người nông dân vẫn tăng lên và ta có $Y_1[D = 1] - Y_0[D = 1] > 0$.

Tương tự, ta cũng không thể coi tác động của chính sách công là khác biệt về kết quả sau khi thi hành chính sách giữa nhóm bị chi phối và nhóm không bị chi phối (tức là, $Y_1[D = 1] - Y_1[D = 0]$). Lý do là nhóm bị chi phối (nhóm xử lý) và nhóm không bị chi phối (nhóm kiểm soát) có thể khác nhau về một số đặc điểm cơ sở. Ví dụ, mặc dù trên thực tế chính sách có thể có tác động làm tăng thu nhập, nhưng vì trước khi thi hành chính sách thu nhập của nhóm kiểm soát đã cao hơn nhiều so với nhóm xử lý nên sau khi thi hành chính sách ta có $Y_1[D = 1] - Y_1[D = 0] < 0$.

Phương pháp DID tính tới hai khác biệt: khác biệt theo thời gian trước và sau khi thi hành chính sách và khác biệt chéo giữa nhóm xử lý và nhóm kiểm soát, và do vậy có tên gọi là khác biệt trong khác biệt.

Đồ thị dưới đây mô tả phương pháp này.



Theo thời gian kết quả của nhóm kiểm soát (ví dụ như thu nhập) thay đổi từ $Y_0[D = 0]$ thành $Y_1[D = 0]$. Vì nhóm kiểm soát không hề chịu chi phối của chính sách công, nên ta có thể coi $Y_1[D = 0] - Y_0[D = 0]$ là thay đổi thu nhập theo xu thế thời gian. Một giả định phải đưa ra để áp dụng phương DID là nếu như không có chính sách công thì theo thời gian thay đổi thu nhập của hai nhóm xử lý và kiểm soát sẽ là như nhau.

Vậy, nếu không có chính sách công thì thay đổi thu nhập của nhóm xử lý cũng sẽ là:

$$Y_1[D = 0] - Y_0[D = 0]$$

Nói một cách khác, nếu không có chính sách công thì thu nhập của nhóm xử lý vào thời điểm $t = 1$ sẽ là:

$$Y_0[D = 1] + (Y_1[D = 0] - Y_0[D = 0])$$

Vì có chính sách công nên thu nhập của nhóm xử lý vào thời điểm $t = 1$ trên thực tế là:

$$Y_1[D = 1]$$

Tác động của chính sách công là:

$$Y_1[D = 1] - \{Y_0[D = 1] + (Y_1[D = 0] - Y_0[D = 0])\}$$

$$= (Y_1[D = 1] - Y_0[D = 1]) - (Y_1[D = 0] - Y_0[D = 0])$$

Ta có thể tóm tắt ước lượng DID trong bảng sau:

	Trước khi thi hành chính sách, $t = 0$	Sau khi thi hành chính sách, $t = 1$	Khác biệt
--	--	--------------------------------------	-----------

Nhóm kiểm soát	$Y_0[D = 0]$	$Y_1[D = 0]$	$Y_1[D = 0] - Y_0[D = 0]$
Nhóm xử lý	$Y_0[D = 1]$	$Y_1[D = 1]$	$Y_1[D = 1] - Y_0[D = 1]$
Khác biệt trong khác biệt			$(Y_1[D = 1] - Y_0[D = 1]) - (Y_1[D = 0] - Y_0[D = 0])$

Lưu ý:

Giả định tối quan trọng của phương pháp DID là nếu như không có chính sách công thì ai nhóm xử lý và nhóm kiểm soát sẽ có cùng xu thế vận động theo thời gian. Điều này có thể đúng hay có thể sai trên thực tế. Giả định này có tên gọi là giả định song song (parallel assumption). Chỉ khi nào giả định này đúng thì ta mới áp dụng được DID.

