

Học nhiều hơn chưa chắc đã tốt hơn:
Bằng chứng thực nghiệm sử dụng biến công cụ
trong cải cách giáo dục ở Việt Nam

Lê Việt Phú
Chương trình Giảng dạy Kinh tế Fulbright

Ngày 29 tháng 12 năm 2014

Table of contents

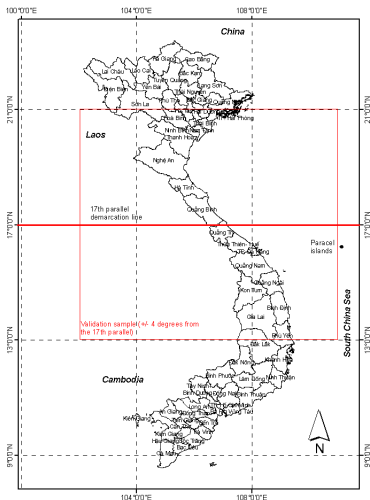
1. Lịch sử cải cách giáo dục và tỷ suất thu nhập của giáo dục
2. Phương pháp ước lượng tỷ suất thu nhập của học tập
3. Sử dụng biến công cụ để ước lượng tính nhân quả của cải cách giáo dục ở Việt Nam
4. Kết quả và kiểm tra độ vững của kết quả
5. Kết luận và hàm ý chính sách

Lịch sử cải cách giáo dục ở Việt Nam

Trong thế kỷ 20 Việt Nam đã thực hiện ba cuộc cải cách giáo dục năm 1950, 1956, và thập kỷ 80 sau mỗi sự kiện lịch sử lớn:

- ▶ Cải cách giáo dục 1950: giáo dục phổ thông giảm từ 11 năm của Pháp xuống còn 9 năm, chia làm 3 cấp học với 4, 3, 2 năm.
- ▶ Cải cách giáo dục 1956: Chuyển từ hệ 9 năm và 12 năm ở miền Bắc và vùng tạm giải phóng thành hệ 10 năm theo hệ của Liên Xô cũ, chia làm ba cấp với 4, 3, và 3 năm.
- ▶ Cải cách giáo dục lần thứ ba (thập kỷ 80): Chuyển từ hệ 10 năm lên 12 năm giống như hệ giáo dục của Mỹ đang thực hiện tại miền Nam.
- ▶ Khó đánh giá chất lượng do thiếu tiêu chí đánh giá và số liệu.

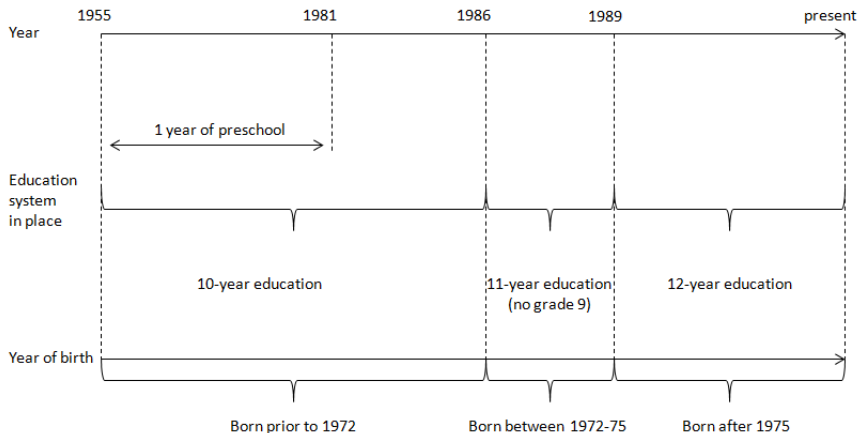
Cải cách giáo dục lần thứ ba trong thập kỷ 80



- ▶ Bối cảnh: tồn tại hai hệ thống giáo dục tại hai miền phía Bắc và Nam của vĩ tuyến 17 sau khi thống nhất đất nước. Miền Bắc theo hệ 10 năm của các nước XHCN cũ. Miền Nam vẫn sử dụng hệ 12 năm của Mỹ.
- ▶ Cải cách giáo dục lần thứ ba hợp nhất hai hệ thống thành hệ thống giáo dục phổ thông chung 12 năm cho cả nước.
- ▶ Hệ thống giáo dục mới học 5 năm tiểu học, 4 năm THCS, 3 năm THPT.

Thời gian thực hiện cải cách giáo dục lần thứ ba

Hệ thống giáo dục ở miền Bắc (dựa vào bảng điều tra VHLSS)



Đặc điểm cải cách giáo dục lần thứ ba

- ▶ Quá trình cải cách ngành giáo dục trái ngược với cải cách kinh tế năm 1986 bởi nó diễn ra rất chậm và bao gồm nhiều bước cải tiến nhỏ từ năm 1979 đến 1996 mới hoàn thành. Cải cách chương trình học và sách giáo khoa diễn ra rất thận trọng.
- ▶ Giáo dục tiểu học vẫn là miễn phí, nhưng trung học cơ sở và trung học phổ thông bắt đầu phải đóng học phí. Tỷ lệ đóng góp của tư nhân ngày càng tăng, đồng thời cho phép mở trường học bán công. Chi phí không thường xuyên và học thêm tràn lan.
- ▶ Đối với chương trình PTCS, thời gian học là 33 tuần học nửa ngày một năm, ngắn hơn ở một số nước đang phát triển.
- ▶ Phương pháp dạy vẫn tập trung vào vai trò người dạy, thiếu tương tác, làm tăng tính học thụ động và học thuộc lòng.

Nguồn dữ liệu và mẫu quan sát

- ▶ Bộ số liệu điều tra mức sống hộ gia đình VHLSS năm 2004, 2006, 2008, và 2010.
- ▶ Tập trung vào nhóm có trình độ học vấn cao nhất không quá cấp 3 (tốt nghiệp phổ thông trung học), được kỳ vọng là đối tượng dễ bị ảnh hưởng nhất bởi chính sách giáo dục, đồng thời cũng là nhóm được kỳ vọng có tỷ suất thu nhập của học hành cao nhất.
- ▶ Chỉ sử dụng nhóm đối tượng làm công ăn lương, phi nông nghiệp, độ tuổi từ 20 trở lên, không đang đi học.
- ▶ Có một số lý do quan trọng khác tại sao không đưa vào mẫu những người học trung cấp, đại học và cao hơn.

