

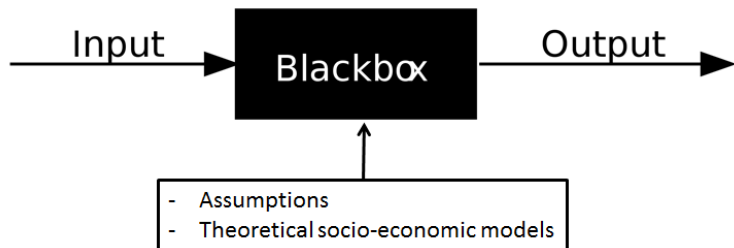
Nhập môn Kinh tế lượng (Introduction to Econometrics)

Lê Việt Phú
Trường Chính sách Công và Quản lý Fulbright

Ngày 24 tháng 11 năm 2018

Mục đích của môn học

- ▶ Hiểu bản chất của các mô hình kinh tế lượng căn bản.
- ▶ Sử dụng Stata để tiến hành các phân tích định lượng.
- ▶ Diễn giải, phân tích, và phê phán các kết quả nghiên cứu thực nghiệm.



Học kinh tế lượng để làm gì?

Để xác lập và lượng hóa các mối quan hệ kinh tế-xã hội-môi trường sử dụng trong nghiên cứu kinh tế và phân tích chính sách. Ví dụ:

- ▶ Tác động của việc đi học đến thu nhập như thế nào?
- ▶ Chương trình xóa đói giảm nghèo có giúp tăng thu nhập của người dân không?
- ▶ Tham nhũng có thực sự cản trở doanh nghiệp phát triển hay không?
- ▶ Biến đổi khí hậu có ảnh hưởng như thế nào đến năng suất mùa màng?
- ▶ Nhân tố nào ảnh hưởng đến hành vi sử dụng phương tiện đi lại (xe buýt, xe máy, xe đạp, đi bộ) của người dân ở các thành phố lớn?
- ▶ Tăng thuế xăng dầu từ 3000 đồng lên 8000 đồng/lít ảnh hưởng như thế nào đến nhu cầu đi lại của người dân?

Các ứng dụng khác của phân tích dữ liệu trong kinh tế và kinh doanh

- ▶ Cảnh báo giao dịch gian lận trong thanh toán online (fraudulent credit card transaction)
- ▶ Giới thiệu mua hàng (recommender system)
- ▶ Quảng cáo bán hàng
- ▶ Giá cả hàng hóa, bất động sản, Uber ridesharing
- ▶ Cho vay tín dụng
- ▶ Dự báo doanh nghiệp phá sản
- ▶ Đầu tư chứng khoán
- ▶ Y tế, chuẩn đoán bệnh tật theo triệu chứng
- ▶ Các chính sách kinh tế xã hội

Các kỹ năng cần thiết

- ▶ Thu thập, phân loại, tổng hợp, tổ chức, phân tích và diễn giải kết quả
- ▶ Xây dựng mô hình kinh tế lượng, diễn giải các giả định, và xử lý các vấn đề khi giả định bị vi phạm
- ▶ Trình bày kết quả và báo cáo

Giáo trình, phần mềm, tài liệu tham khảo

- ▶ Sách giáo trình: Introductory Econometrics: A Modern Approach của Jeffrey Wooldridge, có bản dịch tiếng Việt phiên bản 5th và sách gốc tiếng Anh.
- ▶ Phần mềm thống kê Stata, phiên bản 11 hoặc cao hơn.
- ▶ Phần mềm mã nguồn mở R và R-Studio cho phần khoa học dữ liệu.
- ▶ Tham khảo cách mô phỏng lại các ví dụ tại:
<http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge.html>

Yêu cầu của môn học

- ▶ 2 bài tập về nhà (20%), gồm các bài tập lý thuyết và thực hành trên máy tính.
- ▶ Hoàn thiện đề án môn học theo nhóm (40%).

Nội dung môn học

- ▶ Bài 1: Nhập môn kinh tế lượng (JW Ch1)
- ▶ Bài 2: Hồi quy đơn biến (JW Ch2)
- ▶ Bài 3: Hồi quy đa biến (JW Ch3)
- ▶ Bài 4-5: Hướng dẫn sử dụng Stata và khai thác các bộ dữ liệu kinh tế xã hội.
- ▶ Bài 6: Giả thuyết và kiểm định giả thuyết (JW Ch4).
- ▶ Bài 7: Cấu trúc hàm và lựa chọn mô hình (JW Ch6).
- ▶ Bài 8: Hồi quy với biến định tính (JW Ch7).
- ▶ Bài 9: Phương sai thay đổi và tự tương quan (JW Ch8).
- ▶ Bài 10: Chuẩn đoán và xử lý các vấn đề liên quan đến dạng hàm số và dữ liệu (JW Ch9).
- ▶ Bài 11-13: Giới thiệu khoa học dữ liệu và ứng dụng.

