

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN**

LÊ THÁI SƠN

**CÁC MÔ HÌNH TOÁN KINH TẾ ĐÁNH GIÁ SUẤT
SINH LỜI CỦA GIÁO DỤC VÀ VAI TRÒ PHÁT TÍN HIỆU
CỦA GIÁO DỤC SAU PHỔ THÔNG VIỆT NAM**

**LUẬN ÁN TIẾN SĨ
NGÀNH KINH TẾ HỌC**

HÀ NỘI – NĂM 2020

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN**

LÊ THÁI SƠN

**CÁC MÔ HÌNH TOÁN KINH TẾ ĐÁNH GIÁ SUẤT
SINH LỜI CỦA GIÁO DỤC VÀ VAI TRÒ PHÁT TÍN HIỆU
CỦA GIÁO DỤC SAU PHỔ THÔNG VIỆT NAM**

**Chuyên ngành: Toán Kinh tế
Mã số:9310101**

LUẬN ÁN TIẾN SĨ

Người hướng dẫn khoa học: PGS.TS. Nguyễn Thị Minh

HÀ NỘI – NĂM 2020

LỜI CAM ĐOAN

*Tôi đã đọc và hiểu về các hành vi vi phạm sự trung thực trong học thuật.
Tôi cam kết bằng danh dự cá nhân rằng nghiên cứu này là tôi tự thực hiện và
không vi phạm yêu cầu về sự trung thực trong học thuật.*

Hà Nội, ngày tháng năm 2020

Tác giả luận án

Lê Thái Sơn

LỜI CẢM ƠN

Tác giả xin bày tỏ lòng biết ơn sâu sắc đến PGS.TS. Nguyễn Thị Minh, người hướng dẫn khoa học, đã tận tình hướng dẫn tác giả trong suốt quá trình học tập, nghiên cứu và hoàn thành luận án.

Tác giả xin chân thành cảm ơn các nhà khoa học, các thầy cô giáo công tác trong và ngoài trường Đại học Kinh tế quốc dân đã đóng góp nhiều ý kiến quý báu giúp tác giả hoàn thiện luận án.

Tác giả xin chân thành cảm ơn các thầy cô Khoa Toán kinh tế - Trường Đại học Kinh tế quốc dân đã tạo mọi điều kiện và giúp đỡ tác giả trong suốt quá trình học tập và nghiên cứu.

Tác giả xin trân trọng cảm ơn các thầy, cô và cán bộ Viện Sau đại học - Trường Đại học Kinh tế quốc dân đã nhiệt tình giúp đỡ tác giả trong suốt quá trình học tập và nghiên cứu.

Cuối cùng, xin được bày tỏ lòng biết ơn đến người thân trong gia đình, bạn bè, đồng nghiệp đã động viên và khích lệ tác giả trong suốt quá trình học tập, nghiên cứu và hoàn thành luận án.

Tác giả luận án

Lê Thái Sơn

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC	iii
DANH MỤC BẢNG	vi
DANH MỤC HÌNH	viii
DANH MỤC CÁC CHỮ VIẾT TẮT	ix
PHẦN MỞ ĐẦU	1
CHƯƠNG 1: CƠ SỞ LÝ THUYẾT VÀ TỔNG QUAN NGHIÊN CỨU	7
1.1. Một số khái niệm	7
1.1.1. Giáo dục sau phổ thông	7
1.1.2. Suất sinh lời của giáo dục	7
1.1.3. Đo lường suất sinh lời của giáo dục	8
1.2. Cơ sở lý thuyết của luận án	8
1.2.1. Lý thuyết về đầu tư trong giáo dục – Mô hình Mincer	8
1.2.2. Lý thuyết phát tín hiệu trên thị trường lao động	10
1.2.3. Vai trò của giáo dục trong lý thuyết tín hiệu và lý thuyết nguồn nhân lực	18
1.3. Tổng quan nghiên cứu	19
1.4. Khung phân tích của luận án	26
CHƯƠNG 2: PHƯƠNG PHÁP NGHIÊN CỨU	29
2.1. Hàm tiền lương Mincer	29
2.2. Phương pháp ước lượng điểm thiên hướng (propensity score method) ...	33
2.2.1. Vấn đề khi đánh giá sự tác động của một chương trình	33
2.2.2. Phương pháp ước lượng điểm thiên hướng PSM	35
2.2.3. Ưu nhược điểm của phương pháp PSM	37
2.3. Phương pháp Heckman ước lượng mô hình với biến nội sinh	38
2.4. Phương pháp Lewbels với vấn đề nội sinh	40
2.5. Phương pháp phi tham số hồi quy Kernel	42

CHƯƠNG 3: THỰC TRẠNG VỀ GIÁO DỤC SAU PHỔ THÔNG VÀ THỊ TRƯỜNG LAO ĐỘNG VIỆT NAM.....	45
3.1. Thực trạng về giáo dục sau phổ thông Việt Nam.....	45
3.1.1 Quy mô và mạng lưới cơ sở giáo dục sau phổ thông	45
3.1.2 So sánh tỷ lệ đi học trong giáo dục của Việt Nam và trên thế giới	46
3.1.3. Tổng quan về đầu tư cho giáo dục và đào tạo	47
3.2. Thực trạng về thị trường lao động Việt Nam.....	50
3.2.1 Lực lượng lao động theo nhóm tuổi	50
3.2.2. Lực lượng lao động theo khu vực thành thị nông thôn	51
3.2.3. Lực lượng lao động theo giới tính	52
3.2.4. Lực lượng lao động theo trình độ được đào tạo	52
3.2.5. Tỷ lệ thất nghiệp	53
3.3. Thực trạng về thu nhập người lao động.....	54
3.3.1. Thu nhập bình quân đầu người năm 2010.....	54
3.3.2. Thu nhập bình quân đầu người năm 2014.....	57
CHƯƠNG 4: MÔ HÌNH VÀ CÁC KẾT QUẢ NGHIÊN CỨU	60
4.1. Nguồn số liệu được sử dụng và các biến số sử dụng trong mô hình ước lượng	60
4.1.1. Nguồn số liệu được sử dụng.....	60
4.1.2. Các biến số sử dụng trong các mô hình.....	60
4.1.3. Thống kê mô tả	63
4.2. Nghiên cứu dấu hiệu của vai trò phát tín hiệu trên thị trường lao động Việt Nam bằng phương pháp PSM.....	68
4.2.1. Kết quả ước lượng cho năm 2014	69
4.2.2. So sánh với năm 2010	73
4.3. Ước lượng suất sinh lời và vai trò phát tín hiệu của giáo dục bằng phương pháp hàm kiểm soát tiền lương Heckman	73
4.3.1. Phương pháp Heckman ước lượng suất sinh lợi giáo dục cho năm 2014	75
4.3.2. So sánh với năm 2010	80
4.4. Ước lượng hiệu suất sinh lời và vai trò phát tín hiệu của giáo dục bằng phương pháp Lewbels.....	82

4.4.1. Phương pháp Lewbels ước lượng hiệu suất sinh lời và vai trò phát tín hiệu của giáo dục 2014	83
4.4.2. So sánh với năm 2010	89
4.5. Ước lượng suất sinh lời và vai trò phát tín hiệu của giáo dục bằng phương pháp phi tham số Kernel	90
4.5.1. Phương pháp Kernel ước lượng hàm tiền lương năm 2014	90
4.5.2. So sánh với năm 2010	94
CHƯƠNG 5: KẾT LUẬN VÀ KIẾN NGHỊ.....	97
5.1. Kết luận	97
5.2. Đề xuất một số kiến nghị chính sách	98
5.3. Các kết quả chính của luận án.....	100
5.3.1. Về mặt lý thuyết	100
5.3.2. Về mặt thực tiễn	101
5.4. Những hạn chế của luận án	102
DANH MỤC CÔNG TRÌNH ĐÃ CÔNG BỐ CỦA TÁC GIẢ.....	103
TÀI LIỆU THAM KHẢO	104
PHỤ LỤC.....	110

DANH MỤC BẢNG

Bảng 1.1:	Tóm tắt các các nghiên cứu đánh giá vai trò phát tín hiệu của giáo dục thông qua so sánh hiệu suất sinh lợi của giáo dục qua đặc điểm, khu vực lao động:	22
Bảng 3.1:	Giáo dục đại học	45
Bảng 3.2:	Giáo dục nghề nghiệp	46
Bảng 3.3:	Tỷ lệ Đi học Tiểu học	47
Bảng 3.4:	Tỷ lệ Đi học đại học	47
Bảng 3.5:	Cơ cấu chi NSNN cho giáo dục, đào tạo (%)	49
Bảng 3.6:	Cơ cấu NSNN theo các cấp học (%)	49
Bảng 3.7:	LLLĐ chia theo nhóm tuổi (Đơn vị: nghìn)	51
Bảng 3.8:	LLLĐ chia theo khu vực thành thị nông thôn (Đơn vị nghìn)	51
Bảng 3.9:	LLLĐ chia theo giới tính (Đơn vị nghìn)	52
Bảng 3.10:	LLLĐ chia theo trình độ được đào tạo (%)	53
Bảng 3.11:	Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, thành thị, nông thôn năm 2010 (đơn vị: nghìn đồng)	55
Bảng 3.12:	Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, nam, nữ năm 2010 (đơn vị: nghìn đồng)	56
Bảng 3.13:	Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, thành thị, nông thôn năm 2014 (đơn vị: nghìn đồng)	58
Bảng 3.14:	Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, nam, nữ năm 2014 (đơn vị: nghìn đồng)	58
Bảng 4.1:	Thống kê mô tả giá trị trung bình các biến năm 2010	64
Bảng 4.2:	Thống kê mô tả giá trị trung bình các biến năm 2014	66
Bảng 4.3:	Thống kê mô tả logarit tiền lương theo trình độ giáo dục năm 2010	67
Bảng 4.4:	Thống kê mô tả logarit tiền lương theo trình độ giáo dục năm 2014	68
Bảng 4.5:	Kết quả ước lượng theo phương pháp PSM cho năm 2014	69
Bảng 4.6:	Kết quả ước lượng theo phương pháp PSM cho khu vực thành thị năm 2014	70

Bảng 4.7:	Kết quả ước lượng theo phương pháp PSM cho khu vực nông thôn năm 2014	71
Bảng 4.8:	Kết quả ước lượng theo phương pháp PSM cho lao động nam năm 2014	72
Bảng 4.9:	Kết quả ước lượng theo phương pháp PSM cho lao động nữ năm 2014 .	72
Bảng 4.10	Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng năm 2014	76
Bảng 4.11	Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng khu vực thành thị và nông thôn năm 2014	78
Bảng 4.12	Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng theo lao động nam và nữ năm 2014.....	79
Bảng 4.13:	Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels cho ba nhóm kinh nghiệm năm 2014	84
Bảng 4.14:	Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels cho ba nhóm kinh nghiệm khu vực thành thị và nông thôn năm 2014	86
Bảng 4.15:	Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels ba nhóm kinh nghiệm của lao động nam và nữ năm 2014.....	88
Bảng 4.16:	Phương pháp Kernel ước lượng hàm tiền lương năm 2014	91
Bảng 4.17:	Phương pháp Kernel ước lượng hàm tiền lương khu vực thành thị và nông thôn năm 2014	92
Bảng 4.18:	Phương pháp Kernel ước lượng hàm tiền lương cho lao động nam và nữ năm 2014	93

DANH MỤC HÌNH

Hình 1.1:	Mối quan hệ giữa tiền lương và giáo dục theo lý thuyết tín hiệu.....	15
Hình 1.2:	Sự lựa chọn giáo dục của công nhân tay nghề thấp và tay nghề cao..	16
Hình 1.3:	Khung phân tích của luận án	27
Hình 3.1:	Chi cho giáo dục, đào tạo đối với tổng chi tiêu công của Việt Nam với một số nước khu vực (%)	48
Hình 3.2.	Tỉ lệ thất nghiệp phân theo khu vực thành thị và nông thôn	53
Hình 3.3.	Tỉ lệ thất nghiệp phân theo giới tính người lao động	54
Hình 3.4:	Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo trình độ giáo dục năm 2010 (đơn vị: nghìn đồng).....	54
Hình 3.5:	Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo năm kinh nghiệm năm 2010 (đơn vị: nghìn đồng)	56
Hình 3.6:	Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo trình độ giáo dục năm 2014 (đơn vị: nghìn đồng).....	57
Hình 3.7:	Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo năm kinh nghiệm năm 2014 (đơn vị: nghìn đồng)	59

DANH MỤC CÁC CHỮ VIẾT TẮT

AFQT	Armed Forces Qualification Tests
GDDH	Giáo dục đại học
GDNN	Giáo dục nghề nghiệp
LLLĐ	Lực lượng lao động
NLSY	The National Longitudinal Survey of Youth
PGSS	Polish General Social Survey
PSM	Propensity score matching
NSĐP	Ngân sách địa phương
NSTW	Ngân sách Trung ương
NSNN	Ngân sách Nhà nước
NLSYM	National Longitudinal Surveys of Young Men
NLSYW	National Longitudinal Surveys of Young Women
THCN	Trung học chuyên nghiệp
THPT	Trung học phổ thông
TTBĐX	Thông tin bất đối xứng

PHẦN MỞ ĐẦU

1. Lí do chọn đề tài

Theo lý thuyết về nguồn nhân lực (Becker, 1964) thì giáo dục giúp trang bị kiến thức, kỹ năng người học, và chính nguồn vốn nhân lực này sẽ giúp nâng cao năng suất của người lao động và do đó thu nhập trong tương lai. Vì vậy có thể nói về khía cạnh kinh tế, đầu tư cho giáo dục cũng cần được xem xét như các dạng đầu tư khác (Becker, 1962). Khi đó bài toán phân tích chi phí – lợi nhuận (cost – benefit) được đặt ra để tìm ra lời giải tối ưu cho bài toán đầu tư cho giáo dục: đầu tư bao nhiêu và ai là người đứng ra đầu tư, nhà nước hay cá nhân. Nói một cách cụ thể hơn, xét về khía cạnh kinh tế, con người sẽ đầu tư cho giáo dục cho đến khi đóng góp biên của giáo dục ngang bằng với chi phí biên. Như vậy, để giải được bài toán này, việc cần thiết là phải xác định được đóng góp biên của giáo dục, còn được gọi là suất sinh lời của giáo dục.

Tại Việt Nam, một phần do truyền thống hiếu học có từ xa xưa, nhiều người dân vẫn tìm mọi cách để đầu tư cho con cái đi học mà ít quan tâm đến khía cạnh hiệu quả kinh tế của giáo dục, do đó có xu hướng tập trung vào các trường đại học mà xem nhẹ các trường nghề. Về phía cung, các trường đại học cũng chưa thực sự quan tâm đến vai trò chính của mình là giúp gia tăng vốn nhân lực nhằm nâng cao năng suất cho người lao động, điều này đã được thể hiện qua nhận xét của các nhà tuyển dụng lao động về chất lượng của sinh viên tốt nghiệp. Vì vậy, việc đưa ra các đánh giá định lượng về suất sinh lời của giáo dục sau phổ thông tại Việt Nam là hết sức quan trọng nhằm đưa ra các bằng chứng thực nghiệm để (1) xem xét hoạt động của hệ thống giáo dục Việt Nam trên vai trò của đơn vị cung cấp vốn nhân lực cho người lao động, và (2) giúp các nhà hoạch định chính sách giáo dục có thể đưa ra các quyết định đầu tư hợp lý.

Việc tính toán ước lượng suất sinh lợi của giáo dục không phải là một công việc dễ dàng. Các nghiên cứu về lĩnh vực này đã cho thấy có nhiều vấn đề hóc búa cần phải được giải quyết để có thể thu được các ước lượng đáng tin cậy. Việc ước lượng này thường gặp 2 khó khăn chính: *Thứ nhất là vấn đề chệch do chọn mẫu (sample selection bias)*, trong đó những người lựa chọn đi làm (do đó có lương) có thể có những đặc trưng khác với những người lựa chọn không đi làm (không có lương, do đó không được lựa chọn vào mẫu nghiên cứu), vì vậy việc sử dụng mẫu

chỉ gồm những người có lương có thể làm cho việc đánh giá tác động của việc học lên mức lương trở nên thiếu tin cậy. Thứ hai là, và phức tạp hơn, đó là vai trò phát tín hiệu của giáo dục thường bị lẫn lộn với vai trò cung cấp vốn nhân lực. Theo Spence (1973), ngoài vai trò cung cấp vốn nhân lực nhằm nâng cao năng suất lao động do đó cũng tăng mức lương, thì giáo dục còn đóng vai trò phát tín hiệu của người lao động về năng lực của người lao động. Cụ thể hơn, nếu người lao động có thể trải qua được các kỳ thi và các khóa học, thì đây là một tín hiệu về năng lực nội tại của người lao động. Việc phát tín hiệu này nhằm cung cấp thêm thông tin cho thị trường lao động về năng lực của người lao động, giảm bớt sự không cân xứng về mặt thông tin giữa người lao động và người sử dụng lao động về năng lực của người lao động. Do đó với một người có học vấn thì ngay cả nếu giáo dục không thực sự giúp gia tăng năng suất lao động thì nó vẫn mang lại mức lương cao hơn cho người đó, ít nhất là trong những năm đầu tiên của quá trình lao động. Do khó có thể tách được hai vai trò này của giáo dục nên nếu sử dụng các phương pháp ước lượng thông thường thì việc ước lượng hiệu suất sinh lời của giáo dục sẽ bị thiếu tin cậy.

Trên thế giới đã có khá nhiều nghiên cứu về vai trò phát tín hiệu cũng như suất sinh lời của giáo dục. Như nghiên cứu của Layard và Psacharopolous (1974), Wiles (1974) hay các nghiên cứu gần đây của Monojit Chatterji (2003), John S. Heywood và XiangDong Wei (2004), Jim Kjelland (2008), Baris Kaymaky (2008), các nghiên cứu đều cho thấy rằng vai trò phát tín hiệu của giáo dục là phổ biến trên hầu hết các nền kinh tế, trong đó các thủ tục trước khi tuyển dụng như thử việc, phỏng vấn, xem lý lịch khoa học cũng chỉ giúp làm giảm nhẹ sự mất cân đối thông tin chứ không xóa bỏ được hiện tượng này. Các nghiên cứu cho thấy mức độ của phát tín hiệu là tùy thuộc vào mức độ phát triển của thị trường lao động cũng như của hệ thống giáo dục. Tại Việt Nam đã có một số nghiên cứu thực nghiệm ước lượng suất sinh lời của giáo dục, tuy nhiên chưa có nghiên cứu nào đề cập đến vai trò phát tín hiệu của giáo dục, do không đề cập đến vai trò phát tín hiệu của giáo dục nên cũng chưa nghiên cứu nào phân tích đánh giá vai trò phát tín hiệu của hệ thống giáo dục Việt Nam. Với những lý do nêu trên, tác giả chọn đề tài **“Các mô hình toán kinh tế đánh giá suất sinh lời của giáo dục và vai trò phát tín hiệu của giáo dục sau phổ thông Việt Nam”** làm đề tài luận án tiến sĩ của mình.

2. Mục tiêu nghiên cứu và câu hỏi nghiên cứu của luận án

Mục tiêu nghiên cứu: Mục tiêu nghiên cứu của luận án là nghiên cứu thực nghiệm hai vai trò chính của giáo dục tại Việt nam, bao gồm: vai trò phát tín hiệu và vai trò cung cấp vốn nhân lực cho người học, trong đó quan tâm chủ yếu đến việc định lượng vai trò cung cấp vốn nhân lực sau khi đã bóc tách vai trò phát tín hiệu. Từ đó rút ra các kiến nghị nhằm nâng cao hiệu quả đầu tư trong giáo dục, tăng cường vai trò của giáo dục như là nơi cung cấp vốn nhân lực nhằm nâng cao năng suất lao động.

Các mục tiêu cụ thể:

- Hệ thống hóa lý luận về vai trò cung cấp vốn nhân lực và vai trò phát tín hiệu của giáo dục.
- Phân tích thực trạng giáo dục sau phổ thông và thực trạng thị trường lao động Việt Nam giai đoạn 2010 - 2014
- Xây dựng các mô hình toán kinh tế ước lượng hiệu suất sinh lời và vai trò phát tín hiệu của giáo dục.
- Đề xuất các kiến nghị chính sách cho giáo dục đại học Việt Nam dựa trên các kết quả nghiên cứu đạt được.

Câu hỏi nghiên cứu:

Để thực hiện được mục tiêu này, luận án sẽ trả lời những câu hỏi nghiên cứu sau:

- Vai trò phát tín hiệu của giáo dục trên thị trường lao động Việt Nam là như thế nào?
- Vai trò phát tín hiệu của giáo dục trong và sau khủng hoảng kinh tế thế giới khác biệt như thế nào?
- Suất sinh lời của giáo dục Việt Nam là như thế nào?

3. Đối tượng và phạm vi nghiên cứu của luận án

Đối tượng nghiên cứu:

Luận án nghiên cứu hai vai trò của giáo dục sau phổ thông, bao gồm vai trò phát tín hiệu và vai trò cung cấp nguồn nhân lực cũng như một số mô hình sử dụng trong nghiên cứu suất sinh lời và phát tín hiệu.

Phạm vi nghiên cứu: Về mặt không gian, luận án nghiên cứu cho người lao động trên toàn quốc với số liệu được sử dụng từ bộ số liệu Điều tra mức sống dân cư (VHLSS) của Tổng cục thống kê. Về thời gian, luận án nghiên cứu cho năm 2010 và 2014, đặc trưng cho hai năm trong và sau khủng hoảng kinh tế thế giới.

Trong thời gian khủng hoảng, do có khó khăn về kinh tế, lao động dư thừa nhiều hơn, cầu về lao động ít hơn nên các nhà tuyển dụng có điều kiện tuyển lựa một cách khắt khe hơn, do đó tính bất đối xứng về thông tin có thể giảm đi. Việc nghiên cứu tại năm 2010 và 2014 do đó sẽ cho phép tìm hiểu đầy đủ hơn về vai trò phát tín hiệu của giáo dục cũng như suất sinh lời của giáo dục.

4. Phương pháp nghiên cứu

Luận án sử dụng các phương pháp nghiên cứu bao gồm: phương pháp tổng hợp, phân tích thống kê, các mô hình toán kinh tế bao gồm:

Phương pháp PSM (propensity score method) nhằm xác định sự hiện diện của vấn đề bất đối xứng thông tin – phát tín hiệu trên thị trường lao động Việt Nam.

Các mô hình kinh tế lượng với cơ sở là mô hình Mincer mở rộng, trong đó việc ước lượng được thực hiện theo hai cách tiếp cận: ước lượng tham số và ước lượng phi tham số.

Các phương pháp ước lượng tham số bao gồm phương pháp Heckman và Phương pháp Lewbels nhằm giải quyết vấn đề biến nội sinh – một vấn đề thường gặp khi ước lượng suất sinh lời của giáo dục. Phương pháp phi tham số, với ưu điểm là việc ước lượng không cần dựa trên một giả định về dạng hàm, được sử dụng nhằm để kiểm tra và đảm bảo độ tin cậy của các kết quả thu được từ các phương pháp ước lượng tham số nói trên.

5. Đóng góp của luận án

- *Về mặt lý luận, lý thuyết:*

+) Luận án làm rõ được vai trò phát tín hiệu của giáo dục sau phổ thông Việt Nam thông qua việc xây dựng các mô hình để nghiên cứu vấn đề này tại Việt Nam. Theo hiểu biết chủ quan của NCS, chưa có nghiên cứu nào ở Việt Nam về vấn đề này. Việc phân tách vai trò của phát tín hiệu sẽ giúp cho việc ước lượng suất sinh lời của giáo dục theo nghĩa

cung cấp nguồn nhân lực sẽ được chính xác hơn, giúp cho việc thiết kế chính sách giáo dục hợp lý hơn.

+) Luận án cũng nghiên cứu sự khác biệt của vai trò phát tín hiệu trong thời kỳ khủng hoảng kinh tế thế giới và thời kỳ nền kinh tế phục hồi. Điều kiện thị trường lao động khác nhau giữa hai thời kỳ này sẽ giúp bộc lộ tốt hơn bản chất của vai trò phát tín hiệu cũng như vai trò của từng nhân tố đối với quá trình phát tín hiệu.

- *Những phát hiện từ kết quả nghiên cứu*

1. Nghiên cứu cho thấy tồn tại vai trò phát tín hiệu của giáo dục sau phổ thông ở Việt Nam: sau khi đã kiểm soát các yếu tố ảnh hưởng, kết quả ước lượng từ phương pháp PSM (propensity score method) cho thấy điều này. Điều này cũng đúng cho cả hai thời kỳ nghiên cứu: năm 2010 và năm 2014.

2. Vai trò của phát tín hiệu là khác nhau giữa khu vực nông thôn và thành thị, giữa lao động nam và lao động nữ. Trong đó vai trò của tín hiệu ở thành thị kéo dài hơn so với khu vực nông thôn. Điều này phù hợp với thực tế rằng tại khu vực thành thị là nơi tập trung nhiều lao động tay nghề cao, người sử dụng cần nhiều thời gian để nhận ra năng lực thực sự của họ. Tương tự, vai trò của tín hiệu đối với nam kéo dài hơn so với nữ.

3. Trong giai đoạn nền kinh tế khó khăn, quá trình phát tín hiệu (của giáo dục) ngắn hơn so với thời kỳ nền kinh tế bình thường. Điều này có thể được giải thích như sau: khi sản xuất khó khăn, chủ lao động sẽ khắt khe hơn trong tuyển dụng lao động bằng việc sử dụng thêm các thông tin ngoài bằng cấp về năng lực lao động, chẳng hạn giới thiệu của bạn bè, mức lương trước đó, v.v.

4. Sau khi kiểm soát yếu tố phát tín hiệu của giáo dục, vai trò cung cấp vốn nhân lực của giáo dục gia tăng theo bằng cấp. Trong đó vai trò cung cấp vốn nhân lực của giáo dục ở thành thị lớn hơn so với khu vực nông thôn. Do khu vực thành thị nơi tập trung các ngành nghề cần nhiều kỹ năng hơn nông thôn, bên cạnh đó mức sống của thành thị cũng cao hơn nông thôn. Kết quả tính toán cũng cho thấy vai trò cung cấp vốn nhân lực cho lao động nam cũng lớn hơn lao động nữ.

6. Kết cấu của luận án

Ngoài phần mở đầu, cam kết, mục lục, phụ lục các bảng biểu, luận án được chia thành 5 chương.

- Chương 1: Cơ sở lý thuyết và tổng quan nghiên cứu. Chương 1 trình bày cơ sở lý thuyết đầu tư trong giáo dục của Mincer, lý thuyết phát tín hiệu của giáo dục trên thị trường lao động. Tổng quan các nghiên cứu về ước lượng suất sinh lời của giáo dục và vai trò phát tín hiệu của giáo dục.

- Chương 2: Phương pháp nghiên cứu. Trong chương này trình bày về các phương pháp tham số và phi tham số nhằm ước lượng hiệu suất sinh lời của giáo dục và làm rõ vai trò phát tín hiệu của giáo dục.

- Chương 3: Thực trạng về giáo dục sau phổ thông và thị trường lao động Việt Nam. Phần này trình bày chính sách đầu tư cho giáo dục và thực trạng của thị trường lao động Việt Nam.

- Chương 4: Mô hình và các kết quả nghiên cứu. Các kết quả phân tích thực nghiệm về đánh giá thông tin bất đối xứng và ước lượng hiệu suất sinh lời của giáo dục có tính đến vai trò phát tín hiệu của giáo dục.

- Chương 5: Kết luận và kiến nghị. Dựa trên các kết quả thực nghiệm chương này đề xuất một số chính sách về tiền lương lao động và chính sách về giáo dục sau phổ thông.

CHƯƠNG 1

CƠ SỞ LÝ THUYẾT VÀ TỔNG QUAN NGHIÊN CỨU

1.1. Một số khái niệm

1.1.1. Giáo dục sau phổ thông

Giáo dục sau phổ thông bao gồm giáo dục đại học và giáo dục nghề nghiệp

Giáo dục đại học bao gồm các hình thức giáo dục diễn ra ở các cơ sở học tập bậc sau trung học, cuối khóa học thường được trao văn bằng học thuật hoặc cấp chứng chỉ. Các cơ sở giáo dục đại học không chỉ bao gồm các trường đại học và viện đại học mà còn các trường chuyên nghiệp, trường sư phạm, trường cao đẳng, trường đại học công lập và tư thục hệ hai năm, và viện kỹ thuật. Điều kiện nhập học căn bản đối với hầu hết các cơ sở giáo dục đại học là phải hoàn thành giáo dục trung học, và tuổi nhập học thông thường là khoảng 18 tuổi.

Giáo dục nghề nghiệp là một bậc học của hệ thống giáo dục quốc dân nhằm đào tạo trình độ sơ cấp, trình độ trung cấp, trình độ cao đẳng và các chương trình đào tạo nghề nghiệp khác cho người lao động, đáp ứng nhu cầu nhân lực trực tiếp trong sản xuất, kinh doanh và dịch vụ, được đào tạo theo hai hình thức là đào tạo chính quy và đào tạo thường xuyên.

1.1.2. Suất sinh lời của giáo dục

Đầu tư cho giáo dục về bản chất cũng là một hình thức đầu tư, trong đó đầu tư cho giáo dục đem lại cho người học vốn nhân lực – giúp nâng cao năng suất lao động và do đó cải thiện thu nhập. Giáo dục cũng cung cấp vốn tri thức cho con người, giúp nâng cao chất lượng văn hóa của cuộc sống. Trong đó các nghiên cứu về suất sinh lời của giáo dục chủ yếu tập trung vào khía cạnh kinh tế của đầu tư cho giáo dục, hay nói cách khác, quan tâm đến suất sinh lời của đầu tư cho giáo dục. Suất sinh lời của giáo dục được bao gồm lợi suất ròng của cá nhân và lợi suất ròng của xã hội [54]. Lợi suất ròng cá nhân= lợi ích cá nhân (ví dụ: thu nhập cao hơn, nguy cơ thất nghiệp thấp) - chi phí cá nhân (ví dụ: học phí và sách vở, thu nhập bị từ bỏ).

Lợi suất xã hội ròng= lợi ích công cộng (ví dụ: thu nhập từ thuế cao hơn, giảm các khoản chi phí chuyển khoản xã hội) – chi phí công (ví dụ: bỏ thuế thu nhập, tăng chi cho giáo dục).

1.1.3. Đo lường suất sinh lời của giáo dục

Tuy định nghĩa về suất sinh lời của giáo dục là khá rõ ràng, nhưng việc đo lường nó một cách trực tiếp lại là vấn đề khó khăn. Thứ nhất, giáo dục có ảnh hưởng đến thu nhập không chỉ trong 1 năm mà kéo dài suốt cuộc đời của người lao động, nên việc tính toán trực tiếp ảnh hưởng này là điều hầu như không thể thực hiện được. Thứ hai, việc tính toán lợi suất xã hội luôn là vấn đề khó khăn khi tính toán mọi suất đầu tư, bao gồm cả đầu tư giáo dục. Vì vậy, các nghiên cứu thực nghiệm về lợi suất đầu tư cho giáo dục tập trung đánh giá tác động của giáo dục lên mức lương của người lao động sau khi đã kiểm soát các yếu tố có ảnh hưởng lên mức lương.

1.2. Cơ sở lý thuyết của luận án

1.2.1. Lý thuyết về đầu tư trong giáo dục – Mô hình Mincer

Có thể nói đầu tư cho giáo dục cũng là một dạng đầu tư, trong đó con người chi tiền ra đầu tư với mục đích thu được lợi ích cho hiện tại và tương lai. Lợi ích này có thể bao gồm nhiều khía cạnh: gia tăng hiểu biết làm cho cuộc sống phong phú, có được danh tiếng trong xã hội, .., nhưng có lẽ quan trọng nhất vẫn là lợi ích kinh tế. Do đó nhiều nghiên cứu đã được tiến hành nhằm đánh giá suất sinh lời của đầu tư cho giáo dục. Và các nghiên cứu này thường được dựa trên lý thuyết về đầu tư cho giáo dục, được đề xuất bởi Mincer (Mincer, 1958, 1974). Trong lý thuyết này, lợi suất của đầu tư cho giáo dục được mô hình hóa dựa trên bài toán tối ưu hàm lợi ích của các cá nhân.

Mô hình Mincer dạng cơ bản có thể được trình bày thành hai phần như sau: phần thứ nhất, sử dụng xuất phát điểm là bài toán tối ưu hàm lợi ích phổ dụng, trong đó người tiêu dùng tối ưu hóa hàm lợi ích của mình thông qua tiêu dùng, sau đó chuyển thành bài toán tối ưu trong đầu tư cho giáo dục. Phần thứ hai sẽ dựa trên bài toán tối ưu đầu tư cho giáo dục để xây dựng hàm số cho lợi suất của đầu tư cho giáo dục này.

Phần 1: từ bài toán tối ưu phổ dụng sang bài toán quyết định đầu tư cho giáo dục

Giả sử một người lao động sẽ lao động đến T tuổi, khi đó anh ta sẽ chọn hàm tiêu dùng $c(t)$ sao cho cực đại hàm lợi ích toàn đời như sau:

$$V = \int_0^T u(c(t))e^{-rt} dt \quad (1.1)$$

Trong đó $u(\cdot)$ là hàm lợi ích, thỏa mãn các giả thiết thông thường, là hàm tăng chặt và lõm, r là tỷ lệ chiết khấu thời gian.

Vốn nhân lực, ký hiệu là $h(t)$ tuân theo phương trình sau:

$$\dot{h}(t) = f(t, h(t), s(t)) \quad (1.2)$$

Trong đó $s(t)$ là tỷ lệ thời gian dành cho học vấn ($0 \leq s(t) \leq 1$)

Thu nhập của người đó tại thời điểm t sẽ được xác định bởi hàm sau:

$$TN(t) = w(t)(1 - s(t))h(t) \quad (1.3)$$

Trong đó $w(t)$ là mức lương trả cho một đơn vị lao động cơ bản. Người lao động được giả sử rằng chỉ được tiêu dùng trong giới hạn của thu nhập:

$$\int_0^T u(c(t))e^{-rt} dt \leq \int_0^T w(t)(1 - s(t))h(t)e^{-rt} dt \quad (1.4)$$

Sử dụng định lý tách, có thể chứng minh được rằng (<https://econ.lse.ac.uk/staff/spischke/ec533/Acemoglu%20Autor%20chapter%2001.pdf>) nghiệm của bài toán (1.1) – (1.4) cũng chính là nghiệm của bài toán: tìm s để cực đại hóa hàm:

$$\text{Max}_s \int_0^T w(t)(1 - s(t))h(t)e^{-rt} dt \quad (1.5)$$

Với các ràng buộc (1.2)-(1.4)

Kết quả này cho phép chuyển từ bài toán tối ưu tiêu dùng thành bài toán tối ưu cho giáo dục.

Phần 2: Xây dựng hàm toán học cho lợi suất đầu tư cho giáo dục

Giả sử $s(t) = 1$ trong một khoảng thời gian S , và bằng 0 ngoài khoảng đó. Và sau đó thì người lao động sẽ nhận được một lượng vốn nhân lực $m(S)$, trong đó m là hàm tăng, khả vi liên tục và lõm.

Giả sử vốn nhân lực của người lao động thay đổi (nhờ lao động) theo công thức:

$$\dot{h}(t) = g_h \cdot h(t) \quad (1.6)$$

Và lương thay đổi theo công thức:

$$\dot{w}(t) = g_w \cdot w(t) \quad (1.7)$$

Trong đó g_h và g_w là các hàm dương nào đó.

Từ (1.6) và (1.7), bài toán cực trị (1.5) có thể chuyển thành:

$$\text{Max}_S \frac{m(S)w(0)S.e^{-r-g_w}}{r - g_h - g_w} \quad (1.8)$$

Do $m(S)$ là hàm lõm, điều kiện cần bậc nhất cũng chính là nghiệm tối ưu duy nhất cho bài toán (1.8):

$$\frac{m'(S^*)}{m(S)} = r - g_w \quad (1.9)$$

Từ phần 1 ta có:

$$TN(t) = e^{g_h(t-S^*)} e^{g_w t} w_0 \cdot m(S) \quad (1.10)$$

Lấy logarit hai vế của (1.10) và thay vào (1.9) ta có:

$$\log(TN(t)) = c + (r - g_w)S^* + g_h(t - S^*) + g_w t \quad (1.11)$$

(1.11) cho thấy mối quan hệ giữa học vấn và thu nhập từ lao động, là phương trình cơ bản để xác định lợi suất cho học vấn của người lao động, chính là mô hình Mincer, là mô hình nền tảng cho các nghiên cứu thực nghiệm về lợi suất của học vấn.

1.2.2. Lý thuyết phát tín hiệu trên thị trường lao động

Lý thuyết thông tin bất đối xứng cho rằng có sự mất cân bằng thông tin giữa người mua và người bán, dẫn đến sự không hiệu quả ở một số thị trường nhất định. Ba nhà kinh tế có sự ảnh hưởng đặc biệt trong việc phát triển lý thuyết thông tin bất đối xứng là George Akerlof, Michael Spence và Joseph Stiglitz. Akerlof (1973) lần đầu tiên chỉ ra sự bất cân xứng thông tin bằng minh họa cho thị trường ô tô, trong đó người mua xe nhìn thấy thông tin khác với người bán, tạo cho người bán một động lực bán những chiếc xe có chất lượng thấp hơn mức trung bình của thị trường. Michael Spence (1973) chỉ ra rằng có sự thông tin bất đối xứng trên thị trường lao động, người sử dụng lao động không chắc chắn về năng lực của người lao động khi được tuyển dụng. Ông so

sánh tình huống này như việc mua xổ số, và để làm rõ thông tin người lao động sẽ phát tín hiệu về năng lực của mình đến người sử dụng lao động.

Lý thuyết phát tín hiệu

Trong việc xây dựng lý thuyết tín hiệu, Spence (1973) đã sử dụng thị trường lao động để mô hình hóa chức năng phát tín hiệu của giáo dục. Người sử dụng lao động tiềm năng thiếu thông tin về chất lượng của những ứng cử viên công việc. Do đó những ứng viên sử dụng giáo dục để phát tín hiệu về chất lượng của họ đến người sử dụng và làm giảm thông tin bất đối xứng. Đây có lẽ là tín hiệu xác thực vì các ứng cử viên có chất lượng thấp sẽ không thể chịu đựng được những khắc khe của giáo dục đại học. Mô hình của Spence đi ngược với lý thuyết vốn nhân lực vì ông coi trọng vai trò của giáo dục để tăng năng suất lao động và tập trung sử dụng giáo dục như một phương tiện truyền thông của các đặc tính khác không quan sát được của ứng cử viên công việc.

Kirman và Rao (2000) cung cấp ví dụ tổng quát giúp minh họa cho mô hình cơ bản tín hiệu. Giống như hầu hết các ví dụ về tín hiệu, tác giả phân biệt hai thực thể: doanh nghiệp chất lượng cao và doanh nghiệp chất lượng thấp. Các doanh nghiệp trong ví dụ này biết rõ về chất lượng của mình, và các nhà đầu tư và khách hàng thì không, do đó mỗi doanh nghiệp có cơ hội để đưa tín hiệu hoặc không đưa tín hiệu về chất lượng thật của mình đến với nhà đầu tư và khách hàng. Khi doanh nghiệp chất lượng cao đưa ra tín hiệu, họ sẽ nhận được sự hoàn lại A, khi họ không đưa tín hiệu họ nhận được sự hoàn lại B, đối với doanh nghiệp chất lượng thấp họ nhận được sự hoàn lại C khi đưa tín hiệu, và sự hoàn lại D khi họ không đưa tín hiệu. Tín hiệu báo hiệu một chiến lược khả thi đối với các công ty chất lượng cao khi $A > B$ và khi $D > C$. Trong trường hợp này, các doanh nghiệp chất lượng cao sẽ thúc đẩy để phát tín hiệu và các doanh nghiệp chất lượng thấp thì không dẫn đến sự cân bằng tách biệt. Trong những trường hợp như vậy các nhà đầu tư, khách hàng có thể phân biệt chính xác giữa các doanh nghiệp chất lượng cao và thấp. Ngược lại khi cả hai công ty được hưởng lợi từ tín hiệu (tức là $A > B$ và $C > D$) có một sự cân bằng tổng hợp và người bên ngoài không thể phân biệt giữa hai loại hình doanh nghiệp.

Các nhà kinh tế tài chính đã phát triển một số ví dụ minh họa những mối quan hệ chung này. Họ xác nhận nợ (Ross, 1973) và cổ tức (Bhattacharya) thể hiện chất lượng doanh nghiệp. Theo các mô hình, chỉ có các doanh nghiệp chất lượng cao có khả năng thanh toán lãi suất và cổ tức trong thời gian dài. Ngược lại doanh nghiệp chất lượng thấp sẽ không thể duy trì các khoản thanh toán đó. Do đó các tín hiệu như vậy ảnh hưởng đến người quan sát bên ngoài (nhà đầu tư, cho vay) nhận thức về chất lượng doanh nghiệp.

Hầu hết các mô hình tín hiệu bao gồm chất lượng là đặc trưng phân biệt, khái niệm về chất lượng có thể được giải thích bằng nhiều cách liên quan. Chất lượng đề cập đến khả năng cơ bản không quan sát được của người truyền tín hiệu nhằm đáp ứng nhu cầu của những người bên ngoài. Trong ví dụ cổ điển của Spence, chất lượng đề cập đến khả năng không thể quan sát được của cá nhân, được tín hiệu bằng việc hoàn thành các cấp bậc giáo dục. Ở nghiên cứu của Ross (1973), chất lượng đề cập đến khả năng không nhìn thấy được của tổ chức để thu được dòng tiền mặt tích cực ở tương lai, có thể được biểu hiện bằng cấu trúc tài chính hoặc ưu đãi quản lí. Hoặc như nghiên cứu của Kreps và Wilson (2003) chất lượng là sự chia sẻ một số đặc tính liên quan đến danh tiếng và uy tín.

Lý thuyết về phát tín hiệu trên thị trường lao động

Micheal Spence được biết đến như là cha đẻ của lý thuyết tín hiệu. Trong tác phẩm kinh điển của mình (1973), ông đề xuất một mô hình giáo dục khác đáng kể với lý thuyết nguồn nhân lực. Ông lập luận rằng các giả thiết của Becker là không phù hợp với thực tế dẫn đến việc liên quan đến tuyển dụng lao động là không chắc chắn. Người sử dụng lao động không chắc chắn về các kỹ năng hoặc hiểu biết của người lao động khi tuyển dụng một lao động mới. Hơn thế nữa họ không thể nhận ra năng lực thực sự của người lao động và phải qua một quãng thời gian để nhận biết rõ năng lực của người công nhân. Spence cho rằng để thuê một lao động mới cũng giống như việc mua xổ số, trong đó chi phí mua một vé số được xác định như mức lương của người lao động. Mặc dù doanh nghiệp không thể quan sát được đặc điểm năng suất của người lao động nhưng không có nghĩa là không quan sát các đặc điểm khác của người lao động như giới tính, tuổi, trình độ giáo dục, chủng tộc, và nhiều tính năng khác.

