

# Trường Chính sách Công và Quản lý Fulbright

Học kỳ Hè 2019  
Đánh giá Chính sách

## BÀI TẬP 2

### Ô nhiễm có làm giảm giá bất động sản không?\*

Ngày Phát: Thứ tư 10/07/2019

Ngày Nộp: Thứ tư 17/07/2019

## Giới thiệu

Bài tập này mô phỏng lại kết quả nghiên cứu “Does hazardous waste matter? Evidence from the housing market and the superfund program” của hai tác giả Greenstone và Gallagher (2008). Mục tiêu của nghiên cứu là ước lượng mức sẵn lòng chi trả cho chương trình cải thiện môi trường của chính quyền liên bang Hoa Kỳ có tên là Superfund. Chương trình này hỗ trợ tẩy rửa và cải thiện môi trường ở những khu vực ô nhiễm bị đánh giá là có rủi ro cao nhất. Nghiên cứu này đánh giá liệu giá bất động sản có biến động khi có chương trình can thiệp cải thiện môi trường.

Mô hình đánh giá thụ hưởng (hedonic valuation) thường được sử dụng để đánh giá giá trị của một tài sản. Mô hình này giả định rằng giá nhà chịu ảnh hưởng bởi đặc tính của căn nhà và các thuộc tính môi trường khu vực. Ví dụ, nếu ngôi nhà có thêm một phòng ngủ thì giá nhà sẽ tăng lên một lượng bằng với giá trị biên của việc có thêm một phòng ngủ, trong điều kiện các yếu tố khác không đổi. Trong trường hợp ngôi nhà có vị trí gần với một địa điểm gây ô nhiễm, chúng ta có thể kỳ vọng giá nhà tăng lên nếu có cách nào đó cải thiện môi trường ở khu vực. Trong bài tập này, chúng ta sẽ xem xét liệu giả thuyết này được có được kiểm chứng bằng số liệu thực tế.

Dữ liệu sử dụng trong nghiên cứu này được quan sát ở cấp độ *dãy điều tra* (census tract, đơn vị hành chính nhỏ hơn cấp quận, county). Mỗi dãy điều tra có xấp xỉ 4,000 người và khoảng 1,000 căn nhà. Tổng cộng có khoảng 65,000 dãy điều tra trong thống kê dân số nhà ở của Hoa Kỳ. Dữ liệu về nhà ở được tổng hợp từ Tổng điều tra Nhà ở, thực

---

\*Bài tập này được thiết kế dựa trên nghiên cứu của Greenstone and Gallagher (2008) và bài giảng môn kinh tế lượng ARE 213 của Michael Anderson, University of California at Berkeley.

hiện 10 năm một lần, giai đoạn 1970-2000. Bài nghiên cứu tập trung làm rõ giá nhà trung bình biến động như thế nào giữa những dãy điều tra có địa điểm ô nhiễm được xử lý và không được xử lý.

Chương trình Superfund được áp dụng tại những khu vực ô nhiễm thỏa mãn tiêu chí đánh giá hai bước, bao gồm đánh giá khu vực và đánh giá mẫu phẩm (không khí và đất). Nếu một khu vực bị cho là ô nhiễm nghiêm trọng, các nhà khoa học sẽ áp dụng một chỉ số - gọi là chỉ số xếp hạng rủi ro (Hazardous Ranking System - HRS - index). Số lượng địa điểm phải đánh giá HRS chỉ chiếm khoảng 1% tổng số tất cả các địa điểm có ô nhiễm. Nếu một địa điểm có chỉ số HRS vượt ngưỡng 28.5 (thang đo 100) sẽ được đưa vào danh sách ưu tiên (National Priorities List). Các khu vực nằm trong danh sách ưu tiên bắt buộc phải cải thiện môi trường. Chương trình Superfund bắt đầu năm 1980. Đánh giá theo hệ thống HRS bắt đầu năm 1982. Những địa điểm có chỉ số HRS lớn hơn 28.5 được đưa vào danh sách NPL năm 1983.

Việc áp dụng hệ thống HRS cho phép áp dụng thiết kế hồi quy gián đoạn để đánh giá tác động của chương trình Superfund. Học viên được cung cấp 4 bộ dữ liệu. Ba bộ dữ liệu (allsites.dta, allcovariates.dta, và sitecovariates.dta) được sử dụng trong phần I để phân tích bối cảnh nghiên cứu. Bộ dữ liệu 2miledata.dta sẽ được sử dụng cho nội dung chính bắt đầu từ phần II.

## Phần I

Phần này yêu cầu học viên khám phá quan hệ giữa giá nhà năm 2000 và liệu dãy điều tra có địa điểm gây ô nhiễm được đặt trong danh sách NPL trước năm 2000.