Phân tích định lượng (phân tích thực nghiệm) bao gồm những gì?

- ▶ Thu thập và xử lý dữ liệu: loại dữ liệu gì, đặc tính của dữ liệu
- ▶ Xây dựng mô hình dựa trên các lý thuyết kinh tế để thiết lập quan hệ giữa các biến số
- ▶ Lựa chọn phương pháp ước lượng dựa trên thiết kế nghiên cứu và khả năng cho phép của dữ liệu
- ▶ Diễn giải, kiểm định và giải thích kết quả
- ▶ Chuẩn đoán và xử lý các vấn đề có liên quan: kiểm tra tính vững của kết quả trong các điều kiện khác nhau
- ▶ Nâng cao: Sử dụng các mô hình phân tích, phân nhóm, giảm chiều dữ liệu, khám phá mối quan hệ tiềm ẩn giữa các nhóm biến số – nhằm mục đích diễn tả thực tại và khuyến nghị hoặc dự báo trong mô-đun khoa học dữ liệu.

Thu thập dữ liệu

- ▶ Dữ liệu thử nghiệm/thí nghiệm (experimental data): thu được trong nghiên cứu khoa học cơ bản dựa trên các điều kiện có kiểm soát chặt chẽ trong phòng thí nghiệm.
 - ▶ Ví dụ thử nghiệm một loại thuốc chữa bệnh nào đó. Hai nhóm bệnh nhân có tình trạng bệnh tật như nhau. Một nhóm cho điều trị theo phác đồ (gọi là nhóm điều trị), một nhóm không (gọi là nhóm đối chứng). Sau một thời gian, so sánh mức độ tiến triển của bệnh thì có thể xác định tác động của phác đồ điều trị đối với mỗi nhóm.
- ▶ Dữ liệu phi thử nghiệm: là các dữ liệu quan sát/điều tra (observational/surveyed data). Hầu hết các dữ liệu kinh tế xã hội đều là dữ liệu phi thử nghiệm.
 - ▶ Không đảm bảo được điều kiện các nhóm tương đồng nên khi phân tích chính sách sẽ gặp rất nhiều khó khăn để **thiết lập quan hệ nhân quả** giữa các biến số kinh tế.

Nhân quả (Causation) hay Tương quan (Correlation)?

- ▶ Nhân quả: A gây ra B. Ví dụ: tăng giá điện dẫn đến giảm tiêu thụ.
- ▶ Tương quan: A và B biến động cùng hoặc ngược chiều. Ví dụ: giá điện và thu nhập người dân trong 10 năm qua đều tăng. Nhưng không có quan hệ nhân quả, giá điện tăng không phải do thu nhập tăng hay ngược lại.
- ▶ Cảnh báo với quan hệ nhân quả: giá điện tăng chưa chắc là nguyên nhân dẫn đến việc giảm tiêu thụ điện. Có thể do các nguyên nhân khác như sự xuất hiện của các thiết bị tiết kiệm điện năng, hay thu nhập bị giảm, hay hành vi tiêu dùng thay đổi dẫn đến tiêu thụ điện giảm. Để xác lập quan hệ nhân quả phải kiểm soát và loại trừ các nguyên nhân khác có thể gây ra kết luận sai, dẫn đến đề xuất chính sách sai.

Thể nào là một thiết kế nghiên cứu hợp lý

Các nghiên cứu chính sách vững chắc cần phải dựa trên các thiết kế nghiên cứu (research design) hợp lý để bóc tách được tác động của chính sách can thiệp đến kết quả.

- ▶ Thiết kế nghiên cứu phải có mô hình lý thuyết vững chắc – thay vì chỉ sử dụng các **thuật toán** tính toán để tìm ra mô hình có khả năng dự báo cao nhất.
- ▶ Phải nhận định được các hạn chế của mô hình và dữ liệu, và đề xuất phương án xử lý nhằm đảm bảo độ vững của mô hình khi các giả định căn bản bị vi phạm.

Học phần kinh tế lượng 1 chỉ tập trung vào việc xây dựng mô hình vững chắc khi các giả định căn bản bị vi phạm. Học phần kinh tế lượng 2 tập trung vào các thiết kế nghiên cứu để bóc tách được quan hệ nhân quả trong mô hình kinh tế. Học phần kinh tế lượng 3 tập trung vào một thiết kế nghiên cứu duy nhất nhằm thiết lập quan hệ nhân quả là đánh giá thử nghiệm ngẫu nhiên có kiểm soát (RCT).