Rõ ràng là người lao động có những đặc tính là ngoại sinh (tuổi, giới tính chẳng hạn) và những đặc tính chịu sự ảnh hưởng của người lao động như trình độ giáo dục. Trong lý thuyết có sự phân chia rõ ràng liên quan đến đặc điểm của người lao động những đặc điểm ngoại sinh được gọi là chỉ số (indices) và những đặc tính nội sinh được gọi là tín hiệu (signal). Theo mô hình, người sử dụng lao động sẽ biết được năng suất trung bình của người lao động sau một khoảng thời gian. Trên cơ sở này sẽ đánh giá những người ứng tuyển chính xác hơn. Chỉ số và tín hiệu sẽ ảnh hưởng đến xác suất có điều kiện là người đó là (hoặc không là) người có năng lực cao. Những người lao động cũng có thể quan sát tác động của các chỉ số và tín hiệu về cơ hội việc làm và lựa chọn một mức độ để tối đa hóa mức lương của họ và điều này sẽ ảnh hưởng đến thị trường lao động.

A. Các giả định

Lý thuyết tín hiệu giả định tất cả tổng thể (ít nhất về mặt năng suất) có thể được biểu diễn bởi hệ số θ , đo lường năng suất cá nhân. Mô hình đơn giản nhất là năng lực cao θ_1 hoặc năng lực thấp θ_2 (nó có nghĩa là công việc của những công nhân hiệu quả chỉ cần một giờ và những công nhân không hiệu quả sẽ mất $\frac{\theta_1}{\theta_2}$ giờ cao hơn các công nhân khác). Để thuận tiện xét giá của hàng hóa được sản xuất là 1. Tỷ lệ lao động có năng suất cao trong dân số được ký hiệu là h .

+ Kế tiếp ta giả định rằng mặc dù các công nhân biết rõ năng lực của họ và người sử dụng lao động không sở hữu thông tin này.

+ Giả định rằng các doanh nghiệp hoạt động trong môi trường cạnh tranh hoàn hảo và có rủi ro chia đều (nhằm đơn giản hóa việc tính toán). Hơn nữa người chủ không trực tiếp quan sát được năng suất của người công nhân hoặc chi phí quá cao để biết sự đạt được do tri thức nhưng năng suất trung bình của người công nhân có thể quan sát được và miễn phí.

+ Tiếp theo để đơn giản hóa, giả định rằng không có những yếu tố khác của quá trình sản xuất. Trên cơ bản mô hình được giả định tiền lương người công nhân là không co giãn.

Sự cạnh tranh hoàn hảo giữa các hãng đảm bảo rằng công nhân được trả công theo năng suất của họ. Bởi vì người chủ không thể phân biệt được hai nhóm công nhân, họ tính toán tiền lương của công nhân dựa trên năng suất theo công thức:

$$1. \quad h_1\theta_1 + (1 - h_1)\theta_2 = \theta_{average}$$

Chúng ta có thể thấy $\theta_1 > \theta_{average} > \theta_2$ vì công nhân của nhóm đầu tiên có sự động viên để nhận thấy năng suất cao của họ. Tuy nhiên không thể (hoặc không hiệu quả) trực tiếp sắp xếp họ theo năng suất của họ. Nói cách khác người lao động có thể thuyết phục nhà tuyển dụng rằng họ có năng suất cao hơn, hoặc ít nhất cũng có vẻ như là họ nhiều năng suất hơn. Bằng hai cách:

- Sự khác nhau của các chỉ số tạo nên sự độc lập của các công nhân, người sử dụng lao động không thể ảnh hưởng đến những chỉ số này, làm cho công nhân kém hấp dẫn trên quan điểm tối ưu của người sử dụng lao động.

- Sự khác nhau của các tín hiệu chỉ ra một cấp độ năng suất của các công nhân và các tín hiệu này được xác định bởi quyết định cá nhân dựa trên sự lựa chọn mà ảnh hưởng bởi chi phí cơ hội của cá nhân, ở đó hành vi tối ưu của người quản lý công việc có thể quan sát được.

Spence (1973) cho rằng giáo dục là một loại tín hiệu. Người công nhân có năng suất sẽ cố gắng đầu tư nhiều cho giáo dục để phân biệt chính họ với các công nhân khác có cấp độ năng suất thấp hơn. Vấn đề họ phải đối mặt tối đa hóa sự khác biệt giữa mức lương được mời chào (offered wages) và chi phí liên quan đến việc đạt được một cấp độ giáo dục, đây là một đề cập có giá trị các giả định quan trọng nhất của lý thuyết này. Giả định quan trọng là ít có thể các công nhân đối mặt với chi phí lớn hơn do đi học, bởi vì nếu điều kiện này không đầy đủ, khi đó những người công nhân sẽ có khuyến khích bắt chước công nhân năng suất cao. Chúng ta không nên xem xét chi phí cho tín hiệu không chỉ là tiền mà nó còn bao gồm thời gian và sự cố gắng.

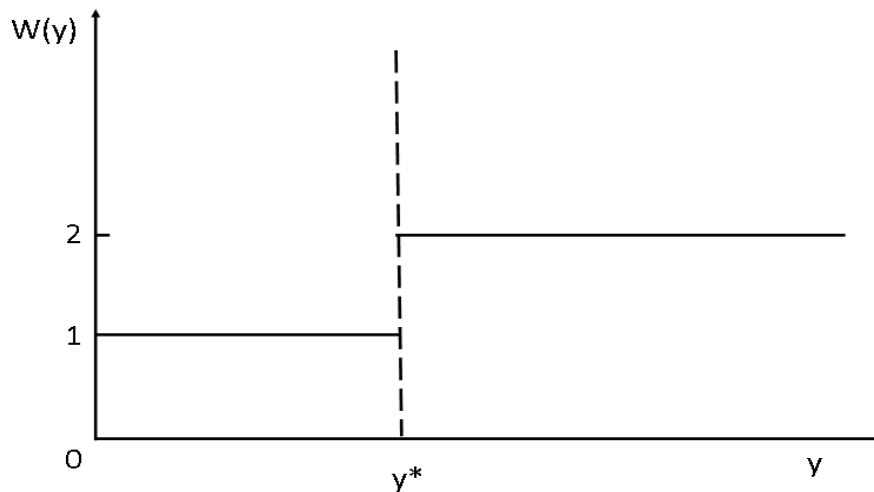
B. Mô hình

Những công nhân năng suất cao sẽ có chi phí giáo dục là c_1 bằng y_1, y_1 là số năm đi học, trong khi đó các công nhân năng suất thấp có chi phí tương ứng cao hơn:

$$2. \quad y_1 = c$$

$$3. \quad y_2 = ac, \text{ trong đó } a > 0$$

Các trạng thái cân bằng phụ thuộc rất nhiều vào niềm tin của người tham gia thị trường. Spencer giả định rằng có nhiều cấp độ giáo dục y^* đảm bảo tách biệt hoàn toàn trên thị trường. Nếu một cá nhân đạt được ít nhất một cấp độ giáo dục ($y \geq y^*$) khi đó với xác suất bằng một, có thể nói rằng cá nhân này là năng lực cao. Ngược lại nếu một cá nhân dành ít thời gian ở nhà trường ($y < y^*$) khi đó người chủ lao động chắc chắn 100% là cá nhân này có năng lực thấp. Với niềm tin có điều kiện với năng suất của những người lao động, người chủ sẽ đưa ra một mức lương $W(y)$ bằng với năng suất cận biên của người lao động mà mình đánh giá (θ_1 là của người năng suất cao, θ_2 là của người năng suất thấp).

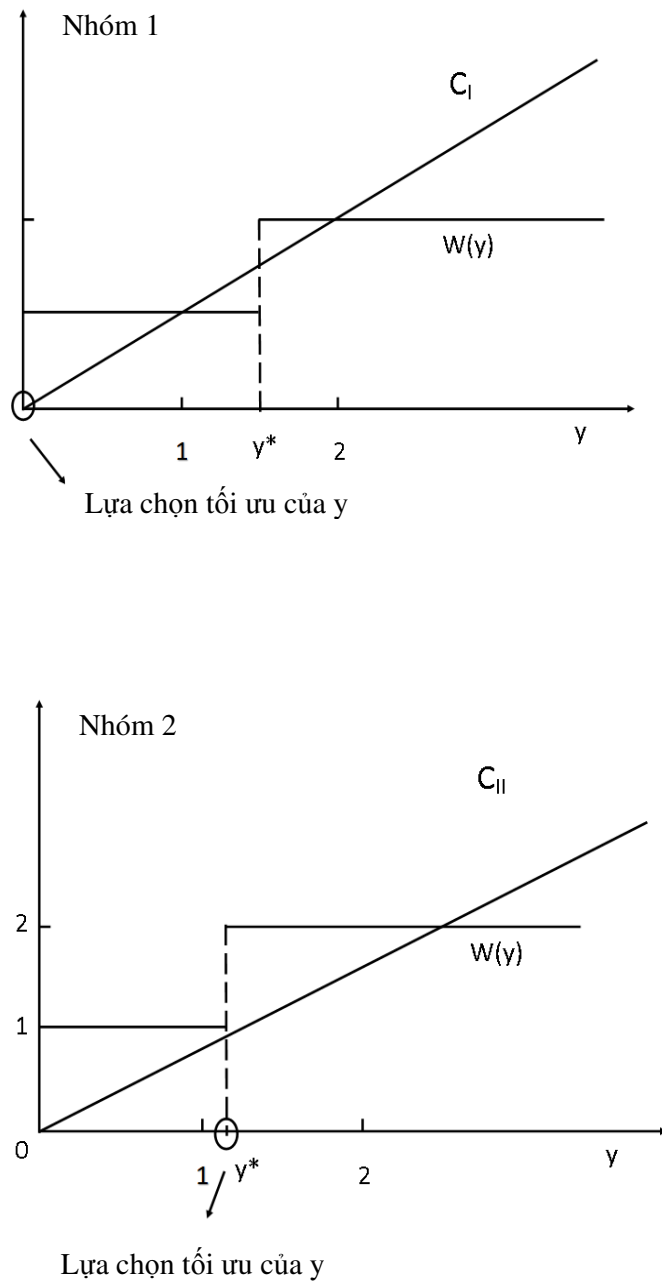


Hình 1.1: Mối quan hệ giữa tiền lương và giáo dục theo lý thuyết tín hiệu

Nguồn: Spence (1973)

Với mức lương đã cho, các thành viên của mỗi nhóm sẽ chọn cấp độ giáo dục tối ưu. Những công nhân của nhóm năng lực cao, cấp độ giáo dục tối ưu là y^* , họ sẽ không bỏ thời gian để học thêm bởi vì giả thiết ban đầu giáo dục không làm tăng năng suất và phần nhận được từ tín hiệu rõ ràng là lãng phí. Cũng giống như vậy, những công nhân năng lực thấp sẽ lựa chọn cấp độ giáo dục là $y=0$. Ta có thể thấy mọi người sẽ đầu tư cho giáo dục ở mức y^* hoặc sẽ không đầu tư vào giáo dục, bởi vì sẽ không tạo năng suất và ảnh hưởng đến niềm tin của người sử dụng

lao động. Người lao động sẽ lựa chọn cấp độ giáo dục dựa trên niềm tin của người sử dụng lao động, do đó sẽ có một trạng thái tín hiệu.



Hình 1.2: Sự lựa chọn giáo dục của công nhân tay nghề thấp và tay nghề cao

Nguồn: Spence (1973)

Nhóm một sẽ học chính xác là y^* nếu chi phí của việc học này sẽ thấp hơn lợi suất từ việc học:

$$4. \theta_1 - \theta_2 > y^*$$

Và những người có năng suất thấp chi phí cho giáo dục phải cao hơn lợi ích

$$5. \theta_1 - ay^* < \theta_2$$

Nếu kết hợp hai điều kiện:

$$6. (\theta_1 - \theta_2) / a < y^* < (\theta_1 - \theta_2)$$

Ở đây chúng ta thấy rằng có một số lượng vô hạn các điểm cân bằng ổn định. Giá trị tối ưu của y là một khoảng, nếu ta giả định rằng biến là liên tục thì có số lượng vô hạn của y sẽ đáp ứng các điều kiện nhất định. Nếu hệ số y^* tăng một lượng Δy đủ nhỏ thì cấp độ giáo dục tối ưu vẫn bị giới hạn trên và dưới, điều này làm cho nhóm 1 (năng suất cao hơn) sẽ bị ảnh hưởng, bởi vì nhóm này phải chi trả thêm để có thêm một lượng Δy giáo dục. Trong khi đó điều này sẽ không ảnh hưởng đến hành vi của nhóm thứ 2 bởi vì nhóm này không cần phải đầu tư thêm cho giáo dục. Tuy nhiên nhóm này phải đối mặt với việc tổn thất nếu người sử dụng lao động phân biệt người lao động có khả năng và người lao động không có khả năng dựa trên trình độ học vấn. Nếu không có sự khác biệt với trình độ giáo dục thì theo (1) tất cả người lao động sẽ được trả lương như nhau.

$$7. \theta_{average} - \theta_2 = h^*(\theta_1 - \theta_2)$$

Việc tổn thất sẽ tăng và làm cho khoảng cách năng suất càng tăng và được chia sẻ cho lao động ở nhóm 1. Vì vậy những người lao động không có kỹ năng sẽ thiệt hại nếu sự khác biệt do giáo dục được thiết lập. Nói cách khác, lao động không có kỹ năng có lẽ sẽ thiệt hại do tín hiệu. Sẽ có lợi ích cho những người lao động trình độ giáo dục cao nếu và chỉ nếu tiền lương của họ có được phải lớn hơn chi phí cho tín hiệu năng lực cao của họ. Kết quả là:

$$8. y^* < \theta_1 - \theta_{average} \rightarrow y^* < (1-h)(\theta_1 - \theta_2)$$

Nếu số lượng của những công nhân không có kỹ năng giảm tương đối so với những người lao động có kỹ năng hoặc nếu khoảng cách tiền lương giảm, thì xác suất của nhóm 1 đạt được từ sự khác biệt chính họ sẽ thấp hơn. Có nghĩa

là tỷ lệ cao những người được đào tạo sẽ làm tăng tiền lương trung bình, trong đó có hai tác động đồng thời: có thể làm cho những người lao động ít đầu tư vào giáo dục để phân biệt mình, những người lao động có năng lực thấp sẽ có xu hướng bắt chước những người có kỹ năng cao.

1.2.3. Vai trò của giáo dục trong lý thuyết tín hiệu và lý thuyết nguồn nhân lực

Quyết định theo đuổi giáo dục đại học và sau đại học có liên quan đến một phân tích không chính thức của cá nhân dựa vào các chi phí bỏ ra so với giá trị kì vọng thu được. Xác định lợi ích kinh tế kì vọng như một sự đầu tư, kết quả của những yếu tố xã hội liên quan, là một chủ đề quan trọng và có ý nghĩa đối với các nhà nghiên cứu kinh tế. Trên thế giới đã thực hiện các nghiên cứu khác nhau và ước tính giá trị thu được từ giáo dục (được đo bằng mức tăng theo tiền lương mỗi tuần và tiền lương mỗi năm) khoảng 10% cho mỗi năm học. Với mỗi quan hệ được thiết lập trên, chúng ta có thể tin chắc rằng kết quả giáo dục bậc càng cao, về trung bình, lại sẽ đem lại một mức lương càng cao trong thị trường lao động. Do đó, một cá nhân sẽ cân nhắc về chi phí vật chất và tinh thần mà đầu tư sao cho phù hợp.

Hai lý thuyết tồn tại trong nỗ lực giải thích mối quan hệ nhân quả giữa giáo dục và thu nhập. Những lý thuyết này đưa ra giả thuyết về các cơ chế cụ thể mà thông qua đó giáo dục tác động đến thu nhập.

Lý thuyết nguồn nhân lực lập luận một cách trực quan rằng giáo dục làm tăng giá trị vốn có của con người thông qua việc tăng năng suất lao động, và do đó làm tăng thu nhập trong thị trường lao động. Lý thuyết thị trường cạnh tranh cho rằng xét cho cùng, đòi hỏi người lao động nhận được một khoản thu nhập tương xứng với sản phẩm biên họ tạo ra.

Thuyết tín hiệu đưa ra một lập luận đối lập: giáo dục chỉ phản ánh tài nghệ vốn có của nhân viên. Nguồn vốn nhân lực này, chứ không phải là chính sự giáo dục, đã làm gia tăng năng suất và dẫn đến việc tăng tiền lương lên mức cao hơn.

Từ quan điểm của một cá nhân thì rất khó để xác định lý thuyết nào là chính xác nhất. Xét cho cùng, các cá nhân không thể làm gì nhiều hơn việc tối ưu hóa lợi ích bản thân với một hệ thống đã có sẵn. Dù giáo dục bậc cao ưu đãi một cá nhân với tài năng vốn có hay biểu hiện một thực tài thì sự thật giáo dục

đã đạt được một chức năng đó là tăng mức tiền lương. Nói cách khác, quyết định của một cá nhân có theo đuổi giáo dục đại học phụ thuộc vào việc thiết lập mối liên hệ tích cực giữa giáo dục và thu nhập, điều đó thì cả lý thuyết nguồn vốn nhân lực và lý thuyết tín hiệu đều đã đề cập.

Từ một quan điểm xã hội rộng hơn, hai lý thuyết khác nhau đáng kể về hiệu quả tác động. Nếu giáo dục đại học chỉ hoạt động như là một tín hiệu và không góp phần độc lập làm phát triển khả năng tự có của con người thì sẽ có rất ít tác động thú vị liên quan, do đó giáo dục là một tín hiệu rất tốn kém tiền bạc và thời gian. Ngoài ra, đó có thể không phải là tín hiệu tốt nhất, phải xem xét khả năng rằng một cá nhân quyết định không theo đuổi giáo dục bậc cao vì sự hạn chế về mặt tài chính hoặc đơn giản hơn là họ nhận định chủ quan không suy tính rằng giá trị mà giáo dục mang lại cho họ thấp hơn so với nhận định những sinh viên khác không được đào tạo ở bậc cao hơn. Điều này có nghĩa giáo dục có thể không phản ánh được những cá nhân tốt nhất, trong trường hợp này sẽ có lợi hơn nếu tìm kiếm một con đường khác phù hợp hơn và ít tốn kém hơn.

1.3. Tổng quan nghiên cứu

Nghiên cứu về vai trò của giáo dục trong việc tăng năng suất lao động cũng như phát tín hiệu trên thị trường lao động là một vấn đề quan trọng trong việc xây dựng các chính sách về giáo dục cũng như đầu tư cho giáo dục, do đó trên thế giới đã có nhiều nghiên cứu về vấn đề này.

Như nghiên cứu của Layard và Psacharopolous (1974) đưa ra một số kết luận: suất sinh lời của giáo dục đem lại cho cá nhân có liên quan đến bằng cấp đạt được chứ không phải là số năm đi học. Có nghĩa là những người sử dụng lao động sẽ quan tâm đến những người lao động có bằng cấp hơn là những người lao động dành nhiều thời gian để đi học. Và khi số năm kinh nghiệm tăng lên thì chênh lệch hiệu suất sinh lợi của giáo dục đem lại cho những người có bằng cấp cao sẽ càng lớn so với những người có bằng cấp thấp hơn.

Một phương pháp được đề xuất bởi Wiles (1974) so sánh tiền lương của người lao động làm việc đúng ngành đào tạo với những người lao động không đúng ngành đào tạo. Phương pháp này được Miller và Volker (1984) tính toán cho thị trường lao động Úc. Theo lý thuyết nguồn nhân lực những người lao

động làm đúng ngành nghề sẽ có tiền lương cao hơn những người lao động làm không đúng ngành nghề. Kết quả cho thấy: (1) Sinh viên nam tốt nghiệp kinh tế có xu hướng để có một mức độ tương tự tiền lương cho dù công việc là phù hợp hay không phù hợp, (2) Đối với sinh viên nữ học ngành kinh tế hoặc khoa học, tiền lương của nữ lao động đúng ngành lại thấp hơn lao động không đúng ngành đào tạo, cho thấy giáo dục có đóng vai trò tín hiệu. Cũng bằng phương pháp tương tự Wiles, nghiên cứu của Chung (1990) dựa trên số liệu lao động của Hồng Kông so sánh tiền lương của những người lao động làm việc đúng lĩnh vực đào tạo và khác lĩnh vực đào tạo. Tính toán cho thấy giáo dục đào tạo đóng vai trò phát tín hiệu trên thị trường lao động Hồng Kông do người lao động làm đúng lĩnh vực được đào tạo sẽ có tiền lương thấp hơn những người lao động làm không đúng lĩnh vực được đào tạo.

Một nghiên cứu khác về vai trò tín hiệu của giáo dục là so sánh tỷ lệ học đại học với tỷ lệ bỏ học ở phổ thông. Nếu người lao động có năng suất cao phải bị giới hạn bởi số năm anh ta có thể học (ví dụ như việc thiếu thốn mạng lưới đại học,...) khi đó người có năng suất thấp sẽ có xu hướng kết thúc THPT hoặc nếu có học đại học thì sẽ có xu hướng bỏ dở. Khi việc gia nhập đại học lúc này đảm bảo những người kỹ năng thấp biết rằng trạng thái cân bằng sẽ được phân rõ, khi đó những người thuộc nhóm này sẽ thiếu động lực, và cho rằng việc học là mất nhiều chi phí. Bedard (2001) tính toán dựa trên bộ số liệu NLSYM (National Longitudinal Surveys of Young Men) và NLSYW (National Longitudinal Surveys of Young Women) cho thấy có thể quan sát tín hiệu trên thị trường lao động bởi do việc vào đại học sẽ làm học sinh bỏ học ở THPT. Một nghiên cứu khác dựa trên ý tưởng của Bedard là nghiên cứu của Jarosław Hornowski (2010) cho thấy bằng chứng người lao động đưa ra tín hiệu về kỹ năng của họ trong thị trường lao động Ba Lan. Ông tìm thấy bằng chứng của một tích cực mối quan hệ lớn giữa cơ sở hạ tầng giáo dục đại học và xác suất mà một cá nhân sẽ bỏ học cao, tác giả sử dụng một mẫu của các PGSS (Polish General Social Survey – Social Diagnosis), bao gồm các câu trả lời cho khoảng 9.000 người trả lời, cung cấp các thông tin như họ đạt được như giáo dục, nơi sống, và cha mẹ giáo dục đạt được, và các biến kiểm soát khác. Hạn chế mẫu để các năm 1975 - 1998 do cải cách hành chính, tác giả sử dụng mô hình logit, đánh giá tác

động của giáo dục đại học lên sự bỏ học ở THPT cho cả đối tượng nam và nữ. Kết quả cho thấy trình độ giáo dục của bố mẹ càng cao thì tỷ lệ bỏ học ở THPT càng giảm và tỷ lệ tham gia đại học càng tăng, tỷ lệ học đại học cao thì tỷ lệ bỏ học phổ thông càng cao.

Lang và Kropp (1986) nghiên cứu sự tác động của đạo luật giáo dục bắt buộc cho nhóm những người tham gia đại học, nghiên cứu dựa trên việc thu thập dữ liệu từ năm 1908 đến 1970 ở Hoa Kỳ. Trong khoảng thời gian này, các tiểu bang khác nhau áp dụng khác nhau một cách đáng kể đạo luật này và cấu trúc này không bất biến theo thời gian. Tác giả sử dụng phương pháp GLS để ước lượng và các kết quả cho thấy giáo dục có đóng vai trò tín hiệu trong giai đoạn này.

Ngoài ra còn có các nghiên cứu đánh giá vai trò phát tín hiệu của giáo dục thông qua so sánh hiệu suất sinh lợi của giáo dục qua đặc điểm khu vực lao động, một trong những cách đánh giá vai trò tín hiệu và tính toán suất sinh lời của giáo dục là phân biệt giữa tác động của giáo dục lên năng suất giữa nhóm người làm thuê và nhóm người tự làm. Với người lao động tự làm, vai trò chủ yếu của giáo dục sẽ là cung cấp kiến thức và kỹ năng nhằm nâng cao hiệu quả kinh doanh, do đó thu nhập của những người tự làm sẽ phản ánh trung thực vai trò của giáo dục như là một nguồn cung cấp vốn nhân lực cho họ. Trong khi đó, với người lao động làm thuê, giáo dục sẽ có cả hai vai trò: vừa cung cấp vốn nhân lực, vừa phát tín hiệu. Chẳng hạn nghiên cứu của Brown và Sessions (1999) so sánh hiệu suất sinh lợi của giáo dục đem lại cho người lao động làm thuê và tự làm, kết quả cho thấy giáo dục không chỉ làm tăng năng suất cá nhân cho người làm thuê nó còn là tín hiệu cho năng lực vốn có của họ.

Trong một nghiên cứu tương tự, John S. Heywood và XiangDong Wei (2004) sử dụng hàm tiền lương Mincer ước lượng hiệu suất sinh lợi của giáo dục cho cả người làm thuê và người tự làm thị trường lao động cạnh tranh ở Hồng Kông trước năm 1997. Kết quả ước lượng cho thấy, hiệu suất sinh lợi của giáo dục mang lại cho những người làm thuê sẽ cao hơn những người tự làm ở mỗi cấp độ giáo dục, người làm thuê đã sử dụng giáo dục như là tín hiệu cho năng suất của mình. Tuy nhiên các nghiên cứu trên chưa tính toán đến tác động của số năm kinh nghiệm làm việc của người lao động đến tiền lương của họ, nếu giáo dục đại học đóng vai trò tín hiệu về năng lực của người lao động thì ở cùng một cấp độ giáo dục người làm thuê sẽ có mức lương trung bình cao

hơn tự làm, sự khác biệt này chỉ đúng khi số năm kinh nghiệm của người lao động còn ít, lúc này người sử dụng lao động chưa nhận ra năng lực thực sự của người lao động làm thuê. Sau khi số năm kinh nghiệm tăng dần thì sự khác biệt đó dần mất đi, lúc này vai trò tín hiệu của giáo dục sẽ không còn nữa. Ngoài ra còn có nhiều nghiên cứu tính toán sự khác biệt của hiệu suất sinh lợi của giáo dục cho người lao động làm việc trong các khu vực kinh tế tư nhân, kinh tế nhà nước, của người lao động làm thuê và người lao động tự làm, ..., như các nghiên cứu của Wolpin (1977) và Riley (1979), Heywood, J & Wei, X (2004).

Tuy nhiên hạn chế của các nghiên cứu trên chưa tính đến việc khi số năm kinh nghiệm tăng lên thì vai trò phát tín hiệu của giáo dục cho người lao động có giảm dần hay là không còn hay không. Vai trò phát tín hiệu của giáo dục thực tế chỉ xuất hiện khi số năm kinh nghiệm của người lao động còn ít, lúc này người sử dụng lao động chưa thể biết được nhiều về năng lực thực sự của người lao động, khi đó giáo dục sẽ được hiểu như tín hiệu về năng lực này. Khi đã làm lâu năm trong công việc, sự bất đối xứng thông tin giữa người lao động và chủ lao động giảm đi, học vấn không còn vai trò phát tín hiệu nữa, năng suất lao động sẽ thể hiện đầy đủ hơn về năng lực của người lao động mà không cần đến tín hiệu nữa.

Bảng 1.1: Tóm tắt các các nghiên cứu đánh giá vai trò phát tín hiệu của giáo dục thông qua so sánh hiệu suất sinh lợi của giáo dục qua đặc điểm, khu vực lao động:

Tác giả	Kết quả	Quốc doanh và Tư nhân	Tự làm và làm thuê	Phương pháp	Giới tính	Quốc gia	Năm
Kugler & Psacharopolous	Có		x	R	Toàn bộ	Argentina	1989
Miller & Volker	Có	X		R	Nam, Nữ	Australia	1984
Gomez-Castellans & Psacharopolous	Có		x	R	Toàn bộ	Equador	1990
	Không	X		R	Toàn bộ	Equador	1990
Lambropoulous	Không	X		R	Toàn bộ, Nam	Greece	1992
Katz & Zimmermann	Có		x	E	Nam	Israel	1980

Tác giả	Kết quả	Quốc doanh và Tư nhân	Tự làm và làm thuê	Phương pháp	Giới tính	Quốc gia	Năm
Zidemann	Không		x	R	Nam	Israel	1992
Brown & Sessions	Có		x	R	Nam	Italy	1999
	Có	X		R	Nam	Italy	1999
Al-Qudsi	Không	X		R	Toàn bộ	Kuwait	1989
Soon	Có		x	R	Nam	Malaysia	1987
Lee	Không	X		R	Nam	Malaysia	1980
Guisinger et al	Có		x	R	Nam	Pakistan	1984
Psacharopoulous et al.	Không	X		R	Toàn bộ	Paraguay	1994
	Không		X	R	Nam	Paraguay	1994
	Có		X	R	Nữ	Paraguay	1994
Alba-ramirez & Segundo	Không		X	R	Nam,Nữ	Spain	1995
	Có	X		R	Nam,Nữ	Spain	1995
Arabshiebani & Rees	Không	X		R	Nam	United Kingdom	1997
Brown & Sessions	Có		X	R	Nam	United Kingdom	1998
Cohnetal	Không		X	E	Nam	United States	1987
	Không		X	R	Nam	United States	1987
	Không	X		R	Nam	United States	1987
Grubb (highschool)	Có		X	R	Nam	United States	1993
(university)	Không		X	R	Nam	United States	1993
Hamilton	Có		X	R	Nam	United States	2000
Tucker	Không		X	R	Toàn bộ	United States	1985
Psacharopolous & Steier	Có		X	R	Nam	Venezuela	1988
	Không	X		R	Nam	Venezuela	1988

X: Ước lượng sự khác biệt của cấp độ giáo dục

R: Ước lượng sự khác biệt hiệu suất sinh lợi của giáo dục

Nguồn: Heywood & Wei (2004)

Một vấn đề cần tính đến khi tính toán hiệu suất sinh lời của giáo dục cần tính đến năng lực bẩm sinh của người lao động, đây cũng là yếu tố góp phần vào tiền lương. Có nhiều nghiên cứu đã tính đến năng lực bẩm sinh của người lao động như nghiên cứu của Jim Kjelland (2008) sử dụng phương pháp hồi qui tiền lương theo các cấp bậc giáo dục, tác giả sử dụng bộ số liệu của Hoa Kỳ là NLSY (The National Longitudinal Survey of Youth) và điểm số của AFQT (Armed Forces Qualification Tests) nhằm đánh giá sự thông minh vốn có của người lao động, điểm số này được đưa vào như một biến tác động đến tiền lương của người lao động. Tính toán cho thấy khi hệ số của điểm số AFQT là dương và làm giảm sự tác động của giáo dục lên tiền lương, đây là dấu hiệu cho thấy giáo dục đóng vai trò là tín hiệu.

Monojit Chatterji và các cộng sự (2003) chỉ ra tín hiệu của giáo dục bằng đo lường sự khác biệt giữa trình độ được yêu cầu và trình độ cần thiết của người lao động phụ thuộc vào qui mô công ty, đặc điểm công việc, địa lý, và các thuộc tính cá nhân. Tác giả sử dụng mô hình Order probit để tính xác suất của việc lựa chọn mức độ quan trọng của các trình độ giáo dục của người công nhân, xác suất này được xem như là biến độc lập trong hàm tiền lương của người công nhân, ước lượng cho thấy biến này làm giảm tác động của số năm đi học của người công nhân lên tiền lương, và đóng góp vào tiền lương của người công nhân là gần 10%.

Nghiên cứu của Baris Kaymaky (2008), trong đó đã nghiên cứu hiệu suất đầu tư cho giáo dục sau phổ thông của người dân Mỹ có sử dụng số liệu điều tra về thanh niên quốc gia. Trong nghiên cứu này, tác giả đã ước lượng vai trò phát tín hiệu của giáo dục thông qua tốc độ khám phá của chủ lao động về năng suất thực của người lao động (speed of learning), kết quả cho thấy rằng vai trò phát tín hiệu chiếm khoảng 22% cho những người có năng lực thấp và khoảng 1% cho những người có năng lực cao, trong đó năng lực được xấp xỉ theo kết quả kiểm tra quốc gia AFQT. Nhiều nghiên cứu khác cũng cho các kết quả tương tự (Farber và Gibbons (1996); Lange (2005), Jed DeVaro và Michael Waldmancho (2012)) cho thấy vai trò phát tín hiệu của giáo dục là phổ biến mặc dù tại các quốc gia được nghiên cứu thì người tuyển dụng đều phải trải qua các kỳ sát hạch hoặc thời kỳ tập sự chặt chẽ.

Ở Việt Nam cũng đã có một số nghiên cứu về hiệu suất sinh lời của giáo dục, trong đó có nghiên cứu của Trần Thị Tuấn Anh (2014) sử dụng hàm tiền lương dạng Mincer ước lượng tiền lương người lao động Việt Nam giai đoạn 2002-2010. Kết quả cho thấy có sự chênh lệch tiền lương giữa người lao động ở khu vực thành thị và nông thôn, giữa nam giới và nữ giới, sự chênh lệch giữa các nhóm có xu hướng giảm dần theo thời gian. Trình độ học vấn thể hiện qua bằng cấp cũng thực sự tác động đến tiền lương. Bằng cấp càng cao thì mức lương nhận được càng lớn. Tác giả sử dụng thủ tục Heckman 2 bước để giải quyết vấn đề ước lượng do chọn mẫu tuy nhiên tính toán vẫn chưa tính các yếu tố không quan sát được có tương quan với giáo dục tác động đến tiền lương của người lao động.

Đáng kể là nghiên cứu của Nguyễn Xuân Thành (2006). Trong nghiên cứu này, tác giả đã sử dụng phương pháp khác biệt kép với hàm Mincer cho số liệu Điều tra mức sống hộ gia đình (VHLSS) năm 2002 để ước lượng suất sinh lợi của giáo dục phổ thông. Phương pháp khác biệt kép giúp loại trừ vấn đề về thiếu biến không quan sát được - năng lực bẩm sinh của người học -, ngoài ra tác giả cũng tận dụng một đặc điểm của hệ thống giáo dục Việt Nam là năm 1988-1989, hệ thống giáo dục phổ thông miền Bắc được chuyển đổi từ hệ 10 năm thành hệ 12 năm. Kết quả tìm được cho thấy giáo dục phổ thông đem lại lợi suất ở mức khá cao: mỗi năm học tập phổ thông giúp tăng năng suất lao động được 11.43%, khá tương đồng với các nghiên cứu khác (Psachropoulos, 1994). Nghiên cứu của Nguyễn Dũng Tiến (2015) ước lượng suất sinh lời của giáo dục trên thị trường lao động Việt Nam bằng phương pháp tác động ngẫu nhiên với hàm Mincer cho bộ số liệu lao động Việt Nam năm 2012, tính toán có tính đến các yếu tố không quan sát được tác động đến tiền lương của công nhân. Hay gần đây có nghiên cứu của hay Tinh Doan và cộng sự (2016). Các tác giả đã ước lượng suất sinh lời của giáo dục cho các năm từ 1998-2014 tính theo số năm đi học và đã tìm thấy xu hướng giảm của suất sinh lời trong giai đoạn nghiên cứu. Trong nghiên cứu này, các năm đi học được xem là có vai trò như nhau trong đóng góp cho suất sinh lời. Tuy nhiên các nghiên cứu trên thế giới đều cho thấy vai trò của một năm học ở bậc phổ thông thường khác biệt so với một năm học sau phổ thông, một năm học ở đại học sẽ khác với một năm học ở

các trường nghề (Heckman, 2003). Các nghiên cứu này đều không xét đến tính nội sinh của học vấn cũng như vấn đề phát tín hiệu của giáo dục.

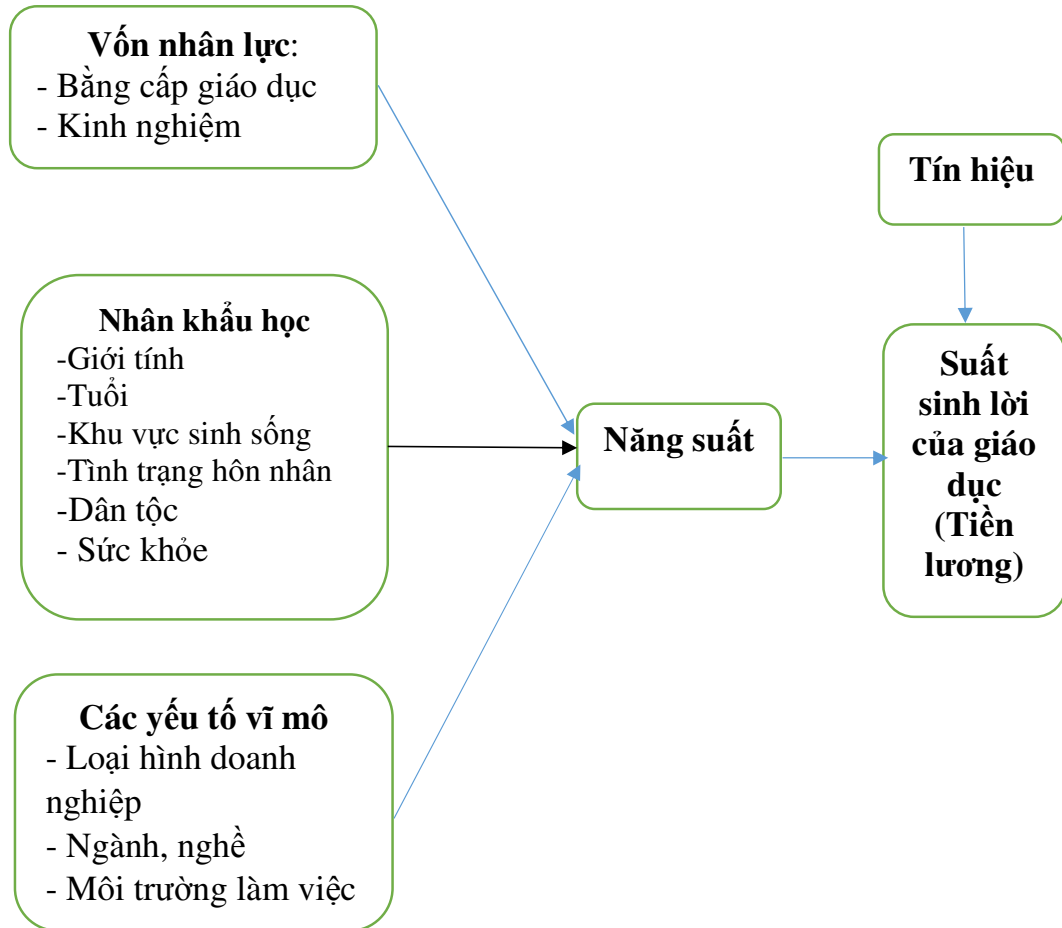
Như vậy, có thể nói các nghiên cứu trên thế giới về suất sinh lời của giáo dục và vai trò phát tín hiệu là khá phong phú, các kết quả đều cho thấy giáo dục không chỉ cung cấp kiến thức nhằm gia tăng năng suất lao động mà còn đóng vai trò phát tín hiệu về năng lực của người lao động, giúp giảm thiểu sự bất đối xứng về mặt thông tin giữa người lao động và người sử dụng lao động. Ở Việt Nam tuy đã có nhiều nghiên cứu về suất sinh lời của giáo dục nhưng các nghiên cứu này không tính đến vấn đề thông tin bất đối xứng. Do đó các khuyến nghị sẽ không còn đảm bảo tính chính xác. Xuất phát từ khoảng trống này, luận án sẽ tập trung nghiên cứu về suất sinh lời của giáo dục trong đó có tính đến thông tin bất đối xứng và vai trò phát tín hiệu của giáo dục Việt Nam. Do giáo dục chỉ có vai trò phát tín hiệu đối với giáo dục sau phổ thông, nên luận án tập trung ở cấp độ này của giáo dục.

Phân tích tổng quan cho thấy tại Việt Nam, đã có một số nghiên cứu về suất sinh lời của giáo dục, tuy nhiên các nghiên cứu này chưa đề cập đến vai trò phát tín hiệu, do đó kết quả ước lượng có thể không đáng tin cậy. Ngoài ra, việc kiểm soát vấn đề biến nội sinh chủ yếu được các tác giả thực hiện bằng phương pháp Heckman. Tuy nhiên phương pháp này khi thực hiện cho thị trường lao động Việt Nam là có một số hạn chế, vì vậy Luận án sử dụng phương pháp Lewbels nhằm giải quyết vấn đề này. Ngoài ra phương pháp phi tham số, với ưu điểm là việc ước lượng không cần dựa trên một giả định về dạng hàm, được sử dụng nhằm để kiểm tra và đảm bảo độ tin cậy của các kết quả thu được từ các phương pháp ước lượng tham số nói trên.

1.4. Khung phân tích của luận án

Phân tích tổng quan cho thấy suất sinh lời của giáo dục (tiền lương) bị ảnh hưởng bởi nhóm nhân tố chính sau: i) bằng cấp giáo dục, kinh nghiệm làm việc; ii) nhân khẩu học: bao gồm tuổi, giới tính, khu vực sinh sống, sức khỏe, tình trạng hôn nhân, dân tộc; iii) các yếu tố vĩ mô: loại hình doanh nghiệp, ngành, nghề, môi trường làm việc; iv) vai trò phát tín hiệu của giáo dục vai trò này xuất phát từ thông tin bất đối xứng trong thị trường lao động. Do vậy các nhóm biến này sẽ được sử dụng chính trong các mô hình phân tích định lượng để đánh giá hiệu suất sinh lời của giáo dục sau phổ thông.

Khung phân tích của luận án được mô tả trong Hình 1.3 dưới đây:



Hình 1.3: Khung phân tích của luận án

Nguồn: Nghiên cứu của tác giả

Theo khung phân tích trên, luận án sẽ xây dựng các mô hình toán kinh tế để đánh giá hiệu suất sinh lời của giáo dục. Trong đó suất sinh lời của giáo dục, theo thông lệ các nghiên cứu cùng chủ đề, được đo bởi tiền lương của người lao động. Trong đó tiền lương được cấu thành từ hai yếu tố chính, là năng suất lao động và tín hiệu về năng suất lao động. Việc đưa yếu tố tín hiệu vào phân tích, như tổng quan đã chỉ ra, là do vấn đề thông tin bất đối xứng giữa người lao động và người chủ lao động về năng suất của người lao động, nhất là trong thời kỳ đầu của hợp đồng lao động.

Các biến độc lập được xem xét đến bao gồm các biến số có tác động trực tiếp đến năng suất lao động, gồm Vốn nhân lực, Nhân khẩu học, Các

yếu tố vĩ mô Bên cạnh biến độc lập chính là biến giáo dục (đo lường vốn nhân lực) thì yếu tố tín hiệu cũng là mối quan tâm chính của luận án. Do cùng với quá trình lao động thì sự bất đối xứng thông tin giữa chủ lao động và lao động sẽ giảm dần, vai trò của tín hiệu trong việc xác định tiền lương sẽ giảm dần, luận án cũng sẽ kiểm soát vấn đề này trong phân tích ở chương 4.

CHƯƠNG 2

PHƯƠNG PHÁP NGHIÊN CỨU

Chương này sẽ giới thiệu các phương pháp ước lượng suất sinh lời của giáo dục, đầu tiên là hàm tiền lương Mincer, là cơ sở của các mô hình kinh tế lượng về tiền lương. Tiếp theo là phương pháp PSM, mô hình Heckman và phương pháp Lewbels để giải quyết vấn đề biến nội sinh do lựa chọn ngược – thường xuất hiện trong các bài toán ước lượng hàm tiền lương, cuối cùng phương pháp phi tham số Kernel là phương pháp là không phụ thuộc vào dạng hàm về mối quan hệ giữa các biến số.