Thống kê mô tả dữ liệu

Miền bắc vĩ tuyến 17 - Trung bình và độ lệch chuẩn (SD)

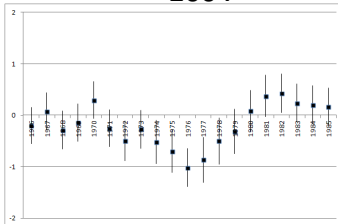
Variable	2004	2006	2008	2010
Age	35.05	35.07	35.71	35.95
SD	(10.98)	(11.43)	(11.13)	(11.24)
Annual Wage (x1000)	7,840.39	10,230.59	14,512.24	22,366.23
SD	(6,173.53)	(7,742.54)	(11,392.26)	(14,105.42)
EDUC	9.82	9.93	9.96	9.76
SD	(2.07)	(2.02)	(2.05)	(2.42)
EXP	7.67	7.68	7.50	—
SD	(7.56)	(7.79)	(7.18)	
Observations	2,074	2,191	2,261	2,307

* EXP: kinh nghiệm làm việc thực tế tại thời điểm khảo sát, khác với mô hình Mincer truyền thống là $EXP = Age - EDUC - b$. b là tuổi học tiểu học bắt buộc. Lý do: nhảy việc khá phổ biến, kinh nghiệm không liên quan đến công việc hiện tại.

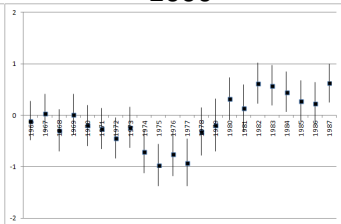
* Tỷ giá chính thức trong giai đoạn này tăng từ USD/VND 15,746 năm 2004 đến 18,613 năm 2010.

Tác động của năm sinh đến số năm học trung bình

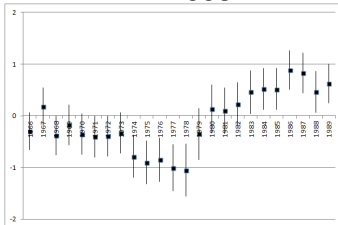
2004



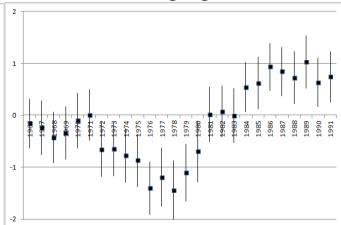
2006



2008



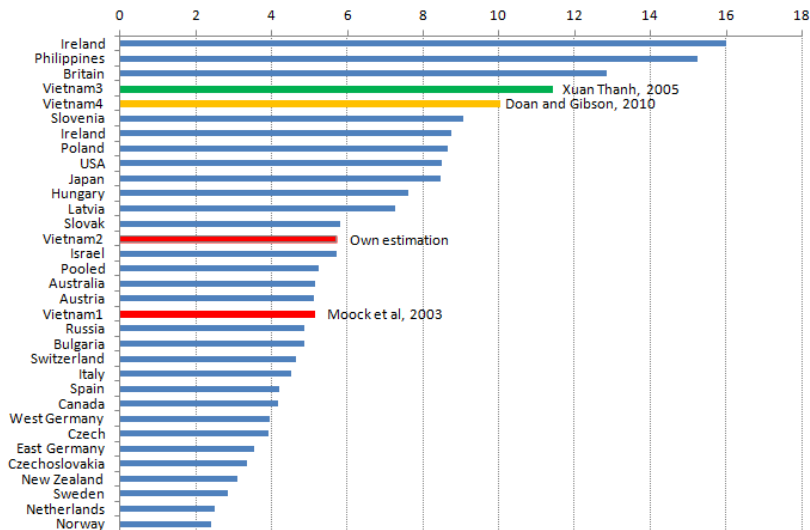
2010



Ước lượng điểm và 95% khoảng tin cậy của tác động của năm sinh lên số năm học bằng hồi quy theo biến giả.

Tỷ suất thu nhập trung bình cho mỗi năm học trên thế giới và ở Việt Nam

Tỷ suất thu nhập trung bình cho một năm học (%)



Phương pháp ước lượng tỷ suất thu nhập của học tập

Dựa vào phương trình Mincer (1974):

$$\log Y_i = \log Y_0 + \beta_1 \times EDUC_i + \beta_2 \times EXP_i + \beta_3 \times EXP_i^2 + \varepsilon_i$$

★ Biến phụ thuộc: $\log Y_i$ là logarit của thu nhập.

★ Biến giải thích:

- ▶ $EDUC_i$ là tổng số năm học. Có thể quy đổi thành các cấp học, vd. học đủ 12 năm theo hệ giáo dục mới thì tương đương với học đủ ba cấp tiểu học, THCS, THPT.
- ▶ EXP_i là số năm kinh nghiệm làm việc.
- ▶ Có thể đưa thêm các biến nhân khẩu học và ngành nghề vào làm biến giải thích.

Phương pháp ước lượng tỷ suất thu nhập của học tập

$$\log Y_i = \log Y_0 + \beta_1 \times EDUC_i + \beta_2 \times EXP_i + \beta_3 \times EXP_i^2 + \varepsilon_i$$

Ý nghĩa của các hệ số:

- ▶ β_1 : Tỷ suất thu nhập trung bình cho một năm học tăng thêm. Ví dụ $\beta_1 = 0.1$ thì mỗi năm học tăng thêm sẽ giúp thu nhập tăng thêm 10%.
- ▶ Kinh nghiệm làm việc có tác động phi tuyến đến thu nhập theo dạng đường Parabol ngược (β_3 âm). Kinh nghiệm có vai trò lớn khi cho người mới làm việc (số năm làm việc ít, EXP nhỏ) nhưng giảm dần với người đã đi làm. Mức thu nhập biên được tính theo công thức $\{\beta_2 + 2\beta_3 \times EXP_i\}$, và giảm dần đến một mức tối ưu $\{EXP^* = -\frac{\beta_2}{2\beta_3}\}$ là đỉnh của hàm parabol. Ngưỡng tối ưu được xác định khoảng 26 năm ở nhiều nước Đông Âu và ở Việt Nam.