Ví dụ, ta có thể nghiên cứu tác động của phương tiện viễn thông đến thu nhập của nông dân trồng cây ăn trái. Hai làng ở hai bờ kênh thuộc tỉnh Vĩnh Long. Các hộ trồng cây ăn trái ở một làng được kết nối điện thoại trong khi làng kia thì không có. Hai làng trồng cùng một loại cây ăn trái, chịu cùng tác động như nhau về khí hậu, cùng sự hỗ trợ của chính quyền,... Do vậy, hoàn toàn hợp lý khi giả định rằng nếu không có điện thoại thì theo thời gian thay đổi thu nhập bình quân hộ trồng cây ăn trái của hai làng sẽ như nhau. Ta có thể dùng phương pháp DID để kiểm chứng bằng số liệu thực tế xem có phải những hộ có điện thoại và do vậy có thông tin tốt hơn về thị trường sẽ có thu nhập tăng lên cao hơn so với các hộ không có điện thoại hay không.

Một ví dụ khác, một người có thể mong muốn sử dụng tình huống nông dân ở Bà Rịa-Vũng Tàu bị tác động bởi bão Durian trong khi nông dân ở Đồng Nai không bị để ước tác động của bão tới đời sống của nông dân bằng phương pháp DID. Tuy nhiên, có nhiều lý do để lập luận rằng cho dù không có bão thì xu hướng thay đổi thu nhập theo thời gian của nông dân ở Đồng Nai và Bà Rịa-Vũng Tàu vẫn khác nhau. Áp dụng phương pháp DID trong trường hợp này là không thích hợp.

Một cách để kiểm định giả định song song là thu thập thêm số liệu tại một thời điểm nữa trước khi thi hành chính sách. Như vậy, ta có hai điểm thời gian trước khi thi hành chính sách công ở đó cả hai nhóm xử lý và kiểm soát đều không chịu tác động của chính sách công. Dựa vào các số liệu này, ta có thể tính thay đổi kết quả của hai nhóm theo thời gian và kiểm định xem sự thay đổi này của hai nhóm có như nhau hay không.

Vậy, ngoài Y_0 , ta thu thập thêm số liệu vào thời điểm trước đó nữa, Y_{-1} .

Thay đổi kết quả từ thời điểm $t = -1$ đến $t = 0$ của nhóm kiểm soát là:

$$Y_0[D = 0] - Y_{-1}[D = 0]$$

Thay đổi kết quả từ thời điểm $t = -1$ đến $t = 0$ của nhóm xử lý là:

$$Y_0[D = 1] - Y_{-1}[D = 1]$$

Vì chưa hề có chính sách công trong thời điểm $t = -1$ cũng như $t = 0$, nên giả định song song mặc định rằng: $Y_0[D = 0] - Y_{-1}[D = 0] = Y_0[D = 1] - Y_{-1}[D = 1]$

2. Kết hợp phương pháp khác biệt trong khác biệt với hồi quy OLS

Ta có thể có được ước lượng DID bằng cách chạy hồi quy OLS sau đây:

$$Y = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 (D \cdot T) + \varepsilon$$

trong đó,

D là biến giả về nhóm xử lý/kiểm soát: $D = 1$ là nhóm xử lý và $D = 0$ là nhóm kiểm soát.

T là biến giả về thời gian: $T = 1$ là sau khi thi hành chính sách và $T = 0$ là trước khi thi hành.

$D \cdot T$ là biến tương tác của hai biến giả D và T .

Đối với nhóm kiểm soát trước khi thi hành chính sách, ta có $D = 0$ và $T = 0$.

$$E[Y_0[D = 0]] = \beta_0$$

Đối với nhóm xử lý trước khi thi hành chính sách, ta có $D = 1$ và $T = 0$.

$$E[Y_0[D = 1]] = \beta_0 + \beta_1$$

Đối với nhóm kiểm soát sau khi thi hành chính sách, ta có $D = 0$ và $T = 1$.

$$E[Y_1[D = 0]] = \beta_0 + \beta_2$$

Đối với nhóm xử lý sau khi thi hành chính sách, ta có $D = 1$ và $T = 1$.