2.1. Hàm tiền lương Mincer

Miner (1958, 1974) là người đầu tiên đưa ra công thức thực nghiệm về tiền lương, trong công thức của ông tại thời điểm t trong cuộc đời của một cá nhân, tiền lương quan sát được có thể được miêu tả như hàm lõm theo kinh nghiệm của người lao động. Giả sử rằng giai đoạn đầu tư cho giáo dục của một cá nhân là S năm và công việc được đào tạo lại giảm tuyến tính theo vòng đời¹, logarit tiền lương là một hàm bậc hai theo kinh nghiệm làm việc.

$$\ln Y_i(t) = a_0 + a_1 S_i + a_2 t_i + a_3 t_i^2 + \varepsilon_i \quad (2.1)$$

Trong đó a_0 liên quan đến khả năng tiền lương ban đầu, a_2 là suất sinh lợi của giáo dục (giả định rằng tất cả các chi phí giáo dục là chi phí cơ hội), a_3 liên quan đến hiệu suất sinh lợi của việc được đào tạo sau khi hoàn thành việc học. Phương trình (2.1) còn được gọi là hàm thu nhập Mincer.

Việc xây dựng hàm thu nhập được Mincer thực hiện như sau:

Gọi C_t là chi phí đầu tư cho vốn nhân lực ở năm t . Khi đó, tiền lương trong một giai đoạn là:

$$E_1 = E_0 + rC_0 \quad (2.2)$$

¹ Giai đoạn ban đầu là kết quả của một giải pháp của quá trình tối đa hóa của việc đầu tư vào chuyên môn thuần túy và được gọi là đi học, trong quá trình này giá trị vốn nhân lực sẽ tăng theo cấp bậc giáo dục. Giai đoạn tiếp theo cá nhân cả làm việc và đầu tư. Giai đoạn này còn được gọi là đầu tư sau khi học (PSI-post school investment) hoặc đào tạo lại (OJT- on the job training). Một điều quan trọng giai đoạn này: thời gian đầu tư giảm đơn điệu, có nghĩa là mỗi cá nhân mỗi năm tạo ra ít nguồn nhân lực hơn bởi vì cá nhân đó sẽ già hơn.

Trong đó E_0 đại diện cho tiền lương tiềm năng của một cá nhân dựa trên năng lực bẩm sinh.

$$\text{Tương tự: } E_2 = E_1 + rC_1 = E_0 + rC_0 + rC_1 \quad (2.3)$$

Bằng cách đệ quy ta có:

$$E_t = E_0 + r \sum_{i=0}^{t-1} C_i \quad (2.4)$$

Bởi vì thực tế rất khó có được dữ liệu về số tiền đầu tư vào vốn nhân lực, Mincer định nghĩa $k_t = \frac{C_t}{E_t}$ đại diện cho tỷ lệ thu nhập của một cá nhân chi cho việc đầu tư vào vốn nhân lực. Tỷ lệ này xấp xỉ phần trăm thời gian của một cá nhân đầu tư vào vốn nhân lực. Thay C_t bởi k_t , ta có

$$E_t = E_0 \prod_{i=0}^{t-1} (1 + rk_i) \quad (2.5)$$

Logarit hai vế ta được:

$$\ln E_t = \ln E_0 + \sum_{i=0}^{t-1} \ln(1 + rk_i) \quad (2.6)$$

Khi rk_i nhỏ thì $\ln(1 + rk_i) \approx rk_i$. Khi đó (2.6) được viết lại như sau:

$$\ln E_t = \ln E_0 + r \sum_{i=0}^{t-1} k_i \quad (2.7)$$

Trong suốt quá trình đi học k_i sẽ bằng 1 bởi vì lúc này đi học về bản chất là sự hoạt động toàn thời gian. Tuy nhiên sau khi việc học kết thúc, k_i sẽ giảm đơn điệu đến 0 khi cá nhân đó bắt đầu đi làm cho đến nghỉ hưu. Điều này ngụ ý rằng k_i có thể chia thành 2 phần: (1) thời gian đi học trong đó $k_i = 1$ với tất cả i , (2) giai đoạn đầu tư sau khi học, trong đó sự đầu tư giảm đơn điệu. Như vậy (2.7) có thể viết lại:

$$\ln E_t = \ln E_0 + r_S S + r_P \sum_{i=0}^{t-1} k_i \cong \ln E_0 + r_S S + r_P \int_0^t k_j dj \quad (2.8)$$

Trong đó r_S hiệu suất sinh lợi của việc đi học mang lại, r_P là hiệu suất sinh lợi của việc đầu tư sau khi đi học, và việc đầu tư sau đi học được thể hiện trong thời gian liên tục.

Mincer thực nghiệm với 4 chi tiết: (2.1) số tiền đầu tư cho đào tạo lại giảm tuyến tính, (2.2) thời gian đầu tư tương đương cho đào tạo lại giảm tuyến tính, (2.3) lũy thừa của số tiền đầu tư cho đào tạo lại giảm tuyến tính, (2.4) lũy thừa thời gian đầu tư tương đương cho đào tạo lại giảm tuyến tính. Trong đó lũy thừa thời gian đầu tư tương đương cho đào tạo lại giảm tuyến tính làm hàm được sử dụng nhiều nhất. Mincer giả định:

$$k_t = k_0 - \frac{k_0}{T}t \quad (2.9)$$

Trong đó k_0 là đầu tư tương đương về thời gian ở thời điểm ban đầu, và T là tổng số thời kì đầu tư. Thay (2.9) vào (2.8) ta có:

$$\ln E_t = \ln E_0 + r_s S + r_p k_0 t - \frac{r_p k_0}{2T} t^2 \quad (2.10)$$

Cho thấy tiền lương tiềm năng của một người là hàm bậc 2 của năm kinh nghiệm. Vấn đề là tiền lương thực tế không bằng tiền lương tiềm năng bởi vì người làm công ăn lương chi tiêu một phần từ tiền lương để đầu tư. Do đó tiền lương thực tế sẽ là:

$$Y_t = (1 - k_t)E_t \quad (2.11)$$

Logarit hai vế ta có:

$$\ln Y_t = \ln E_0 + r_s S + r_p k_0 t - \frac{r_p k_0}{2T} t^2 + \ln(1 - k_t) \quad (2.12)$$

Thay (2.9) vào (2.12) được:

$$\ln Y_t = \ln E_0 + r_s S + r_p k_0 t - \frac{r_p k_0}{2T} t^2 + \ln\left(1 - k_0 + \frac{k_0}{T}t\right) \quad (2.13)$$

Lấy xấp xỉ Taylor bậc 2 của hàm $\ln\left(1 - k_0 + \frac{k_0}{T}t\right)$

Và đặt $a_0 = \ln E_0 - k_0\left(1 + \frac{k_0}{2}\right)$

$$a_1 = r_s$$

$$a_2 = r_p k_0 + \frac{k_0}{T}(1 + k_0)$$

$$a_3 = -\frac{r_p k_0}{2T} - \frac{k_0^2}{2T^2}$$

Thay vào (2.13) ta được phương trình (2.1)

Chức năng của hàm tiền lương Mincer đạt được ít nhất hai hàm ý cơ bản: Thứ nhất nó cho rằng tiền lương liên quan đến đầu tư nguồn nhân lực, điều này có nghĩa là đầu tư cho vốn nhân lực nhiều hơn của một cá nhân làm cho tiền lương sẽ tăng lên. Hệ số của biến học vấn phản ánh hiệu suất sinh lợi của học vấn. Thứ hai hàm tiền lương là hàm lõm. Thu nhập sẽ tăng nhanh cho những người còn trẻ, sau đó tăng trưởng thu nhập sẽ giảm xuống cho nhóm trung niên. Ngày nay hầu hết các hàm tiền lương bao gồm nhiều biến bổ sung bên cạnh các vấn đề về học vấn và số năm kinh nghiệm mà Mincer sử dụng trong hàm tiền lương. Bao gồm chủng tộc, giới tính, các biến giả về khu vực sống, tình trạng sức khỏe, dân tộc, tình trạng hôn nhân, số con trong gia đình, thành viên công đoàn, và nhiều biến số khác. Các biến ngoại sinh này được sử dụng như các biến kiểm soát của mô hình.

Một số vấn đề khi sử dụng hàm tiền lương Mincer

Không đồng nhất do những biến không quan sát được (Unobserved Heterogeneity)

Phân tích hồi qui đa biến, bao gồm cả phương trình ước lượng đồng thời kết hợp với sự điều chỉnh về sự lựa chọn chệch được dựa trên các đặc tính quan sát được của các cá nhân tác động đến tiền lương. Vấn đề này thường có các biến số quan trọng kết hợp, nhưng chúng bị bỏ qua do không có dữ liệu về những đặc tính không thể quan sát được. Ví dụ có thể nói rằng những người thông minh sẽ có nhiều hơn năm học tập so với người khác, nhưng bởi cá nhân thông minh tự nhiên sẽ kiếm thêm thu nhập độc lập với việc học tập. Nếu càng có thể đi học dài hơn, thì hệ số của biến giáo dục sẽ bị chệch và có xu hướng tăng lên bởi vì hiệu suất sinh lợi của giáo dục có một phần đóng góp của năng lực bẩm sinh không đo được. Như vậy mô hình hồi quy sử dụng dữ liệu cá nhân bỏ qua năng lực bẩm sinh không quan sát được của người lao động sẽ bị chệch.

Sự lựa chọn (selectivity)

Việc ước lượng chệch còn do sự lựa chọn mẫu. Điều này là do dữ liệu được sử dụng trong ước lượng không đủ độ ngẫu nhiên. Ví dụ như có nhiều nghiên cứu đánh giá thấp khoảng cách tiền lương theo giới tính, vì có nhiều phụ nữ không tham gia thị trường lao động hơn nam giới. Nếu mẫu của phụ nữ trên

thị trường lao động là ngẫu nhiên thì thực tế sẽ không có ước lượng chệch khi đánh giá khoảng cách tiền lương giữa nam và nữ. Tuy nhiên có thể nghi ngờ rằng việc phụ nữ đi làm không phải là ngẫu nhiên vì một người chỉ làm việc khi mức lương đề nghị cao hơn mức lương kì vọng. Do đó nếu bao gồm những người làm việc dưới mức lương kì vọng thì sẽ có sự khác biệt lớn giữa tiền lương của nam và nữ.

2.2. Phương pháp ước lượng điểm thiên hướng (propensity score method)

2.2.1. Vấn đề khi đánh giá sự tác động của một chương trình

Ở cấp độ cá nhân chúng ta quan sát đầu ra của thị trường lao động của những người được đào tạo và những người không nhận được đào tạo. Để biết được hiệu quả của việc đào tạo về một cá nhân tham gia chúng ta phải so sánh đầu ra với cá nhân không tham gia chương trình. Tuy nhiên thực tế chỉ có một đầu ra được quan sát, và được gọi là đầu ra thực tế hay còn gọi là đầu ra của những cá nhân được tác động bởi chương trình. Và đầu ra không thực tế là kết quả của cá nhân tham gia và không nhận được sự tác động của chương trình. Đầu ra này không thể quan sát được và đây là vấn đề của việc ước lượng, do đó một trong những khó khăn của đánh giá chương trình đó là dữ liệu thiếu.

Không chắc chắn là tất cả các cá nhân sẽ phản ứng lại một cách chính xác sự can thiệp của một chính sách theo cùng một cách. Thay vào đó có sự không đồng nhất trong tác động giữa các cá nhân. Điều này đưa ra hai câu hỏi mà việc đánh giá tác động cần giải quyết. Đầu tiên là những tác động gì chương trình sẽ có trên một cá nhân lấy ngẫu nhiên từ tổng thể, còn gọi là ATE (average treatment effect). Thứ hai là những tác động gì của sự tham gia chương trình đến những cá nhân thực sự tham gia chương trình – ATET (average effect of treatment on the treated). Hai tác động này sẽ giống nhau nếu chúng ta giả định sự phản hồi của cá nhân là như nhau. Tuy nhiên khi thực tế hơn về sự phản hồi khác nhau giữa các cá nhân thì kết quả đo lường sẽ khác nhau. Một ví dụ về một chương trình tự nguyện, chúng ta có thể dự đoán rằng những người tình nguyện sẽ khác tổng thể nếu xét về lợi ích kỳ vọng của họ từ chương trình: bởi vì họ nhận thấy lợi ích từ sự tham gia và sẽ tham gia đầu tiên. Và do đó sẽ không chắc chắn rằng việc ước lượng tác động cho những người tham gia sẽ liên quan đến những người không tham gia hội đủ điều kiện. ATET

có thể chỉ ra lợi ích trung bình của sự tham gia, còn ATE sẽ liên quan đến việc quan tâm chính sách tập trung vào thực hiện một chương trình bắt buộc.

Điều quan trọng là các nhà hoạch định chính sách cần nhận thấy được sự tác động khác nhau của chương trình với 2 lí do. Thứ nhất, khi so sánh các kết quả nghiên cứu, người đọc cần nhận thấy được sự tác động của chương trình mà nghiên cứu đang giải quyết. Nói chung nếu có lợi ích kỳ vọng đem lại lớn nhất khi tham gia, ATE sẽ nhỏ hơn ATET. Thứ hai, các câu hỏi chính sách khác nhau được giải quyết bởi sự tác động khác nhau. Ví dụ, ATET là tham số có thể trả lời câu hỏi chính sách liệu rằng chương trình có nên thực hiện hay không, bởi vì nếu trung bình tác động có lợi dưới mức chi phí mà mỗi người tham gia bỏ ra, thì chương trình cần được loại bỏ. Khi quyết định có đưa ra một chương trình bắt buộc, khi mở rộng cho toàn bộ mẫu, câu hỏi trở thành liệu chương trình bắt buộc có thỏa mãn vấn đề lợi ích và chi phí hay không. Trong trường hợp này tham số quan tâm là ATE, được dùng để tính toán sự tác động không đồng đều của chương trình.

Để làm rõ các vấn đề liên quan đến việc đánh giá một chương trình, ta có thể xem xét một chương trình đào tạo tự nguyện. Cách tiếp cận để đánh giá hiệu quả của chương trình sẽ là so sánh kết quả của người tham gia với người không tham gia chương trình. Tác động của một chương trình đào tạo có thể đặc trưng bởi hai quá trình riêng biệt. Thứ nhất là quyết định tham gia vào chương trình. Thứ hai là tiến trình xác định công việc (đầu ra), và sự quan tâm chính là liệu việc đào tạo của chương trình có ảnh hưởng đến có được một công việc. Và trong hai quá trình, điều quan trọng là phải nhận ra những đặc tính quan sát được và không quan sát được có thể đóng vai trò.

Câu hỏi đặt ra là sự lựa chọn chệch sẽ gia tăng khi các thành phần của quyết định tham gia chương trình có liên quan đến quá trình xác định công việc thành công. Có nghĩa là sự lựa chọn chệch là kết quả của việc tham gia vào chương trình có ảnh hưởng đến đầu ra. Có thể mối liên hệ giữa hai quá trình có thể được tính do các đặc điểm có thể quan sát được. Trong trường hợp này sự lựa chọn chệch có thể tránh được bằng cách bao gồm các biến có liên quan trong phương trình giải thích kết quả đầu ra. Trong một số trường hợp khác các đặc tính không quan sát được ảnh hưởng đến quá trình tham gia chương trình cũng tác động đến đầu ra của chương trình.

Một việc khó khăn trong thực tế là có thể tìm các nhóm có thể so sánh có cùng đặc điểm. Thay vào đó người ta thường quan sát thấy rằng những người tham gia chương trình có những đặc điểm khác nhau với những người không tham gia, nghĩa là sự lựa chọn tham gia chương trình là nội sinh. Vấn đề nội sinh không mất đi hoàn toàn ngay cả khi các đặc điểm quan sát được được kiểm soát, nếu các đặc điểm không quan sát được liên quan chặt chẽ với việc lựa chọn tham gia chương trình và các đặc điểm có thể quan sát được. Ví dụ những người có năng lực và động lực có nhiều khả năng tham gia chương trình đào tạo hơn là những người ít năng lực và động lực. Ngoài ra những người có khả năng cao và động lực cao hơn có thể có xu hướng muốn trình độ học vấn cao hơn và thể hiện năng lực của mình hơn các cá nhân khác trên thị trường lao động. Trong trường hợp này, việc so sánh đơn giản giữa người tham gia và không tham gia chương trình hoặc một hồi qui đơn giản không tính đến vấn đề nội sinh sẽ dẫn đến việc đánh giá không thực sự chính xác hiệu quả của chương trình. Trong chừng mực mà các đặc điểm không thể quan sát được gắn liền với các đặc điểm có thể quan sát, một so sánh đơn giản hoặc một hồi qui đơn giản sẽ dẫn đến một hiệu ứng là chương trình bị đánh giá quá cao.

2.2.2. Phương pháp ước lượng điểm thiên hướng PSM

Việc lựa chọn làm thuê được xem như là việc lựa chọn tham gia một chương trình và việc lựa chọn tự làm như là việc không tham gia chương trình. Để tính toán được tác động của chương trình này một cách đúng đắn thì điều kiện cần thiết là sự tham gia vào chương trình phải là hoàn toàn ngẫu nhiên. Tuy nhiên trong đa số các nghiên cứu kinh tế xã hội, thì việc tham gia hay không tham gia chương trình không phải là một sự lựa chọn ngẫu nhiên, mà thường là có vấn đề về sự lựa chọn ngược. Chẳng hạn một số người lựa chọn tự làm là do họ không đủ năng lực để được tuyển dụng làm thuê, hoặc họ là người có tính cách thích thử thách và muốn tự mình làm việc cho mình. Sự lựa chọn không ngẫu nhiên này có thể dẫn đến việc ước lượng tác động của chương trình bị chệch. Một trong những biện pháp để giảm thiểu vấn đề này là sử dụng phương pháp PSM.

Ý tưởng của phương pháp PSM là như sau: Đó là chỉ ra một nhóm các quan sát của những cá nhân không tham gia chương trình (nhóm control) nhưng có những đặc tính giống như những cá nhân tham gia chương trình (nhóm

treated), sau đó so sánh sự khác biệt giữa các cá thể trong nhóm control với các đối ứng có cùng đặc tính trong nhóm treated để đưa ra kết luận.

Phương pháp PSM dựa trên giả định:

Giả định 1: Cho một tập các biến X có những thuộc tính quan sát được, và X không bị ảnh hưởng bởi chương trình (treatment), sự khác nhau có liên quan giữa những người tham gia chương trình và không tham gia chương trình có được là do các đặc tính quan sát được, và với điều kiện những đặc tính này thì sự lựa chọn tham gia chương trình hay không tham gia chương trình là ngẫu nhiên. Giả định này cũng khẳng định rằng những người lao động có cùng đặc tính giống nhau có một xác suất tham gia chương trình hoặc không tham gia chương trình: $0 < P(T=1|X) < 1$. Trong đó T là biến 0-1, nhận giá trị 1 nếu người lao động tham gia chương trình, và 0 là người không tham gia chương trình.

Giả định 2: Đầu ra độc lập với việc tham gia hay không tham gia chương trình theo các đặc tính của X . Trên thực tế có nhiều biến X dẫn đến khó khăn trong việc quy định tất cả các đặc tính có liên quan. Đề xuất giải quyết vấn đề này Rosenbaum và Rubin (1983) chỉ ra rằng: Nếu đầu ra của chương trình độc lập với việc tham gia chương trình theo các đặc tính X thì đầu ra cũng độc lập với chương trình theo với điều kiện $P(X) = P(T=1|X)$. Xác suất này được gọi là điểm thiên hướng (propensity score). Các cá nhân có thể được lựa chọn dựa trên thiên hướng $P(X)$ của chương trình thay cho điều kiện của X .

Với hai điều kiện trên, phương pháp PSM đánh giá tác động của chương trình thông ước lượng tham số ATT (average treatment effect on treated)

$$ATT = E_{P(X)|T=1} \{E[Y|T=1, P(X)] - [Y|T=0, P(X)]\} \quad (2.14)$$

Về thực nghiệm, công thức ATT được tính toán bởi công thức sau:

$$\hat{ATT} = \frac{1}{n} \sum_{i \in N} \left(Y_i - \frac{1}{m_i} \sum_{j \in M_j} Y_j \right) \quad (2.15)$$

Trong đó, N chỉ nhóm người tham gia chương trình, n là số người tham gia chương trình, M_i chỉ nhóm người không tham gia chương trình nhưng có điểm thiên hướng tương xứng với điểm thiên hướng của người thứ i trong nhóm có tham gia chương trình, m_i là số người trong nhóm M_i .

Với mỗi cá thể i thuộc nhóm N , có nhiều phương pháp để lựa chọn nhóm M_i tương ứng, bao gồm các phương pháp như: 1 chọn 1, phương pháp điểm gần nhất, phương pháp compa,...

+ *Phương pháp điểm gần nhất (nearest neighbor matching)*: với mỗi quán sát được tác động i , lựa chọn ra quán sát không tham gia chương trình j có những đặc tính gần nhất.

$$\min \|P_i - P_j\|$$

+ *Phương pháp compa (Radius matching)*: với mỗi quán sát được tác động i , lựa chọn ra quán sát không tham gia chương trình j rơi vào một bán kính r

$$\|P_i - P_j\| < r$$

+ *Phương pháp kết hợp hạt nhân (Kernel matching)*: với mỗi quán sát được tác động i và một nhóm các quán sát không tham gia chương trình với trọng số tỷ lệ nghịch với khoảng cách giữa các quán sát bị tác động với các quán sát không tham gia chương trình.

Với sự kết hợp dựa trên điểm thiên hướng, trọng số được tính toán bởi công thức:

$$w(i, j) = \frac{K\left(\frac{P_j - P_i}{h}\right)}{\sum_{j=1}^{n_0} K\left(\frac{P_j - P_i}{h}\right)}$$

h được gọi là hệ số băng thông (bandwidth)

2.2.3. Ưu nhược điểm của phương pháp PSM

Nhược điểm: PSM có hai nhược điểm trong phân tích thực nghiệm. Đầu tiên là việc cần đưa ra giả thiết độc lập có điều kiện. Trong trường hợp việc tham gia chương trình là ngẫu nhiên, cần thấy rằng những cá nhân tham gia và không tham giống nhau cả về những đặc tính quán sát được và không quán sát được. Thứ hai trong khi phương pháp PSM chỉ tính toán được sự tác động của chương trình cho một số cá nhân trong số các cá nhân không nhận được sự tác động, trong khi việc tham gia chương trình ngẫu nhiên đảm bảo rằng có một sự hỗ trợ trên toàn bộ mẫu. Những yêu cầu này làm cho các kỹ thuật thực nghiệm khác rõ ràng hơn PSM. Bên cạnh đó PSM không thể trả lời câu hỏi liên quan

đến ảnh hưởng phân phối của chương trình, chẳng hạn như phần trăm những người tham gia chương trình hưởng lợi.

Ưu điểm: Thứ nhất, phương pháp PSM giải quyết được phần nào vấn đề nội sinh trong việc lựa chọn tham gia chương trình. Thứ hai phương pháp PSM không yêu cầu những giả định như các hàm hồi qui thông thường. Phương pháp hồi qui áp đặt hàm số về mối quan hệ (thường là tuyến tính) mà có thể hoặc không có thể chính xác, các biến số trong mô hình hồi qui thông thường hoặc là không phù hợp với lý thuyết kinh tế hoặc là không phù hợp với dữ liệu.

2.3. Phương pháp Heckman ước lượng mô hình với biến nội sinh

Việc tham gia hay không tham gia vào chương trình lao động được xem như là việc người lao động tham gia hay không tham gia vào một chương trình. Do đó hàm tiền lương đối với người lao động tham gia chương trình là:

$$W_{ii} = X_i \gamma_i + \varepsilon_{ii} \quad (2.16)$$

Đối với người lao động không tham gia chương trình:

$$W_{ni} = X_i \gamma_n + \varepsilon_{ni} \quad (2.17)$$

Trong đó $\varepsilon_{ii}, \varepsilon_{ni}$ là sai số ngẫu nhiên và giả định độc lập với X_i

Đặt $T_i = 1$ nếu người lao động tham gia chương trình, $=0$ nếu không tham gia chương trình. Do đó mức lương sẽ được xác định là: $W_i = T_i W_{ii} + (1 - T_i) W_{ni}$

Tuy nhiên việc lựa chọn tham gia hay không tham gia chương trình là không ngẫu nhiên và tồn tại vấn đề biến nội sinh. Do đó để đánh giá một cách khách quan sự tác động của chương trình cần tính đến vấn đề nội sinh. Ta có thể xem xét vấn đề này như sau:

Gọi C_i là chi phí của việc tham gia chương trình, khi đó:

$$C_i = S_i \delta + \varepsilon_{ci} \quad (2.18)$$

Trong đó: S_i là các biến ngoại sinh độc lập với $(\varepsilon_{ii}, \varepsilon_{ni}, \varepsilon_{ci})$, S_i có thể là các biến X_i

Giả định các cá nhân đều muốn tối đa hóa tiền lương. Khi đó quyết định tham gia chương trình sẽ là:

$$T_i = I(W_{it} - W_{ni} - C_i > 0) = I(Z_i\pi + \varepsilon_i > 0) \quad (2.19)$$

Với $\varepsilon_i = \varepsilon_{it} - \varepsilon_{ni} - \varepsilon_{ci}$, Z_i là vecto các nhân tố chứa trong X_i, S_i

Mô hình hàm tiền lương gộp của người lao động làm thuê và tự làm:

$$W_i = \alpha_n + (\alpha_t - \alpha_n)T_i + X_i\gamma + \eta_i \quad (2.20)$$

Trong đó: $\eta_i = T_i\varepsilon_{it} + (1-T_i)\varepsilon_{ni}$. Sự tác động của chương trình đối với mỗi cá nhân là: $W_{it} - W_{ni}$ và $\alpha = \alpha_t - \alpha_n$ và sự tác động trung bình của chương trình cho mỗi cá nhân tham gia chương trình một cách ngẫu nhiên là:

$$E\{W_{it} - W_{ni} | T_i=1, Z_i\} = \alpha + E\{\eta_i | T_i=1, Z_i\} \quad (2.21)$$

Với giả thiết $E\{\eta_i | X_i, T_i\} = 0$ (2.22) thì α có thể ước lượng bằng OLS. Điều kiện cần thiết để xảy ra (2.22) là $\sigma_{te} = \sigma_{ne} = 0$, điều này ngụ ý rằng những nhân tố không quan sát được của tiền lương không liên quan đến quyết định tham gia chương trình, nghĩa là T_i là ngoại sinh với tiền lương. Trong trường hợp (2.22) không xảy ra cần tính đến vấn đề nội sinh của việc lựa chọn chương trình.

Để ước lượng α và giải quyết vấn đề biến nội sinh, Heckman (1979) đề xuất phương pháp CF (control function). Phương pháp này dựa trên trung bình có điều kiện của W_i được cho bởi T_i và Z_i ,

$$E\{W_i | Z_i, T_i\} = \alpha_n + \alpha T_i + X_i\gamma + E\{\eta_i | Z_i, T_i\} \quad (2.23)$$

$$\text{Ta có } E\{\eta_i | Z_i, T_i\} = T_i E\{\varepsilon_{it} | Z_i, T_i=1\} + (1-T_i) E\{\varepsilon_{ni} | Z_i, T_i=0\}$$

$$\text{Trong đó: } E\{\varepsilon_{ji} | Z_i, T_i\} = \sigma_{je} \lambda_j(Z_i\pi) \quad j = t, n \quad (2.24)$$

$$\text{Với } \lambda_j(Z_i\pi) = E\{\varepsilon_{ji} | Z_i, T_i\} = (1-T_i) \frac{-\phi(Z_i\pi)}{\Phi(-Z_i\pi)} + T_i \frac{\phi(-Z_i\pi)}{1-\Phi(-Z_i\pi)} \quad (2.25)$$

Có thể tóm tắt phương pháp CF như sau: dùng mô hình probit tính toán khả năng tham gia chương trình hay không tham gia chương trình để ước lượng $\pi, \hat{\pi}$ nhằm ước lượng λ ở (2.25), sau đó ước lượng hàm tiền lương và xem λ như là một biến trong hàm tiền lương.

2.4. Phương pháp Lewbels với vấn đề nội sinh

Phương pháp được giới thiệu bởi Lewbels (JBIS 2012) nhằm xác định tham số cấu trúc trong mô hình hồi quy trong trường hợp có vấn đề nội sinh hoặc không đo được khi thiếu những thông tin cần thiết.

Cho Y_1, Y_2 là các biến nội sinh quan sát được, X là vecto của các biến phụ thuộc nội sinh quan sát được, và $\varepsilon = (\varepsilon_1, \varepsilon_2)$ là sai số không quan sát được. Xét một mô hình cấu trúc có dạng sau:

$$Y_1 = X' \beta_1 + Y_2 \gamma_1 + \varepsilon_1 \quad (2.26)$$

$$Y_2 = X' \beta_2 + Y_1 \gamma_2 + \varepsilon_2 \quad (2.27)$$

Hệ này là tam giác khi $\gamma_2 = 0$ (hoặc khi $\gamma_1 = 0$). Ngược lại hệ này sẽ hoàn toàn đồng thời. Các sai số $\varepsilon_1, \varepsilon_2$ có thể tương quan với nhau.

Nếu giả định yếu tố ngoại sinh, $E(\varepsilon X) = 0$ đúng, dạng rút gọn được xác định nhưng nếu thiếu các hạn chế xác định, thì các hệ số cấu trúc không được xác định. Những hạn chế này thường liên quan đến việc đặt các yếu tố chắc chắn của β_1 hoặc β_2 bằng 0, phương pháp biến công cụ sẽ được sử dụng.

Trong nhiều trường hợp ứng dụng, công dụng của một công cụ chỉ gián tiếp ảnh hưởng đến biến phản hồi rất khó để xây dựng.

Cách xác định trong cách tiếp cận của Lewbels đạt được bằng việc hạn chế các mối tương quan của $\varepsilon \varepsilon'$ với X .

Các tham số của mô hình cấu trúc sẽ giữ nguyên không được xác định với giả định sai số phương sai thay đổi chuẩn: rằng $E(\varepsilon \varepsilon' | X)$ là một ma trận hằng số. Tuy nhiên, với phương sai thay đổi chuẩn liên quan đến ít nhất một vài thành phần của X . Cách xác định có thể đạt được.

Trong một hệ hoàn toàn đồng thời, giả sử rằng $\text{cov}(X, \varepsilon_j^2) \neq 0, j = 1, 2$ và $\text{cov}(Z, \varepsilon_1 \varepsilon_2) = 0$ cho Z quan sát được thì ta có thể xác định được các tham số cấu trúc. Chú ý rằng Z có thể là một tập con của X vì vậy không có một thông tin bên ngoài nào của mô hình trên là bắt buộc.

Khẳng định quan trọng nhất mà $\text{cov}(Z, \varepsilon_1 \varepsilon_2) = 0$ sẽ tự động được thỏa mãn nếu các quy trình sai số không trung bình là độc lập có điều kiện: $\varepsilon_1 \perp \varepsilon_2 | Z = 0$. Tuy nhiên sự độc lập này không thực sự cần thiết.

Mô hình nhân tố đơn không quan sát được

Một lớp các mô hình thỏa mãn những khẳng định dựa trên cơ sở của phương pháp Lewbels là trong đó phương trình tương quan sai số chéo bởi vì một thừa số chung không quan sát được:

$$Y_1 = X' \beta_1 + Y_2 \gamma_1 + \varepsilon_1, \quad \varepsilon_1 = \alpha_1 U + V_1 \quad (2.28)$$

$$Y_2 = X' \beta_2 + Y_1 \gamma_2 + \varepsilon_2, \quad \varepsilon_2 = \alpha_2 U + V_2 \quad (2.29)$$

Trong đó U, V_1, V_2 không quan sát được, không tương quan với X và không tương quan có điều kiện với nhau khi phụ thuộc X, V_1, V_2 là các sai số đặc trưng, với U là một biến bị bỏ đi mà có thể ảnh hưởng trực tiếp đến cả Y_1, Y_2

Cơ sở chung này gồm cả trường hợp của sai số độ đo cổ điển, khi $\gamma_1 = 0$ và $\alpha_2 = 1$. Trong trường hợp này thừa số chung không quan sát được U là một sai số độ đo trong cả Y_2 .

Những mô hình này cũng bao gồm cách xây dựng khi mà biến bị bỏ đi gây ra những sai lệch và không chính xác. Ví dụ, trong phương trình tiền lương và giáo dục, yếu tố không quan sát được có thể đại diện cho khả năng của một cá nhân và năng lực, điều mà ảnh hưởng đến cả vấn đề giáo dục và năng suất lao động của chính người đó.

Để giải quyết sai số độ đo hoặc các biến bị bỏ sót chúng ta thường đặt ra các hạn chế cho cách xác định mà cung cấp các công cụ. Giả sử rằng không có sẵn một công cụ nào, và cho Z là một vecto của các biến nội sinh quan sát được: một vecto con của X hoặc là chính X . Giả sử X không tương quan với bộ ba (U, V_1, V_2) và Z không tương quan với $(U^2, UV_j, V_1 V_2)$; Z tương quan với V_2^2 (hay trong một hệ đồng thời, với V_1^2).

Với những khẳng định này, có thể thấy rằng:

$$\text{cov}(Z, \varepsilon_1 \varepsilon_2) = 0 \quad (2.30)$$

$$\text{cov}(Z, V_2^2) = 0 \quad (2.31)$$

thì phương pháp này được sử dụng

Tuy nhiên những sai số không thực sự có từ một mô hình thừa số của dạng này; nó có thể là các điều kiện:

$$E(X\varepsilon_j) = 0, j = 1, 2 \quad (2.32)$$

$$\text{cov}(Z, \varepsilon_1 \varepsilon_2) = 0 \quad (2.33)$$

thỏa mãn, cùng với phương sai thay đổi nào đó của ε_j . Cách xác định sẽ luôn đạt được dù cho Z có phải là một vecto con của X hay không.

Ước lượng phương trình đơn

Trong một vài trường hợp rõ ràng việc áp dụng cách tiếp cận bằng các biến công cụ vào một phương trình đơn, nhưng lại thiếu công cụ thích hợp hoặc gặp phải các hạn chế. Phương trình phụ hoặc ‘hồi quy bậc một’ có thể được sử dụng để cung cấp các thành phần cần thiết cho phương pháp Lewbels.

Nói đơn giản nhất về cách tiếp cận này, các công cụ được tạo ra có thể được xây dựng từ phần dư của các phương trình phụ, nhân với mỗi biến nội sinh trong dạng trung bình trung tâm:

$$Z_j = (X_j - \bar{X}) \cdot \varepsilon \quad (2.34)$$

Trong đó ε là một vecto của phần dư ‘hồi quy bậc một’ của mỗi biến phụ thuộc nội sinh trong tất cả các biến phụ thuộc ngoại sinh bao gồm cả vecto hằng số.

Phần dư hồi quy phụ có hiệp phương sai 0 với mỗi biến phụ thuộc được dùng để tạo nên chúng, cho thấy rằng trung bình của các công cụ được tạo ra sẽ là 0. Tuy nhiên, tích của mỗi phần với biến phụ thuộc trung tâm sẽ không là 0 và sẽ chứa các thành phần đáng kể nếu có một bằng chứng rõ ràng của phương sai thay đổi tỉ lệ với các biến phụ thuộc. Khi tỉ lệ phương sai thay đổi trong quy trình sai số càng lớn thì sự tương quan giữa các công cụ được tạo ra với các biến nội sinh mà là biến hồi quy phụ thuộc trong hồi quy phụ càng cao.

2.5. Phương pháp phi tham số hồi quy Kernel

Giả sử rằng nghiên cứu mối quan hệ giữa một biến phụ thuộc Y và một biến giải thích X :

$$E(Y|_x) = m(X) \quad (2.35)$$

Trong đó $m(X)$ là một hàm nào đó theo X . Với cách tiếp cận tham số, cho $m(\cdot)$ là một dạng cụ thể nào đó, ví dụ $m(\cdot)$ có thể là một hàm số tuyến tính:

$$E(Y|_x) = \beta_1 + \beta_2 X \quad (2.36)$$

Các phương pháp tham số như là OLS, ML, GMM có thể được áp dụng cho ước lượng giới hạn. Các ước lượng cho β_1, β_2 từ cách tiếp cận tham số thường rất dễ để diễn tả. Tuy nhiên nếu $m(\cdot)$ không cụ thể thì công thức ước lượng sẽ bị sai khác và mâu thuẫn, dẫn đến một kết luận sai lệch và các hệ quả không chính xác. Trong nhiều trường hợp, đặt một dạng hàm cụ thể cho $m(\cdot)$ có thể sẽ khó, vậy một cách tiếp cận không phụ thuộc tham số là một lựa chọn tốt. Một trong các phương pháp phi tham số được sử dụng là phương pháp hồi quy Kernel để ước lượng (2.35). Đây là một cách tiếp cận hiện đại dựa trên cơ sở hàm Kernel như sau đây:

Ta có:

$$m(x) = m(X |_{X=m}) = E(Y |_{X=x}) = \int_R yf(y|x)dy \quad (2.37)$$

Trong đó $f(y|x)$ là hàm mật độ của Y phụ thuộc vào X . Phương pháp không phụ thuộc vào tham số mà sử dụng hàm mật độ Kernel để ước lượng (2.35) được gọi là phương pháp hồi quy Kernel.

Một vài hàm Kernel trong hồi quy gồm có: Hàm Epanechnikov

$$K(z) = \frac{3}{4}(1-z^2)1(|z| \leq 1)$$

Với $1(|z| \leq 1)$ là hàm đặc trưng hoặc Kernel thông thường:

$$\ddot{u}(\cdot) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \text{ cho các biến liên tục, và Aitchison hay Aitken cho biến}$$

cố định danh.

Hai phương pháp thông thường được sử dụng trong hồi quy Kernel: phương pháp hằng số địa phương và phương pháp tuyến tính địa phương. Phương pháp hằng số địa phương được phát biểu bởi Nadaraya (1964) và Watson (1964) và được biết đến như N-W (Nadaraya-Watson):

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n K_h(x - X_i) Y_i}{\sum_{i=1}^n K_h(x - X_i)} \quad (2.38)$$

Trong đó $K_h(\cdot)$ là hàm mật độ Kernel với băng thông h . Dưới những điều kiện thông thường của Hàm Kernel, Nadanaya (1964) chứng minh rằng (2.38) là công thức ước lượng vững của $m(x)$. Tuy nhiên công thức ước lượng này thường sai lệch ở biên và tại nơi mà phân phối không đồng đều.

Phương pháp tuyến tính địa phương phát biểu bởi Li và Racine (2004) đã giải quyết các vấn đề sai lệch trong phương pháp N-W. Ý tưởng của phương pháp này có thể được tóm tắt như sau: Trong một lân cận của X_0 , nó giả định rằng Y là một hàm tuyến tính của X trong một lân cận nào đó của X_0 thay vì giả định Y là hằng số như trong N-W.

Cụ thể hơn, tại mỗi điểm x ta tìm được các vectơ hệ số $\alpha(x), \beta(x)$ sao cho:

$$\text{Min}_{\alpha(x), \beta(x)} \sum_{i=1 \rightarrow N(x)} \frac{(y_i - \alpha(x) - (x_i - x)\beta(x))^2}{N(x)} K_h(x_i - x) \quad (2.39)$$

Trong đó tổng được lấy trên các $x_i : |x_i - x| \leq h$ với băng thông cho trước h .

CHƯƠNG 3

THỰC TRẠNG VỀ GIÁO DỤC SAU PHỔ THÔNG VÀ THỊ TRƯỜNG LAO ĐỘNG VIỆT NAM

3.1. Thực trạng về giáo dục sau phổ thông Việt Nam

3.1.1 Quy mô và mạng lưới cơ sở giáo dục sau phổ thông

Tính đến năm 2017, cả nước có 319.5 nghìn sinh viên tốt nghiệp đại học, tăng 4.6% so với năm 2016. Năm học 2016-2017, học sinh tốt nghiệp THPT vào đại học, cao đẳng khoảng 41%, vào cao đẳng nghề, trung cấp khoảng 23%, học nghề tại trung tâm đào tạo nghề khoảng 13%, đi làm khoảng 10%.

Bảng 3.1: Giáo dục đại học

	2010	2014	2015	2016	2017
Số trường học (Trường)	188	219	223	235	235
Công lập	138	159	163	170	170
Ngoài công lập	50	60	60	65	65
Số giáo viên (Nghìn người)	51,0	65,7	69,6	72,8	75,0
Phân theo loại hình					
Công lập	43,4	52,7	55,4	57,6	59,3
Ngoài công lập	7,6	13,0	14,2	15,2	15,7
Phân theo giới tính					
Nam	27,7	30,0	36,9	37,7	38,4
Nữ	23,3	35,7	32,7	35,1	36,6
Số sinh viên (Nghìn người)	1435,9	1824,3	1753,2	1767,9	1695,9
Phân theo loại hình					
Công lập	1246,4	1596,7	1520,8	1523,9	1432,6
Ngoài công lập	189,5	227,6	232,4	244,0	263,3
Phân theo giới tính					
Nam	742,7	974,3	821,1	833,4	791,9
Nữ	693,2	850,0	932,1	934,5	904,0
Số sinh viên tốt nghiệp (Nghìn người)	187,4	353,9	352,8	305,6	319,5
Công lập	166,2	302,6	307,8	268,4	282,0
Ngoài công lập	21,2	51,3	45,0	37,2	37,5

Nguồn: Niên giám thống kê năm 2017

Quy mô đào tạo thạc sĩ là 105.801 (tăng 12.8% so với năm học 2015-2016). Quy mô đào tạo tiến sĩ là 15.112 (tăng 21% so với năm học 2015-2016). Quy mô đào tạo tiến sĩ ở các viện NCKH thay đổi theo chiều hướng giảm. Tính đến tháng 6/2017, các Viện NCKH mới tuyển được khoảng 38% NCS so với chỉ tiêu đã đăng ký. Chỉ số phát triển con người (HDI) của nước ta theo bảng xếp loại của chương trình phát triển Liên hợp quốc trong những năm gần đây có những tiến bộ đáng kể: từ 0.688 xếp hạng thứ 109 trong số 174 quốc gia năm 2000, 0.666 xếp thứ 116 trong số 188 quốc gia năm 2015.