Lý thuyết mô hình ước lượng tuyến tính sử dụng phương pháp bình phương tối thiểu - Ordinary Least Square (OLS)

$$y_i = \beta_0 + \beta_1 x_i^1 + \dots + \beta_n x_i^n + \varepsilon_i, \quad i = 1, \dots, N$$

Viết dưới dạng ma trận:

$$Y = X\beta + \varepsilon$$

Ước lượng bằng phương pháp OLS: $\sum_{i=1}^N \varepsilon_i^2 \Rightarrow \min$

$$\beta_{OLS} = [X'X]^{-1}X'Y = \frac{\text{Cov}[X, Y]}{\text{Var}[X]}$$

★ Các giả định để ước lượng OLS là BLUE (Best Linear Unbiased Estimator) - giả định Gauss-Markov:

1. $E[\varepsilon_i] = 0$
2. $\text{Var}[\varepsilon_i] = \sigma^2$
3. $\text{Cov}[\varepsilon_i, \varepsilon_j] = 0$
4. $\text{Cov}[X_i, \varepsilon_i] = 0$
5. Mỗi quan hệ X và Y là tuyến tính

★ Vi phạm giả định Gauss-Markov dẫn đến ước lượng chệch, không hiệu quả, và các kiểm định thống kê sai.

Những vấn đề gặp phải trong mô hình ước lượng truyền thống

- ★ Trong mô hình tỷ suất thu nhập của học tập, có hai vấn đề lớn: biến không được kiểm soát (omitted variables) có thể ảnh hưởng đến thu nhập, và sai số đo đạc (measurement errors).
 - ▶ Biến không được kiểm soát: tổ chất cá nhân ảnh hưởng đến quyết định học hành và số năm học. Tổ chất cá nhân cũng ảnh hưởng đến thu nhập. Nếu không kiểm soát được yếu tố tổ chất cá nhân, ước lượng OLS sẽ chệch (biased) do tương quan giữa biến phụ thuộc và biến dư.
 - ▶ Sai số đo lường: biến giải thích khó quan sát được chính xác. Số liệu được thống kê ít hơn so với thực tế khoảng 10-15% (Angrist and Krueger, Card). Khi đó ước lượng OLS cũng bị chệch.
 - ▶ Ngoài ra, còn có vấn đề biến nội sinh và cấu trúc hàm ước lượng.

Vấn đề biến không được kiểm soát (omitted variables bias)

$$\log Y_i = \log Y_0 + \beta_1 \times EDUC_i + \beta_2 \times EXP_i + \beta_3 \times EXP_i^2 + \gamma \times \mathbf{Ability}_i + \varepsilon_i$$

★ **Ability** đại diện cho tố chất cá nhân, không quan sát được. Khi này (xem Griliches, 1977),

$$E[b] = [X'X]^{-1}X'Y = [X'X]^{-1}X'[X\beta + \gamma \times \mathbf{Ability} + \varepsilon]$$

$$= \underbrace{[X'X]^{-1}X'X\beta}_{=\beta} + \underbrace{\gamma[X'X]^{-1}X'\mathbf{Ability}}_{=\gamma \frac{\text{Cov}[\mathbf{Ability}, EDUC]}{\text{Var}[EDUC]}} + \underbrace{[X'X]^{-1}X'\varepsilon}_{=0}$$

$$\Rightarrow E[b_1] = \beta_1 + \gamma \frac{\text{Cov}[\mathbf{Ability}, EDUC]}{\text{Var}[EDUC]}$$

★ Do tố chất được kỳ vọng có tương quan dương với số năm học, ước lượng tỷ suất thu nhập của việc học hành sẽ bị phóng đại so với thực tế (upward bias).

Vấn đề sai số đo lường (measurement errors)

- ★ Đo lường số năm đi học không chuẩn, chúng ta quan sát được **EDUC*** chứ không phải là biến chính xác EDUC:

$$EDUC = \mathbf{EDUC}^* + \mathbf{errors}$$

$$\log Y_i = \log Y_0 + \beta_1 \times EDUC_i + \beta_2 \times EXP_i + \beta_3 \times EXP_i^2 + \underbrace{\varepsilon_i - \beta_1 \times \mathbf{errors}_i}_{\text{composite residuals}}$$

- ★ Vi phạm giả định Gauss-Markov về không tương quan giữa biến giải thích *EDUC* và biến dư (composite residuals, $\varepsilon - \beta_1 \times \mathbf{errors}$)
⇒ ước lượng chệch.

Ước lượng tỷ suất thu nhập khi có sai số đo lường là:

$$E[b_1] = \beta_1 \frac{\text{Var}[EDUC]}{\text{Var}[EDUC] + \text{Var}[\mathbf{errors}]}$$

- ★ Ước lượng sẽ bị giảm thiểu (attenuation or downward bias) so với thực tế. Mức độ tùy thuộc vào tỷ lệ sai số (signal-to-noise ratio).

Vấn đề cấu trúc hàm ước lượng: tỷ suất thu nhập của việc đi học là tuyến tính hay phi tuyến?

$$\log Y_i = \log Y_0 + \beta_1 \times EDUC_i + \beta_2 \times EXP_i + \beta_3 \times EXP_i^2 + \varepsilon_i$$

- ★ Nếu giả định cấu trúc hàm trên là đúng và các giả định hồi quy OLS được thỏa mãn, khi đó ước lượng tỷ suất thu nhập β_1 là BLUE.
- ★ Một số nghiên cứu sử dụng hàm bậc hai của biến số năm đi học $\beta_1 \times EDUC_i + \beta_2 \times EDUC_i^2$, tương tự như biến số năm kinh nghiệm làm việc.
- ★ Làm sao để biết chính xác cấu trúc hàm ước lượng?
⇒ Hồi quy nội tại có thể giúp quan sát định dạng cấu trúc hàm tỷ suất thu nhập mà không cần bất kỳ giả định nào. Đây là ưu điểm của phương pháp phi tham số - nonparametric approach.