$$E[Y_1[D = 1]] = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

Khác biệt trong khác biệt (DID)

$$\begin{aligned} &= (E[Y_1[D = 1]] - E[Y_0[D = 1]]) - (E[Y_1[D = 0]] - E[Y_0[D = 0]]) \\ &= [(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1)] - [(\beta_0 + \beta_2) - (\beta_0)] \\ &= \beta_3 \end{aligned}$$

Vậy, sau khi chạy hồi quy, ta sẽ có ước lượng tác động của chính sách công theo phương pháp khác biệt trong khác biệt là: $\hat{\beta}_3$.

Ta biết rằng, biến kết quả Y còn có thể chịu tác động của nhiều yếu tố khác, ngoài tác động của chính sách và thời gian. Vì vậy, có thể hữu ích nếu ta đưa thêm các biến giải thích X (ví như đặc điểm cá nhân và địa lý) vào mô hình hồi quy.

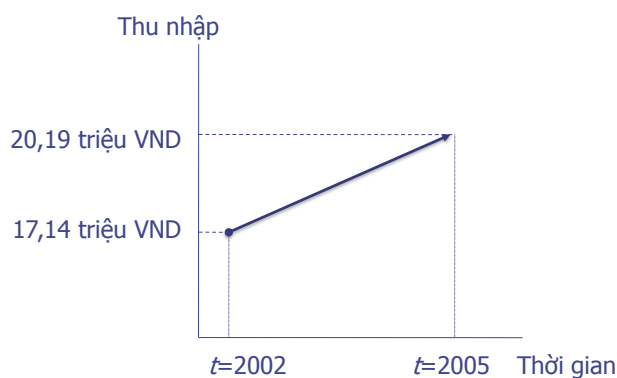
$$Y = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 (D \cdot T) + X' \gamma + \varepsilon$$

$\hat{\beta}_3$ vẫn là ước lượng khác biệt trong khác biệt.

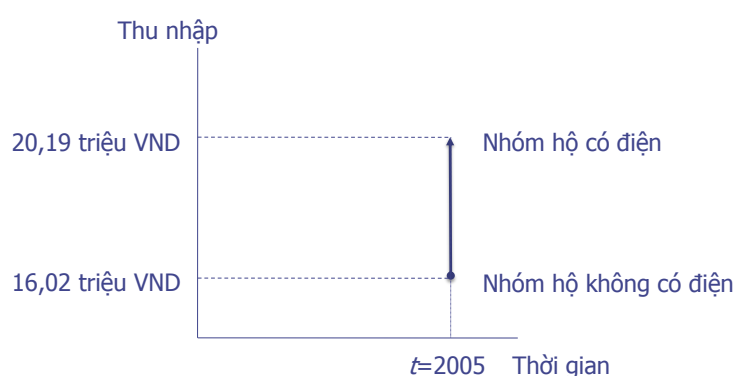
3. Ví dụ minh họa 1

Chương trình điện khí hóa nông thôn do Ngân hàng Thế giới tài trợ diễn ra từ năm 2002 đến 2005. Vào năm 2002, 1120 hộ gia đình không có điện lưới. Vào năm 2005, 828 hộ tiếp tục không có điện lưới, trong khi 292 hộ có điện lưới. Không có sự khác biệt đáng kể giữa nhóm các hộ không có điện lưới trong cả giai đoạn 2002-05 và nhóm hộ đến năm 2005 có điện. Các hộ đều làm nông nghiệp và nói chung là nghèo. Câu hỏi đặt ra là chương trình điện khí hóa nông thôn có giúp gia tăng thu nhập của hộ gia đình được kết nối với điện lưới hay không.