Bảng 3.2: Giáo dục nghề nghiệp

	2016	2017
Số cơ sở GDNN (Cơ sở)	2697	3006
Công lập	1465	1574
Ngoài công lập	1232	1432
Số giáo viên (Nghìn người)	67,7	86,4
Phân theo loại hình		
Công lập	41,6	58,4
Ngoài công lập	26,1	28,0
Phân theo giới tính		
Nam	47,9	56,7
Nữ	19,8	29,7
Số học sinh, sinh viên tốt nghiệp (Nghìn người)	1479,4	1984,0
Phân theo loại hình - By type		
Công lập – <i>Public</i>	965,4	1291,6
Ngoài công lập - <i>Non-public</i>	514,0	692,4

Nguồn: Niên giám thống kê năm 2017

3.1.2 So sánh tỷ lệ đi học trong giáo dục của Việt Nam và trên thế giới

Năm 2014, so sánh giữa 4 tỷ lệ đi học của các trình độ từ trước tiểu học đến đại học cho thấy Việt Nam đạt thành tựu cao nhất ở Tỷ lệ Đi học Tiểu học PEGER (*Pre-Primary Education Gross Enrolment Ratio*) là 109%, so với các nước trên thế giới ở loại khá; so với 8 nước châu Á chỉ kém Hàn Quốc (103%) và hơn các nước còn lại.

Bảng 3.3: Tỷ lệ Đi học Tiểu học

Nước	Lào	Campuchia	Việt Nam	Indonesia	Nhật	Trung Quốc	Hàn Quốc
PEGER%	118	116	109	104	99	99	97
Hạng	11	17	36	73	144	147	176

Nguồn: Ngân hàng thế giới

Tỷ lệ Đi học đại học (*Tertiary Education Gross Enrolment Ratio* TEGER) của 8 nước châu Á và thứ hạng trên thế giới trình bày trong bảng nhỏ sau đây theo thứ tự cao đến thấp năm 2014.

Bảng 3.4: Tỷ lệ Đi học đại học

Nước	Hàn Quốc	Nhật	Indonesia	Philippines	Trung Quốc	Việt Nam	Lào	Campuchia
TEGER%	98	62	32	28	27	25	18	16
Hạng	2	39	80	85	89	92	104	111

Nguồn: Ngân hàng thế giới

Việt Nam đạt thành tựu kém nhất trong các tỷ lệ đi học ở Tỷ lệ Đi học Đại học TEGER (25%). So với nước đứng đầu thế giới là Hy Lạp (117%), đứng cuối thế giới là Seychelles và Malawi (1%) thì ta ở mức dưới trung bình. Tám nước châu Á không có nước nào đạt được một tỷ lệ đi học cao nhất thế giới, nhưng trong 8 nước thì dẫn đầu cả 4 tỷ lệ là Hàn Quốc và Nhật Bản.

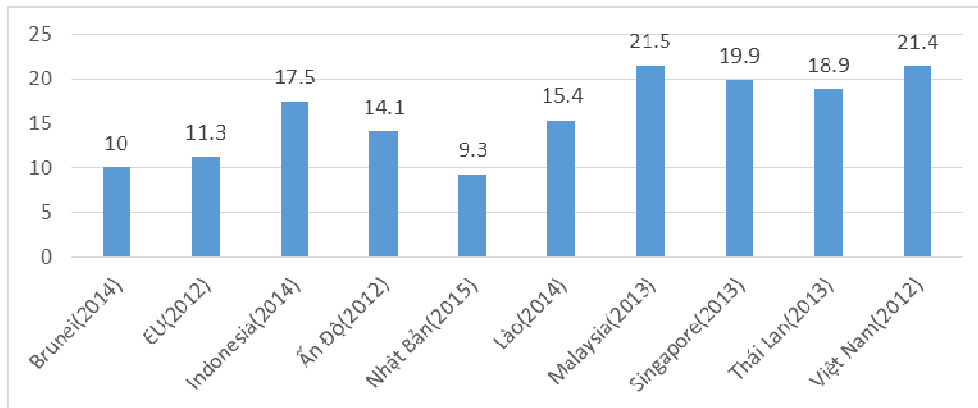
3.1.3. Tổng quan về đầu tư cho giáo dục và đào tạo

Luật giáo dục 2005 (sửa đổi vào 2010) của nước ta cũng đã khẳng định: “Phát triển giáo dục là quốc sách hàng đầu nhằm nâng cao dân trí, đào tạo nhân lực, bồi dưỡng nhân tài” (điều 9). Tại điều 13 cũng nhấn mạnh: “Đầu tư giáo dục là đầu tư phát triển, Nhà nước ưu tiên đầu tư cho giáo dục. Khuyến khích bảo hộ các quyền và lợi ích hợp pháp của tổ chức và cá nhân trong và ngoài nước đầu tư cho giáo dục, trong đó ngân sách Nhà nước giữ vai trò chủ yếu trong tổng nguồn lực đầu tư cho giáo dục”.

Năm 2015, tổng nguồn NSNN dành cho lĩnh vực giáo dục, đào tạo là 224.826 tỷ đồng, chiếm khoảng 20% tổng chi NSNN. Chi thường xuyên lĩnh vực giáo dục đào tạo năm 2015 là 184.070 tỷ đồng. Theo đó, dự toán chi từ ngân sách địa phương (NSDP) là 152.000 tỷ đồng để thực hiện nhiệm vụ giáo dục đào tạo của địa phương; chi từ ngân sách trung ương (NSTW) là 32.070 tỷ đồng.

Trong tổng chi từ NSTW 32.017 tỷ đồng, cũng bố trí 10.398 tỷ đồng hỗ trợ hoạt động thường xuyên của các cơ sở giáo dục đào tạo thuộc các bộ, cơ quan trung ương. Chi đầu tư phát triển giáo dục, đào tạo năm 2015 là 33.756 tỷ đồng; trong đó, chi của NSTW là 14.096 tỷ đồng; chi NSDP là 19.660 tỷ đồng.

So với các nước, trong khu vực, tỷ lệ chi tiêu công cho giáo dục trên GDP của Việt Nam cao hơn hẳn nhiều nước, thậm chí so với các nước có trình độ phát triển kinh tế cao hơn, chẳng hạn như Singapore (3.2% năm 2010), Malaysia (5.1%), Thái Lan (3.8%), Hàn Quốc (5.2% năm 2011). Tính theo GDP, chi tiêu từ nguồn lực nhà nước cho giáo dục, đào tạo ở Việt Nam khá cao so với các nước, khu vực được đem ra so sánh. Số liệu hình 3.1 cho thấy chi tiêu công cho giáo dục, đào tạo/GDP của Việt Nam năm 2012 chiếm 6.3%, cao hơn rất nhiều so với các nước có cùng trình độ phát triển.



Hình 3.1: Chi cho giáo dục, đào tạo đối với tổng chi tiêu công của Việt Nam với một số nước khu vực (%)

Nguồn: Ngân hàng Thế giới

Ngoài ưu tiên chi tiêu ngân sách cho giáo dục, Chính phủ còn có nhiều chính sách hỗ trợ khác như thực hiện chính sách miễn, giảm học phí và hỗ trợ chi phí cho học sinh, sinh viên nghèo; kinh phí hỗ trợ phát triển giáo dục mầm non; kinh phí hỗ trợ phát triển giáo dục đối với con em đồng bào dân tộc thiểu số, vùng sâu, vùng xa; kinh phí hỗ trợ học bổng và hỗ trợ phương tiện, đồ dùng học tập thực hiện chính sách giáo dục đối với người khuyết tật...

Ngoài ra, Chương trình Tín dụng ưu đãi dành sinh viên thông qua Ngân hàng Chính sách xã hội đã tạo điều kiện cho hàng triệu học sinh, sinh viên thuộc hộ nghèo, hộ cận nghèo và hộ có hoàn cảnh khó khăn được vay vốn học tập và lập nghiệp. Đến năm 2016, tổng doanh số cho vay của Chương trình đạt

trên 56 nghìn tỷ đồng, tổng dư nợ gần 21 nghìn tỷ đồng với trên 3,3 triệu lượt học sinh, sinh viên được vay vốn học tập và lập nghiệp.

Hạn mức cho vay đối với sinh viên được điều chỉnh tăng qua từng năm, từ mức vay 8 triệu đồng/sinh viên/năm, năm 2008 về sau lên mức 11 triệu đồng/năm. Đồng thời, mức lãi suất cũng được điều chỉnh từ 0,65%/tháng còn 0,55%/tháng, đáp ứng nhu cầu thực tế của sinh viên và điều kiện kinh tế xã hội hiện nay.

Bảng 3.5: Cơ cấu chi NSNN cho giáo dục, đào tạo (%)

	2008	2009	2010	2011	2012	2015
Tổng chi	100	100	100	100	100	100
Chi xây dựng cơ bản	23,1	17,1	18,4	18	17,7	18,1
Chi thường xuyên	76,9	82,9	81,6	82	82,3	81,9

Nguồn: Bộ Tài chính

Bảng 3.5 cho thấy, tỷ lệ chi thường xuyên chiếm trên dưới 82% tổng chi NSNN cho giáo dục, đào tạo. Trong chi thường xuyên, chi cho con người chiếm 80% tổng chi, còn lại chi cho hoạt động dạy học, nâng cao chất lượng giáo trình. Chi đầu tư xây dựng cơ bản còn thấp so với nhu cầu nâng cao cơ sở trường học, mua sắm thiết bị dạy học, phòng thí nghiệm...

Bảng 3.6: Cơ cấu NSNN theo các cấp học (%)

	2010	2011	2012	2013	2014
Chi NSNN các cấp	100	100	100	100	100
Mầm non	7,9	8,2	8,2	8,2	8,2
Tiểu học	28,5	28,2	28,3	28,3	28,3
Trung học cơ sở	21,5	21,4	21,6	21,6	21,6
THPT	11,8	11,2	11,1	10,9	11,1
Cộng chi giáo dục mầm non và giáo dục phổ thông	69,7	69	69,2	69	29,2
Dạy nghề	9,7	9,9	9,7	9,7	9,7
Trung cấp chuyên nghiệp	3,4	3,6	3,5	3,5	3,5
Cao đẳng, đại học	11,7	12	12,4	12,4	12,4
Giáo dục thường xuyên	1,8	1,7	1,6	1,8	1,6
Giáo dục đào tạo khác	3,7	3,8	3,6	3,6	3,6

Nguồn: Actionaid (2016), Chi tiêu cho y tế giáo dục và giao thông công cộng ở một số tỉnh tại Việt Nam, Một số quan sát và khuyến nghị

Chi tiêu dành cho dạy nghề, trung cấp, cao đẳng, đại học, giáo dục thường xuyên chiếm 30% tổng chi cho giáo dục. Trong đó, chi cho đào tạo cao đẳng và đại học chiếm trên 12%. Năm 2010, tỷ lệ chi ngân sách trên mỗi sinh viên đại học ở Việt Nam bằng xấp xỉ 40% thu nhập bình quân đầu người, trong khi ở Hoa Kỳ, tỷ lệ này là 21%, Singapore là 28%, Hàn Quốc là 13%, Nhật Bản là 25%. Điều này cho thấy, mức ưu tiên mà Chính phủ Việt Nam dành cho giáo dục khá lớn so với khả năng tài chính của quốc gia.

3.2. Thực trạng về thị trường lao động Việt Nam

Trong năm 2016, ngành LĐ, TB & XH đã có nhiều giải pháp huy động nguồn lực trong nước và quốc tế để thực hiện tốt các nhiệm vụ của ngành, hầu hết các chỉ tiêu của năm đều đạt và vượt kế hoạch, các nhiệm vụ phát triển thị trường lao động, tạo việc làm, đào tạo nghề, nâng cao chất lượng nguồn nhân lực, giảm nghèo, đảm bảo an sinh xã hội được triển khai thực hiện đồng bộ, hiệu quả. Trong lĩnh vực lao động – việc làm, năm 2016 đã giải quyết việc làm cho khoảng 1.641 nghìn người, vượt 2.5% so với kế hoạch và tăng 1% so với năm 2015; tỷ lệ thất nghiệp của lao động trong độ tuổi là 2.30%, trong đó khu vực thành thị là 3,18%, khu vực nông thôn là 1.86%

Tính đến 31/12/2016, cả nước có 13,103 triệu người tham gia bảo hiểm xã hội, chiếm 24.1% lao động; gần 11 triệu người tham gia bảo hiểm thất nghiệp. Theo báo cáo của các địa phương, trong năm 2016 có 574.310 người nộp hồ sơ hưởng trợ cấp thất nghiệp, tăng 12.06% so với cùng kỳ năm 2015; 877.718 lượt người được tư vấn, giới thiệu việc làm; 27.642 người được hỗ trợ học nghề, tỷ lệ lao động qua đào tạo khoảng 53%.

3.2.1 Lực lượng lao động theo nhóm tuổi

Bảng 3.7 là kết quả thống kê lực lượng lao động theo nhóm tuổi, độ tuổi lao động từ 15 -19 giảm dần qua các năm từ 2010 đến 2017, ở độ tuổi này có thể xem lao động chưa qua đào tạo nhiều hoặc chưa có bằng cao, trong giai đoạn gần đây đòi hỏi lao động có tay nghề cao là cần thiết nên việc giảm số lượng lao động chưa có tay nghề có thể thấy được. Bên cạnh đó lao động từ độ tuổi 24 trở lên có số lượng tăng qua các năm, đây là nhóm lao động có thể xem là có tay nghề, có kinh nghiệm làm việc và đạt được bằng cấp giáo dục cao hơn.

Bảng 3.7: LLLĐ chia theo nhóm tuổi (Đơn vị: nghìn)

Năm	2010	2011	2012	2013	2014	2015	2016	2017
15-19	3.170,9	2.924,7	2.550,9	2.601,5	2.395,4	2.415,2	2.119,5	2.195,6
20-24	5.422,1	5.096,7	4.904,8	4.826,4	4.714,9	5.033,7	4.832,7	4.816,8
25-29	6.618,6	6.449,4	6.258,0	6.119,1	6.121,1	6.082,4	6.178,5	6.007,0
30-34	6.342,2	6.301,9	6.232,3	6.352,2	6.514,6	6.851,0	6.823,0	6.705,4
35-39	6.225,4	6.387,9	6.520,9	6.449,6	6.456,7	6.389,8	6.471,9	6.635,0
40-44	5.778,4	6.067,1	6.412,3	6.462,2	6.525,2	6.348,2	6.480,1	6.614,6
45-49	5.418,4	5.835,2	6.212,2	6.122,7	6.085,9	5.822,6	5.992,3	6.194,4
50+	10.072,5	11.289,2	12.331,0	13.274,2	13.930,7	13.897,1	14.404,8	14.534,6

Nguồn: Niên giám thống kê năm 2017

3.2.2. Lực lượng lao động theo khu vực thành thị nông thôn

Có thể thấy số lượng lao động ở khu vực tăng lên đáng kể qua các năm theo bảng 3.8, khu vực thành thị là nơi tập trung lao động có tay nghề cao, và có thể có được mức lương cao hơn khu vực nông thôn. Mặc dù số lượng lao động khu vực thành thị gần đây có tăng nhanh tuy nhiên số lượng lao động khu vực nông thôn vẫn chiếm tỉ trọng lớn hơn.

Bảng 3.8: LLLĐ chia theo khu vực thành thị nông thôn (Đơn vị nghìn)

Năm	Thành thị		Nông thôn	
	LLLĐ	Cơ cấu dân số (%)	LLLĐ	Cơ cấu dân số (%)
2010	13.531,4	51.0	35.517,1	58.8
2011	14.732,5	52.8	35.619,5	59.4
2012	15.412,0	54.5	36.010,4	59.5
2013	15.509,0	53.7	36.698,8	60.3
2014	16.009,0	53.3	36.735,5	60.5
2015	16.374,8	52.6	36.465,2	60.2
2016	16.923,6	52.9	36.379,2	59.9
2017	17.116,7	52.0	36.586,7	60

Nguồn: Niên giám thống kê năm 2017

3.2.3. Lực lượng lao động theo giới tính

Bảng 3.9 cho thấy từ năm 2010 đến 2017 lao động nam tăng đều đặn theo các năm chiếm lần lượt trong cơ cấu dân số là 58.9% cho năm 2010, 59.9% cho năm 2011, 60.3% cho năm 2012, các năm tiếp theo từ 2014 đến 2017 có giảm theo tỉ lệ %, cụ thể là 60.4% cho năm 2013, cùng tỉ lệ với năm 2014, 60.2% cho năm 2015, 60 % cho năm 2016, và 2017 cũng ở tỉ lệ này.

Bảng 3.9: LLLĐ chia theo giới tính (Đơn vị nghìn)

Năm	Nam		Nữ	
	LLLĐ	Cơ cấu LLLĐ (%)	LLLĐ	Cơ cấu LLLĐ (%)
2010	25.305,9	58.9	23.742,6	54.0
2011	26.024,7	59.9	24.327,3	54.8
2012	26.499,2	60.3	24.923,2	55.6
2013	26.830,2	60.4	25.377,6	56.1
2014	27.025,8	60.4	25.718,7	55.9
2015	27.216,7	60.2	25.623,3	55.1
2016	27.442,8	60.0	25.860,0	55.0
2017	27.7813.7	60.0	25.889.7	54.8

Nguồn: Niên giám thống kê năm 2017

3.2.4. Lực lượng lao động theo trình độ được đào tạo

Bảng 3.10 cho thấy năm 2010 tổng số lao động được đào tạo là 14.6% tỉ lệ này tăng dần đến năm 2017 cụ thể là 15.4%, 16.55%, 17.9%, 18.2%, 19.9%, 20.6% và 21.4% lần lượt cho các năm 2011 đến 2017. Điều này cho thấy xã hội cần ngày càng có nhu cầu lao động có tay nghề. Đối với đào tạo nghề, trung cấp chuyên nghiệp, cao đẳng cũng tăng qua các năm. Đặc biệt là với bậc đại học trở lên, tỉ lệ lao động được đào tạo nghề tăng nhanh từ năm 2010 đến 2017, tương ứng là 5.7%, 6.1%, 6.4%, 6.9%, 7.6%, 8.5%, 9% và 9.3% cho năm 2017.

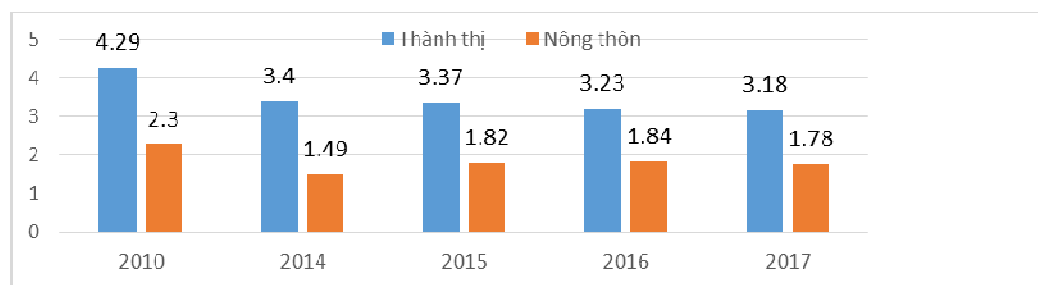
Bảng 3.10: LLLĐ chia theo trình độ được đào tạo (%)

	2010	2011	2012	2013	2014	2015	2016	2017
Tổng số	14.6	15.4	16.6	17.9	18.2	19.9	20.6	21.4
Dạy nghề	3.8	4.0	4.7	5.3	4.9	5.0	5.0	5.3
TCCN	3.4	3,7	3.6	3.7	3.7	3.9	3.9	3.7
Cao đẳng	1.7	1.7	1.9	2.0	2.1	2.5	2.7	2.7
Đại học trở lên	5.7	6.1	6.4	6.9	7.6	8.5	9.0	9.3

Nguồn: Niên giám thống kê năm 2017

3.2.5. Tỷ lệ thất nghiệp

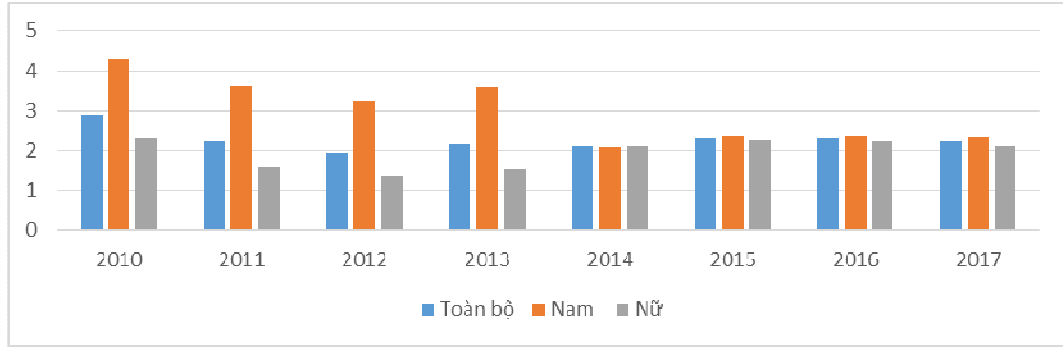
Hình 3.2 cho thấy Tỷ lệ thất nghiệp khu vực thành thị cao hơn ở nông thôn, và tỉ lệ này giảm qua các năm cụ thể là 4.29% cho năm 2010, 3.4% cho năm 2014, 3.37% cho năm 2015, 3.23% cho năm 2016 và 3.18% là của năm 2017. Tỷ lệ này tương ứng cho khu vực nông thôn là 2.3%, 1.49%, 1.82%, 1.84% và 1.78%. Có thể thấy năm 2010 là năm khủng hoảng kinh tế, việc tuyển dụng lao động trở nên khắt khe hơn dẫn đến tỉ lệ thất nghiệp tăng cao cho cả hai khu vực thành thị và nông thôn. Khi kinh tế dần hồi phục tỉ lệ thất nghiệp cũng giảm theo.



Hình 3.2. Tỷ lệ thất nghiệp phân theo khu vực thành thị và nông thôn

Nguồn: Niên giám thống kê năm 2017

Hình 3.3 cho thấy tỉ lệ thất nghiệp của lao động nam là khá cao so với toàn bộ lao động và lao động nữ trong những năm 2010 đến 2013, tỉ lệ này có giảm so với những năm 2014 đến 2017 và khá cân bằng so với tỉ lệ thất nghiệp của toàn bộ lao động và lao động nữ.



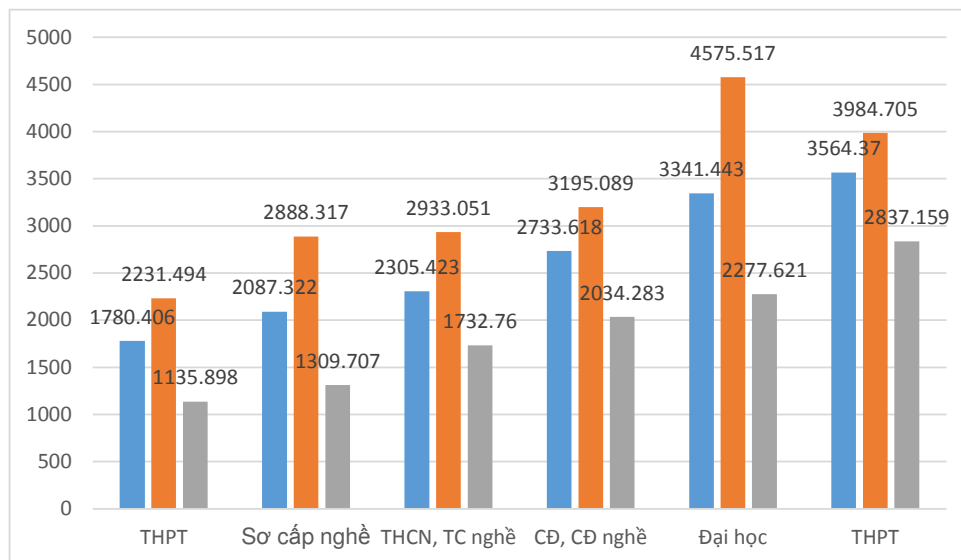
Hình 3.3. Tỷ lệ thất nghiệp phân theo giới tính người lao động

Nguồn: Niên giám thống kê năm 2017

3.3. Thực trạng về thu nhập người lao động

3.3.1. Thu nhập bình quân đầu người năm 2010

Thu nhập bình quân đầu người theo tháng năm 2010 tăng theo cấp độ giáo dục, trong đó nhóm có trình độ THPT có mức thu nhập là 1780.406 nghìn đồng/tháng, sơ cấp nghề là 2087.322 nghìn đồng/tháng, THCN và TC nghề là 2305.423 nghìn đồng/tháng, CD và CD nghề là 2733.618 nghìn đồng/tháng và nhóm đại học là 3341.443 nghìn đồng/tháng. Năm 2010 lao động làm thuê có thu nhập cao hơn lao động tự làm ở tất cả các cấp độ giáo dục.



Hình 3.4: Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo trình độ giáo dục năm 2010 (đơn vị: nghìn đồng)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2010

Bảng 3.11 cho thấy số liệu về thu nhập bình quân đầu người khu vực thành thị và nông thôn. Năm 2010 thu nhập bình quân đầu người khu vực thành thị tăng theo cấp độ giáo dục, lao động có trình độ đại học cao hơn nhiều so với các nhóm còn lại điều này là vì khu vực thành thị tập trung nhiều ngành nghề đòi hỏi kỹ năng cao nên những lao động có bằng cấp cao sẽ được trả lương cao hơn. Ở khu vực nông thôn thu nhập bình quân đầu người thấp hơn khu vực thành thị, có ít sự khác biệt tiền lương giữa các cấp độ giáo dục. Chênh lệch thu nhập giữa lao động làm thuê và lao động tự làm tăng theo các cấp độ giáo dục cho cả hai khu vực thành thị và nông thôn.

Bảng 3.11: Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, thành thị, nông thôn năm 2010 (đơn vị: nghìn đồng)

Trình độ giáo dục	THPT	Sơ cấp nghề	THCN, TC nghề	CD, CD nghề	Đại học
Thành thị					
Bình quân	2010.954	2402.485	2528.704	2806.285	5210.954
Làm thuê	2513.732	3189.812	3250.765	3618.506	6411.977
Tự làm	1375.269	1626.583	1680.298	1860.226	1951.176
Nông thôn					
Bình quân	1584.698	1715.132	1838.972	2087.871	2204.065
Làm thuê	1824.045	1944.88	2126.969	2467.679	2558.194
Tự làm	984.546	1297.703	1392.5	1671.889	1747.542

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2010

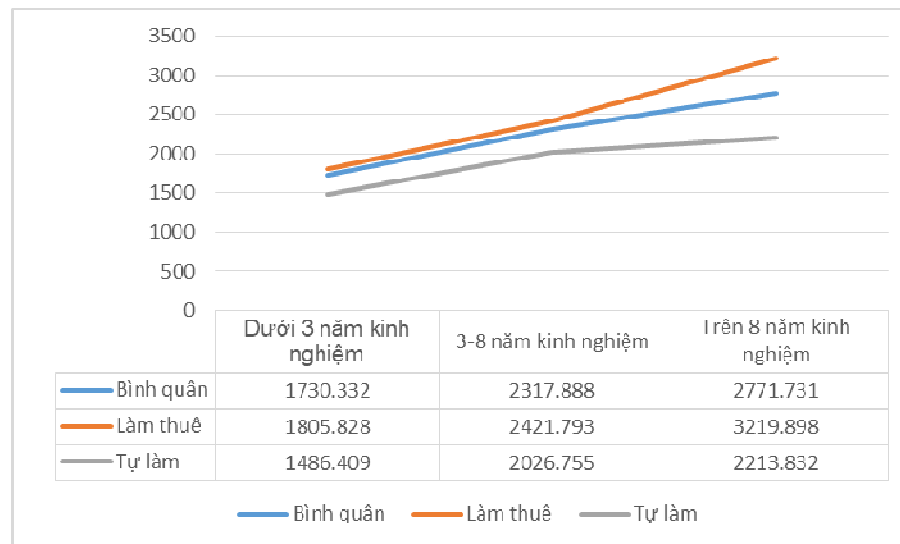
Số liệu tính toán ở bảng 3.12 chỉ ra thu nhập bình quân của nam cao hơn nữ ở các cấp độ giáo dục. Thu nhập bình quân tăng theo trình độ giáo dục. Cả lao động nam và nữ đều có sự chênh lệch thu nhập giữa lao động làm thuê và tự làm. Chênh lệch này càng cao ở cấp giáo dục càng cao. Đặc biệt là nhóm có trình độ đại học.

Bảng 3.12: Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, nam, nữ năm 2010 (đơn vị: nghìn đồng)

Trình độ giáo dục	THPT	Sơ cấp nghề	THCN, TC nghề	CĐ, CĐ nghề	Đại học
Nam					
Bình quân	2136.877	2313.401	2414.084	2792.255	4267.087
Làm thuê	2485.963	3125.92	3187.994	3485.724	5563.824
Tự làm	1346.994	1606.31	1706.231	1966.818	2766.818
Nữ					
Bình quân	1440.627	1506.704	1922.438	2174.692	2784.081
Làm thuê	1962.943	2084.456	2486.089	2801.117	3716.598
Tự làm	934.4595	1049.597	1221.71	1252.417	1789

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2010

Hình 3.5 cho thấy năm 2010 lao động có thu nhập bình quân đầu người tăng theo số năm kinh nghiệm, đối với nhóm dưới 3 năm kinh nghiệm chênh lệch thu nhập giữa lao động làm thuê và tự làm lớn hơn với nhóm 3 đến 8 năm kinh nghiệm, điều này là do ngoài việc cung cấp vốn nhân lực giáo dục còn cung cấp vai trò tín hiệu cho nhóm lao động có ít năm kinh nghiệm.

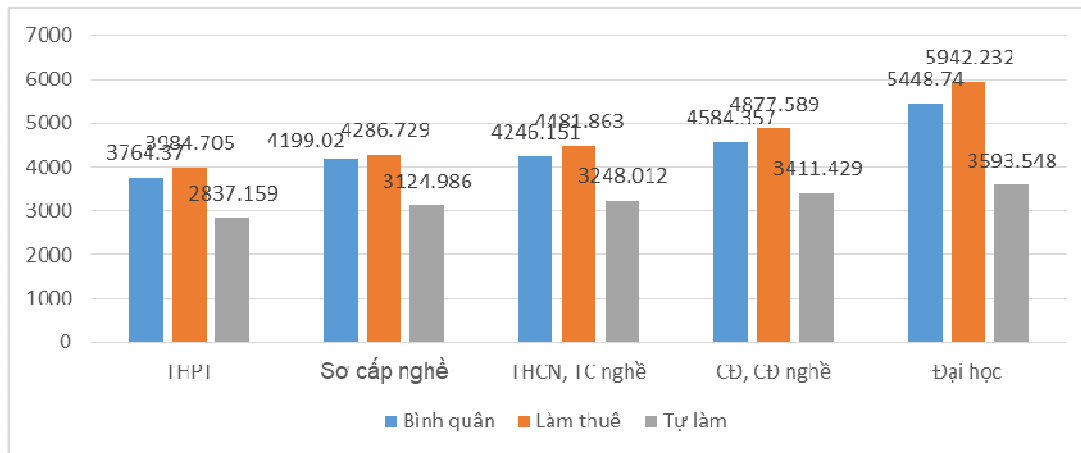


Hình 3.5: Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo năm kinh nghiệm năm 2010 (đơn vị: nghìn đồng)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2010

3.3.2. Thu nhập bình quân đầu người năm 2014

Thu nhập bình quân đầu người năm 2014 tăng theo cấp độ giáo dục, và tăng cao hơn so với năm 2010. Thu nhập bình quân theo tháng cho lao động có trình độ THPT là 3764.37 nghìn đồng, trình độ sơ cấp nghề là 4199.02 nghìn đồng, trình độ THCN và TC nghề là 4246.151 nghìn đồng, cao đẳng là 4584.357 nghìn và đại học là 5448.74 nghìn đồng trên tháng. Chênh lệch thu nhập giữa lao động làm thuê và lao động tự làm là khá lớn. Đặc biệt là nhóm đại học.



Hình 3.6: Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo trình độ giáo dục năm 2014 (đơn vị: nghìn đồng)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Thu nhập bình quân đầu người khu vực thành thị và nông thôn năm 2014 tăng theo trình độ giáo dục, số liệu bảng 3.13 cho thấy khu vực thành thị thu nhập bình quân theo trình độ giáo dục tăng nhiều hơn nông thôn. Chênh lệch tiền lương của lao động làm thuê và tự làm năm 2014 tăng theo trình độ giáo dục và chênh lệch này cao hơn so với năm 2010, đặc biệt là nhóm đại học. Năm 2010 là năm khủng hoảng kinh tế người sử dụng lao động trả lương thấp hơn năm 2014, là những năm kinh tế được phục hồi.

Bảng 3.13: Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, thành thị, nông thôn năm 2014 (đơn vị: nghìn đồng)

Trình độ giáo dục	THPT	Sơ cấp nghề	THCN, TC nghề	CD, CD nghề	Đại học
Thành thị					
Bình quân	3597.808	4146.978	4272.677	4482.256	5893.537
Làm thuê	4343.758	4611.387	4671.512	5076.376	6997.038
Tự làm	2867.088	2954	3282.95	3996.875	4458.235
Nông thôn					
Bình quân	3299.486	3460.699	3696.864	3912.398	4121.694
Làm thuê	3626.807	3811.313	3880.053	4206.871	4651.489
Tự làm	2550.094	3173.7	3236.557	3453.261	3536.182

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

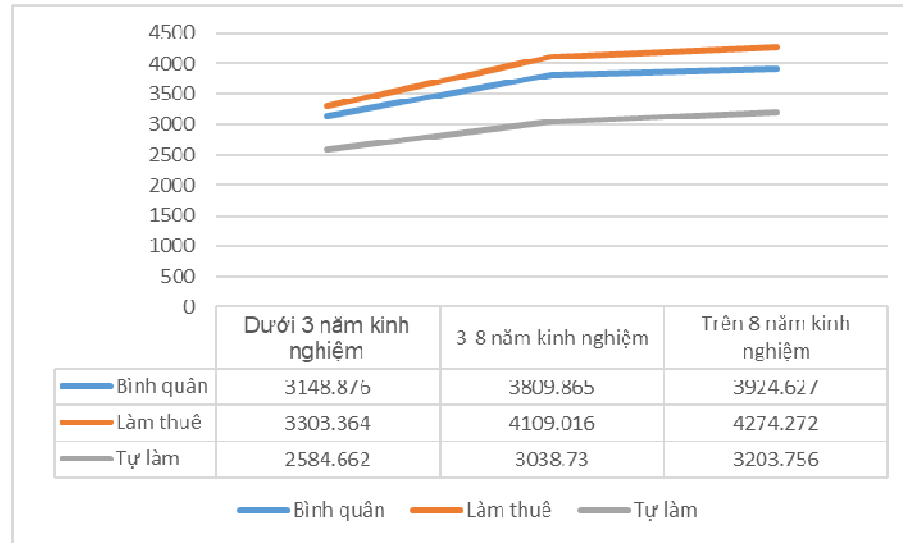
Thống kê ở bảng 3.14 cho thấy lao động nam năm 2014 có thu nhập tăng theo trình độ giáo dục, lao động nam làm thuê có tiền lương cao hơn tự làm là đáng kể, tuy nhiên sự chênh lệch này so với năm 2010 là ít khác biệt. Đối với lao động nữ năm 2014 thu nhập bình quân tăng theo cấp độ giáo dục, nhóm đại học có tiền lương khác biệt nhiều hơn so với nhóm giáo dục còn lại. Tương tự như lao động nam, lao động nữ làm thuê có tiền lương cao hơn lao động nữ tự làm theo trình độ giáo dục.

Bảng 3.14: Thu nhập bình quân đầu người theo tháng chia theo trình độ giáo dục, nam, nữ năm 2014 (đơn vị: nghìn đồng)

Trình độ giáo dục	THPT	Sơ cấp nghề	THCN, TC nghề	CD, CD nghề	Đại học
Nam					
Bình quân	3961.746	4030.423	4321.124	4497.192	5006.978
Làm thuê	4264.628	4870.981	4885.268	5192.773	6595.694
Tự làm	3015.148	3309.848	3570	3624.388	3639.706
Nữ					
Bình quân	3002.669	3205.698	3318.405	3580.611	4339.603
Làm thuê	3629.51	3677.358	3820.055	3913.768	5238.854
Tự làm	2628.365	2681.905	2754.969	2914.524	3537.5

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Năm 2014 lao động làm thuê có thu nhập bình quân cao hơn lao động tự làm ở nhóm dưới 3 năm kinh nghiệm. Chênh lệch này tiếp tục nói rộng ra ở nhóm 3 đến 8 năm kinh nghiệm. Và khi số năm kinh nghiệm tăng trên 8 năm thì chênh lệch này được thu hẹp. Có thể nhận định ngoài vai trò cung cấp vốn nhân lực giáo dục còn đóng vai trò phát tín hiệu, thời gian phát tín hiệu của năm 2014 kéo dài hơn năm 2010, là năm kinh tế khủng hoảng.



Hình 3.7: Thu nhập bình quân đầu người theo tháng của lao động làm thuê và tự làm theo năm kinh nghiệm năm 2014 (đơn vị: nghìn đồng)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Chương 3 trình bày về thực trạng giáo dục sau phổ thông, thực trạng thị trường lao động, thực trạng về tiền lương, thu nhập của người lao động gắn kết với các đặc tính khác. Mục tiêu chương 3 cung cấp một cái nhìn khái quát về mối quan hệ giữa giáo dục sau phổ thông và tiền lương, thu nhập trước khi đi vào ước lượng hiệu suất sinh lời của giáo dục sau phổ thông.

CHƯƠNG 4

MÔ HÌNH VÀ CÁC KẾT QUẢ NGHIÊN CỨU

Chương 4 sẽ sử dụng ba phương pháp chính khác nhau để tiếp cận thực nghiệm vấn đề đặt ra của Luận án. Thứ nhất là phương pháp PSM nhằm xác định sự hiện diện của vấn đề bất đối xứng thông tin trên thị trường lao động Việt Nam. Tiếp theo, Luận án thực hiện ước lượng suất sinh lời của giáo dục có tính đến sự hiện diện của TTBDX bằng phương pháp tham số và phương pháp phi tham số. Phương pháp tham số bao gồm phương pháp Heckman và phương pháp Lewbels – là các phương pháp giúp giải quyết vấn đề biến nội sinh thường gặp phải khi ước lượng suất sinh lời của giáo dục. Phương pháp phi tham số Kernel, với ưu điểm là việc ước lượng không cần dựa trên một giả định về dạng hàm, được sử dụng nhằm để kiểm tra và đảm bảo độ tin cậy của các ước lượng tham số. Việc ước lượng được thực hiện cho năm 2014 và cho năm 2010 nhằm so sánh sự khác biệt về suất sinh lời cũng như vai trò phát tín hiệu của giáo dục trong hai thời kỳ khác nhau của nền kinh tế.

4.1. Nguồn số liệu được sử dụng và các biến số sử dụng trong mô hình ước lượng

4.1.1. Nguồn số liệu được sử dụng

Số liệu của cuộc điều tra mức sống dân cư (VHLSS- Việt Nam Household Living Standard Survey) năm 2010 và 2014 được sử dụng trong nghiên cứu này. Đây là bộ số liệu được điều tra bởi Tổng Cục Thống kê, thực hiện hai năm một lần trên phạm vi toàn quốc. Bộ số liệu bao gồm các thông tin về thu nhập, học vấn, tình trạng ốm đau và sử dụng các dịch vụ y tế, tình trạng làm việc, thời gian làm việc, điều kiện sinh sống, tham gia các chương trình tín dụng, xóa đói giảm nghèo và các biến số nhân khẩu học khác như: tuổi, giới tính, dân tộc, tình trạng hôn nhân của người dân từ hơn 9398 hộ.

Các nghiên cứu luận án chỉ chọn các cá nhân từ 18 đến 65 tuổi, và loại bỏ những quan sát là công chức nhà nước với lý do mức lương của các công chức nhà nước không được xác định bởi năng suất của họ.

4.1.2. Các biến số sử dụng trong các mô hình

Trong chương 4 luận án sử dụng 4 phương pháp ước lượng hiệu suất sinh lời và vai trò phát tín hiệu của giáo dục. Tùy thuộc vào mỗi phương pháp sẽ sử dụng biến số trong mô hình như sau:

Tên biến	Đo bằng	Vai trò
Educ	Biến giả =1 cho tốt nghiệp phổ thông, =2 cho sơ cấp nghề, = 3 cho trung học chuyên nghiệp và trung cấp nghề, =4 cho tốt nghiệp cao đẳng và cao đẳng nghề và = 5 cho có bằng đại học và sau ĐH	Cung cấp vốn nhân lực cho người lao động, phát tín hiệu về năng lực bản thân => có tác động đến lương
Age	Tuổi, tính theo năm	Đại diện cho sức khỏe và các yếu tố xã hội liên quan đến năng suất lao động
Age Square	Bình phương của tuổi	Nhằm thể hiện tác động của quy luật cận biên giảm dần trong NSLĐ theo tuổi, khi tuổi của người lao động càng gia tăng thì năng suất lao động, do đó mức lương, cũng gia tăng, tuy nhiên mức tăng này sẽ giảm dần khi tuổi càng lớn
Female	Giới tính, = 1 nếu là nữ, = 0 nếu là nam	Đặc trưng về giới tính có ảnh hưởng đến năng suất lao động
Urban	=1 nếu ở nông thôn, =0 nếu ở thành thị	Sự khác biệt giữa mức lương giữa nông thôn và thành thị do điều kiện sống và điều kiện thị trường khác nhau
Married	=1 nếu còn độc thân,=0 đã lập gia	Người độc thân có năng

Tên biến	Đo bằng	Vai trò
	đình và khác	suất lao động cao hơn do đó có tiền lương cao hơn người đã lập gia đình
Ethnic	=1 nếu là dân tộc Kinh, =0 nếu là dân tộc khác	Đặc trưng về văn hóa xã hội có ảnh hưởng đến mức lương
Healthy	=1 nếu có bảo hiểm xã hội, = 0 nếu không có	Đặc trưng của công ty có thể tương quan với mức lương
Experience	Kinh nghiệm làm việc, biến giả: =1 cho những người có <3 năm làm việc, =2 cho người từ 3 – 8 năm, =3 cho từ 8 năm trở lên	Đại diện cho các kỹ năng và kiến thức học được trong quá trình làm việc và là nhân tố kiểm soát tính bất đối xứng thông tin giữa chủ lao động và lao động
Feduc	Học vấn của bố, biến giả, =1 với người có bằng THCS, =2 với THPT, =3 với đại học, = 4 trên ĐH	Biến công cụ cho biến học vấn của con.
Employed	=1 nếu là lao động làm thuê, =0 nếu là lao động tự làm	Việc làm thuê hay tự làm có tác động đến tiền lương
Occupation	=1 nếu lao động nhóm tay nghề cao, =0 lao động nhóm tay nghề thấp	Đại diện cho môi trường làm việc của người lao động

4.1.3. Thống kê mô tả

Bảng 4.1 cho thấy thống kê mô tả của các biến số sử dụng trong mô hình năm 2010. Đối với tổng thể, độ tuổi trung bình của lao động là 35, lao động nữ chiếm 37%, lao động khu vực thành thị là 34.5%, lao động chưa lập gia đình là 0.7%, lao động mang dân tộc Kinh là 91.3%. Lao động có thẻ bảo hiểm y tế là 8.2%. Đối với trình độ giáo dục của người lao động cho thấy, lao động có bằng THPT là 12.03%, 6.6% lao động được đào tạo sơ cấp nghề, 7.8% lao động có bằng THCN và trung cấp nghề, 2% lao động có bằng cao đẳng và cao đẳng nghề, và 9.6% lao động có bằng đại học và sau đại học. Trong toàn bộ mẫu có 68% lao động là làm thuê và 32% lao động là tự làm. Mức lương trung bình theo tháng của lao động là 2129 nghìn đồng, trung bình của logarit tiền lương là 7.33.