Sử dụng hồi quy phi tham số nội tại (nonparametric local regression) để loại bỏ mọi giả định về cấu trúc hàm

- ▶ Bước một, lọc biến động dư \tilde{Y} ra khỏi các nhân tố khác bằng cách bỏ biến học tập $EDUC_i$ ra khỏi hàm thu nhập:

$$\log Y_i = \log Y_0 + \beta_1 \times \cancel{EDUC_i} + \beta_2 \times EXP_i + \beta_3 \times EXP_i^2 + \tilde{Y}_i$$

Khi này, \tilde{Y} sẽ chứa biến động thu nhập do học tập và các ảnh hưởng khác không được kiểm soát.

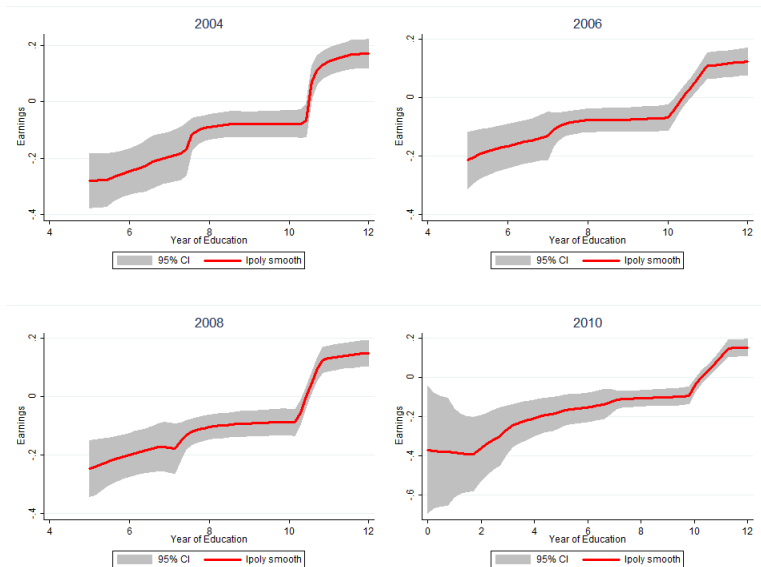
- ▶ Bước hai, ước lượng hàm thu nhập \tilde{Y} theo một biến giải thích duy nhất là số năm học:

$$\tilde{Y}_i = g(EDUC_i) + \varepsilon_i$$

$g(\cdot)$ là hàm đa thức (polynomial) tùy ý, được ước lượng cho mỗi khoảng giá trị (neighborhood) xung quanh mỗi quan sát.

- ▶ Do đặc tính phi tham số, chúng ta chỉ dùng đồ thị để giải thích kết quả. Phương pháp này còn có tên là LOESS hay LOWESS (locally weighted scatterplot smoothing).

Bằng chứng cho thấy tác động phi tuyến của năm học đến tỷ suất thu nhập



Cách xử lý đối với vấn đề biến không được kiểm soát và sai số đo lường trong nghiên cứu giáo dục

Nếu không xử lý được hai vấn đề kể trên, ước lượng OLS sẽ bị chệch với mức độ không xác định được.

- ▶ Đối với tổ chất cá nhân: đưa các biến đại diện (proxy variables) như IQ hay điểm số thi cử. Tuy nhiên hầu hết các bộ dữ liệu không có biến này.
- ▶ Diff-in-Diff: Sử dụng số liệu bảng với quan sát lặp lại. Những nhân tố không thay đổi giữa hai kỳ quan sát sẽ bị loại bỏ. Khác biệt về thu nhập và giáo dục giữa hai kỳ quan sát có thể chỉ do thay đổi giáo dục tạo ra. Tuy nhiên khả năng áp dụng phương pháp này rất hạn chế vì chỉ áp dụng được đối với nhóm đối tượng vừa làm vừa học, thường có mức độ học vấn thấp và có tỷ suất thu nhập từ học hành cao, dẫn đến vấn đề lựa chọn mẫu (selection bias). Khả năng áp dụng phương pháp này khá hạn chế.

Một cách khác nữa là sử dụng cặp sinh đôi. Các cặp sinh đôi thường có chung nhiều đặc điểm gia đình, tổ chất, do đó khác biệt về thu nhập có thể do khác biệt về giáo dục tạo ra.

Xử lý vấn đề biến không kiểm soát được và sai số đo lường trong nghiên cứu giáo dục

Sử dụng biến công cụ có liên quan đến số năm học nhưng không ảnh hưởng trực tiếp đến thu nhập. Một số trường hợp sử dụng biến công cụ nổi tiếng:

- ▶ Tháng và quý sinh có thể liên quan đến số năm học. Ở Mỹ có luật yêu cầu học sinh phải học đến một độ tuổi nhất định dựa vào tháng và năm sinh. Những người sinh sớm trong năm có thể được ra trường sớm hơn, do đó thời gian học ngắn hơn người sinh muộn. Tháng hay quý sinh được kỳ vọng không liên quan đến thu nhập.
Việt Nam không có quy định về độ tuổi, tuy nhiên có thể xảy ra trường hợp là các bạn sinh vào đầu năm (ví dụ tháng 1) được bố mẹ cho đi học cùng với các bạn sinh năm trước. Nếu có dữ liệu cụ thể có thể dùng biến công cụ là ngày sinh rơi vào tuần đầu của năm.
- ▶ Trình độ học vấn của vợ hoặc chồng dựa trên giả định hôn nhân thường xảy ra giữa các cặp vợ chồng không kém xa nhau về học vấn. Hoặc trình độ học vấn của cha mẹ cũng có thể được dùng làm biến công cụ cho số năm học của con cái. Trình độ học vấn của người thân cũng được kỳ vọng không liên quan đến thu nhập của người đó.