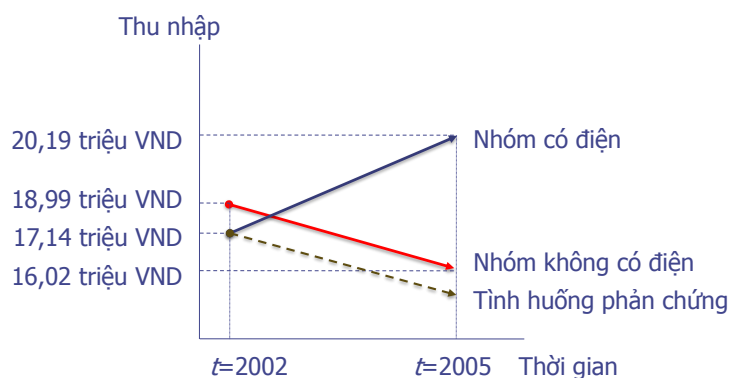
Thu nhập thay đổi theo thời gian: Thu nhập bình quân của 292 hộ không có điện lưới năm 2002, nhưng đến 2005 thì có.



Thu nhập thay đổi giữa nhóm: Thu nhập bình quân vào năm 2005 của nhóm hộ có điện so với nhóm hộ không có điện



Khác biệt trong khác biệt (DID)



Theo thời gian, nhóm kiểm soát có thu nhập bình quân giảm 2,96 triệu VND. Vậy, nếu không có chương trình điện khí hóa thì, thu nhập các hộ sẽ giảm đi 2,96 triệu VND. Khi có chương trình, các hộ được kết nối điện lưới có thu nhập tăng lên 3,04 triệu VND.

Tác động của chương trình là: $3,04 - (-2,96) = 6,00$ triệu VND.

	Thu nhập b/q năm 2002	Thu nhập b/q năm 2005	Khác biệt
Nhóm kiểm soát: không có điện lưới trong suốt 2002-05	18,99	16,02	-2,96
Nhóm xử lý: không có điện lưới trong 2002 và có điện trong 2005	17,14	20,19	3,04
Khác biệt trong khác biệt			6,00

4. Ví dụ minh họa 2

Đọc bài nghiên cứu:

Card, D. & A. B. Krueger (1994), “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,” *American Economic Review*, vol. 84, 772-793.

Lý thuyết kinh tế vi mô cổ điển cho rằng tăng mức lương tối thiểu sẽ làm giảm số lượng việc làm, đặc biệt là đối với các lao động cần kỹ năng giản đơn vốn có mức lương bằng hoặc chỉ cao hơn lương tối thiểu một chút ít. Card & Krueger (1994) sử dụng phương pháp khác biệt trong khác biệt và số liệu trong ngành cửa hàng ăn nhanh để nghiên cứu tác động của việc tăng lương tối thiểu ở Bang New Jersey đến việc làm.

Vào năm 1992, mức lương tối thiểu của Bang New Jersey được tăng từ 4,25 USD/giờ lên 5,05 USD/giờ, trong khi mức lương tối thiểu của miền Đông Bang Pennsylvania, giáp với New Jersey, vẫn được giữ nguyên ở mức 4,25 USD. Đây là tình huống rất hấp dẫn đối với phương pháp DID.

Phân tích số liệu

Vì việc tăng lương tối thiểu ở New Jersey không có ảnh hưởng tới các cửa hàng ăn nhanh ở Pennsylvania, nên các cửa hàng ăn nhanh ở miền Đông Pennsylvania được đưa vào nhóm kiểm soát, trong khi các cửa hàng ăn nhanh ở New Jersey thuộc nhóm xử lý.

Các bảng số liệu dưới đây so sánh hai nhóm về loại cửa hàng, lương, việc làm và một số đặc điểm khác. Ta thấy rằng phân phối các cửa hàng ăn nhanh ở hai nhóm khá giống nhau. Mức khác biệt ở tất cả các đặc điểm đều không có ý nghĩa thống kê. Trước khi tăng lương tối thiểu, các cửa hàng ăn nhanh ở Đông Pennsylvania và New Jersey tương tự nhau về lương khởi điểm, giờ mở cửa. Tuy nhiên, ta nhận thấy có khác biệt đáng kể về số việc làm và giá đồ ăn. Sau khi tăng lương tối thiểu, việc làm ở các cửa hàng ăn nhanh tại New Jersey tăng lên. Ở Pennsylvania, việc làm lại giảm xuống. Những đặc điểm khác, như giá đồ ăn và số giờ mưa cửa, không thay đổi nhiều ở cả nhóm kiểm soát lẫn xử lý.