Với lao động nam, có tuổi trung bình là 35, độ tuổi trung bình của lao động nữ là 34. Lao động nam làm việc ở khu vực thành thị là 32.75%, so với 37.5% của lao động nữ. Tỷ lệ chưa lập gia đình ở lao động nam là 0.63% và 0.84% đối với nữ. 90% lao động nam là dân tộc Kinh và tỷ lệ này đối với nữ là 92%. Tỷ lệ 6.4% lao động nam có thẻ bảo hiểm y tế, 10.7% là đối với lao động nữ. Xét về trình độ giáo dục, 10.69% lao động nam có bằng THPT và tỷ lệ này với nữ là 14.3%. Lao động nam có bằng sơ cấp nghề là 7.9% và lao động nữ là 4.4%. Có bằng trung cấp nghề và trung học chuyên nghiệp với lao động nam là 8% và nữ là 7.4%. 1.9% lao động nam có bằng cao đẳng và tỷ lệ này đối với nữ là 2.3%. Lao động nam có bằng đại học là 8.7% và với lao động nữ là 7.3%. Mức lương tháng trung bình của nam là 2325 nghìn đồng, của nữ là 1798 nghìn đồng.

Bảng 4.1: Thống kê mô tả giá trị trung bình các biến năm 2010

Tên biến	Toàn bộ	Giới tính		Khu vực	
		LĐ nam	LĐ nữ	Thành thị	Nông thôn
Age	34.69	35.04	34.10	35.91	34.05
Age_square	1336.35	1359.61	1296.82	1426.21	1288.97
Female	0.37	0.00	1.00	0.40	0.35
Urban	0.34	0.32	0.37	1.00	0.00
Married	0.007	0.006	0.008	0.007	0.006
Ethnic	0.91	0.90	0.92	0.95	0.89
Healthy	0.082	0.06	0.107	0.07	0.09
Edu1	0.12	0.10	0.14	0.14	0.10
Edu2	0.06	0.07	0.04	0.09	0.053
Edu3	0.078	0.08	0.074	0.122	0.05
Edu4	0.02	0.018	0.023	0.035	0.012
Edu5	0.096	0.087	0.112	0.16	0.062
Employed	0.681	0.65	0.73	0.89	0.56
Income_m	2129.822	2325.038	1798.011	3025.697	1657.478
Log(income_m)	7.336	7.451	7.140	7.668	7.161
Số quan sát	5718	3600	2118	1974	3744

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2010

Năm 2010 độ tuổi trung bình khu vực thành thị là 36, nông thôn là 34. 40% lao động thành thị là nữ, 35.5 % là tỉ lệ cho khu vực nông thôn Lao động chưa lập gia đình ở thành thị là 0.75%, tỉ lệ này ở nông thôn là 0.69%. 95% lao động ở thành thị là dân tộc Kinh, 89% là của nông thôn. 7.1% lao động thành thị có thẻ bảo hiểm y tế, và 9% lao động nông thôn có thẻ bảo hiểm y tế. Trình độ giáo dục của lao động ở thành thị là 14.5% có bằng THPT, 9.1% có bằng sơ cấp nghề, 12.2% có bằng THCS và trung cấp nghề, 3.5% có bằng cao đẳng, và 16% có bằng đại học. Tỉ lệ này lần lượt ở nông thôn là 10.68%, 5.3%, 5.5%, 1.2% và 6.2%. Tiền lương trung bình cho lao động thành thị là 3025 nghìn đồng/tháng và 1657 nghìn đồng/tháng là của lao động nông thôn.

Bảng 4.2 thống kê mô tả cho các biến số ước lượng cho năm 2014. Đối với toàn bộ mẫu độ tuổi trung bình của người lao động là 43 (với lao động nam là 42, lao động nữ là 44). Lao động nữ ở năm 2014 chiếm 52%. 28.9% lao động làm việc ở thành thị (tỉ lệ này với lao động nam là 28.3% và nữ là 29.5%). Tỉ lệ chưa lập gia đình với toàn bộ mẫu là 18% (lao động nam tương ứng là 23% và lao động nữ là 14.6%). Đối với toàn bộ mẫu lao động dân tộc Kinh là 81.45% (lao động nam là 80.09%, lao động nữ là 81.8%). 11.2% lao động có thẻ bảo hiểm y tế (tỉ lệ này với lao động nam là 9.3%, với nữ là 12.8%). Trình độ giáo dục của lao động là 14.6% có bằng THPT, 3.2% có bằng sơ cấp nghề, 4.3% có bằng THCS và trung cấp nghề, 1.9% có bằng cao đẳng, và 10.6% có bằng đại học. Trình độ giáo dục tương ứng cho lao động nam là 15.5% có bằng THPT, 4.7% có bằng sơ cấp nghề, 5,3% có bằng THCS và trung cấp nghề, 1.9% có bằng cao đẳng, và 8.4% có bằng đại học. Trình độ giáo dục tương ứng cho lao động nữ là 13.8% có bằng THPT, 1,8% có bằng sơ cấp nghề, 3.5% có bằng THCS và trung cấp nghề, 1.8% có bằng cao đẳng, và 12.6% có bằng đại học. Tiền lương trung bình cho toàn bộ mẫu là 4035.776 nghìn đồng/tháng, của lao động nam là 3994.047 nghìn đồng/tháng, lao động nữ là 3323.349 nghìn đồng/tháng.

Bảng 4.2: Thống kê mô tả giá trị trung bình các biến năm 2014

Tên biến	Toàn bộ	Giới tính		Khu vực	
		LĐ nam	LĐ nữ	Thành thị	Nông thôn
Age	43.15	41.76	44.43	43.59	42.97
Age_square	2158.227	2016.72	2288.811	2195.041	2143.208
Female	0.52	0.00	1.00	0.53	0.51
Urban	0.28	0.28	0.29	1.00	0.00
Married	0.18	0.21	0.14	0.21	0.170
Ethnic	0.81	0.80	0.81	0.92	0.76
Healthy	0.11	0.09	0.12	0.09	0.11
Edu1	0.14	0.15	0.13	0.19	0.12
Edu2	0.032	0.047	0.018	0.04	0.026
Edu3	0.04	0.053	0.03	0.07	0.031
Edu4	0.019	0.019	0.018	0.03	0.013
Edu5	0.10	0.08	0.12	0.12	0.09
Employed	0.24	0.25	0.18	0.41	0.15
Income_m	4035.776	3994.047	3323.349	4486.241	3292.706
Log(income_m)	8.053	8.138	7.92	8.227	7.95
Số quan sát	6145	2308	3837	2122	4023

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Năm 2014 độ tuổi trung bình của lao động khu vực thành thị là 43.5, nông thôn là 42.9. Tỷ lệ 34.40% lao động thành thị là nữ, 35.5% này ở nông thôn. Lao động chưa lập gia đình ở thành thị là 21%, tỷ lệ này ở nông thôn là 17%. Có 92.8% lao động ở thành thị là dân tộc Kinh, 76.8% là của nông thôn. 9.8% lao động thành thị có thẻ bảo hiểm y tế, và 11.8% lao động nông thôn có thẻ bảo hiểm y tế. Trình độ giáo dục của lao động ở thành thị là 19,5% có bằng THPT, 4,7% có bằng sơ cấp nghề, 7,5% có bằng THCN và trung cấp nghề, 3,2% có bằng cao đẳng, và 12,4% có bằng đại học. Tỷ lệ này lần lượt ở nông thôn là 12,6%, 2,6%, 3,1%, 1,3% và 9,8%. Tiền lương trung bình cho lao động thành thị là 4486 nghìn đồng/tháng và 3292 nghìn đồng/tháng là của lao động nông thôn.

Bảng 4.3: Thống kê mô tả logarit tiền lương theo trình độ giáo dục năm 2010

Trình độ giáo dục	Toàn bộ mẫu		Làm thuê		Tự làm	
	Trung bình	S.D	Trung bình	S.D	Trung bình	S.D
THPT	7.21	0.75	7.34	0.76	6.75	0.87
Sơ cấp nghề	7.50	0.88	7.59	0.88	7.00	0.68
TCN, THCN	7.59	0.84	7.61	0.82	7.26	0.61
CD, CD nghề	7.67	0.71	7.77	0.70	7.36	0.82
ĐH và SDH	7.76	0.73	7.85	0.67	7.55	0.23
Số quan sát	5718		3894		1824	

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2010

Bảng 4.3 mô tả logarit tiền lương theo tháng của người lao động làm thuê và tự làm năm 2010. Đối với toàn bộ mẫu logarit tiền lương tăng theo cấp độ giáo dục, đối với nhóm THPT, sơ cấp nghề, TCN và THCN, CD, và Đại học lần lượt là 7.21, 7.50, 7.59, 7.67 và 7.76. Logarit tiền lương của lao động làm thuê và tự làm cũng tăng theo cấp độ giáo dục. Logarit tiền lương của lao động làm thuê cao hơn tự làm.

Bảng 4.4 cho thấy năm 2014 logarit tiền lương theo tháng của người lao động cao hơn năm 2010, lần lượt cho các cấp độ giáo dục là: 7.90, 8.127, 8.209, 8.32 và 8.33, có thể thấy logarit tiền lương tăng theo cấp độ giáo dục, logarit tiền lương người lao động làm thuê cao hơn tự làm.

Bảng 4.4: Thống kê mô tả logarit tiền lương theo trình độ giáo dục năm 2014

Trình độ giáo dục	Toàn bộ mẫu		Làm thuê		Tự làm	
	Trung bình	S.D	Trung bình	S.D	Trung bình	S.D
THPT	7.90	0.59	8.029	0.58	7.69	0.69
Sơ cấp nghề	8.12	0.68	8.24	0.63	7.70	0.55
TCN, THCN	8.20	0.67	8.26	0.68	7.81	0.42
CD, CD nghề	8.32	0.58	8.37	0.59	7.94	0.65
Đại học, SDH	8.33	0.56	8.43	0.49	8.10	0.47
Số quan sát	6,145		4,180		1,965	

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

4.2. Nghiên cứu dấu hiệu của vai trò phát tín hiệu trên thị trường lao động Việt Nam bằng phương pháp PSM

Ở đây, phương pháp PSM chủ trương so sánh thu nhập giữa những người tự làm và những người làm thuê có cùng điểm thiên hướng. Với những người tự làm thì thu nhập phản ánh trung thực năng lực của họ - không có vấn đề về TTBDX. Với người làm thuê thì có sự bất đối xứng thông tin giữa chủ lao động và người làm thuê về năng lực thực sự của người lao động. Do đó nếu kiểm soát được các yếu tố ảnh hưởng đến thu nhập, thì sự khác biệt trong thu nhập giữa hai nhóm người này sẽ là minh chứng cho sự hiện diện của TTBDX. Phương pháp PSM thực hiện việc kiểm soát này thông qua việc xác định điểm thiên hướng cho mỗi người lao động, và so sánh sự khác biệt thu nhập giữa những người có cùng mức điểm thiên hướng.

Do cùng với thời gian, mức độ thông tin bất đối xứng giữa người lao động và chủ lao động giảm dần, nên luận án thực hiện cho người lao động theo 3 nhóm kinh nghiệm: dưới 3 năm kinh nghiệm, từ 3 đến 8 năm kinh nghiệm, và trên 8 năm kinh nghiệm. Việc phân chia này do điều kiện số liệu. Việc tính toán sẽ được thực hiện cho năm 2010 và năm 2014, phân tích theo khu vực thành thị nông thôn và theo giới tính.

4.2.1. Kết quả ước lượng cho năm 2014

Bảng 4.5 cho thấy ATT cho toàn bộ mẫu năm 2014, ATT cho nhóm lao động làm thuê là 4964.95 và cho nhóm tự làm là 4050.43 tiền lương trung bình theo tháng của lao động làm thuê cao hơn lao động tự làm là 914.51 ngàn đồng, tỷ lệ phần trăm là 18.41%, với ý nghĩa thống kê là 3.69. Bảng chứng cho thấy thông tin bất đối xứng hiện diện đối với năm 2014.

Bảng 4.5: Kết quả ước lượng theo phương pháp PSM cho năm 2014

	Sample	Treated	Controls	Difference	% khác biệt	T-stat
Toàn bộ	Unmatched	4964.95	2630.69	2334.25	47.01	17.81
	ATT	4964.95	4050.43	914.51	18.41	3.69
Dưới 3 năm kinh nghiệm	Unmatched	4311.38	2421.35	1890.02	43.83	7.07
	ATT	4311.38	3228.06	1083.31	25.12	2.60
3 đến 8 năm kinh nghiệm	Unmatched	5367.79	2973.92	2393.87	44.59	7.10
	ATT	5367.79	3587.15	1780.64	33.17	3.33
Trên 8 năm kinh nghiệm	Unmatched	5229.18	2604.42	2624.76	50.19	15.30
	ATT	5229.18	4518.06	711.11	13.59	0.58

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Năm 2014, chênh lệch ATT của nhóm dưới 3 năm kinh nghiệm là 1083.31, và nhóm 3 đến 8 năm kinh nghiệm là 1780.64 (tỷ lệ phần trăm lần lượt là 25.12% và 33.17%) và đều có ý nghĩa thống kê, cho thấy thông tin bất đối xứng xuất hiện trên hai nhóm này. Đối với nhóm trên 8 năm kinh nghiệm có chênh lệch ATT là 711.11 và không có ý nghĩa thống kê. Có thể cho rằng

đối với năm 2014 thông tin bất đối xứng tồn tại với những lao động dưới 8 năm kinh nghiệm. Với lao động trên 8 năm kinh nghiệm, người sử dụng lao động đã nhận ra năng lực thực sự của người làm thuê và đưa ra mức lương tương xứng.

Kết quả ước lượng cho thành thị và nông thôn

Bảng 4.6 và 4.7 là kết quả tính toán ATT năm 2014 cho lao động làm việc ở khu vực thành và nông thôn. Chênh lệch ATT của khu vực thành thị và nông thôn lần lượt là 2293.56 và 854.73 và có ý nghĩa thống kê. Đối với nhóm lao động dưới 3 năm kinh nghiệm, chênh lệch ATT theo tỷ lệ phần trăm của khu vực thành thị là 46.81%, tỷ lệ này cao hơn nhiều so với khu vực nông thôn là 19.59%.

Bảng 4.6: Kết quả ước lượng theo phương pháp PSM cho khu vực thành thị năm 2014

	Sample	Treated	Controls	Difference	% khác biệt	T-stat
Toàn bộ	Unmatched	6078.69	3517.31	2561.38	42.13	5.45
	ATT	6078.69	3785.12	2293.56	37.73	4.29
Dưới 3 năm kinh nghiệm	Unmatched	4393.79	2646.96	1746.82	39.75	2.59
	ATT	4393.79	2336.94	2056.85	46.81	4.07
3 đến 8 năm kinh nghiệm	Unmatched	5950.90	3903.27	2047.62	34.40	1.65
	ATT	5950.90	4372.90	1577.99	26.51	1.65
Trên 8 năm kinh nghiệm	Unmatched	6043.71	3391.72	2651.99	43.88	6.12
	ATT	6043.71	5688.40	355.312	5.87	0.59

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Chênh lệch ATT của nhóm lao động trên 3 đến 8 năm kinh nghiệm có ý nghĩa thống kê đối với khu vực thành thị (chênh lệch phần trăm tương ứng là 26.51%) và không có ý nghĩa thống kê đối với khu vực nông thôn. Tuy nhiên với nhóm 8 năm kinh nghiệm trở lên chênh lệch ATT của khu vực thành thị cũng không còn nữa. Điều này cho thấy, thời gian tồn tại thông tin bất đối xứng ở khu vực thành thị năm 2014 là dưới 8 năm kinh nghiệm, và dưới 3 năm kinh nghiệm đối với khu vực nông thôn.

Bảng 4.7: Kết quả ước lượng theo phương pháp PSM cho khu vực nông thôn năm 2014

	Sample	Treated	Controls	Difference	% khác biệt	T-stat
Toàn bộ	Unmatched	6078.69	3517.31	2561.38	42.13	5.45
	ATT	6078.69	3785.12	2293.56	37.73	4.29
Dưới 3 năm kinh nghiệm	Unmatched	4393.79	2646.96	1746.82	39.75	2.59
	ATT	4393.79	2336.94	2056.85	46.81	4.07
3 đến 8 năm kinh nghiệm	Unmatched	5950.90	3903.27	2047.62	34.40	1.65
	ATT	5950.90	4372.90	1577.99	26.51	1.65
Trên 8 năm kinh nghiệm	Unmatched	6043.71	3391.72	2651.99	43.88	6.12
	ATT	6043.71	5688.40	355.312	5.87	0.59

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Kết quả ước lượng theo giới tính người lao động

Thời gian để người chủ lao động có thể nhận ra năng lực của người lao động còn phụ thuộc giới tính của người lao động. Chẳng hạn lao động nam có thể chủ động hơn khi thực hiện công việc nên mức độ thông tin bất đối xứng cũng có thể khác nhau giữa nam và nữ.

Bảng 4.8 và 4.9 là kết quả tính toán ATT cho nhóm lao động nam và nữ 2014. Với nhóm dưới 3 năm kinh nghiệm chênh lệch ATT cho nam giới là 915.83 với nữ giới là 1000.55 (tương ứng phần trăm là 23.21% và 25.92%). Đối với nhóm 3 năm kinh nghiệm trở lên chênh lệch ATT cho lao động nam giới đều có ý nghĩa thống kê, cho thấy thông tin bất đối xứng với lao động nam là trên 8 năm kinh nghiệm. Với lao động nữ, ATT cho nhóm trên 3 năm kinh nghiệm không có sự chênh lệch. Năm 2014 người sử dụng lao động cần nhiều thời gian để hiểu rõ năng lực thực sự của lao động nam so với lao động nữ (trên 8 năm đối với lao động nam và dưới 3 năm đối với lao động nữ).

Bảng 4.8: Kết quả ước lượng theo phương pháp PSM cho lao động nam năm 2014

	Sample	Treated	Controls	Difference	% khác biệt	T-stat
Toàn bộ	Unmatched	5336.99	2612.50	2724.48	51.04	15.32
	ATT	5336.99	3503.99	1833.00	34.34	6.42
Dưới 3 năm kinh nghiệm	Unmatched	3945.10	2458.68	1486.41	37.67	7.25
	ATT	3945.10	3029.27	915.83	23.21	3.47
3 đến 8 năm kinh nghiệm	Unmatched	5760.28	2843.56	2916.71	50.63	5.44
	ATT	5760.28	2931.79	2828.48	49.10	1.95
Trên 8 năm kinh nghiệm	Unmatched	5640.50	2676.94	2963.56	52.54	13.80
	ATT	5640.50	3276.33	2364.17	41.91	3.44

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Bảng 4.9: Kết quả ước lượng theo phương pháp PSM cho lao động nữ năm 2014

	Sample	Treated	Controls	Difference	% khác biệt	T-stat
Toàn bộ	Unmatched	4539.18	2658.53	1880.64	41.43	9.79
	ATT	4539.18	3567.48	971.69	21.40	2.29
Dưới 3 năm kinh nghiệm	Unmatched	3859.30	2633.34	1225.95	31.76	4.27
	ATT	3859.30	2858.75	1000.55	25.92	2.53
3 đến 8 năm kinh nghiệm	Unmatched	5022.83	3172.88	1849.94	36.83	4.68
	ATT	5022.83	4517.60	505.23	10.05	0.53
Trên 8 năm kinh nghiệm	Unmatched	4616.41	2488.58	2127.83	46.09	7.58
	ATT	4616.41	5492.12	-875.71	-18.96	-0.33

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

4.2.2. So sánh với năm 2010

Chênh lệch tiền lương giữa lao động làm thuê và lao động tự làm năm 2010 của nhóm dưới 3 năm kinh nghiệm là 692.98, và nhóm 3 đến 8 năm kinh nghiệm là 1459.26 (tỷ lệ phần trăm lần lượt là 32.42% và 47.29%) cho thấy thời gian thông tin bất đối xứng là dưới 8 năm (Phụ lục 1).

Có thể thấy thời gian thông tin bất đối xứng ở hai năm 2010 và 2014 không có sự khác biệt theo khu vực làm việc, dưới 8 năm đối với khu vực thành thị và dưới 3 năm đối với khu vực nông thôn. Tuy nhiên chênh lệch tiền lương của lao động làm thuê và lao động tự làm năm 2014 là cao hơn nhiều so với năm 2010 (cả khu vực thành thị và nông thôn), có thể giải thích là do mức sống năm 2014 cao hơn năm 2010 và do trong giai đoạn khủng hoảng năm 2010 người sử dụng lao động khắt khe hơn trong việc tuyển dụng và chi trả lương cho người lao động so với năm 2014 (Phụ lục 2 và Phụ lục 3).

Thời gian thông tin bất đối xứng của lao động nam năm 2010 và năm 2014 là như nhau (trên 8 năm kinh nghiệm), thời gian thông tin bất đối xứng của lao động nữ năm 2010 dài hơn năm 2014 (dưới 3 năm đối với năm 2014 và dưới 8 năm đối với năm 2010) (Phụ lục 4 và Phụ lục 5).

Như vậy có thể thấy có sự hiện diện của thông tin bất đối xứng trên thị trường lao động Việt Nam và đây là dấu hiệu cho thấy vai trò phát tín hiệu của giáo dục trên thị trường lao động. Do đó bước tiếp theo, luận án sẽ thực hiện ước lượng suất sinh lời của giáo dục cũng như vai trò phát tín hiệu của giáo dục.

4.3. Ước lượng suất sinh lời và vai trò phát tín hiệu của giáo dục bằng phương pháp hàm kiểm soát tiền lương Heckman

Trong phần này chúng tôi sẽ ước lượng suất sinh lời dựa trên sự mở rộng của hàm tiền lương Mincer (1974) bằng phương pháp Heckman. Việc sử dụng phương pháp Heckman thay vì hồi quy thông thường nhằm để kiểm soát vấn đề biến nội sinh do vấn đề thiếu biến không quan sát được. Hàm tiền lương của Mincer cơ bản có dạng:

$$\ln w = \alpha + \beta_1 z + \beta_2 z^2 + \beta_3 e + \beta X + \varepsilon \quad (1)$$

Trong đó:

W: tiền lương của người lao động

Z: số năm kinh nghiệm của người lao động, số năm kinh nghiệm của người lao động được kỳ vọng là có tác động thuận chiều lên mức lương. Điều này là do kinh nghiệm làm việc sẽ giúp người lao động nâng cao năng suất, do đó mức lương của họ.

Z^2 : bình phương số năm kinh nghiệm. Biến này được đưa vào nhằm tính đến tác động biên giảm dần của kinh nghiệm lên mức lương

E: học vấn của người lao động. Học vấn của người lao động là yếu tố quan trọng tạo ra vốn nhân lực, do đó được kỳ vọng tác động thuận chiều đến mức lương

X: các biến kiểm soát khác

Khi đó hệ số β_3 sẽ cho biết hiệu suất sinh lời của giáo dục.

Nếu giáo dục hàm chứa cả vai trò phát tín hiệu, khi đó mô hình trên có thể được biến đổi thành:

$$\ln w = \alpha + \beta_1 z + \beta_2 z^2 + \beta_3 e + c(e) + \beta X + \varepsilon \quad (2)$$

Trong đó β_3 thể hiện đóng góp của học vấn lên năng suất lao động và do đó lên lương của người lao động, và $c(e)$ thể hiện vai trò phát tín hiệu của giáo dục lên người sử dụng lao động. Hay nói một cách khác, $c(e)$ là mức lương mà người sử dụng lao động sẵn sàng chi trả thêm cho người lao động có học vấn cao hơn so với người không có học vấn, kể cả khi các kiến thức học được không giúp nâng cao năng suất lao động, do việc có được học vấn này là minh chứng cho năng lực của người lao động. Vấn đề là ở chỗ ta không thể đồng thời ước lượng hai tham số này trong mô hình (2).

Tuy nhiên có thể phân tích như sau: nếu người lao động là làm thuê thì học vấn sẽ có đồng thời hai tác dụng như đã phân tích ở trên, còn nếu người lao động là tự làm, thì bản thân họ biết năng lực của họ là bao nhiêu, và không có vấn đề về bất đối xứng thông tin, do đó vai trò phát tín hiệu của giáo dục là không tồn tại, hay nói cách khác, (2) có thể viết như sau:

$$Lnw = \alpha + \beta_1 z + \beta_2 z^2 + \beta_3 e + c(e) + \beta X + \varepsilon \text{ với người làm thuê}$$

$$Lnw = \alpha + \beta_1 z + \beta_2 z^2 + \beta_3 e + \beta X + \varepsilon \text{ với người tự làm}$$

Hay gộp vào trong mô hình:

$$Lnw = \beta_0 + \beta_1 z + \beta_2 z^2 + \beta_3 e + \alpha Employed + \beta X + \varepsilon \quad (3)$$

Trong đó $Employed = 1$ nếu người lao động làm thuê, $=0$ nếu người lao động tự làm.

Việc áp dụng trực tiếp các phương pháp ước lượng chuẩn tắc lên mô hình (3) vẫn có thể có những sai lệch đáng kể, do vấn đề selection bias, là vấn đề mà trong đó sự khác biệt giữa những người làm thuê và những người tự làm là không phải do ngẫu nhiên mà do sự khác biệt nội tại giữa hai nhóm này. Chẳng hạn có thể cho rằng những người tự làm là họ chọn tự làm do có tính cách thích mạo hiểm hơn so với những người làm thuê. Vì vậy chúng tôi sẽ sử dụng phương pháp hàm kiểm soát (control function) được đề xuất bởi Heckman (Heckman, 1979).

4.3.1. Phương pháp Heckman ước lượng suất sinh lợi giáo dục cho năm 2014

Có thể cho rằng, với những người lao động làm thuê thì vai trò phát tín hiệu chỉ có tác động lên lương trong một thời gian ban đầu, khi mà người chủ lao động chưa thể biết được nhiều về năng lực thực sự của người lao động, khi đó giáo dục sẽ được hiểu như tín hiệu về năng lực này. Điều này có thể làm cho tiền lương của người làm thuê sẽ cao hơn người tự làm trong giai đoạn ban đầu. Khi đã làm lâu năm trong công việc, sự bất đối xứng thông tin giữa người lao động và chủ lao động giảm đi, học vấn không còn vai trò phát tín hiệu nữa, năng suất lao động sẽ thể hiện đầy đủ hơn về năng lực của người lao động mà không cần đến tín hiệu nữa, khi đó sự khác biệt giữa người làm thuê và tự làm sẽ bị triệt tiêu. Do đó chúng tôi chia mẫu ra làm 3 nhóm theo số năm kinh nghiệm, trong đó nhóm 1 dành cho những người dưới 3 năm kinh nghiệm, nhóm 2 từ 3 năm đến 8 năm kinh nghiệm, và nhóm thứ 3 là trên 8 năm kinh nghiệm.

Bảng 4.10 cho thấy kết quả hồi quy hàm tiền lương của người lao động năm 2014: nhóm 1 có dưới 3 năm kinh nghiệm và nhóm 2 từ 3 đến 8 năm kinh nghiệm, nhóm 3 trên 8 năm kinh nghiệm.

Đối với lao động nhóm 1 và 2 hệ số của biến $Employed$ là dương và có ý nghĩa thống kê, hệ số này không có ý nghĩa thống kê đối với nhóm 3. Về định lượng, có thể cho rằng vai trò phát tín hiệu là khá đáng kể. Từ bảng 4.6 có thể

thấy rằng, với nhóm người ít kinh nghiệm làm việc thì học vẫn có đóng vai trò phát tín hiệu. Cụ thể ở đây, hệ số của biến này bằng 0.85, điều này có nghĩa là mức lương theo tháng trung bình của người làm thuê sẽ cao hơn người tự làm là 133.96% ($= (e^{0.85} - 1) \cdot 100\%$) đối với nhóm 1, và 33.4% ($= (e^{0.28} - 1) \cdot 100\%$) đối với nhóm 2. Với nhóm 3 hệ số của biến Employed là -0.02 và không có ý nghĩa thống kê, nghĩa là số năm kinh nghiệm càng tăng thì chênh lệch tiền lương của người làm thuê và tự làm sẽ dần mất đi, mức lương lúc này phản ánh năng lực thực sự của người lao động làm thuê.

Bảng 4.10 Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng năm 2014

<i>Variable</i>	Nhóm 1 (dưới 3 năm kinh nghiệm)	Nhóm 2 (3 đến 8 năm kinh nghiệm)	Nhóm 3 (trên 8 năm kinh nghiệm)
Age	-0.34 (-0.86)	0.15 (0.28)	0.07*** (6.92)
Age square	0.009 (0.93)	-0.0029 (-0.26)	-0.001*** (-9.59)
Urban	-0.05 (-1.02)	0.08 (1.59)	0.37*** (7.69)
Ethnic	0.03 (0.54)	0.28*** (2.93)	0.32*** (6.99)
Married	-0.12** (-2.33)	-0.10** (-2.24)	-0.05 (-0.99)
Female	-0.12** (-2.51)	-0.16*** (-4.2)	-0.26*** (-10.38)
Edu_2	0.17* (1.59)	0.03 (0.46)	0.19*** (3.62)
Edu_3	0.10 (1.13)	0.09* (1.44)	0.10** (2.33)
Edu_4	-0.002 (-0.03)	0.04 (0.87)	0.25*** (3.38)
Edu_5	0.18** (1.98)	0.33*** (6.82)	0.31*** (8.13)
Employed	0.85*** (7.54)	0.28 (1.42)	-0.02 (-0.22)
Constant	10.69*** (2.57)	5.55 (0.74)	6.65*** (26.64)
Lambda	-0.34	-0.032	0.16
N	530	627	2350

(* , ** , *** có ý nghĩa với mức 10%, 5%, 1%. Giá trị trong ngoặc là tỷ số t)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Sau khi đã kiểm soát yếu tố phát tín hiệu của giáo dục, các hệ số của các biến edu lúc này phản ánh lợi suất của giáo dục trong vai trò cung cấp vốn nhân lực. Kết quả hồi quy trong bảng 4.6 cho thấy năm 2014 đóng góp của giáo dục gia tăng cùng với bằng cấp đào tạo của họ: cụ thể với nhóm ít kinh nghiệm thì người lao động có trình độ sơ cấp nghề, trung cấp nghề và trung cấp chuyên nghiệp, cao đẳng nghề và cao đẳng mức lương không có sự khác biệt so với nhóm tốt nghiệp THPT, bậc đại học có mức lương cao hơn nhóm THPT là 19.86%. Đối với nhóm 2 bậc đại học này có mức lương cao hơn bậc THPT là 40.03%. Với nhóm 3 chênh lệch tiền lương giữa các bậc giáo dục cao hơn bậc THPT lần lượt là 19.7%, 10.7%, 25.7% và 31.7%.

Như vậy có thể cho rằng giả thuyết nêu ở phần trên là xác thực, trong đó vai trò phát tín hiệu của giáo dục là tồn tại, và chỉ tồn tại với nhóm ít kinh nghiệm mà thôi, nó sẽ mất dần đi do dần dần sẽ không còn sự mất cân xứng thông tin giữa người làm thuê và người chủ lao động, và người lao động làm thuê sử dụng bằng cấp giáo dục như là tín hiệu về năng lực của mình đến người sử dụng lao động. Điều này là phù hợp với suy luận ở phần trên.

Kết quả ước lượng cho khu vực thành thị và nông thôn

Để chi tiết hơn, luận án sẽ ước lượng theo khu vực. Bảng 4.11 cho thấy kết quả ước lượng hàm tiền lương khu vực thành thị và nông thôn năm 2014. Ở khu vực thành thị người lao động dưới 3 năm kinh nghiệm sử dụng giáo dục như vai trò phát tín hiệu, tiền lương của lao động làm thuê cao hơn 2.38 lần so với lao động tự làm, và chênh lệch giảm còn 0.77 lần ở nhóm 3 đến 8 năm kinh nghiệm, với nhóm lao động trên 8 năm kinh nghiệm vai trò phát tín hiệu của giáo dục không còn nữa. Về vai trò cung cấp vốn nhân lực, lao động có bằng đại học có mức lương cao hơn lao động tốt nghiệp THPT là 21.33% ở nhóm 1 và 53.48% ở nhóm 2. Đối với nhóm 3 lao động sơ cấp nghề, cao đẳng và đại học có mức lương cao hơn lao động tốt nghiệp THPT lần lượt là 13.23%, 23.76% và 67.41%.

Khu vực nông thôn vai trò phát tín hiệu chỉ tồn tại ở nhóm 1, chênh lệch tiền lương giữa lao động làm thuê và tự làm do tín hiệu tạo ra là 75.68%. Hệ số biến Employed không có ý nghĩa thống kê ở nhóm 2, và 3. Về tác động cung cấp vốn nhân lực của giáo dục, nhóm 1 không có sự khác biệt tiền lương của lao động theo cấp độ giáo dục, ở nhóm 2 chỉ có những lao động có trình độ giáo dục đại học trở lên có mức lương cao hơn nhóm lao động tốt nghiệp THPT là 17.7%. Ở nhóm 3 có sự khác biệt tiền lương giữa nhóm sơ cấp nghề, trung cấp nghề và cao đẳng so với nhóm THPT lần lượt 30.86%, 22.39% và 40.61%.

Năm 2014 thời gian thể hiện vai trò phát tín hiệu của giáo dục ở khu vực thành thị dài hơn khu vực nông thôn (dưới 8 năm kinh nghiệm ở khu vực thành thị và dưới 3 năm ở khu vực nông thôn).

Bảng 4.11 Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng khu vực thành thị và nông thôn năm 2014

Variable	Nhóm 1 (dưới 3 năm kinh nghiệm)		Nhóm 2 (3 đến 8 năm kinh nghiệm)		Nhóm 3 (trên 8 năm kinh nghiệm)	
	Thành thị	Nông thôn	Thành thị	Nông thôn	Thành thị	Nông thôn
Age	-2.88** (-1.76)	0.12 (0.1)	1.53** (1.73)	-0.52 (-0.72)	0.09*** (6.27)	0.013 (0.83)
Age square	0.06** (1.81)	-0.002 (-0.09)	-0.029** (-1.71)	0.01 (0.72)	-0.001*** (-7.93)	-0.0004** (-2.36)
Ethnic	-0.02 (-0.14)	0.13 (1.31)	0.16 (0.9)	0.28*** (2.95)	0.17** (2.15)	0.38*** (5.62)
Married	0.05 (0.65)	-0.16** (-2.37)	-0.06 (-0.97)	-0.08* (-1.37)	-0.09 (-1.27)	-0.029 (-0.39)
Female	-0.10* (-1.37)	-0.16** (-2.53)	-0.21*** (-3.61)	-0.10** (-2.07)	-0.20*** (-5.62)	-0.25*** (-7.09)
Edu_2	0.19 (1.22)	0.25** (2.06)	0.25** (1.82)	-0.08 (-0.8)	0.12** (1.72)	0.26*** (3.18)
Edu_3	0.15 (1.19)	0.03 (0.27)	0.13* (1.33)	0.08 (1.03)	0.057 (0.97)	0.20*** (2.71)
Edu_4	0.11 (0.93)	0.02 (0.24)	0.0009 (0.01)	0.08 (1.1)	0.21** (2.32)	0.34** (2.41)
Edu_5	0.19** (1.62)	0.17 (1.43)	0.42*** (6.22)	0.16** (2.22)	0.51*** (7.85)	-0.09 (-1.59)
Employed	1.22*** (10.41)	0.56** (2.24)	0.57** (1.72)	0.29 (1.5)	0.65 (1.44)	-0.32 (-1.58)
Constant	37.17** (2.1)	6.08 (0.46)	-12.59 (-1.1)	14.49 (1.52)	5.86*** (9.25)	8.007*** (20.67)
Lambda	-0.53	-0.19	-0.17	-0.028	-0.14	0.32
N	250	275	254	371	1051	1277

(* , ** , *** có ý nghĩa với mức 10%, 5%, 1%. Giá trị trong ngoặc là tỷ số t)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Kết quả ước lượng theo giới tính người lao động

Kết quả ước lượng tiền lương theo giới tính của năm 2014 được trình bày ở bảng 4.12.

Đối với lao động nam giáo dục đóng vai trò phát tín hiệu ở cả nhóm 1 và nhóm 2. Chênh lệch tiền lương do tín hiệu tương ứng là 129.39 % cho nhóm 1 và 78.61 % cho nhóm 2. Về vai trò cung cấp vốn nhân lực, không có sự chênh lệch tiền lương giữa các cấp độ giáo dục ở nhóm 1, ở nhóm 2 chênh lệch tiền lương giữa các nhóm trung cấp, cao đẳng và đại học so với THPT lần lượt là

28.09%, 15.33% và 53.12%. Chênh lệch tiền lương tăng theo cấp độ giáo dục ở nhóm 3 tương ứng là 29.6%, 20.29%, 35.72% và 38.36%.

Đối với nữ giới, lao động làm thuê có mức lương trung bình cao hơn lao động tự làm là 1.28 lần, không có sự khác biệt tiền lương giữa các cấp độ giáo dục ở nhóm 1. Đối với nhóm 2 chỉ tồn tại chênh lệch tiền lương giữa lao động nữ có bằng cấp đại học trở lên với nhóm THPT là 30.44%. Và đối với nhóm 3 chênh lệch tiền lương giữa nhóm lao động có bằng cấp cao đẳng, đại học trở lên với THPT lần lượt là 24.94% và 34.41%.

Bảng 4.12 Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng theo lao động nam và nữ năm 2014

Variable	Nhóm 1 (dưới 3 năm kinh nghiệm)		Nhóm 2 (3 đến 8 năm kinh nghiệm)		Nhóm 3 (trên 8 năm kinh nghiệm)	
	Nam	Nữ	Nam	Nữ	Nam	Nữ
Age	0.008 (0.01)	-0.54 (0.96)	-1.21 (-1.39)	1.37** (1.85)	0.094*** (7.35)	0.039*** (2.73)
Age square	0.00038 (0.03)	0.013 (1.01)	0.023 (1.39)	-0.02** (-1.82)	-0.001*** (-9.34)	-0.0007*** (-4.22)
Urban	-0.08 (-1.06)	-0.029 (-0.41)	0.10 (1.41)	0.03 (0.55)	0.33*** (5.3)	0.39*** (5.38)
Ethnic	-0.04 (-0.55)	0.18** (1.87)	-0.018 (-0.14)	0.57*** (5.93)	0.38*** (5.89)	0.23*** (3.48)
Married	-0.15 (-1.75)	-0.08 (-1.24)	-0.17*** (-2.93)	-0.01 (-0.15)	-0.18** (-2.53)	0.08 (1.12)
Edu_2	0.23 (1.45)	0.10 (0.71)	0.17 (1.38)	-0.09 (-0.98)	0.25*** (3.9)	0.10 (1.11)
Edu_3	0.03 (0.24)	0.12 (1.05)	0.24** (2.56)	-0.03 (-0.43)	0.18*** (3.22)	0.006 (0.08)
Edu_4	0.10 (0.76)	-0.07 (-0.66)	0.14* (1.56)	-0.04 (-0.54)	0.30*** (3.03)	0.22** (1.91)
Edu_5	0.16 (1.27)	0.18 (1.43)	0.42*** (5.29)	0.26*** (4.53)	0.32*** (6.2)	0.29*** (5.14)
Employed	0.83*** (5.35)	0.82*** (5.03)	0.58** (2.37)	0.23 (0.85)	-0.08 (-0.54)	0.09 (0.62)
Constant	7.17 (1.22)	12.54** (2.15)	23.4** (2.07)	-10.75 (-1.13)	6.08*** (19.81)	7.13*** (16.69)
Lambda	-0.34	-0.32	-0.12	-0.059	0.20	0.08
N	244	286	308	319	1356	994

(* , ** , *** có ý nghĩa với mức 10%, 5%, 1%. Giá trị trong ngoặc là tỷ số t)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

4.3.2. So sánh với năm 2010

Năm 2010, vai trò phát tín hiệu của giáo dục là dưới 3 năm kinh nghiệm, điều này có thể thấy do hệ số biến Employed của nhóm này là dương và có ý nghĩa thống kê. Chênh lệch tiền lương giữa lao động làm thuê và tự làm do tín hiệu là 124.29% ($= (e^{0.8077706} - 1) \cdot 100\%$). Đối với nhóm trên 3 năm kinh nghiệm vai trò phát tín hiệu của giáo dục không còn nữa. Như vậy vai trò phát tín hiệu của giáo dục trong giai đoạn 2014 là dài hơn so với năm 2010 là giai đoạn khủng hoảng kinh tế.

Về vai trò cung cấp nguồn nhân lực của giáo dục, không có sự khác biệt tiền lương giữa các bậc giáo dục với bậc THPT đối với nhóm 1. Đối với nhóm 2 tiền lương của người lao động có trình độ sơ cấp nghề, trung cấp nghề và trung cấp chuyên nghiệp, cao đẳng nghề và cao đẳng, đại học có tiền lương cao hơn nhóm THPT lần lượt là 13.57%, 16.33%, 18.55% và 24.34%. Chênh lệch này đối với lao động nhóm 3 lần lượt là 21.12 %, 16.62%, 16.64% và 26.63% (Phụ lục 6).

Năm 2010 vai trò phát tín hiệu của giáo dục ở khu vực thành thị là hiện diện với lao động dưới 3 năm kinh nghiệm, tiền lương người lao động làm thuê cao hơn lao động tự làm là 1.09 lần. Đối với lao động trên 3 năm kinh nghiệm vai trò này không còn nữa. Về vai trò cung cấp vốn nhân lực, lao động có bằng cấp đại học có tiền lương cao hơn lao động tốt nghiệp THPT là 0.63 lần ở nhóm 1. Lao động có bằng cấp cao đẳng và đại học có tiền lương cao hơn THPT lần lượt là 0.31 lần và 0.93 lần ở nhóm 2. Đối với nhóm 3 lao động sơ cấp nghề, trung cấp nghề, cao đẳng và đại học có mức lương cao hơn THPT lần lượt là 0.35, 0.14, 0.44 và 0.9 lần (Phụ lục 7).

Kết quả tính toán cho khu vực nông thôn, cho thấy năm 2010 vai trò phát tín hiệu của giáo dục thể hiện ở nhóm 1 và nhóm 2, hệ số Employed của hai nhóm này đều dương và có ý nghĩa thống kê (mức lương của lao động làm thuê cao hơn tự làm là 2.96 lần với nhóm 1, và 1.8 lần với nhóm 2). Vai trò phát tín hiệu cũng không còn ở nhóm 3, lúc này người sử dụng lao động trả lương cho người lao động theo năng lực thực sự, nhóm sơ cấp nghề, trung cấp và cao đẳng có mức lương cao hơn nhóm lao động tốt nghiệp THPT lần lượt là 0.5, 0.24 và 0.4 lần (Phụ lục 8).