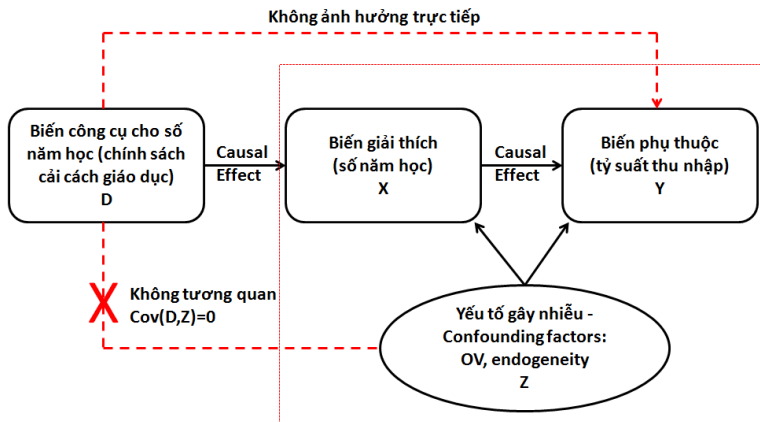
Tóm tắt phần đầu

- ▶ Tỷ suất thu nhập trung bình ở Việt Nam ngang với thế giới.
- ▶ Tỷ suất thay đổi qua các cấp học khác nhau, với tỷ suất cao nhất cho tiểu học và trung học phổ thông, thấp nhất cho trung học cơ sở.
- ▶ **Trong cùng một cấp học thì có vẻ như là thời gian học không ảnh hưởng đến thu nhập. Vd: tốt nghiệp PTTH thì 10 hay 12 năm, hoặc tốt nghiệp PTCS thì 8 hay 9 năm là như nhau.**
⇒ Bằng chứng về giá trị bằng cấp rất rõ rệt. Tiếp theo chúng ta sẽ sử dụng mô hình hồi quy để xác minh liệu việc kéo dài thời gian học một năm nhưng vẫn giữ nguyên cơ cấu bằng cấp liệu có tạo ra khác biệt gì cho thu nhập không.

Phần tiếp theo: dùng biến công cụ để xác định tính nhân quả của thời gian học lớp 9 và tỷ suất thu nhập

- ▶ Câu hỏi thường gặp: Tại sao phải dùng biến công cụ? Thay vì dùng những phương pháp phức tạp, tại sao chúng ta lại không lọc những người học lớp 9 và những người không học rồi dùng biến giả (dummy) đại diện cho nhóm bị ảnh hưởng và hồi quy với biến giả đó?
 - ▶ Câu trả lời là chúng ta không biết chính xác ai học lớp 9. Trong bảng câu hỏi chúng ta biết người học có tốt nghiệp PTTH hay không, từ 10 đến 12 năm. Một số người học đúng nên phải học thêm một năm. Một số người học sớm hơn một năm nên không phải học lớp 9. Một số người học muộn nên phải học thêm một năm.
 - ▶ Nếu có biết, việc học thêm lớp 9 cũng là biến nội sinh và cũng không tránh được vấn đề biến không kiểm soát được dẫn đến ước lượng chệch.

Khung phân tích



- ▶ Dùng biến công cụ để loại ảnh hưởng của các yếu tố gây nhiễu.
- ▶ Yêu cầu biến công cụ là ngoại sinh, có tương quan với số năm học nhưng không ảnh hưởng trực tiếp đến thu nhập.

Đánh giá hai điều kiện để sử dụng cải cách giáo dục làm biến công cụ cho số năm đi học

- ▶ Tính ngẫu nhiên (random assignment) của biến công cụ là thời điểm thực hiện chính sách cải cách giáo dục:
 - ▶ Không ai biết trước được sẽ có cải cách vào thời điểm đó để tránh bằng cách xin đi học sớm, hay di cư từ vùng này sang vùng khác qua vĩ tuyến 17 để tránh phải học lớp 9.
 - ▶ Và cũng không có lý do nào cho thấy người dân tìm cách học hay không học lớp 9 bằng cách xin đi học sớm, hay chuyển từ vùng này sang vùng khác để tránh học lớp 9.

Đánh giá hai điều kiện để sử dụng cải cách giáo dục làm biến công cụ cho số năm đi học

- ▶ Điều kiện loại trừ (exclusion restriction): học chương trình cải cách hay chương trình trước cải cách không tác động trực tiếp đến thu nhập.
 - ▶ Đúng khi cấu trúc chương trình học chưa thay đổi nhiều, đặc biệt những người sinh vào năm đệm giữa hai hệ thống giáo dục, khi sách giáo khoa và chương trình được giữ nguyên. Thay đổi duy nhất là thêm thời gian học. Nhóm bị ảnh hưởng là nhóm sinh ra quanh thời điểm thống nhất đất nước (72-78), và bắt đầu học lớp 9 năm 1990.
 - ▶ Không áp dụng được cho những người học sau này khi chương trình học đã thay đổi.

Dùng biến công cụ cho số năm học

2 biến công cụ được sử dụng:

- ▶ Biến công cụ thứ nhất là năm sinh, chỉ dùng cho mẫu quan sát ở phía bắc vĩ tuyến 17:

$$\mathbf{D}_1 = \begin{cases} 1 & 1972 \leq \text{year of birth} \leq 1978 \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Biến công cụ mở rộng là tương tác giữa biến trạng thái cho vĩ tuyến 17 và năm sinh, dùng cho số liệu cả nước:

$$\mathbf{N} = \begin{cases} 1 & \text{for Provinces north of the 17th parallel line} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{D}_2 = \mathbf{D}_1 \times \mathbf{N}$$

Sử dụng biến công cụ cho số năm học trong hồi quy hai bước - 2SLS

$$EDUC_i = \alpha_0 + \alpha_1 D + \alpha_2 Exp_i + \alpha_3 Exp_i^2 + X_i \alpha_4 + \eta_i$$

$$\log Y_i = \log Y_0 + \beta_1 \times \widehat{EDUC}_i + \beta_2 Exp_i + \beta_3 Exp_i^2 + X_i \beta_4 + \varepsilon_i$$

- ▶ Bước một, ước lượng tác động của cải cách lên số năm học. Trong đó D là biến công cụ, X là vector biến ngoại sinh khác, η_i là biến dư.
- ▶ Trong bước hai, ước lượng tác động của số năm học lên thu nhập qua biến công cụ \widehat{EDUC}_i trong mô hình thu nhập Mincer.