Phân phối loại của hàng ăn nhanh (%)

	Nhóm kiểm soát Đông Pennsylvania	Nhóm xử lý New Jersey	Khác biệt
Burger King	0.443	0.411	-0.032 (0.062)
KFC	0.152	0.205	0.053 (0.046)
Roy Rogers	0.215	0.248	0.033 (0.052)
Wendy	0.190	0.136	-0.054 (0.048)
Cá thể	0.354	0.341	-0.013 (0.060)

Sai số chuẩn mạnh (tức đã hiệu chỉnh cho khả năng có phương sai của sai số thay đổi) là số trong ngoặc.

* Có ý nghĩa ở mức 10%; ** Có ý nghĩa ở mức 5%; *** Có ý nghĩa ở mức 1%.

Trước khi tăng lương tối thiểu

	Nhóm kiểm soát Đông Pennsylvania	Nhóm xử lý New Jersey	Khác biệt
Việc làm quy đổi tương đương	23.33	20.44	-2.89** (1.44)
Lương khởi điểm	4.63	4.61	-0.02 (0.04)
Tỷ lệ % có lương 4.25 USD	0.33	0.31	-0.02 (0.06)
Giá một suất ăn	3.04	3.35	0.31*** (0.08)
Số giờ mở cửa hàng ngày	14.52	14.41	-0.11 (0.37)

Sai số chuẩn mạnh (tức đã hiệu chỉnh cho khả năng có phương sai của sai số thay đổi) là số trong ngoặc.

* Có ý nghĩa ở mức 10%; ** Có ý nghĩa ở mức 5%; *** Có ý nghĩa ở mức 1%.

Trước khi tăng lương tối thiểu

	Nhóm kiểm soát Đông Pennsylvania	Nhóm xử lý New Jersey	Khác biệt
Việc làm quy đổi tương đương	21.17	21.03	-0.14 (1.07)
Lương khởi điểm	4.62	5.08	0.46*** (0.04)
Tỷ lệ % có lương 4.25 USD	0.25	0.00	-0.25 -
Tỷ lệ % có lương 5.05 USD	0.01	0.85	0.84*** (0.02)
Giá một suất ăn	3.03	3.41	0.39*** (0.08)
Số giờ mở cửa hàng ngày	14.65	14.42	-0.23 (0.36)

Sai số chuẩn mạnh (tức đã hiệu chỉnh cho khả năng có phương sai của sai số thay đổi) là số trong ngoặc.

* Có ý nghĩa ở mức 10%; ** Có ý nghĩa ở mức 5%; *** Có ý nghĩa ở mức 1%.

Ước lượng khác biệt trong khác biệt đơn giản

Do tác động của chu kỳ kinh tế tới việc làm ở New Jersey và Đông Pennsylvania là tương tự nhau, ta có thể giả định rằng việc làm ở các cửa hàng ăn nhanh ở hai bang sẽ thay đổi theo mức độ tương đương nhau nếu mức lương tối thiểu vẫn giữ nguyên. Với giả định này, ta có được ước lượng DID theo bảng dưới đây.

Ước lượng DID

Việc làm tương đương	Trước tăng lương tối thiểu	Sau tăng lương tối thiểu	Khác biệt
Nhóm kiểm soát Đông Pennsylvania	23,33	21.17	-2.16
Nhóm xử lý New Jersey	20,44	21.02	0.58
Khác biệt trong khác biệt			2,74

5. Ví dụ minh họa 3

Đọc bài nghiên cứu “Ước lượng suất sinh lợi của việc đi học ở Việt Nam: Phương pháp khác biệt trong khác biệt”.