Khu vực thành thị thời gian phát tín hiệu của giáo dục năm 2014 dài hơn năm 2010 (dưới 8 năm kinh nghiệm đối với năm 2014, dưới 3 năm kinh nghiệm đối với năm 2010). Đối với khu vực nông thôn thời gian phát tín hiệu của giáo dục năm 2014 ngắn hơn năm 2010 (dưới 3 năm đối với năm 2014 và dưới 8 năm đối với 2010). Khu vực thành thị là nơi tập trung ngành nghề lao động cần kỹ năng cao, để nhận ra năng lực thật sự của người lao động người sử dụng lao động cần nhiều thời gian, tuy nhiên trong thời kỳ khủng hoảng việc tuyển dụng được thắt chặt dẫn đến thời gian nhận biết năng lực thực sự của người lao động ngắn hơn trong thời kỳ kinh tế ổn định.

Năm 2010 thời gian phát tín hiệu của giáo dục cho lao động nam và lao động nữ là như nhau (dưới 3 năm kinh nghiệm), mức độ chênh lệch tiền lương giữa lao động làm thuê và tự làm do vai trò tín hiệu là 51.55% của lao động nam và 90.87% là của lao động nữ (Phụ lục 9 và Phụ lục 10).

Đối với lao động nam, vai trò cung cấp nguồn nhân lực của giáo dục được thể hiện qua chênh lệch tiền lương giữa nhóm sơ cấp nghề, trung cấp nghề và cao đẳng, đại học so với nhóm THPT lần lượt là 26.64%, 19.59%, 28.05% và 51.55%, 17.34% ở nhóm 1. Với nhóm 2, vai trò cung cấp nhân lực tăng lên theo cấp độ giáo dục tương ứng là 24.04%, 24.99% và 63.71% (không có sự khác biệt tiền lương giữa lao động nam có trình độ cao đẳng và THPT). Ở nhóm 3, chênh lệch tiền lương giữa các cấp độ giáo dục và cấp độ THPT là 41.51%, 16.50%, 38.78% và 53.91%,

Đối với lao động nữ không có sự khác biệt tiền lương giữa các cấp giáo dục ở nhóm dưới 3 năm kinh nghiệm. Trong nhóm 2 và 3 chỉ có sự chênh lệch giữa đại học và THPT lần lượt là 60.33% và 48.46%.

Như vậy có thể thấy rằng thời gian phát tín hiệu của giáo dục theo giới tính người lao động năm 2014 dài hơn năm 2010. Thời gian phát tín hiệu là dưới 8 năm đối với lao động nam và dưới 3 năm đối với lao động nữ năm 2014, dưới 3 năm đối với lao động nam và nữ năm 2010.

Qua các phân tích từ phương pháp ước lượng Heckman trên cho thấy sự hiện diện của vai trò phát tín hiệu của giáo dục sau phổ thông. Thông tin bất đối xứng xuất hiện ở nhóm lao động dưới 3 năm kinh nghiệm, từ 3 đến 8 năm kinh nghiệm tùy vào khu vực thành thị, nông thôn và giới tính của

người lao động, giáo dục đóng vai trò phát tín hiệu cho hai nhóm kinh nghiệm này. Thông tin không đối xứng sẽ không hiện diện đối với những lao động trên 8 năm kinh nghiệm.

Ngoài ra có thể thấy rằng mức chênh lệch giữa các nhóm theo trình độ đào tạo là khác nhau giữa các nhóm kinh nghiệm, trong đó nhóm người trên 8 năm kinh nghiệm có mức chênh lệch cao hơn so với nhóm người dưới 8 năm kinh nghiệm. Điều này cũng phù hợp với một thực tế lịch sử của Việt Nam: trong những năm gần đây, việc xã hội hóa giáo dục, đặc biệt là ở bậc đại học được thực hiện quá nhanh, làm cho sở hữu được một bằng cấp là khó khăn hơn rất nhiều so với thời gian gần đây, khó đảm bảo được chất lượng đào tạo.

4.4. Ước lượng hiệu suất sinh lời và vai trò phát tín hiệu của giáo dục bằng phương pháp Lewbels

Trong phần này mô hình Mincer, được sử dụng để ước lượng hiệu suất sinh lời của giáo dục. Mô hình Mincer căn bản được viết như sau (Lemieux, 2006):

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 x + \beta_3 x^2 + u$$

Trong đó wage là mức lương, educ: học vấn, x: số năm kinh nghiệm của người lao động.

Phương pháp xử lý hai vấn đề đã nêu trong mục giới thiệu như sau:

1. Xử lý vấn đề biến nội sinh: Do biến *học vấn* thường bị nội sinh, chúng tôi sử dụng phương pháp của Lewbels (Lewbels, 2012) để giải quyết. Thông thường, khi có vấn đề biến công cụ, người ta thường sử dụng phương pháp ước lượng sử dụng biến công cụ, tuy nhiên trong rất nhiều trường hợp, việc tìm biến công cụ không phải là việc dễ dàng. Thêm vào đó, nếu biến công cụ không tốt thì phương pháp này không giúp cải thiện được vấn đề. Phương pháp Lewbels mà một phương pháp mới có thể khắc phục nhược điểm này bằng cách tự tạo ra biến công cụ sử dụng bản thân các biến ngoại sinh trong mô hình, đây cũng chính là ưu điểm vượt trội của nó, đặc biệt là trong các trường hợp mà việc tìm biến công cụ bên ngoài là khó khăn. Ngoài ra, chúng tôi cũng bổ sung thêm biến (không nằm trong mô hình) là biến “*học vấn của bố*”, đây cũng là một biến số thường được sử dụng làm biến công cụ cho học vấn của người lao động. Biến này được chọn do nó thỏa mãn hai điều kiện căn bản của biến công cụ: (1): học vấn của bố thường tương quan với học vấn của con, do người bố có học

vấn cao thường coi trọng việc học hành và do đó tập trung ưu tiên cho học hành của con cái, và (2): học vấn của bố không có ảnh hưởng trực tiếp đến tiền lương của người con, do đó nó không có mặt trong sai số ngẫu nhiên u của mô hình.

2. *Xem xét vai trò phát tín hiệu của giáo dục*: Để xem xét vai trò này, chúng tôi chia các quan sát thành các nhóm theo kinh nghiệm làm việc. Qua thời gian, năng suất lao động của mỗi người sẽ được bộc lộ ra rõ ràng hơn do chủ lao động ngày càng có thông tin đầy đủ hơn về năng suất của người lao động, nên vai trò của tín hiệu sẽ giảm đi cùng với thời gian. Do đó nếu với nhóm nhiều năm kinh nghiệm, đóng góp biên của học vấn lên năng suất lao động giảm thì đây là chứng cứ để cho rằng có vai trò phát tín hiệu của giáo dục (Lange, 2007), (Kjelland, 2008). Còn nếu ngược lại thì chưa có chứng cứ để cho rằng giáo dục có vai trò phát tín hiệu. Để thực hiện điều này, chúng tôi lần lượt ước lượng cho từng nhóm theo kinh nghiệm làm việc, trong đó quan tâm đến đóng góp biên của học vấn, có nghĩa là quan tâm đến sự khác biệt của hệ số biến học vấn giữa nhóm người được học cao đẳng, đại học với những người không được học.

4.4.1. Phương pháp Lewbels ước lượng hiệu suất sinh lời và vai trò phát tín hiệu của giáo dục 2014

Kết quả ước lượng mô hình theo ba nhóm kinh nghiệm được báo cáo trong Bảng 4.13:

Đối với năm 2014, hệ số biến $educ1 - educ5$ trên cột “dưới 3 năm kinh nghiệm”: với những người có thời gian làm việc < 3 năm, không có sự khác biệt tiền lương giữa nhóm sơ cấp nghề và THPT, những lao động có bằng THCN, TC nghề, cao đẳng và đại học trở lên có lương cao hơn THPT lần lượt là $15.5\% (=e^{0.14}-1)$ và $27.36\% (=e^{0.24}-1)$ và $51.5\% (=e^{0.41}-1)$. Với nhóm người có từ 3-8 năm kinh nghiệm, những người có bằng đại học có lương cao hơn 32.9% so với người có bằng phổ thông $(= e^{0.28}-1)$. Với nhóm có trên 8 năm kinh nghiệm, những lao động có bằng đại học, cao đẳng, trung cấp và sơ cấp nghề có lương cao hơn so với THPT lần lượt là 21.1% , 35.7% , 17% và 14.7% .

Việc so sánh hệ số của biến $educ1-educ5$ của ba nhóm cho thấy: đóng góp biên của việc học đại học so với học phổ thông của nhóm 2 là thấp hơn nhóm 1, nhóm 3 thấp hơn nhóm 2. Điều này có thể lý giải như sau: trong thời gian 8 năm, đa phần mọi người đều ở mức độ thử việc nên sự chênh lệch về mức lương là không quá lớn. Tới mức lương kí hợp đồng thì mức chênh lệch

được gia tăng do lúc này vẫn tồn tại đồng thời hai vai trò của giáo dục: giúp gia tăng năng suất lao động và giúp chủ lao động nhận biết năng lực. Tuy nhiên khi làm việc quá lâu (trên 8 năm) thì vai trò phát tín hiệu giảm đi, nên ước lượng của tác động biên của giáo dục đại học là giảm đi. Điều này cũng đúng với tác động biên của giáo dục cao đẳng, trung cấp và sơ cấp nghề. Vậy kết quả ước lượng cho thấy giáo dục có vai trò phát tín hiệu trong việc xác định mức lương.

Bảng 4.13: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels cho ba nhóm kinh nghiệm năm 2014

<i>Variable</i>	Nhóm 1 (dưới 3 năm kinh nghiệm)	Nhóm 2 (3 đến 8 năm kinh nghiệm)	Nhóm 3 (trên 8 năm kinh nghiệm)
<i>Educl</i>	(cơ sở)	(cơ sở)	(cơ sở)
2	0.17 (0.74)	0.57* (1.43)	0.13** (2.35)
3	0.14** (2.21)	0.109 (0.61)	0.15** (2.62)
4	0.24*** (3.11)	0.109 (1.1)	0.18*** (4.58)
5	0.41*** (4.3)	0.28*** (2.76)	0.19*** (3.65)
<i>Age</i>	-1.71 (-0.85)	0.08 (0.26)	0.045*** (3.03)
<i>Age_square</i>	0.037 (0.86)	-0.0015 (-0.27)	-0.0007*** (-4.04)
<i>Urban</i>	0.05 (1.16)	0.102** (2.8)	0.21*** (9.98)
<i>Ethnic</i>	0.08 (1.00)	0.32*** (5.1)	0.28*** (7.93)
<i>Married</i>	-0.079** (-1.79)	-0.09** (-2.44)	0.033 (0.82)
<i>Female</i>	-0.15*** (-3.81)	-0.17*** (-5.28)	-0.33*** (-15.46)
<i>Occupation</i>	0.011 (0.49)	0.02 (1.32)	0.002 (0.28)
<i>Constant</i>	27.44 (1.18)	6.67** (1.52)	7.159853*** (22.51)
<i>r2</i>	0.16	0.08	0.18
<i>N</i>	510	956	3113

(* , ** , *** có ý nghĩa với mức 10%, 5%, 1%. Giá trị trong ngoặc là tỷ số t)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Để định lượng vai trò của phát tín hiệu, nếu cho rằng vai trò phát tín hiệu là bằng 0 sau 3 năm, thì kết quả trong bảng 4.13 cho thấy vai trò phát tín hiệu của giáo dục Việt Nam: kết quả so sánh nhóm 1 với nhóm 2 cho thấy vai trò phát tín hiệu của giáo dục đại học bằng 36.17% trong vai trò của giáo dục (= $(51.5\% - 32.9\%) / 51.5\%$. Và 35.6% (= $(32.9\% - 21.18\%) / 32.9\%$) là giữa nhóm 2 và nhóm 3. Còn với đào tạo cao đẳng, THCN và TC nghề thì vai trò của phát tín hiệu là 27.36% và 15.5% (chênh lệch giữa nhóm 1 và nhóm 2).

Các biến còn lại là các biến kiểm soát, và các hệ số ước lượng đều có ý nghĩa thống kê và có dấu phù hợp với kỳ vọng.

Kết quả ước lượng cho khu vực thành thị và nông thôn

Kết quả ước lượng hàm tiền lương cho khu vực thành thị và nông thôn năm 2014 được trình bày ở bảng 4.14. Khu vực thành thị năm 2014 vai trò phát tín hiệu của trình độ đại học là 8 năm, chênh lệch tiền lương do tín hiệu giữa nhóm 2 và nhóm 3 là 36.7%. Tiền lương do bằng đại học đem lại cao hơn do bằng THPT cho người lao động ở nhóm 1 là 67.88%, ở nhóm 2 là 107.02% và 89.5% ở nhóm 3. Sự khác biệt tiền lương giữa nhóm CD và CD nghề, TC nghề và THCN, sơ cấp nghề so với THPT ở nhóm kinh nghiệm 1 lần lượt là 28.6%, 46.1% và 38.2%. Ở nhóm kinh nghiệm 2 không có sự khác biệt tiền lương ở các nhóm này. Cho thấy thời gian phát tín hiệu của nhóm giáo dục này là dưới 3 năm. Ở nhóm 3 chênh lệch tiền lương giữa các nhóm này lần lượt là 58.53%, 20.08% và 24%.

Ở khu vực nông thôn thời gian để người sử dụng lao động nhận biết được năng lực thực sự của người lao động có bằng cấp đại học là dưới 8 năm kinh nghiệm, hệ số biến educ1 có ý nghĩa thống kê ở cả 3 nhóm kinh nghiệm tạo nên chênh lệch tiền lương của nhóm lao động có bằng đại học so với THPT lần lượt là 65.8%, 88.4% và 83.02%. Lao động có bằng CD và THCN và TC nghề có thời gian phát tín hiệu là dưới 3 năm kinh nghiệm.

Bảng 4.14: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels cho ba nhóm kinh nghiệm khu vực thành thị và nông thôn năm 2014

Variable	Nhóm 1 (dưới 3 năm kinh nghiệm)		Nhóm 2 (3 đến 8 năm kinh nghiệm)		Nhóm 3 (trên 8 năm kinh nghiệm)	
	Thành thị	Nông thôn	Thành thị	Nông thôn	Thành thị	Nông thôn
<i>Educ1</i>	(cơ sở)	(cơ sở)	(cơ sở)	(cơ sở)	(cơ sở)	(cơ sở)
2	0.32** (2.06)	0.32* (1.45)	0.13 (0.94)	0.49*** (3.22)	0.076*** (2.84)	0.27*** (3.48)
3	0.37** (1.96)	0.55** (2.58)	0.19** (2.42)	0.32*** (2.71)	0.084** (2.23)	0.29*** (3.31)
4	0.25** (1.99)	0.33*** (3.29)	-0.08 (-0.5)	0.13* (1.38)	0.46*** (4.59)	0.45*** (3.47)
5	0.51*** (3.42)	0.50*** (4.43)	0.72*** (5.32)	0.63*** (4.68)	0.63*** (10.44)	0.60*** (5.52)
<i>Age</i>	1.49 (0.39)	-1.83 (-0.66)	-0.13 (-0.23)	0.079 (0.2)	0.08*** (3.23)	0.037** (2.26)
<i>Age_square</i>	-0.03 (-0.39)	0.04 (0.67)	0.002 (0.25)	-0.0013 (-0.18)	-0.001*** (-3.67)	-0.0006*** (-3.31)
<i>ethnic</i>	0.137 (0.8)	0.057 (0.55)	0.25 (1.17)	0.32*** (4.62)	0.24*** (2.97)	0.27*** (6.31)
<i>married</i>	-0.019 (-0.25)	-0.12** (-2.02)	-0.12*** (-18.2)	-0.084** (-1.76)	0.07 (1.31)	-0.062 (-1.03)
<i>female</i>	-0.125** (-1.84)	-0.16*** (-2.99)	-0.24*** (-4.55)	-0.17*** (-4.1)	-0.26*** (-7.12)	-0.36*** (-13.29)
<i>Occupation</i>	0.011 (0.92)	-0.01 (-0.26)	0.001 (0.11)	0.016 (0.21)	0.035*** (4.25)	-0.02** (-1.89)
<i>constant</i>	-9.32 (-0.21)	28.56 (0.89)	9.69 (1.26)	6.59 (1.18)	6.37*** (11.26)	7.44*** (20.5)
<i>r2</i>	0.08	0.1	0.14	0.1	0.24	0.2
<i>N</i>	178	335	326	618	1161	1941

(* , ** , *** có ý nghĩa với mức 10%, 5%, 1%. Giá trị trong ngoặc là tỷ số t)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Kết quả ước lượng theo giới tính người lao động

Phân tích cho năm 2014 dựa trên bảng 4.15, khác biệt tiền lương do trình độ đại học đem lại cho người lao động nam so với trình độ THPT ở nhóm 1, nhóm 2 và 3 lần lượt là 66.04%, 58.65% và 30.35%. Điều này cũng có nghĩa là ở năm 2014 người sử dụng lao động cần đến 8 năm để nhận ra năng lực thực sự của lao động nam có bằng đại học, chênh lệch tiền lương do tín hiệu tạo ra là 46.29% ($= (66.04\% - 30.35\%) / 66.04\%$). Đối với lao động nam có trình độ CĐ và CĐ nghề, THCN và TC nghề, chênh lệch tiền lương với lao động trình độ THPT là 33.8% và 17.66% ở nhóm kinh nghiệm 1, không có sự khác biệt ở nhóm kinh nghiệm 2. Ở nhóm 3 hiệu suất sinh lời của giáo dục của các nhóm CĐ và CĐ nghề, THCN và TC nghề, và sơ cấp nghề lần lượt là 29.19%, 20.4% và 27.8%.

Năm 2014 chênh lệch tiền lương cho lao động nữ có bằng đại học so với bằng THPT lần lượt theo nhóm kinh nghiệm 1, 2 và 3 là 73.1%, 64.38% và 57.38%. Kết quả này cũng cho thấy vai trò phát tín hiệu của bằng cấp đại học cho lao động nữ có thời gian là 8 năm, chênh lệch là 21.56% ($((73.1\% - 57.38\%) / 73.1\%)$). Sự khác biệt tiền lương của lao động nữ do trình độ CĐ và CĐ nghề, và sơ cấp nghề đem lại so với trình độ THPT chỉ có ở nhóm 3 là 54.67% và 34.18%.

**Bảng 4.15: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels
ba nhóm kinh nghiệm của lao động nam và nữ năm 2014**

Variable	Nhóm 1 (dưới 3 năm kinh nghiệm)		Nhóm 2 (3 đến 8 năm kinh nghiệm)		Nhóm 3 (trên 8 năm kinh nghiệm)	
	Nam	Nữ	Nam	Nữ	Nam	Nữ
<i>Educl</i>	(cơ sở)	(cơ sở)	(cơ sở)	(cơ sở)	(cơ sở)	(cơ sở)
2	0.16 (0.68)	0.2 (1.11)	0.51*** (4.65)	0.18 (1.29)	0.24*** (4.83)	0.29*** (3.05)
3	0.26** (2.53)	0.25 (0.79)	0.18 (0.8)	-0.16 (-0.63)	0.18*** (3.28)	0.22 (1.31)
4	0.29*** (3.48)	0.2 (1.24)	0.04 (0.4)	0.23 (1.32)	0.20*** (5.43)	0.43*** (3.4)
5	0.50*** (3.77)	0.54*** (3.07)	0.46*** (3.22)	0.49*** (3.2)	0.3*** (5.49)	0.45*** (3.61)
<i>Age</i>	-1.90 (-0.75)	0.020 (0.01)	-0.27 (-0.65)	0.60 (1.12)	0.05** (2.86)	0.034 (1.17)
<i>Age_square</i>	0.042 (0.77)	-0.0004 (-0.01)	0.005 (0.65)	-0.011 (-1.13)	-0.0007*** (-3.58)	-0.0006** (-1.87)
<i>Urban</i>	0.047 (0.77)	0.089 (1.22)	0.13*** (2.84)	0.13** (2.26)	0.20*** (7.91)	0.29*** (6.78)
<i>Ethnic</i>	0.09 (0.91)	0.02 (0.12)	0.29*** (4.09)	0.55*** (3.88)	0.32*** (7.97)	0.27*** (3.65)
<i>Married</i>	-0.08* (-1.38)	-0.084 (-1.21)	-0.10** (-2.32)	-0.026 (-0.33)	-0.10** (-2.07)	0.16*** (2.37)
<i>Occupation</i>	0.005 (0.15)	0.0048 (0.28)	0.016 (0.8)	0.026* (1.45)	0.024** (2.78)	0.03* (1.42)
<i>constant</i>	29.25 (1.00)	7.62 (0.17)	11.53** (1.99)	-0.91 (-0.12)	6.90*** (17.65)	7.06*** (11.1)
<i>r2</i>	0.04	0.02	0.12	0.1	0.18	0.16
<i>N</i>	289	228	588	390	2156	1059

(* , ** , *** có ý nghĩa với mức 10%, 5%, 1%. Giá trị trong ngoặc là tỷ số t)

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

4.4.2. So sánh với năm 2010

Năm 2010 lao động có trình độ đại học trở lên có mức lương cao hơn lao động có trình độ THPT theo nhóm dưới 3 năm kinh nghiệm, 3 đến 8 năm kinh nghiệm và trên 8 năm kinh nghiệm lần lượt là 58.96%, 46.95% và 64.91%. Lao động có trình độ cao đẳng có mức lương cao hơn lao động THPT là 36.1% đối với nhóm dưới 3 năm kinh nghiệm, và 44% ở nhóm 3 đến 8 năm kinh nghiệm, và 54.69% ở nhóm trên 8 năm kinh nghiệm. Vai trò phát tín hiệu của giáo dục bậc thông qua so sánh nhóm 1 và nhóm 2 là 19.9% $(=(58.96\% - 46.95\%)/58.96\%)$ cho lao động có trình độ đại học. Như vậy thời gian phát tín hiệu của giáo dục bậc đại học năm 2010 ngắn hơn năm 2014 là như nhau. Ngoài ra chênh lệch tiền lương trung bình giữa các nhóm THCN, TC nghề và sơ cấp nghề so với bậc THPT lần lượt 28.4% và 30% (Phụ lục11).

Ở khu vực thành thị chỉ có giáo dục đại học đóng vai trò phát tín hiệu, thời gian phát tín hiệu là dưới 3 năm kinh nghiệm. Chênh lệch tiền lương giữa bằng cấp đại học và THPT là 92.46% ở nhóm 1, 46.62% ở nhóm 2 và ở nhóm 3 là 166.64%. Sự khác biệt tiền lương của lao động có bằng cao đẳng và lao động có bằng THPT dưới 8 năm kinh nghiệm là 15.28% và trên 8 năm kinh nghiệm là 75.05%. Như vậy thời gian phát tín hiệu của giáo dục khu vực thành thị năm 2010 ngắn hơn năm 2014 (Phụ lục12).

Ở khu vực nông thôn, lao động có bằng cấp đại học có tiền lương cao hơn THPT lần lượt theo 3 nhóm kinh nghiệm là: 48.75%, 50.04% và 137%, thời gian phát tín hiệu của bằng cấp cao đẳng là dưới 8 năm kinh nghiệm, do hệ số của biến educ2 và educ3 tăng từ nhóm 1 đến nhóm 2, sau đó giảm ở nhóm 3. Theo phương pháp Lewbels thời gian phát tín hiệu của giáo dục khu vực nông thôn năm 2010 dài hơn năm 2014 (Phụ lục 13).

Năm 2010, tiền lương của lao động nam giới có trình độ đại học không có sự khác biệt với tiền lương lao động nam có trình độ THPT cho nhóm 1, với nhóm 2 sự khác biệt này là 63.50% và 139.14% cho nhóm 3. Chênh lệch tiền lương cho lao động nam có trình độ cao đẳng với THPT ở nhóm 1 là 41,13, và nhóm 2 không có sự chênh lệch, và ở nhóm 3 là 88.79%. Ở nhóm 1, lao động

nam có trình độ THCS và TC nghề có tiền lương cao hơn THPT là 49,74% và giảm xuống 18% ở nhóm 2, không có sự khác biệt ở nhóm 3 (Phụ lục 14).

Lao động nữ có bằng đại học có tiền lương cao hơn THPT ở nhóm 1 là 104% và giảm xuống 74.4% ở nhóm 2 và không còn sự khác biệt ở nhóm 3. Đối với lao động nữ có bằng cao đẳng, THPT và TC nghề có chênh lệch tiền lương với lao động nữ có bằng THPT là 54.14% và 26.76% ở nhóm 2. Theo kết quả tính toán cho thấy thời gian phát tín hiệu cho lao động nữ năm 2010 và năm 2014 là như nhau (Phụ lục 15).

4.5. Ước lượng suất sinh lời và vai trò phát tín hiệu của giáo dục bằng phương pháp phi tham số Kernel

Mô hình ước lượng có dạng:

$$\ln income_m = m(age, agesquare, married, female, urban, ethnic, educ)$$

Phương pháp ước lượng cho 3 nhóm lao động: nhóm 1 dưới 3 năm kinh nghiệm, nhóm 2 từ 3 đến 8 năm kinh nghiệm, và nhóm 3 là trên 8 năm kinh nghiệm. Đây là phương pháp chỉ ra sự trung bình tác động của các cấp giáo dục lên tiền lương của người lao động. Sự tác động trung bình của tín hiệu giáo dục sẽ không còn nữa khi số năm kinh nghiệm tăng lên.

4.5.1. Phương pháp Kernel ước lượng hàm tiền lương năm 2014

Bảng 4.16 là kết quả ước lượng phương pháp Kernel cho năm 2014 theo 3 nhóm kinh nghiệm, kết quả ước lượng đối với nhóm 1 (dưới 3 năm kinh nghiệm) cho thấy giá trị trung bình của $\log(\text{income_m})$ là 8.11 và trung bình tác động của bằng cấp giáo dục sơ cấp nghề, trung cấp nghề và trung học chuyên nghiệp, cao đẳng nghề và cao đẳng, đại học trở lên tiền lương cao hơn trung bình tác động của bằng cấp THPT lần lượt là 0.09, 0.11, 0.19, và 0.20. Ở nhóm 2 thì trung bình tác động của bằng cấp giáo dục đều giảm so với nhóm 1, các trung bình tác động của các cấp độ giáo dục so với của trình độ THPT lên tiền lương lần lượt là 0.46, 0.09, 0.12 và 0.17 (giảm phần trăm tương ứng so với nhóm 1 là 98.77%, 31.72%, 52.88% và 16.16%). Thời gian phát tín hiệu của giáo dục năm 2014 là dưới 3 năm kinh nghiệm.

Bảng 4.16: Phương pháp Kernel ước lượng hàm tiền lương năm 2014

	Nhóm 1 (dưới 3 năm kinh nghiệm)	Nhóm 2 (3 đến 8 năm kinh nghiệm)	Nhóm 3 (trên 8 năm kinh nghiệm)
Mean			
Lnincome_m	8.11	8.24	8.22
Age	0.14	0.33	0.046
Age square	-0.002	-0.0061	-0.0007
Married	-0.008	-0.095	0
Female	-0.19	-0.18	-0.31
Urban	0.14	0.18	0.35
Ethnic	0.22	0.34	0
Educ			
(2 vs 1)	0.09	0.046	0.053
(3 vs 1)	0.11	0.09	0.097
(4 vs 1)	0.19	0.12	0.22
(5 vs 1)	0.20	0.17	0.24
R_square	0.34	0.37	0.4
N	367	622	1143

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Năm 2014, tác động trung bình của nhóm sơ cấp nghề, trung cấp nghề và THPT, cao đẳng và cao đẳng nghề, đại học lớn hơn tác động trung bình tác động của THPT lần lượt là 0.11, 0.15, 0.17, 0.20 cho nhóm 1, và 0.11, 0.19, 0.28, 0.26 cho nhóm 2. Trung bình log(income_m) cho nhóm 1 là 7.12 và nhóm 2 là 7.25.

Kết quả ước lượng theo khu vực thành thị và nông thôn

Bảng 4.17 trình bày kết quả ước lượng cho khu vực thành và nông thôn năm 2014. Khu vực thành thị năm 2014 có thời gian phát tín hiệu cho nhóm sơ cấp nghề, trung cấp nghề và trung học chuyên nghiệp và cao đẳng là dưới 8 năm kinh nghiệm, riêng nhóm đại học là trên 8 năm. Trung bình tác động lên tiền lương của cấp độ giáo dục trên cao hơn so với trung bình tác động của THPT ở nhóm 1 là 0.11, 0.15, 0.22, 0.25. Ở nhóm 2 là 0.04, 0.09, 0.15 và 0.26 (giảm tương ứng 176.68% của nhóm sơ cấp nghề, 18.45% của nhóm trung cấp nghề và trung học chuyên nghiệp, 32.62% của cao đẳng và cao đẳng nghề, và tăng 4,4% của đại học trở lên. Chênh lệch trung bình tác động lên tiền lương

của các cấp độ giáo dục so với nhóm THPT ở nhóm 3 là 0.07, 0.18, 0.32, 0.43, tăng tương ứng nhóm 3 hơn nhóm 2 là 45.59%, 49.9%, 53% và 38.73%. Trung bình $\log(\text{income_m})$ của 3 nhóm lần lượt là 8.22, 8.38, 8.39.

Bảng 4.17: Phương pháp Kernel ước lượng hàm tiền lương khu vực thành thị và nông thôn năm 2014

	Nhóm 1 (dưới 3 năm kinh nghiệm)		Nhóm 2 (3 đến 8 năm kinh nghiệm)		Nhóm 3 (trên 8 năm kinh nghiệm)	
	Thành thị	Nông thôn	Thành thị	Nông thôn	Thành thị	Nông thôn
Mean						
L _n income_m	8.22	8.05	8.38	8.161	8.39	7.98
Age	1.007	-1.05	0.14	0.37	0.12	-0.03
Age square	-0.02	0.02	-0.002	-0.006	-0.001	0.00009
Married	-0.013	-0.01	-0.04	-0.14	-0.058	0
Female	-0.17	-0.21	-0.22	-0.13	-0.28	-0.38
Ethnic	0	0.24	0.48	0.32	0.38	0.36
Educ						
(2 vs 1)	0.11	0.07	0.04	0.04	0.076	0.02
(3 vs 1)	0.15	0.09	0.09	0.07	0.18	-0.01
(4 vs 1)	0.22	0.17	0.15	0.10	0.32	0.11
(5 vs 1)	0.25	0.18	0.26	0.09	0.43	0.02
R_square	0.23	0.36	0.37	0.32	0.35	0.39
N	152	224	265	370	647	519

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Bảng 4.17 cũng cho thấy vai trò phát tín hiệu của giáo dục ở khu vực nông thôn năm 2014 là dưới 3 năm kinh nghiệm làm việc của người lao động. Điều này được minh chứng bằng trung bình tác động của các cấp giáo dục so với cấp THPT ở nhóm 1 giảm đi ở nhóm 2. Cụ thể ở nhóm 1 là: 0.07 của nhóm sơ cấp nghề, 0.09 của nhóm trung cấp nghề và THPT, 0.17 của nhóm cao đẳng và cao đẳng nghề, và 0.18 và của nhóm đại học trở lên. Tương ứng với nhóm 2 là 0.04, 0.07, 0.10, và 0.09. Phần trăm giảm tương ứng của nhóm 2 so với nhóm 1 là: 75.07, 32.31%, 69.03% và 89.9%, đây cũng là chênh lệch trung bình tác động lên tiền lương theo cấp độ giáo dục do tín hiệu tạo ra. Trung bình tác động lên tiền lương của các cấp giáo dục so với THPT ở nhóm 3 là 0.02, -0.01, 0.11, 0.02. Trung bình $\log(\text{income_m})$ của 3 nhóm là 8.20, 8.30, 8.30.

Như vậy các kết quả ước lượng của phương pháp Kernel cũng cho thấy vai trò phát tín hiệu của giáo dục khu vực thành thị 2014 lâu hơn khu vực nông thôn 2014. Thời gian phát tín hiệu của giáo dục khu vực thành thị năm 2014 là trên 8 năm cho bậc đại học, dưới 8 năm cho bậc cao đẳng trở xuống. Và cho khu vực nông thôn là dưới 3 năm cho tất cả các bậc giáo dục.

Kết quả ước lượng theo giới tính người lao động

Vai trò phát tín hiệu của giáo dục nam năm 2014 là dưới 3 năm kinh nghiệm. Điều này được chứng tỏ khi phân tích tác động trung bình của giáo dục lên tiền lương của các nhóm kinh nghiệm. Sự tác động trung bình này giảm ở nhóm 2 ở các cấp giáo dục với phần trăm tương ứng giảm là: 93.21% của sơ cấp nghề, 64.38% của trung cấp nghề và trung học chuyên nghiệp, 40% của cao đẳng và cao đẳng nghề và 69.5% của đại học.

Bảng 4.18: Phương pháp Kernel ước lượng hàm tiền lương cho lao động nam và nữ năm 2014

	Nhóm 1 (dưới 3 năm kinh nghiệm)		Nhóm 2 (3 đến 8 năm kinh nghiệm)		Nhóm 3 (trên 8 năm kinh nghiệm)	
	Nam	Nữ	Nam	Nữ	Nam	Nữ
Mean						
Lnnincome_m	8.20	8.03	8.30	8.19	8.30	8.04
Age	0.19	0.08	0.31	0.74	0.06	0.00002
Age square	-0.003	-0.0007	-0.005	-0.01	-0.0009	-0.00045
Married	-0.03	0.043	-0.16	0.006	-0.20	0
Urban	0.12	0.17	0.21	0.11	0.31	0
Ethnic	-0.0094	0	0.34	0.41	0.54	0
Educ						
(2 vs 1)	0.09	0.09	0.049	0.04	0.06	0.049
(3 vs 1)	0.13	0.10	0.08	0.10	0.11	0.089
(4 vs 1)	0.21	0.18	0.109	0.14	0.18	0.29
(5 vs 1)	0.27	0.15	0.16	0.22	0.17	0.36
R_square	0.31	0.31	0.31	0.44	0.36	0.45
N	184	191	367	270	763	405

Nguồn: Tính toán của tác giả từ số liệu Điều tra mức sống dân cư năm 2014

Năm 2014 vai trò tín hiệu của giáo dục cho lao động nữ là nhóm sơ cấp nghề, trung cấp nghề và THPT, cao đẳng cao nghề là dưới 3 năm, chênh lệch phần trăm trung bình tác động của giáo dục của nhóm 1 cao hơn nhóm 2 là: 92.39%, 3.8% và 25.27%. Vai trò phát tín hiệu của nhóm đại học là dưới 8 năm, chênh lệch phần trăm trung bình tác động của giáo dục của nhóm 3 thấp hơn nhóm 2 là: 31.9%. Trung bình $\log(\text{income}_i)$ cho nhóm 1, 2 và 3 lần lượt là: 8.03, 8.19, 8.04.

Có thể thấy thời gian phát tín hiệu của giáo dục cho lao động nữ năm 2014 và nam năm 2014 là như nhau (dưới 3 năm kinh nghiệm).

4.5.2. So sánh với năm 2010

Năm 2010, đối với nhóm 1 (dưới 3 năm kinh nghiệm) có giá trị trung bình của $\log(\text{income}_m)$ là 7.23 và trung bình tác động của bằng cấp giáo dục sơ cấp nghề, trung cấp nghề và trung học chuyên nghiệp, cao đẳng nghề và cao đẳng, đại học trở lên tiền lương cao hơn trung bình tác động của bằng cấp THPT lần lượt là 0.10, 0.09, 0.14, và 0.19, đến nhóm 2 (từ 3 đến 8 năm kinh nghiệm) các trung bình tác động theo bằng cấp giáo dục trên tăng theo năm kinh nghiệm lần lượt 0.11, 0.19, 0.29, 0.36 (phần trăm tăng tương ứng là 12.4%, 52.14%, 52.44% và 45.68%), tuy nhiên khi số năm kinh nghiệm tăng lên trên 8 năm kinh nghiệm trung bình tác động này đã giảm xuống là 0.10, 0.15, 0.23, 0.26, phần trăm giảm tương ứng 12.06%, 23.15%, 26.22%, và 37% đây cũng chính là chênh lệch trung bình tác động của giáo dục lên tiền lương do tín hiệu tạo ra. Trung bình $\log(\text{income}_m)$ ở nhóm 2 và nhóm 3 năm 2010 là 7.43 và 7.31. Như vậy theo phương pháp ước lượng Kernel năm 2010 thời gian phát tín hiệu của giáo dục là dưới 8 năm kinh nghiệm (Phụ lục 16).

Khu vực thành thị năm 2010 có trung bình $\log(\text{income}_m)$ cho nhóm 1 và nhóm 2 lần lượt là 7.49 và 7.71. Chênh lệch trung bình tác động của nhóm sơ cấp nghề, trung học CN và trung cấp nghề, cao đẳng và cao đẳng nghề, và đại học trở lên so với trung bình tác động lên tiền lương của nhóm tốt nghiệp THPT cho tiền lương là 0.11, 0.06, 0.12, và 0.22 cho nhóm kinh nghiệm 1, và tác động này tiếp tục tăng ở nhóm 2 tương ứng là 0.18, 0.34, 0.44, và 0.65. Kết quả thực hiện cho nhóm 3 không ước lượng được do hạn chế về số liệu, chưa khẳng định được thời gian phát tín hiệu của giáo dục của khu vực thành thị năm 2010 (Phụ lục 17).

Kết quả cho thấy khu vực nông thôn năm 2010 có thời gian phát tín hiệu của các bậc giáo dục là dưới 8 năm. Trung bình tác động của các cấp giáo dục lên tiền lương so với THPT tăng từ nhóm 1 lên nhóm 2 và giảm đi ở nhóm 3. Cụ thể là chênh lệch trung bình tác động ở nhóm 1 là 0.10 của nhóm sơ cấp nghề, 0.13 của nhóm trung cấp nghề và THPT, 0.18 của nhóm cao đẳng, cao đẳng nghề, và 0.26 của nhóm đại học. Tương ứng nhóm 2 là 0.13, 0.23, 0.37, và 0.41 (phần trăm tăng từ nhóm 1 đến nhóm 2 là 20.26%, 40.98%, 49,84% và 37.12%). Tương ứng nhóm 3 là 0.08, 0.11, 0.15, 0.14 (phần trăm giảm so với nhóm 2 là 62.91%, 100.5%, 132.5% và 193.92%). Trung bình $\log(\text{income}_m)$ của 3 nhóm kinh nghiệm năm 2010 lần lượt là 7.15, 7.31, 7.15 (Phụ lục 18).

Năm 2010, tiền lương lao động nam được các cấp giáo dục tác động trung bình cao hơn bậc THPT theo nhóm 3 nhóm kinh nghiệm lần lượt gồm: 0.09, 0.07, 0.12, 0.21 cho nhóm 1, 0.13, 0.23, 0.33 và 0.46 cho nhóm 2, và giảm ở nhóm 3: 0.13, 0.21, 0.32, 0.38. Như vậy vai trò phát tín hiệu của giáo dục cho lao động nam năm 2010 là dưới 8 năm kinh nghiệm, và chênh lệch tác động trung bình của giáo dục cho tiền lương ở nhóm 2 và nhóm 3 giảm theo các cấp giáo dục trung cấp nghề và THPT, cao đẳng và cao đẳng nghề, và đại học là: 93.38%, 39% và 20.15% (Phụ lục 19).

Đối với lao động nữ, tác động trung bình của giáo dục lên tiền lương theo cấp độ giáo dục ở nhóm 1 lần lượt là 0.11, 0.15, 0.17 và 0.20. Cho nhóm 2 là: 0.11, 0.19, 0.28, và 0.26. Chưa thể ước lượng được tác động trung bình của giáo dục lên tiền lương của người lao động ở nhóm 3 (Phụ lục 20).

Kết luận:

Các kết quả của các phương pháp ước lượng cho thấy thời gian phát tín hiệu của giáo dục năm 2014 là dài hơn năm 2010. Năm 2010 là giai đoạn khủng hoảng kinh tế, nền kinh tế gặp nhiều khó khăn, việc tuyển dụng trở nên khắt khe hơn đặc biệt là khu vực thành thị nơi tập trung nhiều lao động tay nghề cao, công việc đòi hỏi độ phức tạp hơn nông thôn, người sử dụng lao động càng có những biện pháp thắt chặt tiền lương cũng những làm rõ năng lực thực sự của người lao động đặc biệt là khu vực thành thị, dẫn đến thời gian nhận ra năng lực thực sự của người lao động thành thị diễn ra nhanh hơn nông thôn.

Nghiên cứu cũng cho thấy sự bất bình đẳng trong tiền lương ở giới tính cũng được thể hiện qua vai trò phát tín hiệu của giáo dục. Cả năm 2010 và 2014 vai trò phát tín hiệu của giáo dục cho lao động nam kéo dài hơn so với lao động nữ. Đặc biệt là với bằng cấp giáo dục đại học, thời gian phát tín hiệu của trình độ đại học thường kéo dài và cao hơn so với các trình độ giáo dục còn lại.

Sau khi kiểm soát yếu tố phát tín hiệu của giáo dục, vai trò cung cấp vốn nhân lực của giáo dục gia tăng theo bằng cấp. Trong đó vai trò nhân lực của giáo dục ở thành thị lớn hơn so với khu vực nông thôn. Do khu vực thành thị nơi tập trung các ngành nghề cần nhiều kỹ năng hơn nông thôn, bên cạnh đó mức sống của thành thị cũng cao hơn nông thôn. Kết quả tính toán cũng cho thấy vai trò cung cấp vốn nhân lực của giáo dục cho lao động nam cũng cao hơn lao động nữ.

CHƯƠNG 5

KẾT LUẬN VÀ KIẾN NGHỊ

5.1. Kết luận

Vấn đề ước lượng hiệu suất sinh lời của giáo dục sau phổ thông luôn là mối quan tâm của các nhà hoạch định chính sách giáo dục. Giáo dục ngoài việc cung cấp kỹ năng làm việc cho người lao động còn có vai trò phát tín hiệu đến người sử dụng lao động. Tuy nhiên các nghiên cứu ở Việt Nam còn bỏ qua vai trò này của giáo dục. Luận án đã xây dựng các mô hình để nghiên cứu đồng thời 2 vai trò này của giáo dục: lợi suất của giáo dục và phát tín hiệu của giáo dục. Điều này không chỉ giúp nghiên cứu được vai trò phát tín hiệu của giáo dục sau phổ thông ở Việt Nam mà còn giúp cho việc ước lượng hiệu suất sinh lời của giáo dục được chính xác hơn.

Để thực hiện được mục đích nghiên cứu, luận án đã sử dụng một số phương pháp khác nhau. Trước hết luận án sử dụng phương pháp PSM để tìm hiểu vấn đề thông tin bất đối xứng trên thị trường lao động và vai trò phát tín hiệu của giáo dục. Tiếp theo luận án sử dụng hai phương pháp ước lượng tham số là phương pháp Heckman và phương pháp Lewbels để ước lượng hiệu suất sinh lời của giáo dục có tính đến vai trò phát tín hiệu. Đây là các phương pháp thường được sử dụng để giải quyết vấn đề biến nội sinh là vấn đề thường gặp khi ước lượng hiệu suất sinh lời của giáo dục. Ngoài ra để đảm bảo độ tin cậy của kết quả ước lượng, luận án đã sử dụng phương pháp phi tham số Kernel. Trong đó phương pháp phi tham số có ưu điểm là không phụ thuộc vào dạng hàm về mối quan hệ giữa các biến số. Kết quả cho thấy một sự thống nhất giữa các kết quả ước lượng bằng các phương pháp khác nhau.