Kết quả ước lượng chính

Bước một: hồi quy số năm học dựa vào biến công cụ

Survey → Variables	2004 [1] [2]	2006 [1] [2]	2008 [1] [2]	2010 [†] [1] [2]				
INSTRUMENTS								
Cohort 1972-78	-.2085 (.1847)	-.3511*** (.1305)	-.5836*** (.1908)	-.7550*** (.1916)				
Cohort 1972-78 interacts North of the 17th Parallel		-.3085** (.1279)	-.5393*** (.1904)	-.7288*** (.2041)				
SE								
CONTROLS								
North of the 17th Parallel		1.0415*** (.1482)	1.1074*** (.1366)	1.9869*** (.1950)				
SE								
Experience	.0177 (.0206)	-.0117 (.0182)	-.0145 (.0249)	-.0397** (.0195)	-.0434** (.0170)	-.0708*** (.0141)	-.0883** (.0403)	-.1135*** (.0315)
SE								
Experience-squared	.0002 (.0007)	.0018*** (.0007)	.0008 (.0008)	.0019*** (.0007)	.0017** (.0007)	.0029*** (.0006)	.0008 (.0005)	.0010*** (.0004)
SE								
Constant	9.6951*** (.1095)	8.7773*** (.1397)	10.0045*** (.1180)	9.0157*** (.1500)	10.2035*** (.0877)	9.1555*** (.1149)	11.9343*** (.6926)	10.4926*** (.5465)
SE								
Observations	2,074	4,532	2,191	4,777	2,261	4,838	2,307	5,671

* Cột [1] tương ứng với mẫu quan sát ở phía Bắc vĩ tuyến 17. Cột [2] sử dụng số liệu cả nước.

* [*], [**], [***] thể hiện tham số có ý nghĩa thống kê ở mức 90, 95, và 99%.

* Mẫu quan sát bao gồm những người làm công ăn lương, từ 20-70 tuổi.

Giải thích kết quả

Hồi quy bước một cho kết quả như dự đoán:

- ▶ Sinh ra vào thời điểm cải cách ở miền Bắc có thời gian học trung bình thấp hơn từ 0.35-0.75 năm và có ý nghĩa thống kê trong ba mẫu quan sát 2006, 2008, và 2010.
- ▶ Có khác biệt giữa số năm học trung bình ở hai vùng miền: phía Bắc số năm học trung bình cao hơn khoảng 1-2 năm so với phía Nam, mặc dù học trong hệ thống có thời gian học ngắn hơn. Phản ánh mức độ đầu tư vào học hành của người miền Bắc.
- ▶ Sự khác biệt càng về số năm học trung bình càng tăng khi hệ thống giáo dục ở miền Bắc cho phép học 12 năm thay vì 10 và 11 năm như trước.

Kết quả ước lượng chính

Bước hai: hồi quy thu nhập dựa vào biến công cụ

Survey → Variables	2004		2006		2008		2010†	
	[1]	[2]	[1]	[2]	[1]	[2]	[1]‡	[2]
↓ Schooling	-.7060	-.8664	-.5099**	-.5531**	-.2148**	-.2344**	.0577	.0107
SE	(.6909)	(1.1006)	(.2198)	(.2638)	(.0939)	(.1060)	(.0540)	(.0435)
Experience	.0723***	.0425**	.0476**	.0299	.0384***	.0354***	.1127***	.0932***
SE	(.0250)	(.0209)	(.0201)	(.0203)	(.0132)	(.0114)	(.0152)	(.0106)
Experience-squared	-.0008	-.0007	-.0004	.0002	-.0004	-.0001	-.0015***	-.0012***
SE	(.0006)	(.0020)	(.0007)	(.0008)	(.0005)	(.0005)	(.0002)	(.0001)
North of the 17th Parallel		.6093		.3851		.1120		.0139
SE		(1.0847)		(.2958)		(.1627)		(.0973)
Constant	15.1916**	16.2143*	13.7528***	13.8572***	11.2378***	11.3151***	7.2913***	8.0566***
SE	(6.6863)	(9.6455)	(2.2251)	(2.4028)	(.9695)	(.9739)	(.7874)	(.5406)
Observations	2,074	4,532	2,191	4,777	2,261	4,838	2,307	5,671

* Cột [1] tương ứng với mẫu quan sát ở phía Bắc vĩ tuyến 17. Cột [2] sử dụng số liệu cả nước.

* [*], [**], [***] thể hiện tham số có ý nghĩa thống kê ở mức 90, 95, và 99%.

* Mẫu quan sát bao gồm những người làm công ăn lương, từ 20-70 tuổi.