- (a) Sử dụng bộ dữ liệu allsites.dta. Đây là bộ dữ liệu đã được xử lý và loại bỏ các quan sát không có giá nhà. Trước hết, hồi quy giá nhà năm 2000 theo biến giải thích liệu dãy điều tra có địa điểm nằm trong danh sách NPL 2000. Sử dụng robust standard errors trong tất cả các mô hình hồi quy. Sau đó, lần lượt sử dụng thêm biến kiểm soát là giá nhà năm 1980 (1); đưa thêm các đặc tính nhà (2); và các biến liên quan đến kinh tế và nhân khẩu học (3). Cuối cùng, đưa thêm biến kiểm soát là nhân tố cố định ở cấp độ tiểu bang (state-level fixed effects). Một số biến đã được lựa chọn trong Bảng 1. Học viên có thể cân nhắc đưa thêm các biến khác. Giải thích ngắn gọn các mô hình trên. Với điều kiện gì thì tham số ước lượng của biến NPL 2000 không chệch?
- (b) Chúng ta sẽ so sánh các biến kiểm soát của hai nhóm hưởng lợi và đối chứng. Trước tiên, sử dụng bộ dữ liệu allcovariates.dta để so sánh sự cân bằng của các biến kiểm soát sử dụng trong câu hỏi [I.a] của các dãy điều tra nằm trong danh sách NPL 2000 và không nằm trong danh sách này. Bộ dữ liệu này chứa tất cả các quan sát, kể cả

Bảng 1: Các biến trong bộ dữ liệu allsites.dta

Variable	Definition
lnmdvalhs0	log median housing values 2000
lnmeanhs8	log mean housing values 1980
npl2000	NPL Before 2000
firestoveheat80	% Units Fire Stove Port No Heat 80
noaircond80	% Housing Units No AC 80
nofullkitchen80	% Housing Units No Full Kitchen 80
zerofullbath80	% Housing Units No Full Bath 80
bedrms0_80occ	percent owner-occupied housing units 0 bedrooms 1980
bedrms1_80occ	percent owner-occupied housing units 1 bedrooms 1980
shrbk8	Share of Pop Black 1980
shrhsp8	Share of Pop Hispanic 1980
shrfor8	Share of Pop Forgn Born 1980
smhse8	% Pop in Same House 5 Yrs Ago 80
unemprt8	% Pop > 16 Unemployed 80
povrat8	% Pop < Poverty Line 80
welfare8	% HHolds Pub Ass Recipients 80
avhhin8	average HHold Income 80

các dãy dữ liệu không có giá nhà. Sau đó, sử dụng bộ dữ liệu sitecovariates.dta để so sánh sự cân bằng của các biến kiểm soát ở những khu vực có đánh giá HRS. Bộ dữ liệu này chỉ chứa các quan sát có đánh giá HRS. Cụ thể, so sánh sự cân bằng của nhóm có chỉ số HRS < 28.5 với nhóm có HRS > 28.5. Sau đó, so sánh các nhóm trong khoảng dữ liệu có HRS giới hạn từ 16.5 đến 40.5. Từ ba so sánh trên, bạn rút ra được kết luận gì?

## Phần II

Học viên sẽ thực hành xây dựng thiết kế hồi quy gián đoạn để ước lượng tác động của chương trình Superfund lên giá nhà. Chúng ta sẽ sử dụng bộ dữ liệu 2miledata.dta. Đây là bộ dữ liệu tổng hợp với giá nhà và các đặc tính trung bình trong bán kính 2 dặm (3.62km) từ tâm điểm là khu vực bị ô nhiễm.

- (a) Để sử dụng chỉ số HRS làm biến can thiệp trong thiết kế RDD, chúng ta cần giả định gì? Xem xét và giải thích một trong những nhân tố sau tác động như thế nào đến sự hợp lý của thiết kế RDD:
  - i. Cục quản trị môi trường (EPA) cho rằng ngưỡng 28.5 được lựa chọn vì nó hạn chế số lượng khu vực cần xử lý trong giới hạn có thể quản lý được.