Các kết quả phân tích cho thấy giáo dục sau phổ thông có vai trò phát tín hiệu. Điều này được thể hiện ở sự khác biệt về thu nhập giữa những người tự làm và làm công ở các mức kinh nghiệm khác nhau. Người lao động làm thuê sử dụng bằng cấp giáo dục với hai vai trò cung cấp vốn nhân lực và phát tín hiệu, đối với người lao động tự làm giáo dục chỉ có vai trò cung cấp vốn nhân lực. Khi người lao động bắt đầu tham gia thị trường lao động, vai trò phát tín hiệu của giáo dục giúp lao động làm thuê có tiền lương cao hơn người tự làm,

nhưng khi lao động làm việc nhiều năm, vai trò tín hiệu của giáo dục dần mất đi, mức chênh lệch giữa tiền lương người lao động và tự làm sẽ giảm dần.

Vai trò phát tín hiệu của giáo dục cũng khác nhau giữa khu vực nông thôn và khu vực thành thị, vai trò phát tín hiệu của giáo dục ở khu vực thành thị cao hơn nông thôn, bên cạnh đó thời gian phát tín hiệu của giáo dục ở khu vực thành thị cũng dài hơn nông thôn, khu vực thành thị là nơi tập trung nhiều lao động tay nghề cao, người sử dụng cần nhiều thời gian để nhận ra năng lực thực sự của họ. Các kết quả ước lượng cũng cho thấy vai trò phát tín hiệu của giáo dục cho lao động nam lớn và lâu hơn đối với lao động nữ.

Sau khi kiểm soát yếu tố phát tín hiệu của giáo dục, vai trò cung cấp vốn nhân lực của giáo dục gia tăng theo bằng cấp. Trong đó vai trò nhân lực của giáo dục ở thành thị lớn hơn so với khu vực nông thôn. Do khu vực thành thị nơi tập trung các ngành nghề cần nhiều kỹ năng hơn nông thôn, bên cạnh đó mức sống của thành thị cũng cao hơn nông thôn. Kết quả tính toán cũng cho thấy vai trò cung cấp vốn nhân lực cho lao động nam cũng lớn hơn lao động nữ.

Các phân tích trên cho thấy giáo dục ở Việt Nam không chỉ có vai trò cung cấp vốn nhân lực mà còn có vai trò phát tín hiệu. Kết quả này phù hợp với lý thuyết phát tín hiệu cũng như với các nghiên cứu thực nghiệm trên thế giới. Thêm vào đó, vai trò phát tín hiệu là rất lớn so với vai trò tổng thể của giáo dục. Điều này có thể ngụ ý rằng vai trò của giáo dục Việt Nam như nguồn cung cấp vốn nhân lực giúp gia tăng năng suất lao động là chưa lớn. Điều này cũng phù hợp với ý kiến trên nhiều phương tiện thông tin đại chúng và nhận xét của nhiều nhà tuyển dụng về chất lượng lao động Việt Nam.

5.2. Đề xuất một số kiến nghị chính sách

1. Với các cơ quan quản lý nhà nước về giáo dục: Phân tích của luận án cho thấy vai trò phát tín hiệu của giáo dục sau phổ thông ở Việt Nam là đáng kể và khá phổ biến. Do đó quản lý nhà nước cần nghiên cứu các biện pháp để giảm thiểu vai trò này và nâng cao vai trò cung cấp vốn nhân lực cho người lao động. Luận án đề xuất:

- Đẩy nhanh tiến độ thực hiện kiểm định độc lập: có như vậy, các cơ sở giáo dục mới phát triển tuy chất lượng tốt cũng có thể bắt lợi so với các cơ sở giáo dục đã có thương hiệu trong việc thu hút đầu vào có chất

lượng. Do đó quản lý nhà nước về giáo dục có thể thực hiện các cơ chế để có được sự cạnh tranh bình đẳng hơn giữa các cơ sở giáo dục đại học, nhằm giúp giảm bớt vai trò phát tín hiệu của giáo dục. Một biện pháp có thể giúp thực hiện điều này đó là đẩy nhanh tiến độ kiểm định độc lập, do kết quả kiểm định độc lập chính là một nguồn tín hiệu đáng tin cậy về chất lượng đào tạo của cơ sở đại học.

- Song song với việc kiểm định độc lập, cần yêu cầu – khuyến khích và kiểm tra việc công khai kết quả khảo sát hàng năm về tỷ lệ có việc làm của các cơ sở giáo dục đại học. Đây cũng là một kênh thông tin hữu ích cho người sử dụng dịch vụ giáo dục cũng như chủ lao động về chất lượng lao động

2. Với các cơ sở giáo dục:

- Với các cơ sở giáo dục, đặc biệt là cơ sở giáo dục có thương hiệu, cần phổ biến rộng rãi các thông tin phản ánh năng lực và kết quả đào tạo của mình. Tuy nhiên, về căn bản, việc tiếp tục duy trì và nâng cao chất lượng đào tạo vẫn là điều kiện then chốt cho sự phát triển dài hạn của cơ sở giáo dục. Vì sự bất đối xứng thông tin với mỗi cá thể sẽ giảm dần theo thời gian, và cuối cùng thì năng lực thực sự sẽ được nhận diện
- Với các cơ sở giáo dục mới phát triển: song song với việc nâng cao năng lực đào tạo, cần có những biện pháp hiệu quả để cung cấp thông tin cho người học, chẳng hạn quan tâm đầy đủ hơn đến việc được công nhận và kiểm định bởi các cơ sở kiểm định có uy tín trong nước và quốc tế

3. Với người sử dụng dịch vụ giáo dục và doanh nghiệp

- Đối với doanh nghiệp, để hạn chế vấn đề thông tin bất đối xứng và đánh giá tốt hơn năng lực của người lao động, cần có những phản hồi đến các cơ sở đào tạo về yêu cầu của thị trường cũng như về chất lượng nguồn lao động.
- Các kết quả phân tích cho thấy việc đóng góp cho vốn nhân lực của giáo dục giai đoạn qua của Việt Nam là chưa cao, thể hiện qua việc vai trò phát tín hiệu của giáo dục là khá lớn so với vai trò cung cấp vốn nhân lực. Kết quả này cũng phù hợp với đánh giá của các nhà tuyển dụng đối

với sinh viên. Điều đó cho thấy rằng nội dung chương trình đào tạo ở các trường Đại học, Cao đẳng cũng như các đơn vị đào tạo khác cần phải cải thiện sao cho kiến thức học được góp phần quan trọng hơn nữa trong việc nâng cao năng suất lao động. Nhất là trong bối cảnh của cách mạng công nghiệp 4.0, khi mà sự vận động của khoa học kỹ thuật, của công nghệ, của nền kinh tế diễn ra nhất nhanh, thì việc nội dung – chương trình học tại các cơ sở giáo dục lại càng cần được cải thiện hơn nữa.

- Về người sử dụng dịch vụ giáo dục: về lâu dài, tín hiệu sẽ không có tác dụng. Do đó việc học ở trường có uy tín cao hay có uy tín vừa phải thì vấn đề quan trọng vẫn là nâng cao năng lực thực sự của bản thân. Nhất là trong bối cảnh hiện nay, khi mà mức độ cạnh tranh trên thị trường giáo dục rất cao, thì chất lượng đào tạo của các cơ sở giáo dục chưa chắc đã được phản ánh bởi thương hiệu.

Cuối cùng là, với việc minh chứng về vai trò phát tín hiệu của giáo dục, các nghiên cứu thực nghiệm về lợi suất cho giáo dục tại Việt Nam cần tính đến vai trò này để có thể đưa ra các ước lượng chính xác hơn.

5.3. Các kết quả chính của luận án

5.3.1. Về mặt lý thuyết

Luận án giới thiệu cơ sở và lý thuyết phát tín hiệu của giáo dục sau phổ thông, đây là vấn đề còn mới mẻ trong các nghiên cứu về hiệu suất sinh lời của giáo dục tại Việt Nam. Việc tính đến lý thuyết phát tín hiệu trong mô hình hóa không chỉ giúp đánh giá vai trò phát tín hiệu của giáo dục và còn giúp cho việc ước lượng hiệu suất sinh lời của giáo dục một cách chính xác hơn so với các nghiên cứu hiện thời tại Việt Nam.

Bên cạnh đó, Luận án cũng sử dụng đồng thời các phương pháp tham số và phi tham số trong nghiên cứu. Điều này giúp cho việc đảm bảo độ tin cậy của kết quả nghiên cứu: trong khi các phương pháp tham số cho phép thực hiện các kiểm định thống kê thì các phương pháp phi tham số lại có ưu điểm là không phụ thuộc vào dạng hàm được định sẵn về mối quan hệ của các biến số. Ngoài ra, Luận án cũng đã sử dụng các phương pháp khác nhau để xử lý vấn đề biến nội sinh một cách thấu đáo.

5.3.2. Về mặt thực tiễn

Ngoài những đóng góp về mặt lý thuyết, luận án còn đưa ra các kết quả thực nghiệm:

Một là cung cấp bằng chứng thực nghiệm về sự hiện diện của sự bất đối xứng thông tin trên thị trường lao động Việt Nam giai đoạn 2010, 2014 qua đó cho thấy dấu hiệu của vai trò phát tín hiệu của giáo dục lên tiền lương của người lao động, phương pháp PSM (propensity score method) so sánh tiền lương của hai nhóm lao động làm thuê và tự làm cho các năm kinh nghiệm. Có sự thông tin bất đối xứng ở khu vực thành thị và nông thôn, đối với khu vực thành thị sự hiện diện của thông tin bất đối xứng lâu hơn khu vực nông thôn. Thông tin bất đối xứng cũng xảy ra khi chia lao động theo giới tính của người lao động. Chênh lệch tiền lương của nhóm lao động làm thuê cao hơn nhóm tự làm tăng theo cấp độ giáo dục, cho thấy vai trò phát tín hiệu của giáo dục tăng theo cấp độ. Phương pháp hàm kiểm soát (control function) do Heckman đề xuất cũng được sử dụng để kiểm soát vấn đề ước lượng chệch do lựa chọn việc làm thuê hay tự làm, kết quả cho thấy có sự chênh lệch tiền lương giữa lao động làm thuê và lao động tự làm, điều này cho thấy vai trò phát tín hiệu trên thị trường lao động Việt Nam, lao động làm thuê sử dụng bằng cấp giáo dục như là tín hiệu đến người sử dụng lao động, tín hiệu này sẽ dần mất đi khi số năm kinh nghiệm tăng lên, dẫn đến chênh lệch tiền lương của người làm thuê và tự làm không còn nữa.

Hai là, để giải quyết vấn đề biến nội sinh trong việc ước lượng suất sinh lời của giáo dục, một phương pháp mới được sử dụng là phương pháp Lewbels, là phương pháp mới có thể khắc phục nhược điểm này bằng cách tự tạo ra biến công cụ sử dụng bản thân các biến ngoại sinh trong mô hình, đây cũng chính là ưu điểm vượt trội của phương pháp này, đặc biệt là trong các trường hợp mà việc tìm biến công cụ bên ngoài là khó khăn. Các kết quả ước lượng của phương pháp Lewbels cho thấy đóng góp biên của giáo dục tăng theo cấp độ giáo dục, và giảm dần khi năm kinh nghiệm của người lao động tăng lên, đây là dấu hiệu cho thấy giáo dục đóng vai trò phát tín hiệu cho người lao động, vai trò phát tín hiệu sẽ mất đi khi số năm kinh nghiệm tăng.

Ba là, luận án sử dụng phương pháp phi tham số Kernel ước lượng trung bình tác động của giáo dục lên tiền lương của người lao động, ưu điểm của

phương pháp Kernel so với các phương pháp tham số khác khi ước lượng hàm tiền lương là nếu dạng hàm không cụ thể thì các ước lượng tham số sẽ dễ bị sai lệch.

Bốn là, luận án đề xuất một giải pháp về giáo dục và lao động ở Việt Nam từ các kết quả phân tích thực nghiệm.

5.4. Những hạn chế của luận án

Bên cạnh những đóng góp của luận án về những phân tích thực nghiệm về hiệu suất sinh lời và vai trò phát tín hiệu của giáo dục sau phổ thông ở Việt Nam giai đoạn 2010-2014 bằng các phương pháp tham số và phi tham số, luận án còn một số các hạn chế sau: Lao động tự làm và lao động làm thuê có thể có những đặc điểm khác biệt không quan sát được. Luận án đã sử dụng một số phương pháp khác nhau, chẳng hạn như sử dụng biến trình độ học vấn của cha mẹ làm biến công cụ để kiểm soát các khác biệt này. Tuy nhiên các phương pháp này chỉ có thể kiểm soát tốt về năng lực của người lao động chứ chưa tính đến được sự khác nhau về tính cách, chẳng hạn người lao động thích tự do, sẵn lòng chấp nhận rủi ro có thể thích tự làm hơn là đi làm thuê. Đây là hạn chế về mặt sẵn có của số liệu mà tác giả chưa khắc phục được.

DANH MỤC CÔNG TRÌNH ĐÃ CÔNG BỐ CỦA TÁC GIẢ

1. Lê Thái Sơn, Trần Bá Phi (2017), “Ước lượng hiệu suất sinh lời của giáo dục trong điều kiện thông tin không đối xứng và hàm ý cho chính sách”, *Tạp chí Quản lý Kinh tế*, Số 80, tháng 1+2/2017, tr.28-35.
2. Lê Thái Sơn (2017), “Sử dụng phương pháp PSM nghiên cứu sự hiện diện của vấn đề thông tin bất đối xứng trên thị trường lao động Việt Nam”, *Kỷ yếu hội thảo quốc gia: Thực trạng và giải pháp thúc đẩy tăng trưởng hài hòa trong các khu vực doanh nghiệp ở Việt Nam*, NXB Lao động, tr.37-45.
3. Lê Thái Sơn (2017), The two roles of higher education in Vietnamese labor market-An empirical analysis, *The second Vietnam International Applies Mathematics Conference*, Hội Toán học & Ứng dụng, tr.177-186.
4. Lê Thái Sơn, Nguyễn Thị Minh (2018), “Hiệu suất sinh lời của giáo dục sau phổ thông và vai trò phát tín hiệu của giáo dục sử dụng phương pháp Lewbels – nghiên cứu thực nghiệm tại Việt Nam”, *Tạp chí Kinh tế và Phát triển*, Số 253, tháng 7 năm 2018, tr.10-19.

TÀI LIỆU THAM KHẢO

1. Akerlof, George (1970), 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *The Quarterly Journal of Economics*, No 84, Vol 3, pp.488-500.
2. Baris Kaymar (2012), 'Quantifying the Signaling Role of Education', *Preliminary Draft prepared for 2012 SOLE Meetings*.
3. Bauer, T. K., & John P. H.D (2001), 'Employer learning and the returns to schooling', *Labour Economics*, No 2, Vol 8, pp. 1691-180.
4. Baum, C.F., Lewbels, A., Schaffer, M.E., Talavera, O. (2013), 'Instrumental variables estimation using heteroskedasticity – based instruments', *German Stata Users Group*, Potsdam.
5. Becker, G. S (1962), 'Investment in Human Capital: A Theoretical Analysis', *Journal of Political Economy*, No 70, pp.9-49.
6. Becker, G. S (1964), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* Chicago, University of Chicago Press, ISBN 978-0-226-04120-9
7. Becker, S. O., Ichino, A (2002), 'Estimation of average treatment effects based on propensity scores', *The Stata Journal*, No 2, Vol 4, pp.358-377.
8. Bedard, K (2001), 'Human Capital versus Signaling Models: University Access and High School', *The Journal of Political Economy*, No 4, pp.749-775
9. Brown, S. & Sessions, J (1999), 'Education and employment status: a test of the strong screening hypothesis in Italy', *Economics of Education Review*, No 18, pp. 397-404.
10. Bryson, A (2002), 'The union membership wage premium: an analysis using propensity score matching', *Discussion Paper*, Centre for Economic Performance, London, No 530.
11. Bùi thế Huy (2013), *Phân tích suất sinh lợi của giáo dục Việt nam: Tiếp cận theo phương pháp Clustered Data*, Luận văn thạc sỹ, ĐH Kinh tế thành phố Hồ Chí Minh.

12. Card, D (1995), *Using Geographic Variation in College Proximity to Estimate the Return to Schooling,* in *Aspects of Labour Market Behavior*, Essays in Honour of John Vanderkamp, ed, by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky, Toronto: University of Toronto Press.
13. Chen, M.F (2007), 'Consumer attitudes and purchase intentions in relation to organic foods in Taiwan: Moderating effects of food-related personality traits', *Food Quality and Preference*, No 1, Vol 7, pp.1008-1021.
14. Chih-Ching Teng and Yu-Mei Wang (2015), 'Decisional factors driving organic food consumption: Generation of consumer purchase intentions', *British Food Journal*, No 117, Vol 3, pp.1066-1081.
15. Christofides, E., Grant, K., & Robert, S (1995), *Using Geographic Variation in College Proximity to Estimate the Return to Schooling, Aspects of Labor Market Behaviour*, Essays in Honour of John Vanderkamp, Toronto Press, pp. 201-222.
16. Chung, Y.P (1990), 'Educated mis-employment in Hong Kong: earnings effects of employment in unmatched fields of work', *Economics of Education Review*, No 9, Vol 4, pp.343-350.
17. Doan, T., Tran. Q. T & Le Q (2016), 'Lost in Transition? Declining Returns to Education in Vietnam', *Department of Economics Working Paper in Economics*, University of Waikato.
18. Farber, H. S., & Gibbons, R (1996), 'Learning and wage dynamics', *Quarterly Journal of Economics*, No 111, pp.1007-1047.
19. George A. Akerlof (1970), 'The Market for "Lemon": Quality Uncertainty and the Market Mechanism', *The Quarterly Journal of Economics*, No 84, Vol 3, pp.488-500.
20. Gracia, A. and Magistris, T.D (2008), 'The demand for organic foods in the south of Italy: a discrete choice model', *Food Policy*, No 3, Vol 5, pp.386-396.
21. Heckman, J.J (1979), 'Sample selection bias as a specification error', *Econometrica: Journal of the econometric society*, pp.153-161.

22. Heckman, J. J., Lochner, J. & Todd, P.E (2003), 'Fifty Years of Mincer Earnings Functions', *Working Paper 9732*, Cambridge, MA, NBER.
23. Heckman, J., Ichimura, H., Todd, P. E (1997), 'Matching as an econometric evaluation estimator: evidence from evaluating a job training programme', *Review of Economic Studies*, No 64, pp.605-654.
24. Heywood, J. S. & Wei, XiangDong (2004), 'Education and Signaling: Evidence from a Highly competitive Labor Market', *Education Economics*, No 12, pp.1-16.
25. Hsiao E. Y, Conley.A, Howell. D.A, Sullivan.A, Pritchett. C.J, Carlberg. R.G, Nugent. P.E, Phillips. M.M (2007), 'K-Corrections and Spectral Templates of Type Ia Supernovae', *The Astrophysical Journal*, No 663, Vol 2, pp.1187-1200.
26. Janssen, M. and Hamm, U. (2012), 'Product labelling in the market for organic food: consumer preferences and willingness-to-pay for different organic certification logos', *Food Quality and Preference*, No 25, Vol 1, pp.9-22.
27. Jim Kjelland (2008), 'Economics Return to Higher Education: Signaling v. Human Capital Theory :An Analysis of Competing Theories', *Journal of Economics*, No 84, pp.488–500.
28. Kaymak, B. (2008), *Essays on Education, Selection and Wage Dynamics*, PhD thesis, University of Rochester.
29. Lange, F (2007), 'The Speed of Employer Learning', *Journal of Labor Economics*, No 1, Vol 25, pp.1-36.
30. Lê Thái Sơn & Trần Bá Phi (2017), 'Ước lượng hiệu suất sinh lời của giáo dục trong điều kiện thông tin không đối xứng và hàm ý cho chính sách', *Tạp chí Quản lí kinh tế*, Số 80, tr.28-35.
31. Lemieux, T. (2006), 'The 'Mincer equation' Thirty Years after Schooling, Experience, and Earnings in Jacob Mincer: A Pioneer of Modern Labor Economics', *Shoshanna Grossbard, ed., Springer: New York*, pp.127–145.
32. Li, Q. and Racine, J.S (2004), 'Cross- validated local linear nonparametric regression', *Statistica Sinica*, No 14, pp.485-512.

33. LoFan (2006), *Education and Signaling: Evidence from a Highly Competitive Labor Market in 2001*, An Honours Degree Project Submitted to the School of Business in Partial fulfillment Of the Graduation Requirement for the Degree of Bachelor of Business Administration, Hong Kong Baptist University.
34. Lucas, Jr., R. E (1988), 'On the mechanics of economic development', *Journal of Monetary Economics*, No 22, pp.3–42.
35. McCluskey, J.J. (2000), 'A Game Theoretic Approach to Organic Foods: An Analysis of Asymmetric Information and Policy', *Agricultural and Resource Economics Review*, No 29, Vol 1, pp.1-9.
36. Mergenthaler, M., Weinberger, K. and Qaim, M. (2009), 'The role of consumers' perceptions in the valuation of food safety and convenience attributes of vegetables in Vietnam', *presentation at the International Association of Agricultural Economists Conference, IAAE*, August 16th-22th.
37. Michael Spence (1973). 'Job Market Signaling', *The Quarterly Journal of Economics*, No 87, pp.353-374.
38. Mincer, J. A. (1974), *Schooling, Experience, and Earnings*, NBER (distributed by Columbia University Press), New York.
39. Minh, N.T, Anh. L.T, Hue. T.N (2017), 'Vietnamese consumers' behavior and willingness to pay for safe vegetable in the presence of information asymmetry – study for Hanoi', *Vietnam Journal of Science, Technology and Engineering*, No 19, pp.43-48.
40. Nadaraya, E.A. (1964), 'On Estimating Regression', *Theory of Probability and its Applications*, No 9, Vol 1, pp.141-200.
41. Nguyen Thi Minh and Hoang Bich Phuong (2012), 'The impact of asymmetric information in Vietnam's health insurance', *Journal of Economics and Development*, No 14, Vol 3, pp.5-21.
42. Nguyễn Thị Minh, Hoàng Bích Phương (2013), 'Rủi ro đạo đức trong bảo hiểm y tế tự nguyện và nhân khẩu học tại Việt Nam giai đoạn 2008-2010', *Tạp chí Kinh tế và Phát triển*, No 179, pp.58-63

43. Nguyen Thi Minh, Hoang Bich Phuong, Nguyen Thi Thao (2012), ‘The Impact of Asymmetric of Information In VietNam’s Health Insurance: An Empirical Analysis’, *Journal of Economics and Development*, No 3, pp.5-21.
44. Nguyễn Xuân Thành (2006). *Ước lượng suất sinh lợi của việc đi học ở Việt Nam: Phương pháp khác biệt trong khác biệt*, bài giảng Fullbright.
45. Psacharopoulos, George & Patrinos, Harry Anthony (2004), ‘Returns to investment in education: a further update’, *Education Economics, Taylor & Francis Journals*, vol 12(2), pp 111-134
46. Psacharopoulos, George (1994), ‘Returns to investment in education: A global update’, *World Development Elsevier*, No 22, Vol 9, pp.1325-1343.
47. Roosen J., Lusk, J.L. and Fox, J.A (2003), ‘Consumer Demand for and Attitudes Toward Alternative Beef Labeling Strategies in France, Germany, and the UK’, *Agribusiness*, No 19, Vol 1, pp.77–90
48. Spence, A. M (2002), ‘Signaling in Retrospect and the Informational Structure of Markets’, *The American Economic Review*, No 3, Vol 92, pp.434-459.
49. Spence, M. (1973), ‘Job market signaling’, *Quarterly Journal of Economics*, No 87, pp.355-374.
50. Soon, L.-Y. (1987), ‘Self-employment vs. wage employment: estimates of earnings functions in LDCs’, *Economics of Education Review*, No 6, pp.81-89.
51. Stiglitz, Joseph E. (1975), ‘The Theory of “Screening”, Education and Distribution of Incomes’, *The American Economic Review*, No 65, Vol 3, pp.283-300.
52. Tien, N. D (2014), *An analysis of labour market returns to education in Vietnam: Evidence from the National Labour Force Survey 2012*, Turin School of Development Working Paper, No. 3, International Training Centre of the ILO, Turin, Italy.
53. Tran Thi Tuan Anh (2014), ‘Ước lượng hàm hồi qui tiền lương ở Việt Nam giai đoạn 2002-2010 bằng thủ tục Heckman hai bước’, *Tạp chí Phát triển kinh tế*, Số 284, tr.137-150.
54. The Conference Board of Canada (2013), *Return on investment in tertiary*

education, truy cập ngày 20 tháng 3 năm 2018, từ <https://www.conferenceboard.ca/hcp/Details/education/tertiary.aspx?AspxAutoDetectCookieSupport=1>

55. Uta Schoenberg (2007), 'Testing for Asymmetric Employer Learning', *Journal of Labor Economics*, No 4, Vol 25.
56. Varo, J.D., Waldmancho, M. (2012), 'The Signaling Role of Promotions: Further Theory and Empirical Evidence', *Journal of Labor Economics*, No 1, Vol 30, pp. 91-147.
57. Watson, G.S. (1964), 'Smooth regression analysis', *The Indian Journal of Statistics*, No 26, Vol 4, pp.359-372.
58. Xu, L and Wu, L. (2010), 'Food safety and consumer willingness to pay for certified traceable food in China. J', *Sci. Food Agric*, No 90, pp.1368-1373.
59. Zimmerman, S.D. (2014), 'The Returns to College Admission for Academically Marginal Students', *Journal of Labor Economic*, No 4, pp.711-735

PHỤ LỤC

Phụ lục 1: Kết quả ước lượng theo phương pháp PSM cho năm 2010

```
. psmatch2 treat female honnhan dantoc suckhoe khuvuc tuoi i.educ , outcome(income_m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 13 obs not used
```

```
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 6 obs not used
```

```
Probit regression                               Number of obs =      2038
                                                LR chi2(11)      =    487.79
                                                Prob > chi2      =     0.0000
Log likelihood = -939.5216                     Pseudo R2       =     0.2061
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.2112135	.0696746	3.03	0.002	.0746539	.3477731
honnhan	0	(omitted)				
dantoc	.7758141	.0972787	7.98	0.000	.5851513	.9664769
suckhoe	-.0817514	.1000552	-0.82	0.414	-.277856	.1143531
khuvuc	1.012691	.0775809	13.05	0.000	.8606353	1.164747
tuoi	-.0206373	.0030064	-6.86	0.000	-.0265297	-.0147449
educ						
2	.1940569	.1242773	1.56	0.118	-.0495221	.437636
3	.3823469	.1411103	2.71	0.007	.1057759	.6589179
4	.4840182	.1582107	3.06	0.002	.1739311	.7941054
5	.3897955	.5209976	0.75	0.454	-.631341	1.410932
6	.5624919	.2518714	2.23	0.026	.0688331	1.056151
7	.8056954	.1680009	4.80	0.000	.4764196	1.134971
8	0	(empty)				
_cons	.1511545	.1400654	1.08	0.281	-.1233686	.4256777

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2861.02367	1365.88425	1495.13941	81.6231523	18.32
	ATT	2861.02367	1941.29471	919.728953	125.779467	7.31

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor		Total
Untreated	545	545	
Treated	1,493	1,493	
Total	2,038	2,038	

```
. psmatch2 treat female honnhan dantoc khuvuc tuoi i.educ if tuoi>=16 & tuoi<=65 & kn==1 , outcome(income_m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 16 obs not used
```

```
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 1 obs not used
```

```
Probit regression                               Number of obs =       789
                                                LR chi2(10)      =      117.37
                                                Prob > chi2      =       0.0000
Log likelihood = -348.55486                    Pseudo R2       =       0.1441
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.2579546	.1157414	2.23	0.026	.0311056	.4848037
honnhan	0 (omitted)					
dantoc	.613797	.1645302	3.73	0.000	.2913237	.9362703
khuvuc	1.064503	.1620754	6.57	0.000	.7468413	1.382165
tuoi	.0166328	.0488636	0.34	0.734	-.0791381	.1124036
educ						
2	.2645161	.1858145	1.42	0.155	-.0996736	.6287059
3	-.1460624	.2192581	-0.67	0.505	-.5758004	.2836756
4	.8188447	.3026079	2.71	0.007	.2257441	1.411945
5	.5469866	.5953085	0.92	0.358	-.6197965	1.71377
6	.5891393	.4062454	1.45	0.147	-.207087	1.385366
7	1.046875	.4802523	2.18	0.029	.1055974	1.988152
8	0 (empty)					
_cons	-.5386937	1.165176	-0.46	0.644	-2.822397	1.74501

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2136.94654	1305.73154	831.215006	102.715099	8.09
	ATT	2136.94654	1443.95994	692.986609	234.156212	2.96

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	167	167
Treated	622	622
Total	789	789

```
. psmatch2 treat female honnhan dantoc suckhoe khuvuc tuoi i.educ if tuoi>=16 & tuoi<=65 & kn==2 , outcome(inco
> me_m)
```

```
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 1 obs not used
```

```
note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 4 obs not used
```

```
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 2 obs not used
```

```
Probit regression                               Number of obs   =       448
                                                LR chi2(10)      =       122.00
                                                Prob > chi2      =       0.0000
Log likelihood = -163.73236                    Pseudo R2       =       0.2714
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.1875538	.1702521	1.10	0.271	-.1461342	.5212418
honnhan	0	(omitted)				
dantoc	1.180144	.2017578	5.85	0.000	.784706	1.575582
suckhoe	-.1099004	.3364134	-0.33	0.744	-.7692585	.5494578
khuvuc	.9864458	.2051583	4.81	0.000	.5843429	1.388549
tuoi	-.0301614	.0566921	-0.53	0.595	-.1412759	.0809531
educ						
2	.0900933	.3109913	0.29	0.772	-.5194385	.6996251
3	.3185807	.2931445	1.09	0.277	-.255972	.8931333
4	.5178922	.4333888	1.19	0.232	-.3315342	1.367319
5	0	(empty)				
6	-.1128965	.4083043	-0.28	0.782	-.9131582	.6873653
7	1.099303	.4597106	2.39	0.017	.1982866	2.000319
8	0	(empty)				
_cons	.2132439	1.592371	0.13	0.893	-2.907746	3.334234

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	3085.34357	1314.99259	1770.35098	195.124748	9.07
	ATT	3085.34357	1626.07821	1459.26536	276.134329	5.28

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On support	Total
Untreated	90	90
Treated	358	358
Total	448	448

```
. psmatch2 treat female honnhan dantoc khuvuc i.educ if kn==3 , outcome(income_m)
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 3 obs not used
```

```
Probit regression                               Number of obs =      2883
                                                LR chi2(10)    =      593.95
                                                Prob > chi2    =      0.0000
Log likelihood = -1644.8466                    Pseudo R2     =      0.1529
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.1869134	.0546348	3.42	0.001	.0798312	.2939956
honnhan	.2203574	.8869545	0.25	0.804	-1.518042	1.958756
dantoc	.3085012	.102596	3.01	0.003	.1074166	.5095857
khuvuc	1.071651	.0586159	18.28	0.000	.9567664	1.186536
educ						
2	.3750481	.1107841	3.39	0.001	.1579152	.5921811
3	.6731138	.1696881	3.97	0.000	.3405312	1.005696
4	.4375264	.1607411	2.72	0.006	.1224797	.7525731
5	-.0993631	.6332619	-0.16	0.875	-1.340534	1.141807
6	.5446018	.3047359	1.79	0.074	-.0526697	1.141873
7	.9652151	.1893211	5.10	0.000	.5941525	1.336278
8	0	(empty)				
_cons	-.5288585	.101922	-5.19	0.000	-.728622	-.3290951

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2537.75266	1488.93294	1048.81971	53.5513812	19.59
	ATT	2537.75266	1743.41233	794.34033	547.687768	1.45

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	1,157	1,157
Treated	1,726	1,726
Total	2,883	2,883

Phụ lục 2: Kết quả ước lượng theo phương pháp PSM cho khu vực thành thị năm 2010

```
. psmatch2 treat female honnhan dantoc tuoi i.educ if tuoi>=16 & tuoi<=65 & khuvuc==1 , outcome(income_
> m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 6 obs not used

note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 6 obs not used
```

```
Probit regression                               Number of obs   =       1657
                                                LR chi2(9)      =       84.08
                                                Prob > chi2     =       0.0000
Log likelihood = -535.74289                    Pseudo R2      =       0.0728
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.0438461	.0886606	0.49	0.621	-.1299254	.2176176
honnhan	0	(omitted)				
dantoc	.4377433	.1652974	2.65	0.008	.1137665	.7617202
tuoi	-.0214837	.0041056	-5.23	0.000	-.0295305	-.0134369
educ						
2	.2836588	.1464091	1.94	0.053	-.0032978	.5706155
3	.3769156	.2032836	1.85	0.064	-.0215129	.775344
4	.7381878	.2201077	3.35	0.001	.3067845	1.169591
5	.172294	.541328	0.32	0.750	-.8886894	1.233277
6	.3716527	.2900925	1.28	0.200	-.1969182	.9402236
7	.9647381	.2154088	4.48	0.000	.5425446	1.386932
8	0	(empty)				
_cons	1.449038	.2308092	6.28	0.000	.9966601	1.901415

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2890.99587	1687.52627	1203.4696	135.225667	8.90
	ATT	2890.99587	2470.39827	420.5976	178.574192	2.36

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On support	Total
Untreated	184	184
Treated	1,473	1,473
Total	1,657	1,657

```
. psmatch2 treat female honnhan dantoc i.educ if tuoi>=16 & tuoi<=65 & kn==1 & khuvuc==1 , outcome(income_m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 4 obs not used

note: dantoc != 1 predicts success perfectly
      dantoc dropped and 7 obs not used

note: 2.educ != 0 predicts success perfectly
      2.educ dropped and 19 obs not used

note: 4.educ != 0 predicts success perfectly
      4.educ dropped and 17 obs not used

note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 3 obs not used

note: 6.educ != 0 predicts success perfectly
      6.educ dropped and 8 obs not used

note: 7.educ != 0 predicts success perfectly
      7.educ dropped and 31 obs not used

note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 1 obs not used
```

```
Probit regression                               Number of obs =          155
                                                LR chi2(2)          =           1.10
                                                Prob > chi2         =           0.5776
Log likelihood = -36.529827                    Pseudo R2          =           0.0148
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	-.3222936	.317249	-1.02	0.310	-.9440903	.2995032
honnhan	0	(omitted)				
dantoc	0	(omitted)				
educ						
2	0	(empty)				
3	.1307997	.516601	0.25	0.800	-.8817197	1.143319
4	0	(empty)				
5	0	(empty)				
6	0	(empty)				
7	0	(empty)				
8	0	(empty)				
_cons	1.666467	.2386711	6.98	0.000	1.19868	2.134254

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2233.9569	1433.08334	800.873563	429.493423	1.86
	ATT	2233.9569	697.931043	1536.02586	218.036537	7.04

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	10	10
Treated	145	145
Total	155	155

```
. psmatch2 treat female honnhan dantoc tuoi i.educ if tuoi>=16 & tuoi<=65 & kn==2 & khuvuc==1 , outcome(income_
> m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 2 obs not used

note: 4.educ != 0 predicts success perfectly
      4.educ dropped and 31 obs not used

note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 3 obs not used

note: 7.educ != 0 predicts success perfectly
      7.educ dropped and 58 obs not used

note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 2 obs not used
```

```
Probit regression                               Number of obs   =       272
                                                LR chi2(6)      =         6.05
                                                Prob > chi2     =       0.4172
Log likelihood = -70.92787                    Pseudo R2      =       0.0409
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.1530081	.2453664	0.62	0.533	-.3279012	.6339173
honnhan	0	(omitted)				
dantoc	.7540236	.4193556	1.80	0.072	-.0678983	1.575946
tuoi	-.0047955	.0844905	-0.06	0.955	-.1703939	.1608029
educ						
2	-.243332	.3288542	-0.74	0.459	-.8878743	.4012104
3	.4659383	.4599183	1.01	0.311	-.4354851	1.367362
4	0	(empty)				
5	0	(empty)				
6	.0898934	.5129928	0.18	0.861	-.9155541	1.095341
7	0	(empty)				
8	0	(empty)				
_cons	.7861341	2.422429	0.32	0.746	-3.961739	5.534007

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2550.86952	1492.30158	1058.56793	295.819848	3.58
	ATT	2550.86952	1604.57171	946.297808	246.163558	3.84

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On support	Total
Untreated	21	21
Treated	251	251
Total	272	272

```
. psmatch2 treat female honnhan dantoc i.educ if tuoi>=16 & tuoi<=65 & kn==3 & khuvuc==1 , outcome(income_m)
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 3 obs not used
```

note: honnhan omitted because of collinearity

```
Probit regression                               Number of obs   =    1053
                                                LR chi2(8)      =    35.16
                                                Prob > chi2     =    0.0000
Log likelihood = -418.85237                    Pseudo R2      =    0.0403
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.0641149	.1018351	0.63	0.529	-.1354783	.2637081
honnhan	0	(omitted)				
dantoc	.3856903	.1879215	2.05	0.040	.0173708	.7540097
educ						
2	.3945275	.1679974	2.35	0.019	.0652586	.7237964
3	.3652541	.2573001	1.42	0.156	-.1390448	.869553
4	.5761045	.2301633	2.50	0.012	.1249927	1.027216
5	-.0905251	.638823	-0.14	0.887	-1.342595	1.161545
6	.4570833	.3722428	1.23	0.219	-.2724993	1.186666
7	.8826858	.2306152	3.83	0.000	.4306884	1.334683
8	0	(empty)				
_cons	.5152977	.1872307	2.75	0.006	.1483323	.8822631

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2945.93954	1730.95207	1214.98747	150.970529	8.05
	ATT	2945.93954	1919.41593	1026.52361	811.381572	1.27

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	153	153
Treated	900	900
Total	1,053	1,053

Phụ lục 3: Kết quả ước lượng theo phương pháp PSM cho khu vực nông thôn năm 2010

```
. psmatch2 treat female honnhan dantoc tuoi i.educ if khuvuc==0 , outcome(income_m)
```

```
Probit regression                               Number of obs   =    2985
                                                LR chi2(10)    =    257.45
                                                Prob > chi2    =    0.0000
Log likelihood = -1931.6774                    Pseudo R2      =    0.0625
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.2280402	.0509324	4.48	0.000	.1282146	.3278659
honnhan	1.078668	.5410627	1.99	0.046	.0182049	2.139132
dantoc	.4412414	.0843676	5.23	0.000	.275884	.6065988
tuoi	-.0234416	.0023523	-9.97	0.000	-.028052	-.0188312
educ						
2	.372629	.1056349	3.53	0.000	.1655885	.5796695
3	.4302539	.132818	3.24	0.001	.1699354	.6905724
4	.5046966	.1551955	3.25	0.001	.200519	.8088742
5	1.001409	.5429483	1.84	0.065	-.06275	2.065568
6	.6879277	.271781	2.53	0.011	.1552467	1.220609
7	1.133609	.2304302	4.92	0.000	.6819746	1.585244
_cons	.3831281	.1134161	3.38	0.001	.1608367	.6054196

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2090.98111	1431.29369	659.687422	39.6660066	16.63
	ATT	2090.98111	1602.26344	488.717675	118.313209	4.13

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	1,379	1,379
Treated	1,606	1,606
Total	2,985	2,985

```
. psmatch2 treat female honnhan dantoc suckhoe tuoi i.educ if tuoi>=16 & tuoi<=65 & kn==1 & khuvuc==0 , outcome
> (income_m)
```

```
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 9 obs not used
```

```
note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 3 obs not used
```

```
Probit regression                               Number of obs =      231
                                                LR chi2(9)         =      43.49
                                                Prob > chi2        =      0.0000
Log likelihood = -109.50927                    Pseudo R2         =      0.1657
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.291248	.1998057	1.46	0.145	-.1003639	.68286
honnhan	0	(omitted)				
dantoc	1.30191	.2424783	5.37	0.000	.8266612	1.777159
suckhoe	.0032039	.3655202	0.01	0.993	-.7132026	.7196104
tuoi	-.0556659	.087881	-0.63	0.526	-.2279094	.1165776
educ						
2	.02226	.3177153	0.07	0.944	-.6004505	.6449706
3	-.2251334	.3376947	-0.67	0.505	-.8870029	.4367362
4	.0745301	.373626	0.20	0.842	-.6577634	.8068237
5	0	(empty)				
6	.5505206	.5660225	0.97	0.331	-.5588632	1.659904
7	.4290808	.5497669	0.78	0.435	-.6484424	1.506604
_cons	.7976185	2.072299	0.38	0.700	-3.264013	4.85925

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2145.12645	1481.93645	663.189999	160.369564	4.14
	ATT	2145.12645	1820.92298	324.203469	308.184358	1.05

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On support	Total
Untreated	59	59
Treated	172	172
Total	231	231

```
. psmatch2 treat female honnhan dantoc i.educ if tuoi>=16 & tuoi<=65 & kn==2 & khuvuc==0 , outcome(income_m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 1 obs not used
```

```
note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 4 obs not used
```

```
Probit regression                               Number of obs =          605
                                                LR chi2(7)          =          35.66
                                                Prob > chi2         =          0.0000
Log likelihood = -377.6051                     Pseudo R2          =          0.0451
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.1310323	.1136627	1.15	0.249	-.0917426	.3538072
honnhan	0	(omitted)				
dantoc	.5910728	.1623264	3.64	0.000	.2729189	.9092266
educ						
2	.5311323	.2426169	2.19	0.029	.0556119	1.006653
3	.224833	.2303876	0.98	0.329	-.2267183	.6763844
4	.3830135	.3100739	1.24	0.217	-.2247202	.9907472
5	0	(empty)				
6	.4326083	.4983658	0.87	0.385	-.5441707	1.409387
7	1.314482	.4686738	2.80	0.005	.3958986	2.233066
_cons	-.2970292	.1566248	-1.90	0.058	-.6040081	.0099497

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2221.76916	1431.96292	789.806243	91.5466022	8.63
	ATT	2221.76916	1707.46254	514.306622	500.322984	1.03

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	218	218
Treated	387	387
Total	605	605

```
. psmatch2 treat female honnhan dantoc i.educ if tuoi>=16 & tuoi<=65 & kn==3 & khuvuc==0 , outcome(income_m)
```

```
Probit regression                Number of obs   =    1796
                                LR chi2(8)         =    51.02
                                Prob > chi2        =    0.0000
Log likelihood = -1208.4474      Pseudo R2       =    0.0207
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	.2394771	.064953	3.69	0.000	.1121716	.3667826
honnhan	.2451956	.8870678	0.28	0.782	-1.493425	1.983817
dantoc	.2601101	.1225202	2.12	0.034	.0199749	.5002454
educ						
2	.3528981	.149691	2.36	0.018	.0595092	.646287
3	.9026085	.2265994	3.98	0.000	.4584818	1.346735
4	.2694617	.2392489	1.13	0.260	-.1994575	.7383808
6	.6677816	.5225259	1.28	0.201	-.3563504	1.691914
7	1.050769	.3290876	3.19	0.001	.4057691	1.695769
_cons	-.5053058	.1200594	-4.21	0.000	-.7406179	-.2699936