Giải thích kết quả

Kết quả hồi quy bước hai:

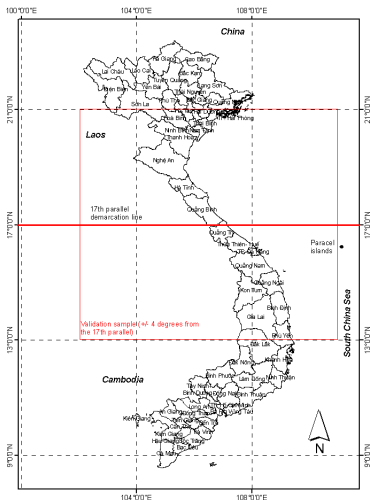
- ▶ Số năm học không có ý nghĩa thống kê, thậm chí là âm \Rightarrow học nhiều hơn cũng không giúp tăng thu nhập, hay thậm chí còn làm giảm thu nhập.
- ▶ Kết quả hoàn toàn trái ngược với mô hình tuyến tính sử dụng OLS, nhưng lại hợp lý với mô hình hồi quy phi tuyến.
 - ▶ Hồi quy dùng OLS là tỷ suất thu nhập trung bình cho một năm học, như nhau cho mọi năm.
 - ▶ Phương pháp hồi quy phi tuyến cho thấy thời gian học thêm lớp 9 không làm tăng thu nhập.
 - ▶ Ước lượng sử dụng biến công cụ là tác động của việc học thêm lớp 9 - còn gọi là tác động nội tại trung bình (Local Average Treatment Effects-LATE), không liên quan đến tỷ suất thu nhập trung bình của tất cả mọi người đi học.

Giải thích kết quả

So sánh chéo giữa ba ước lượng của các năm 2006, 08, 10 cho thấy có một xu hướng rõ rệt là tỷ suất thu nhập cho người học lớp 9 tăng dần, điều này thể hiện thay đổi về bản chất chương trình lớp 9 cho những người sinh sau này và học lớp 9 khi đã có thay đổi về chất.

- ▶ Bộ dữ liệu 2004 có nhiều người học lớp 9 lúc mới cải cách hơn (sinh 76, học lớp 9 năm 1990 lúc 15 tuổi, và 28 tuổi lúc được khảo sát).
- ▶ Bộ dữ liệu 2010 có nhiều người trẻ hơn và học lớp 9 khi chương trình mới đã được triển khai nhiều năm.

Kiểm tra độ vững của kết quả



- ★ Giới hạn mẫu quan sát ở các tỉnh miền Trung, trong vòng 400km ở hai phía của vĩ tuyến 17.
- ★ Các tỉnh ngay sát vĩ tuyến 17 được kỳ vọng không có khác biệt nhiều về điều kiện kinh tế.
- ★ Tránh ảnh hưởng của hai thành phố lớn là Hà Nội và TP HCM.

Kiểm tra độ vững của kết quả

Bước một: hồi quy số năm học dựa vào biến công cụ

Survey → Variables	[1]	2004 [2]	[1]	2006 [2]	[1]	2008 [2]	[1]	2010† [2]
<u>INSTRUMENTS</u>								
Cohort 1972-78								
SE								
Cohort 1972-78 interacts North of the 17th Parallel								
SE								
<u>CONTROLS</u>								
North of the 17th Parallel								
SE								
Experience								
SE								
Experience-squared								
SE								
Constant								
SE								
Observations								

★ Mẫu chỉ gồm các tỉnh miền Trung bao gồm Thanh Hóa, Nghệ An, Hà Tĩnh, Quảng Bình ở phía Bắc vĩ tuyến 17, và Quảng Trị, Thừa Thiên Huế, Đà Nẵng, Quảng Nam, Quảng Ngãi, Bình Định, Phú Yên, và Khánh Hòa ở phía Nam vĩ tuyến 17. Các tỉnh cao nguyên như Gia Lai và Kon Tum không được sử dụng.

★ Cột [1] chỉ sử dụng các tỉnh phía Bắc của vĩ tuyến 17. Cột [2] sử dụng tất cả các tỉnh.

Kiểm tra độ vững của kết quả

Bước hai: hồi quy thu nhập dựa vào biến công cụ

Survey → Variables	2004		2006		2008		2010†	
	[1]	[2]	[1]	[2]	[1]	[2]	[1]	[2]
↓ Schooling	-.0407	-.0539	-.3803	-.4303	-.1002	-.2184	.0462	.0272
SE	(.0781)	(.0861)	(.2324)	(.3164)	(.1245)	(.1754)	(.2365)	(.1519)
Experience	.0538*	.0524***	.0743***	.0105	.0932***	.0424	.0986*	.0980***
SE	(.0297)	(.0125)	(.0179)	(.0439)	(.0231)	(.0368)	(.0542)	(.0244)
Experience-squared	-.0008	-.0007	-.0010	.0010	-.0023***	-.0003	-.0012*	-.0012***
SE	(.0010)	(.0005)	(.0009)	(.0018)	(.0007)	(.0014)	(.0007)	(.0002)
North of the 17th Parallel		-.0994		.0477		-.0803		-.1774
SE		(.0834)		(.2569)		(.1590)		(.1970)
Constant	8.7574***	8.9785***	12.1467***	12.8452***	9.6930***	11.1167***	7.3612**	7.6800***
SE	(.7573)	(.8009)	(2.3323)	(3.0321)	(1.3565)	(1.7353)	(3.3199)	(1.8073)
Observations	265	887	293	954	249	899	337	1,105

★ Mẫu chỉ gồm các tỉnh miền Trung bao gồm Thanh Hóa, Nghệ An, Hà Tĩnh, Quảng Bình ở phía Bắc vĩ tuyến 17, và Quảng Trị, Thừa Thiên Huế, Đà Nẵng, Quảng Nam, Quảng Ngãi, Bình Định, Phú Yên, và Khánh Hòa ở phía Nam vĩ tuyến 17. Các tỉnh cao nguyên như Gia Lai và Kon Tum không được sử dụng.

★ Cột [1] chỉ sử dụng các tỉnh phía Bắc của vĩ tuyến 17. Cột [2] sử dụng tất cả các tỉnh.