- ii. Không ai trong số những chuyên gia phân tích và đánh giá chỉ số HRS năm 1982 biết trước được ngưỡng can thiệp.
  - iii. Tài liệu của EPA nhấn mạnh rằng chỉ số HRS là một tiêu chí đánh giá không hoàn hảo.
  - iv. Sau năm 1982 thì ngưỡng can thiệp được giữ nguyên ở mức 28.5.
- (b) Các nhân tố trên ảnh hưởng như thế nào đến hiệu lực ngoại vi (external validity) của nghiên cứu sử dụng chỉ số HRS đo năm 1982? và hiệu lực ngoại vi của nghiên cứu sử dụng chỉ số HRS đo trước năm 2000?
- (c) Vẽ đồ thị phân phối của chỉ số  $hrs_{1982}$ . Sau đó, ước lượng hàm hồi quy sau để xác định liệu có tác động nào của chương trình can thiệp tại ngưỡng can thiệp  $HRS = 28.5$  hay không. Cụ thể, học viên sẽ ước lượng mô hình sau bằng thiết kế Sharp RDD:

$$\log(Y) = \beta_0 + \beta D + \beta_2 \tilde{X} + \beta_3 D * \tilde{X} + u \tag{1}$$

với  $Y$  là biến giá nhà trung bình hoặc trung vị năm 2000,  $D = 1$  nếu chỉ số HRS đo năm 1982 lớn hơn 28.5, và  $\tilde{X} = hrs_{82} - 28.5$ . Ước lượng mô hình (1) với toàn bộ dữ liệu, sau đó kiểm định độ nhạy với nhiều bandwidth khác nhau xung quanh ngưỡng can thiệp. Từ các kết quả trên, bạn có nhận xét gì về tác động của chương trình Superfund và thiết kế SRDD?

### Phần III

Học viên sẽ sử dụng thiết kế Fuzzy RDD để ước lượng tác động của chương trình Superfund.

- (a) Trước tiên, học viên cần giải thích tại sao thiết kế SRDD có thể không phù hợp trong tình huống nghiên cứu này.
- (b) Để xây dựng thiết kế FRDD, học viên cần giải thích tại sao ngưỡng can thiệp  $hrs_{82} = 28.5$  có thể được sử dụng làm biến công cụ cho việc được lựa chọn đưa vào danh sách NPL 2000. Mô tả cách thức ước lượng tác động của chương trình can thiệp bằng thiết kế FRDD với hồi quy 2SLS và biến công cụ.
- (c) Ước lượng giai đoạn 1 của hồi quy 2SLS/IV đã mô tả trong câu hỏi [III.b]. Cụ thể, học viên sẽ ước lượng mô hình sau đây:

$$npl2000 = \pi_0 + \pi_1 Z + \sum_j \pi_j X_j + u \tag{2}$$

với  $Z = 1$  nếu  $hrs\_82 > 28.5$ , và  $X_j$  là các biến kiểm soát khác đã sử dụng ở phần I. Học viên cũng có thể đưa thêm bình phương của biến  $hrs\_82$  vào phương trình (2) (lý do kiểm tra cấu trúc hàm này là gì?). Đầu tiên, học viên ước lượng phương trình (2) với toàn bộ dữ liệu. Sau đó, giới hạn mẫu ở các quan sát có chỉ số HRS nằm trong khoảng 16.5 - 40.5. Diễn giải kết quả thu được.

- (d) Vẽ đồ thị scatter plot của biến nằm trong danh sách NPL 2000 ( $npl2000$ ) theo biến chỉ số HRS 1982 ( $hrs\_82$ ). Diễn giải đồ thị, sự cần thiết của thiết kế FRDD, và sự phù hợp của biến công cụ.
- (e) Vẽ đồ thị scatter plot của biến giá nhà trung bình năm 1980 ( $meanhs8$ ) theo biến chỉ số HRS 1982 ( $hrs\_82$ ). Diễn giải đồ thị và sự phù hợp của biến công cụ.

## Phần IV

Học viên sẽ ước lượng giai đoạn 2 của thiết kế FRDD nêu ở phần III. Cụ thể, học viên sẽ ước lượng mô hình sau:

$$\log(Y) = \beta_0 + \beta_1 \widehat{npl2000} + \sum_j \beta_j X_j + \varepsilon \quad (3)$$

với  $X_j$  là các biến kiểm soát khác đã sử dụng trong bài. Học viên có thể ước lượng phương trình (2) và (3) tuần tự bằng OLS, hoặc ước lượng đồng thời bằng 2SLS sẽ cho kết quả chính xác hơn. Diễn giải ý nghĩa của kết quả và so sánh với nghiên cứu gốc.

## Phần V

Tổng kết các phương pháp đã học để thiết lập quan hệ nhân quả, ưu nhược điểm, tình huống áp dụng, tác động ước lượng được, và khả năng ứng dụng. Điền thông tin vào bảng 2.

## Tài liệu tham khảo

Greenstone, Michael, and Justin Gallagher 2008. Does hazardous waste matter? Evidence from the housing market and the superfund program. *The Quarterly Journal of Economics*, 951-1003.

Bảng 2: Tổng kết các phương pháp đã học

	RCT	Regression Adjustments	PSM	DiD	SRDD	FRDD/IV
Giả định						
Dữ liệu						
Tác động						
Ưu điểm						
Nhược điểm						