Cấu trúc của học phần định lượng

Fall Semester

- Ordinary Least Squares
- Functional forms
- Qualitative and Quantitative Variables
 - Heteroskedasticity
 - Diagnostics
- *Introduction to Data Science*



Spring Semester

- Regression with Panel Data
- Two-stage Least Squares
- Simultaneous Equations Model
- Model with Limited Dependent Variables
 - *Applied Data Science*



Summer Semester

- Impact Evaluation
- Randomized Control Trials (RCT)
- Propensity Score Matching
- Regression Discontinuity Design

Giới thiệu các cấu trúc dữ liệu phổ biến

- ▶ Dữ liệu chéo (cross-sectional data): thông tin của cá nhân, hộ gia đình, công ty, đơn vị địa lý tại một thời điểm.
- ▶ Dữ liệu chuỗi thời gian (time-series data): các quan sát được theo dõi lặp lại qua nhiều kỳ, tập trung vào đặc tính tần suất và độ trễ của tác động.
- ▶ Dữ liệu chéo gộp (chéo kết hợp - pooled cross-sectional data): quan sát qua nhiều kỳ nhưng của các cá thể khác nhau.
- ▶ Dữ liệu bảng (panel/longitudinal data): các cá thể kinh tế được theo dõi lặp lại ít nhất là 2 kỳ, tập trung vào đặc tính thay đổi giữa các kỳ quan sát.
- ▶ Dữ liệu phi cấu trúc (big data, neural network, text data).

Bộ dữ liệu chéo WAGE1.dta về tiền lương bình quân theo giờ và đặc điểm cá nhân

Household	Var1	Var2	Var3	...
1
2
3

- ▶ Mỗi dòng dữ liệu là một quan sát (cá thể). Thứ tự dữ liệu không ảnh hưởng đến kết quả phân tích.
- ▶ Dễ phân tích nhất nhưng cũng gặp nhiều vấn đề nhất.
- ▶ *Câu hỏi: So sánh tiền lương giữa phụ nữ và nam giới, giữa người có gia đình và độc thân, người có trình độ giáo dục và số năm kinh nghiệm khác nhau.*

Dữ liệu chuỗi thời gian PHILLIPS.dta về tỷ lệ lạm phát và thất nghiệp ở Mỹ

Year	Var1	Var2	Var3	...
1
2
3

- ▶ Thứ tự của dữ liệu rất quan trọng. Cần lưu ý đến tần suất thu dữ liệu và tính chu kỳ.
- ▶ Khó phân tích nhất trong các loại dữ liệu.
- ▶ *Câu hỏi: vẽ đồ thị xu hướng biến động của tỷ lệ thất nghiệp và lạm phát theo thời gian. Dữ liệu trên có phù hợp với lý thuyết về đường Phillips không?*

Dữ liệu chéo gộp HPRICE3.dta

Household	Year	Var2	Var3
1	1
2	1
3	1
4	2
5	2
6	2

- ▶ Hai hoặc nhiều bộ dữ liệu chéo gộp lại thành một bộ dữ liệu lớn, có thêm biến thời gian để xác định thời điểm thu thập dữ liệu. Các quan sát không nhất thiết phải trùng lặp lại giữa các thời kỳ.
- ▶ Có thể cung cấp nhiều thông tin hơn dữ liệu chéo thuần túy, nhưng không mạnh như dữ liệu bảng.
- ▶ *Câu hỏi: các nhân tố nào ảnh hưởng đến giá nhà? và các nhân tố đó thay đổi theo thời gian như thế nào?*

Dữ liệu bảng CRIME2.dta

Household	Year	Var2	Var3
1	1
2	1
3	1
1	2
2	2
3	2

- ▶ Mỗi cá thể được thu dữ liệu trong ít nhất 2 kỳ, kỳ trước và kỳ sau.
- ▶ Cho phép thực hiện các phân tích tốt nhất trong các loại dữ liệu. Tuy nhiên yêu cầu cao về kỹ thuật.

Sử dụng bảng biểu thống kê mô tả và đồ thị để mô tả dữ liệu

- ▶ Các đại lượng về giá trị trung bình, trung vị, mode
- ▶ Các đại lượng về phân phối: Khoảng phân phối, các phân vị, độ lệch chuẩn, hệ số CV
- ▶ Độ lệch
- ▶ Đồng phương sai, hệ số tương quan
- ▶ Đồ thị phân phối, đồ thị điểm, đồ thị tương quan, đồ thị hộp (boxplot)...

Câu hỏi cuối ngày: Chỉ sử dụng các thống kê mô tả và đồ thị bảng biểu liệu bạn có thể nhận định gì về nguyên nhân gây ra tác động trong các bộ dữ liệu đã cung cấp?