Kiểm định sự phù hợp của biến công cụ

Các kiểm định thống kê

Survey →	2004		2006		2008		2010	
	[1]	[2]	[1]	[2]	[1]	[2]	[1]	[2]
Weak Instrument								
Cragg-Donald Wald F statistic	2.99	1.36	8.88	5.25	24.61	16.64	27.20	18.77
Kleibergen-Paap Wald rk F statistic	1.23	.71	7.00	5.73	9.05	7.90	15.00	12.56
Under-Identification								
Kleibergen-Paap rk LM statistic $\chi^2(1)$	1.22	.70	5.61	4.55	6.22	5.47	11.74	10.62
p-value	.2698	.4033	.0179	.0329	.0126	.0193	.0006	.0011
Angrist-Pischke								
Under-identification $\chi^2(1)$	1.28	0.72	7.25	5.83	9.37	8.03	15.56	12.77
p-value	.2585	.3956	.0071	.0158	.0022	.0046	.0001	.0004

Kiểm định biến công cụ yếu và khả năng ước lượng mô hình: ngoại trừ bộ số liệu 2004, còn lại đều bác bỏ giả thuyết biến công cụ không tương quan với biến giải thích.

Bản chất của ước lượng sử dụng biến công cụ là ước lượng tác động nội tại trung bình (IV-LATE)

Instrumental Variables - Local Average Treatment Effects:

$$plim\beta_{IV} = \frac{E[\beta_i * \Delta S_i]}{E[\Delta S_i]}$$

β_i là tỷ suất nhu thập, ΔS_i là số năm học tăng thêm do tác động của cải cách giáo dục lên cho cá nhân i .

- ▶ Chỉ liên quan đến những quan sát bị ảnh hưởng bởi chính sách. Những người bỏ học trước lớp 9 sau cải cách sẽ không bị ảnh hưởng và không có tác động đến kết quả ước lượng.
- ▶ Những người học hết lớp 9 chịu ảnh hưởng lớn nhất và do đó tác động đến ước lượng trung bình nhiều nhất thông qua trọng số ΔS_i . Những người bỏ học khi đang học lớp 9 sẽ tác động đến kết quả trung bình ít hơn.
- ▶ Ước lượng này không liên quan đến tỷ suất thu nhập trung bình cho toàn bộ người đi học.

Tại sao tỷ suất thu nhập cho học lớp 9 bằng không ở Việt Nam?

- ▶ Ước lượng IV-LATE của nhóm quan sát bị ảnh hưởng bởi chính sách thay đổi chỉ là những người đã đạt được số năm học trung bình tương đối cao, ít nhất 8 năm.
- ▶ Những người này có tỷ suất thu nhập thấp biên hơn nhiều so với những người mới chỉ học 3-4 năm.
- ▶ Kết quả sử dụng IV một số nước khác cho kết quả dương và khá cao (lên đến 20%) là do định hướng chính sách áp dụng vào những thành phần dân số nghèo và ít học. Trái ngược với Việt Nam lại áp dụng cho những người đã học tương đối nhiều.

Tại sao tỷ suất thu nhập cho học lớp 9 bằng không ở Việt Nam?

Các nguyên nhân khác:

- ▶ Hệ thống lương cứng nhắc (lương theo hệ số - wage grids) đối với những người làm cho nhà nước.
- ▶ Cải cách chậm rãi do chế độ cầm quyền lo ngại mất quyền kiểm soát. Do đó chương trình học thay đổi rất chậm trong những năm đầu cải cách.
- ▶ Giá trị của bằng cấp quan trọng hơn là chất lượng học.

Tỷ suất thu nhập bằng không cũng được quan sát ở một số nước khác

- ▶ Pischke and von Wachter (2005): nâng số năm học bắt buộc từ 8 lên 9 năm sau thế chiến thứ hai làm tăng số năm học trung bình đạt được khoảng 0.17 năm. Tuy nhiên tỷ suất thu nhập tăng thêm cho thời gian tăng thêm này gần như là bằng không.
 - ▶ Các kỹ năng cần thiết liên quan đến công việc được học trước lớp 9.
 - ▶ Có hệ thống đào tạo nghề (apprenticeship training) ngoài giáo dục phổ thông để bù đắp cho kiến thức không được học trong trường.
 - ▶ Giá trị của bằng cấp ("sheepskin effects") đối với thu nhập.

Kết luận

- ▶ Thu nhập tăng phi tuyến theo số năm học.
- ▶ Giá trị của bằng cấp rất rõ rệt. Thời gian học không ảnh hưởng nhiều đến thu nhập cho mỗi loại bằng cấp đạt được.
- ▶ Ba lý do khiến thu nhập không tăng theo mô hình cổ điển: lương cố định; chương trình học không hiệu quả, không tăng năng suất lao động; hoặc không có cách đánh giá năng lực.
- ▶ Bằng cấp giả là hậu quả của việc trả lương theo bằng cấp.

Khuyến nghị chính sách

- ▶ Cắt giảm chương trình học hoặc tăng mức độ tích hợp giữa các môn sẽ không thay đổi thu nhập.
- ▶ Cải thiện việc trả lương theo năng lực, không theo bằng cấp.
- ▶ Bỏ thi cử cuối cấp, thay vào đó tập trung thời gian để đào tạo kỹ năng làm việc. Thi tốt nghiệp PTCS đã được loại bỏ gần đây. Sắp tới là bỏ thi tốt nghiệp THPT?
- ▶ Chính sách có giá trị lớn nhất chỉ khi nó được áp dụng vào thành phần dân số ít học nhất, do đó có tỷ suất thu nhập biên cao nhất. Tại Việt Nam, đó là các tỉnh thành không thuộc trung ương, miền núi, dân tộc thiểu số. Giảm chi tiêu cho giáo dục đại học và tăng chi cho cấp học thấp hơn (hiện nay mới có tiểu học bắt buộc, sắp tới có thể tiến tới trung học cơ sở).

Lưu ý

- ▶ Tỷ suất thu nhập trong nghiên cứu này là ước lượng cho mỗi cá nhân (chỉ dựa vào thu nhập cá nhân), không phải là tỷ suất đầu tư của xã hội. Tỷ suất thu nhập của lớp 9 có thể bằng không với cá nhân, nhưng với xã hội có thể là tích cực hoặc tiêu cực, cần nghiên cứu thêm.
- ▶ Không tính tác động tổng hợp (general equilibrium effects): tăng thời gian học làm giảm cung lao động, từ đó tăng mức lương cân bằng nếu thị trường lao động vận hành hiệu quả.