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2113.94806	1457.51814	656.429926	52.6820494	12.46
	ATT	2113.94806	1527.27452	586.673542	543.273242	1.08

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	997	997
Treated	799	799
Total	1,796	1,796

Phụ lục 4: Kết quả ước lượng theo phương pháp PSM cho lao động nam năm 2010

```
. psmatch2 treat honnhan dantoc suckhoe khuvuc tuoi i.educ if female==0 , outcome(income_m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 9 obs not used
```

```
note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 6 obs not used
```

```
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 4 obs not used
```

```
Probit regression                Number of obs   =       1163
                                LR chi2(9)          =       335.44
                                Prob > chi2         =       0.0000
                                Pseudo R2           =       0.2326
Log likelihood = -553.46437
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0	(omitted)				
dantoc	.9377538	.1215307	7.72	0.000	.699558	1.17595
suckhoe	-.1228254	.1338397	-0.92	0.359	-.3851463	.1394955
khuvuc	1.020244	.1022302	9.98	0.000	.8198761	1.220611
tuoi	-.0197626	.0038515	-5.13	0.000	-.0273114	-.0122137
educ						
2	.5266617	.1588558	3.32	0.001	.21531	.8380134
3	.4561434	.1568901	2.91	0.004	.1486444	.7636423
4	.7311031	.2338359	3.13	0.002	.2727932	1.189413
5	0	(empty)				
6	.4889321	.3202427	1.53	0.127	-.1387321	1.116596
7	.7903176	.2003837	3.94	0.000	.3975727	1.183063
8	0	(empty)				
_cons	-.0642919	.172143	-0.37	0.709	-.401686	.2731021

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	3187.73398	1371.2942	1816.43978	109.232035	16.63
	ATT	3187.73398	2027.07708	1160.6569	181.131306	6.41

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	362	362
Treated	801	801
Total	1,163	1,163

```
. psmatch2 treat honnhan suckhoe khuvuc tuoi i.educ if tuoi>=16 & tuoi<=65 & kn==1 & female==0, outcome(income_
> m)
```

```
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 7 obs not used
```

```
note: khuvuc != 0 predicts success perfectly
      khuvuc dropped and 49 obs not used
```

```
note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 2 obs not used
```

```
Probit regression                               Number of obs =       121
                                                LR chi2(7)         =        3.81
                                                Prob > chi2        =       0.8013
Log likelihood = -74.879935                    Pseudo R2         =       0.0248
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0 (omitted)					
suckhoe	-.0843234	.4966844	-0.17	0.865	-1.057807	.8891601
khuvuc	0 (omitted)					
tuoi	-.0816547	.1069008	-0.76	0.445	-.2911764	.127867
educ						
2	.2075758	.3879051	0.54	0.593	-.5527042	.9678558
3	-.0827289	.3682476	-0.22	0.822	-.804481	.6390231
4	.8919849	.5768229	1.55	0.122	-.2385672	2.022537
5	0 (empty)					
6	.0568037	.7701009	0.07	0.941	-1.452566	1.566174
7	.3021559	.6956402	0.43	0.664	-1.061274	1.665586
_cons	2.285986	2.49437	0.92	0.359	-2.60289	7.174862

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2339.16254	1420.80209	918.360449	217.505187	4.22
	ATT	2339.16254	1138.5998	1200.56273	486.614341	2.47

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support		Total
	On suppor		
Untreated	40		40
Treated	81		81
Total	121		121

```
. psmatch2 treat honnhan dantoc suckhoe khuvuc i.educ if tuoi>=16 & tuoi<=65 & kn==2 & female==0, outcome(incom
> e_m)
```

note: honnhan != 0 predicts success perfectly
honnhan dropped and 1 obs not used

note: 4.educ != 0 predicts success perfectly
4.educ dropped and 11 obs not used

note: 5.educ != 0 predicts success perfectly
5.educ dropped and 2 obs not used

note: 7.educ != 0 predicts success perfectly
7.educ dropped and 34 obs not used

note: 8.educ != 0 predicts success perfectly
8.educ dropped and 1 obs not used

```
Probit regression                               Number of obs   =       191
                                                LR chi2(6)      =       58.39
                                                Prob > chi2     =       0.0000
Log likelihood = -87.223704                    Pseudo R2      =       0.2508
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0 (omitted)					
dantoc	1.391376	.2684346	5.18	0.000	.8652538	1.917498
suckhoe	.0006088	.7634135	0.00	0.999	-1.495654	1.496872
khuvuc	.8027521	.2852187	2.81	0.005	.2437338	1.36177
educ						
2	.3334489	.3742196	0.89	0.373	-.4000081	1.066906
3	.3164221	.3060054	1.03	0.301	-.2833375	.9161818
4	0 (empty)					
5	0 (empty)					
6	-.3734082	.52674	-0.71	0.478	-1.4058	.6589833
7	0 (empty)					
8	0 (empty)					
_cons	-.8083556	.2321313	-3.48	0.000	-1.263325	-.3533866

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	3012.92349	1380.7924	1632.13109	219.36987	7.44
	ATT	3012.92349	1792.76118	1220.16231	458.936826	2.66

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support		Total
	On suppor		
Untreated	57		57
Treated	134		134
Total	191		191

```
. psmatch2 treat honnhan dantoc suckhoe khuvuc i.educ if tuoi>=16 & tuoi<=65 & kn==3 & female==0, outcome(incom
> e_m)
```

```
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 1 obs not used
```

```
note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 1 obs not used
```

```
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 3 obs not used
```

```
Probit regression                               Number of obs   =       738
                                                LR chi2(8)      =       197.98
                                                Prob > chi2     =       0.0000
Log likelihood = -380.47613                    Pseudo R2      =       0.2065
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0	(omitted)				
dantoc	.5755028	.1592824	3.61	0.000	.2633149	.8876906
suckhoe	-.1830834	.157404	-1.16	0.245	-.4915896	.1254229
khuvuc	1.013994	.1154832	8.78	0.000	.787651	1.240337
educ						
2	.59731	.1931147	3.09	0.002	.2188122	.9758078
3	.7191638	.2340766	3.07	0.002	.260382	1.177946
4	.6142138	.2684613	2.29	0.022	.0880394	1.140388
5	0	(empty)				
6	1.169796	.5735854	2.04	0.041	.0455892	2.294003
7	.7363576	.2308461	3.19	0.001	.2839076	1.188808
8	0	(empty)				
_cons	-.701999	.1524456	-4.60	0.000	-1.000787	-.403211

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	3284.69829	1370.02714	1914.67115	133.538824	14.34
	ATT	3284.69829	1912.77186	1371.92644	365.649381	3.75

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	261	261
Treated	477	477
Total	738	738

Phụ lục 5: Kết quả ước lượng theo phương pháp PSM cho lao động nữ năm 2010

```
. psmatch2 treat honnhan dantoc suckhoe khuvuc i.educ if tuoi>=16 & tuoi<=65 & kn==3 & female==0, outcome(incom
> e_m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 1 obs not used

note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 1 obs not used

note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 3 obs not used
```

```
Probit regression                               Number of obs   =       738
                                                LR chi2(8)      =     197.98
                                                Prob > chi2     =       0.0000
Log likelihood = -380.47613                    Pseudo R2      =       0.2065
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0	(omitted)				
dantoc	.5755028	.1592824	3.61	0.000	.2633149	.8876906
suckhoe	-.1830834	.157404	-1.16	0.245	-.4915896	.1254229
khuvuc	1.013994	.1154832	8.78	0.000	.787651	1.240337
educ						
2	.59731	.1931147	3.09	0.002	.2188122	.9758078
3	.7191638	.2340766	3.07	0.002	.260382	1.177946
4	.6142138	.2684613	2.29	0.022	.0880394	1.140388
5	0	(empty)				
6	1.169796	.5735854	2.04	0.041	.0455892	2.294003
7	.7363576	.2308461	3.19	0.001	.2839076	1.188808
8	0	(empty)				
_cons	-.701999	.1524456	-4.60	0.000	-1.000787	-.403211

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	3284.69829	1370.02714	1914.67115	133.538824	14.34
	ATT	3284.69829	1912.77186	1371.92644	365.649381	3.75

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support		Total
	On suppor		
Untreated	261		261
Treated	477		477
Total	738		738

```
. psmatch2 treat honnhan dantoc suckhoe khuvuc tuoi i.educ if female=1 , outcome(income_m)
=exp not allowed
r(101);
```

```
.
. psmatch2 treat honnhan dantoc khuvuc i.educ if tuoi>=16 & tuoi<=65 & kn==1 & female==1 , outcome(income_m)
note: honnhan != 0 predicts success perfectly
      honnhan dropped and 6 obs not used

note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 4 obs not used

note: 6.educ != 0 predicts success perfectly
      6.educ dropped and 14 obs not used

note: 7.educ != 0 predicts success perfectly
      7.educ dropped and 27 obs not used

note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 1 obs not used
```

```
Probit regression                               Number of obs   =       268
                                                LR chi2(5)      =       14.94
                                                Prob > chi2     =       0.0106
Log likelihood = -115.40488                    Pseudo R2      =       0.0608
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0	(omitted)				
dantoc	.0723131	.3183064	0.23	0.820	-.551556	.6961822
khuvuc	.6945563	.2367099	2.93	0.003	.2306134	1.158499
educ						
2	-.2512567	.3142801	-0.80	0.424	-.8672343	.3647209
3	-.011787	.4331201	-0.03	0.978	-.8606868	.8371128
4	.5810497	.3803525	1.53	0.127	-.1644274	1.326527
5	0	(empty)				
6	0	(empty)				
7	0	(empty)				
8	0	(empty)				
_cons	.6824463	.3065932	2.23	0.026	.0815348	1.283358

There are observations with identical propensity score values.
 The sort order of the data could affect your results.
 Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	1793.8157	1227.63769	566.178012	140.463663	4.03
	ATT	1793.8157	996.332593	797.483104	373.76707	2.13

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	46	46
Treated	222	222
Total	268	268

```

. psmatch2 treat honnhan dantoc suckhoe khuvuc i.educ if tuoi>=16 & tuoi<=65 & kn==2 & female==1, outcome(incom
> e_m)
note: 3.educ != 0 predicts success perfectly
      3.educ dropped and 8 obs not used

note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 2 obs not used

note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 1 obs not used

note: honnhan omitted because of collinearity

Probit regression                               Number of obs   =       204
                                                LR chi2(7)      =       37.46
                                                Prob > chi2     =       0.0000
Log likelihood = -71.558213                    Pseudo R2      =       0.2074

```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0	(omitted)				
dantoc	.8649337	.3315875	2.61	0.009	.2150342	1.514833
suckhoe	.0486336	.3911135	0.12	0.901	-.7179348	.815202
khuvuc	1.210012	.3184205	3.80	0.000	.585919	1.834105
educ						
2	-.5824795	.5659919	-1.03	0.303	-1.691803	.5268443
3	0	(empty)				
4	.0587226	.4959484	0.12	0.906	-.9133183	1.030764
5	0	(empty)				
6	.2911748	.7145284	0.41	0.684	-1.109275	1.691625
7	.4747369	.5566152	0.85	0.394	-.6162089	1.565683
8	0	(empty)				
_cons	-.1721975	.3136418	-0.55	0.583	-.7869242	.4425293

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2734.35233	1201.33838	1533.01395	290.04712	5.29
	ATT	2734.35233	1561.61208	1172.74026	331.514187	3.54

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On support	Total
Untreated	33	33
Treated	171	171
Total	204	204

```
. psmatch2 treat honnhan dantoc suckhoe khuvuc i.educ if tuoi>=16 & tuoi<=65 & kn==2 & female==1, outcome(incom
> e_m)
```

```
note: 3.educ != 0 predicts success perfectly
      3.educ dropped and 8 obs not used
```

```
note: 5.educ != 0 predicts success perfectly
      5.educ dropped and 2 obs not used
```

```
note: 8.educ != 0 predicts success perfectly
      8.educ dropped and 1 obs not used
```

```
note: honnhan omitted because of collinearity
```

```
Probit regression                               Number of obs   =       204
                                                LR chi2(7)      =       37.46
                                                Prob > chi2     =       0.0000
Log likelihood = -71.558213                    Pseudo R2      =       0.2074
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0	(omitted)				
dantoc	.8649337	.3315875	2.61	0.009	.2150342	1.514833
suckhoe	.0486336	.3911135	0.12	0.901	-.7179348	.815202
khuvuc	1.210012	.3184205	3.80	0.000	.585919	1.834105
educ						
2	-.5824795	.5659919	-1.03	0.303	-1.691803	.5268443
3	0	(empty)				
4	.0587226	.4959484	0.12	0.906	-.9133183	1.030764
5	0	(empty)				
6	.2911748	.7145284	0.41	0.684	-1.109275	1.691625
7	.4747369	.5566152	0.85	0.394	-.6162089	1.565683
8	0	(empty)				
_cons	-.1721975	.3136418	-0.55	0.583	-.7869242	.4425293

There are observations with identical propensity score values.

The sort order of the data could affect your results.

Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2734.35233	1201.33838	1533.01395	290.04712	5.29
	ATT	2734.35233	1561.61208	1172.74026	331.514187	3.54

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	33	33
Treated	171	171
Total	204	204

```
. psmatch2 treat honnhan dantoc suckhoe khuvuc i.educ if tuoi>=16 & tuoi<=65 & kn==3 & female==1, outcome(incom
> e_m)
note: 5.educ != 0 predicts failure perfectly
      5.educ dropped and 1 obs not used
```

note: honnhan omitted because of collinearity

```
Probit regression                               Number of obs =      464
                                                LR chi2(8)         =      79.57
                                                Prob > chi2        =      0.0000
Log likelihood = -231.56191                    Pseudo R2         =      0.1466
```

treat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
honnhan	0	(omitted)				
dantoc	.2973391	.2410923	1.23	0.217	-.1751932	.7698714
suckhoe	-.224415	.2202625	-1.02	0.308	-.6561215	.2072915
khuvuc	1.005487	.1474802	6.82	0.000	.7164307	1.294542
educ						
2	-.5033072	.2747146	-1.83	0.067	-1.041738	.0351234
3	-.1343754	.468482	-0.29	0.774	-1.052583	.7838325
4	.2997364	.2926568	1.02	0.306	-.2738604	.8733333
5	0	(empty)				
6	.2361239	.6451756	0.37	0.714	-1.028397	1.500645
7	.8547977	.4400816	1.94	0.052	-.0077463	1.717342
_cons	-.0833552	.2358409	-0.35	0.724	-.5455949	.3788846

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
income_m	Unmatched	2503.08604	1367.34723	1135.73882	152.891017	7.43
	ATT	2503.08604	1711.36981	791.716239	834.352708	0.95

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support	
	On suppor	Total
Untreated	126	126
Treated	338	338
Total	464	464

Phụ lục 6: Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng năm 2010

```
. qui:treatreg lnincome_m tuoi tuoiibp khuvuuc dantoc honnhan female i.educn if kn==1, treat (employed=tuoi tuoi
> bp i.educn khuvuuc dantoc honnhan female )
Warning: variance matrix is nonsymmetric or highly singular

. est store u_20

. qui:treatreg lnincome_m tuoi tuoiibp khuvuuc dantoc honnhan female i.educn if kn==2, treat (employed=tuoi tuoi
> bp i.educn khuvuuc dantoc honnhan female )

. est store u_30

. qui:treatreg lnincome_m tuoi tuoiibp khuvuuc dantoc honnhan female i.educn if kn==3, treat (employed=tuoi tuoi
> bp i.educn khuvuuc dantoc honnhan female )

. est store u_40

. est table u_10 u_20 u_30 u_40 , p stats(r2 N)
```

Variable	u_10	u_20	u_30	u_40
lnincome_m				
tuoi	.08459903	.33056232	-.77413109	.04391865
	0.0000	0.9453	0.0928	0.0015
tuoiibp	-.00117371	-.00805712	.01763072	-.00069373
	0.0000	0.9491	0.0701	0.0000
khuvuuc	.33643839	-.15058132	.18781039	.45171667
	0.0000	0.2490	0.0018	0.0000
dantoc	.58307756	.25564058	.76061622	.53917354
	0.0000	0.4186	0.0000	0.0000
honnhan	-.65216486	-.37146407	-.52258925	-1.0632924
	0.0000	0.2735	0.0001	0.0203
female	-.22472874	-.14819325	-.11177283	-.27645624
	0.0000	0.2068	0.0226	0.0000
educn				
2	.17647027	.07555904	.12727564	.19162579
	0.0001	0.5785	0.0750	0.0005
3	.15347271	-.47097303	.15127544	.15380149
	0.0004	0.0856	0.0160	0.0052
4	.18996925	-.39140208	.17017034	.15394637
	0.0063	0.3483	0.0603	0.1022
5	.21077787	-.26175665	.21786746	.23617379
	0.0000	0.2742	0.0041	0.0000
employed	.32704455	.80777059	.08842536	.17048859
	0.0736	0.0158	0.6425	0.2799
_cons	5.2440906	2.9996818	15.031523	6.1681756
	0.0000	0.9480	0.0057	0.0000

employed				
tuoi	-.0936128	-1.0444889	-1.0257445	-.11819474
	0.0006	0.9413	0.4710	0.0002
tuoi bp	.00080733	.03242286	.0211058	.00113544
	0.0206	0.9305	0.4828	0.0016
educn				
2	.20128967	-.36854453	.02683403	.24055959
	0.0543	0.3076	0.8935	0.0434
3	.31709621	4.265025	-.10219623	.44784952
	0.0021	0.9977	0.5688	0.0003
4	.38951512	3.2922064	.37333505	.37505
	0.0425	.	0.2422	0.1171
5	.56025199	1.6639666	.50205347	.62547096
	0.0000	0.1466	0.0820	0.0000
khuvuc	1.1383883	.88791072	1.4138817	1.0885913
	0.0000	0.0879	0.0000	0.0000
dantoc	.59209535	2.0953661	.78260987	.49611733
	0.0000	0.0088	0.0004	0.0004
honnhan	5.1967984	4.5223735	5.0104989	5.4759528
	0.9940	0.9999	0.9974	0.9986
female	.09306369	.36130626	.21827926	.03938594
	0.2395	0.3224	0.1388	0.6676
_cons	1.7778323	6.8082427	12.36222	2.2848308
	0.0006	0.9598	0.4613	0.0006
athrho				
_cons	.17067713	-.75589439	.37057742	.32200511
	0.3198	0.0786	0.0520	0.0273
lnsigma				
_cons	-.48012067	-.57643703	-.54333124	-.44358285
	0.0000	0.0000	0.0000	0.0000
Statistics				
r2				
N	1846	102	614	1320

legend: b/p

Phụ lục 7: Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng khu vực thành thị năm 2010

```
. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=22 & tuoi<=65 & khuvc==1, treat
> (employed=tuoi tuoibp i.educn dantoc honnhan female )

. est store r_10

. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=16 & tuoi<=65 & kn==1 & khuvc==
> 1, treat (employed=tuoi tuoibp i.educn honnhan female )
Warning: variance matrix is nonsymmetric or highly singular

. est store r_20

. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=16 & tuoi<=65 & kn==2 & khuvc==
> 1, treat (employed=tuoi tuoibp i.educn dantoc honnhan female suckhoe)

. est store r_30

. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=16 & tuoi<=65 & kn==3 & khuvc==
> 1, treat (employed=tuoi tuoibp i.educn dantoc honnhan female suckhoe)

. est store r_40

. est table r_10 r_20 r_30 r_40 , p stats(r2 N)
```

Variable	r_10	r_20	r_30	r_40
lnincome_m				
tuoi	.06512981	7.690182	-.02950171	.10970318
	0.0000	0.0220	0.9660	0.0000
tuoibp	-.00088265	-.16639138	.00136449	-.00139507
	0.0000	0.0226	0.9138	0.0000
dantoc	.18751313	-.04230277	.38562663	.31096628
	0.0013	0.8034	0.0364	0.0016
honnhan	-.69205355	-1.01545	-.24654804	(omitted)
	0.0004	0.0000	0.4560	
female	-.22347279	-.23120158	-.20731237	-.20267866
	0.0000	0.0008	0.0005	0.0000
educn				
2	.29554218	.07432838	.18609128	.26420617
	0.0000	0.5716	0.1311	0.0002
3	.22971812	.11995691	.25787702	.18831288
	0.0000	0.2376	0.0010	0.0045
4	.41694627	.09899995	.27279965	.56042236
	0.0000	0.5580	0.0281	0.0000
5	.59054221	.49075957	.65813133	.6319159
	0.0000	0.0000	0.0000	0.0000
employed	.14674298	.73980284	-.21450541	.28839463
	0.2889	0.0620	0.5078	0.2441
_cons	6.2236692	-81.748466	7.3619041	5.0940474
	0.0000	0.0343	0.4380	0.0000

employed				
tuoi	-.10755087	-12.43521	-5.4126127	-.1245514
	0.0005	0.4856	0.2628	0.1428
tuoi p	.00108759	.27218751	.09726003	.00127718
	0.0042	0.4834	0.2677	0.1683
educn				
2	.2754297	4.7217417	-.36629735	.31869408
	0.0611	.	0.4732	0.1563
3	.52151991	-.11987906	.46679966	.50850116
	0.0006	0.7983	0.3344	0.0196
4	.28305296	4.3048867	-.11889132	.49609825
	0.2805	.	0.8578	0.3148
5	.56718254	11.209698	.92657285	.39122863
	0.0002	.	0.2051	0.0586
dantoc	.49734424		1.5959748	.51960104
	0.0015		0.0021	0.0334
honnhan	3.2408738	4.0660889	4.1391698	(omitted)
	0.9775	.	.	
female	.01749243	-.53657753	.25262193	.01952225
	0.8386	0.1985	0.4828	0.8898
suckhoe			4.0934174	-.49868696
			0.9968	0.0248
_cons	3.0089648	143.87219	75.132065	3.4540234
	0.0000	0.4817	0.2582	0.0724
athrho				
_cons	.14253215	-.73244606	.47981931	.12596667
	0.3084	0.1911	0.1900	0.6090
lnsigma				
_cons	-.67246461	-.78640922	-.78196124	-.637954
	0.0000	0.0000	0.0000	0.0000
Statistics				
r2				
N	1732	191	246	593

legend: b/p

Phụ lục 8: Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng khu vực nông thôn năm 2010

```
. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=22 & tuoi<=65 & khuvc==0, treat
> (employed=tuoi tuoibp i.educn khuvc dantoc honnhan female )

. est store t_10

. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=16 & tuoi<=65 & kn==1 & khuvc==
> 0, treat (employed=tuoi tuoibp i.educn dantoc honnhan female suckhoe)

. est store t_20

. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=16 & tuoi<=65 & kn==2 & khuvc==
> 0, treat (employed=tuoi tuoibp i.educn dantoc honnhan female suckhoe)

. est store t_30

. qui:treatreg lnincome_m tuoi tuoibp dantoc honnhan female i.educn if tuoi>=16 & tuoi<=65 & kn==3 & khuvc==
> 0, treat (employed=tuoi tuoibp i.educn dantoc honnhan female suckhoe)

. est store t_40

. est table t_10 t_20 t_30 t_40 , p stats(r2 N)
```

Variable	t_10	t_20	t_30	t_40
lnincome_m				
tuoi	.06858586	2.609154	.86238058	.10798012
	0.0000	0.0764	0.2987	0.0017
tuoibp	-.00101878	-.0617269	-.01653587	-.0014462
	0.0000	0.0914	0.3123	0.0002
dantoc	.47099146	-.11639996	.26459715	.17091036
	0.0000	0.5479	0.3218	0.0647
honnhan	-.94740176	-1.4532767	-1.2424113	-2.0017835
	0.0000	0.0000	0.0000	0.0312
female	-.41380302	.02814217	-.22216031	-.42280825
	0.0000	0.8231	0.0049	0.0000
educn				
2	.34543885	.45337648	.16574373	.46792468
	0.0000	0.0439	0.2481	0.0023
3	.27561494	.50289934	.13239668	.09590936
	0.0000	0.0812	0.2449	0.4832
4	.37338163	-.14453721	.057652	-.1361895
	0.0009	0.7066	0.7511	0.6818
5	-.06780472	-.58127815	-.10100644	-.31073528
	0.2904	0.1277	0.4090	0.0058
employed	.19772174	1.3775405	1.0301819	1.6789917
	0.3096	0.0000	0.0700	0.0000
_cons	5.740775	-21.327917	-4.6666274	4.4088998
	0.0000	0.1483	0.6583	0.0000

employed				
tuoi	-.11589622	2.2044096	-2.276276	-.13400767
	0.0000	0.3506	0.1467	0.0036
tuoiibp	.00119266	-.05330952	.04576877	.00141592
	0.0000	0.3651	0.1384	0.0073
educn				
2	.35401612	-.35130471	.37427118	-.04735422
	0.0008	0.3030	0.2038	0.8163
3	.38439725	-.34754611	.32978391	.27022851
	0.0002	0.4390	0.1508	0.1199
4	.68987652	6.7080754	.36438698	.68112698
	0.0043	.	0.3275	0.1566
5	.62282464	1.0354388	.29991449	.49240413
	0.0000	0.1765	0.2172	0.0009
khuvuc	(omitted)			
dantoc	.49708197	1.2797443	1.2125354	.41593113
	0.0000	0.0000	0.0000	0.0007
honnhan	1.0796459	8.3848888	5.1261869	55.819166
	0.0333	.	0.9874	.
female	.20813721	.22669078	.25063678	.1968682
	0.0000	0.2569	0.1024	0.0297
suckhoe		-.04983412	-.33534961	-.11643537
		0.8787	0.2328	0.2963
_cons	2.0023723	-23.001701	27.556193	2.5777659
	0.0000	0.3299	0.1648	0.0085
athrho				
_cons	.12220427	-.97323808	-.42556837	-1.1595904
	0.4960	0.0000	0.4300	0.0000
lnsigma				
_cons	-.40602486	-.11133657	-.39532291	-.07801967
	0.0000	0.0895	0.0001	0.1255
Statistics				
r2				
N	3247	237	371	821

legend: b/p

Phụ lục 9: Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng theo lao động nam năm 2010

```
. qui:treatreg lnincome_m tuoi tuoiibp khuvuc dantoc honnhan i.educn if tuoi>=22 & tuoi<=65 & female==0, treat
> (employed=tuoi tuoiibp i.educn khuvuc dantoc honnhan female )

.

. est store n_10

.

. qui:treatreg lnincome_m tuoi tuoiibp khuvuc dantoc honnhan i.educn if tuoi>=16 & tuoi<=65 & kn==1 & female==
> 0, treat (employed=tuoi tuoiibp i.educn khuvuc dantoc female )
Warning: variance matrix is nonsymmetric or highly singular

.

. qui:treatreg lnincome_m tuoi tuoiibp khuvuc dantoc honnhan i.educn if tuoi>=16 & tuoi<=65 & kn==2 & female==
> 0 , treat (employed=tuoi tuoiibp i.educn khuvuc dantoc honnhan female suckhoe)

.

. est store n_30

.

. qui:treatreg lnincome_m tuoi tuoiibp khuvuc dantoc honnhan i.educn if tuoi>=16 & tuoi<=65 & kn==3 & female==0
> , treat (employed=tuoi tuoiibp i.educn khuvuc dantoc honnhan female suckhoe)

.

. est store n_40

.

. est table n_10 n_20 n_30 n_40 , p stats(r2 N)
```

Variable	n_10	n_20	n_30	n_40
lnincome_m				
tuoi	.06167026	2.4267802	-.06348739	.0819192
	0.0000	0.3043	0.9228	0.0038
tuoiibp	-.00089473	-.05261212	.00181739	-.00116573
	0.0000	0.3056	0.8791	0.0002
khuvuc	.20904813	.10441145	.26282066	.31702236
	0.0000	0.1950	0.0002	0.0001
dantoc	.3881823	.04536784	.6984057	.39866122
	0.0000	0.6672	0.0000	0.0000
honnhan	-.29215624	-.55833797	.06813263	.47075486
	0.0376	0.0002	0.8961	0.4753
educn				
2	.30507481	.23622961	.21551197	.34724724
	0.0000	0.0032	0.0356	0.0001
3	.14532382	.17896174	.22313729	.15279484
	0.0002	0.0320	0.0051	0.0648
4	.26224827	.24729445	.17643615	.32774257
	0.0006	0.0602	0.2868	0.0657
5	.31659278	.15994872	.49297569	.43120526
	0.0000	0.1089	0.0000	0.0000
employed	-.01294247	.41575713	-.27094618	.11317923
	0.9010	0.0543	0.1108	0.5398
_cons	6.1444813	-20.828788	7.4936333	5.6045405
	0.0000	0.4430	0.4028	0.0000

employed				
tuoi	-.11499628	-7.7397839	2.6548996	-.14044836
	0.0000	0.3071	0.2250	0.0203
tuoi p	.00117924	.17045082	-.04930409	.00146698
	0.0000	0.3008	0.2163	0.0267
educn				
2	.48452738	.263727	.37409695	.57314858
	0.0000	0.3488	0.2889	0.0032
3	.46608896	.22363528	.25779946	.67164123
	0.0000	0.4390	0.3106	0.0002
4	.50914725	-.25166816	-.33205256	1.1884886
	0.0286	0.5790	0.5007	0.0424
5	.80155678	.15703579	1.8242372	.70267929
	0.0000	0.7055	0.0002	0.0001
khuvuc	1.1412865	1.6117999	.90878735	1.0441616
	0.0000	0.0000	0.0003	0.0000
dantoc	.57081963	.95253052	1.6463129	.55941617
	0.0000	0.0000	0.0000	0.0001
honnhan	.79963371		3.3673503	5.7265094
	0.1369		0.9982	0.9992
female	(omitted)	(omitted)	(omitted)	(omitted)
suckhoe			-.73292325	-.20344009
			0.1260	0.1827
_cons	1.9004608	87.307514	-36.579143	2.544301
	0.0000	0.3160	0.2218	0.0588
athrho				
_cons	.2497285	-.15190524	.60920876	.22917609
	0.0294	0.5809	0.0071	0.1700
lnsigma				
_cons	-.59183021	-.78475433	-.70198251	-.42970932
	0.0000	0.0000	0.0000	0.0000
Statistics				
r2				
N	3147	371	305	797

legend: b/p

Phụ lục 10: Kết quả hồi quy theo phương pháp Heckman cho hàm tiền lương Mincer mở rộng theo lao động nữ năm 2010

```
. qui:treatreg lnincome_m tuoi tuoibp khuvuuc dantoc honnhan i.educn if tuoi>=22 & tuoi<=65 & female==0, treat
> (employed=tuoi tuoibp i.educn khuvuuc dantoc honnhan female )

. est store m_10

. qui:treatreg lnincome_m tuoi tuoibp khuvuuc dantoc honnhan i.educn if tuoi>=16 & tuoi<=65 & kn==1 & female==
> 1, treat (employed=tuoi tuoibp i.educn khuvuuc dantoc honnhan female suckhoe)
Warning: variance matrix is nonsymmetric or highly singular

. est store m_20

. qui:treatreg lnincome_m tuoi tuoibp khuvuuc dantoc honnhan i.educn if tuoi>=16 & tuoi<=65 & kn==2 & female==
> 1, treat (employed=tuoi tuoibp i.educn khuvuuc dantoc honnhan female suckhoe)

. est store m_30

. qui:treatreg lnincome_m tuoi tuoibp khuvuuc dantoc honnhan i.educn if tuoi>=16 & tuoi<=65 & kn==3 & female==
> 1, treat (employed=tuoi tuoibp i.educn khuvuuc dantoc honnhan female suckhoe)

. est store m_40

. est table m_10 m_20 m_30 m_40 , p stats(r2 N)
```

Variable	m_10	m_20	m_30	m_40
lnincome_m				
tuoi	.05515144	4.1323577	.29461546	.01845846
	0.0000	0.3430	0.6988	0.6484
tuoibp	-.00083625	-.08790817	-.00554699	-.00042586
	0.0000	0.3534	0.6889	0.3330
khuvuuc	.35472169	.18410989	.3135026	.46079485
	0.0000	0.0937	0.0004	0.0028
dantoc	.47861898	.32889283	.33393195	.44646194
	0.0000	0.0566	0.0245	0.0000
honnhan	-.2608317	-.58906581	-.60771912	(omitted)
	0.0996	0.0103	0.1325	
educn				
2	.35654397	.24763525	.02257918	.10147518
	0.0000	0.2318	0.9061	0.4992
3	.21154558	-.01613077	.10221965	-.07245682
	0.0000	0.8969	0.3332	0.4910
4	.35747844	-.25498439	.19826194	.16737244
	0.0000	0.2385	0.1825	0.4552
5	.51293946	.13993523	.47211763	.39512576
	0.0000	0.2697	0.0000	0.0000
employed	-.37135429	.6464487	.05014041	.02320771
	0.0000	0.0570	0.8558	0.9593
_cons	6.3871576	-41.962225	3.0947811	6.6718869
	0.0000	0.4022	0.7664	0.0000

employed				
tuo	-.10336982	-7.595127	2.0172513	-.14299087
	0.0000	0.5960	0.3996	0.0614
tuoibp	.00106113	.16341598	-.03655762	.00148223
	0.0000	0.5996	0.4007	0.0785
educn				
2	.45276714	-.59522748	-.10761046	-.48420993
	0.0000	0.2475	0.8287	0.0722
3	.44035659	-.26977931	.61798084	.08182227
	0.0000	0.5219	0.1673	0.7340
4	.60246101	4.24698	1.1083162	-.19980628
	0.0090	0.9837	0.1303	0.6821
5	1.041205	.00216936	.44808632	.23006657
	0.0000	0.9963	0.2415	0.3844
khuvuc	1.1403404	.9288289	1.0951978	1.0032516
	0.0000	0.0150	0.0001	0.0000
dantoc	.58065134	.72463772	1.1792839	.16220862
	0.0000	0.0752	0.0000	0.4343
honnhan	.60380284	4.6687872	3.6444259	(omitted)
	0.2028	0.9936	0.9893	
female	(omitted)	(omitted)	(omitted)	(omitted)
suckhoe		-.16638843	-.00471431	-.21570243
		0.7211	0.9884	0.3292
_cons	1.6250282	88.384429	-28.198925	3.3214416
	0.0000	0.5914	0.3910	0.0560
athrho				
_cons	.65304427	-.09415037	.33358431	.30591729
	0.0000	0.7853	0.2454	0.4565
lnsigma				
_cons	-.46818139	-.65875053	-.58889702	-.38032549
	0.0000	0.0000	0.0000	0.0000
Statistics				
r2				
N	3190	152	280	541

legend: b/p

Phụ lục 11: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels cho ba nhóm kinh nghiệm năm 2010

```
. qui:ivreg2h lncincome_m tuoi tuoibp khuvuc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=18 & tuoi<=61
> & kn==1,robust

.
. est store u_10

.
. qui:ivreg2h lncincome_m tuoi tuoibp khuvuc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=18 & tuoi<=61
> & kn==2,robust
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities

.
. est store u_20

.
. qui:ivreg2h lncincome_m tuoi tuoibp khuvuc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=18 & tuoi<=61
> & kn==3,robust
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities

.
. est store u_30

.
. est table u_10 u_20 u_30 , p stats(r2 N)
```

Variable	u_10	u_20	u_30
edein1	.46182054 0.0303	.3849836 0.0000	.50027541 0.0000
edein2	.30823434 0.0002	.36467775 0.0000	.43628077 0.0000
edein3	.20213118 0.1231	.25006237 0.0000	.05898378 0.3161
edein4	.39392858 0.0001	.24607931 0.0004	.29105588 0.0000
edein5	(omitted)	(omitted)	(omitted)
tuoi	.75196368 0.4086	-.33907845 0.5756	.02617002 0.1732
tuoi bp	-.01587405 0.4120	.00643134 0.5530	-.00048984 0.0289
khuvuc	.14034012 0.0008	.1856202 0.0000	.28300142 0.0000
dantoc	.18788303 0.0218	.25136873 0.0004	.35755605 0.0000
honnhan	-.25362461 0.2141	-.00074361 0.9929	.48381822 0.0000
female	-.22020935 0.0000	-.28761903 0.0000	-.43354011 0.0000
idnghe	.02838654 0.6459	.05286305 0.0914	-.00848218 0.4302
_cons	-1.6684493 0.8758	11.649939 0.1677	6.9269994 0.0000
r2	.11538452	.20931771	.22076746
N	661	783	2486

legend: b/p

Phụ lục 12: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels cho ba nhóm kinh nghiệm khu vực thành thị năm 2010

```
. qui:ivreg2h lnincome_m tuoi tuoiibp dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61 & kn==
> 1 & khuvuuc==1,robust

.
. est store u_10

.
. qui:ivreg2h lnincome_m tuoi tuoiibp dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61 & kn==
> 2 & khuvuuc==1,robust
Warning: estimated covariance matrix of moment conditions not of full rank.
overidentification statistic not reported, and standard errors and
model tests should be interpreted with caution.
Possible causes:
singleton dummy variable (dummy with one 1 and N-1 0s or vice versa)
partial option may address problem.
Warning: estimated covariance matrix of moment conditions not of full rank.
overidentification statistic not reported, and standard errors and
model tests should be interpreted with caution.
Possible causes:
singleton dummy variable (dummy with one 1 and N-1 0s or vice versa)
partial option may address problem.
Warning: estimated covariance matrix of moment conditions not of full rank.
overidentification statistic not reported, and standard errors and
model tests should be interpreted with caution.
Possible causes:
singleton dummy variable (dummy with one 1 and N-1 0s or vice versa)
partial option may address problem.

.
. est store u_20

.
. qui:ivreg2h lnincome_m tuoi tuoiibp dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61 & kn==
> 3 & khuvuuc==1,robust

.
. est store u_30

.
. est table u_10 u_20 u_30 , p stats(r2 N)
```

Variable	u_10	u_20	u_30
edein5	(omitted)	(omitted)	(omitted)
edein1	.65567882 0.0000	.38270481 0.0009	.77317795 0.0000
edein2	.10613408 0.4010	.14227637 0.2715	.55990794 0.0000
edein3	.09790379 0.4591	.19065149 0.0407	.06028032 0.5994
edein4	.22061337 0.0342	.21054982 0.1092	.34162173 0.0002
tuoi	1.7893786 0.4207	-2.2784945 0.0577	.14219244 0.0010
tuoiibp	-.0373855 0.4291	.0409285 0.0569	-.00181 0.0004
dantoc	.66178742 0.1041	.58786213 0.0011	.60485752 0.0035
honnhan	-1.1375122 0.0000	-.04855496 0.6631	(omitted)
female	-.28004708 0.0084	-.16536332 0.0376	-.3141861 0.0000
idnghe	-.02815438 0.3098	.2811152 0.0030	.01346344 0.2160
_cons	-14.307437 0.5831	38.335746 0.0217	4.4081419 0.0000
r2	.29402629	.21808411	.28892389
N	112	170	443

legend: b/p

Phụ lục 13: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels cho ba nhóm kinh nghiệm khu vực nông thôn năm 2010

```
. qui:ivreg2h lnincome_m tuoi tuoiibp dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=16 & tuoi<=61 & kn==
> 3 & khuvc==0,robust
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities
warning: -ranktest- error in calculating underidentification test statistics;
         may be caused by collinearities
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities

. est store u_30

. est table u_10 u_20 u_30 , p stats(r2 N)
```

Variable	u_10	u_20	u_30
edein1	.39715501 0.0160	.40814721 0.0065	.86433029 0.0000
edein2	.28692642 0.0021	.49316683 0.0000	-.03632922 0.9007
edein3	.20159496 0.0893	.16962843 0.0121	.1095612 0.2375
edein4	.3592992 0.0001	.21334922 0.0232	.33209111 0.0000
edein5	(omitted)	(omitted)	(omitted)
tuoi	.19913698 0.8501	.08746569 0.9082	.00225443 0.9228
tuoiibp	-.0037592 0.8668	-.00141243 0.9170	-.00020512 0.4484
dantoc	.16410784 0.0565	.2596057 0.0010	.33765564 0.0000
honnhan	-.16575104 0.4444	.01929437 0.9016	.4628075 0.0000
female	-.1829753 0.0001	-.32430809 0.0000	-.47212603 0.0000
idnghe	.10746047 0.0456	.0274784 0.2805	-.04373357 0.0006
_cons	4.5225028 0.7141	5.9324516 0.5753	7.5142971 0.0000
r2	.12908387	.16176874	.17524152
N	463	508	1592

legend: b/p

Phụ lục 14: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels ba nhóm kinh nghiệm của lao động nam 2010

```
. qui:ivreg2h lnincome_m tuoi tuoiibp khuvuvc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61
> & kn==1 & female==0,robust

. est store u_10

. qui:ivreg2h lnincome_m tuoi tuoiibp khuvuvc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61
> & kn==2 & female==0,robust
warning: -ranktest- error in calculating weak identification test statistics;
        may be caused by collinearities
warning: -ranktest- error in calculating weak identification test statistics;
        may be caused by collinearities

. est store u_20

. qui:ivreg2h lnincome_m tuoi tuoiibp khuvuvc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61
> & kn==3 & female==0,robust
Warning: estimated covariance matrix of moment conditions not of full rank.
        overidentification statistic not reported, and standard errors and
        model tests should be interpreted with caution.
Possible causes:
        singleton dummy variable (dummy with one 1 and N-1 0s or vice versa)
partial option may address problem.
Warning: estimated covariance matrix of moment conditions not of full rank.
        overidentification statistic not reported, and standard errors and
        model tests should be interpreted with caution.
Possible causes:
        singleton dummy variable (dummy with one 1 and N-1 0s or vice versa)
partial option may address problem.
Warning: estimated covariance matrix of moment conditions not of full rank.
        overidentification statistic not reported, and standard errors and
        model tests should be interpreted with caution.
Possible causes:
        singleton dummy variable (dummy with one 1 and N-1 0s or vice versa)
partial option may address problem.

. est store u_30

. est table u_10 u_20 u_30 , p stats(r2 N)
```

Variable	u_10	u_20	u_30
edein1	.36321204 0.0987	.49167734 0.0001	.87190633 0.0000
edein2	.34454239 0.0235	.19025844 0.0847	.63548814 0.0000
edein3	.40378653 0.0137	.16613358 0.0428	.15023675 0.0878
edein4	.44661129 0.0018	.22554955 0.0160	.31940163 0.0000
edein5	(omitted)	(omitted)	(omitted)
tuoi	-1.2848541 0.4574	.6661977 0.5511	.15756664 0.0003
tuoiibp	.02692185 0.4622	-.01116055 0.5785	-.0020364 0.0001
khuvuvc	.19172552 0.0106	.23461206 0.0007	.34628115 0.0000
dantoc	.17419143 0.1283	.44113664 0.0322	.3402588 0.0200
honnhan	-.16896161 0.5677	.01770201 0.8381	.5748927 0.0000
female	(omitted)	(omitted)	(omitted)
idnghe	.11758758 0.1527	.05984893 0.0004	.01199112 0.2907
_cons	22.314714 0.2731	-2.8443975 0.8544	4.1396505 0.0000
r2	.08304855	.23630869	.3128624
N	192	233	564

Phụ lục 15: Kết quả ước lượng hàm tiền lương bằng phương pháp Lewbels ba nhóm kinh nghiệm của lao động nữ năm 2010

```
. qui:ivreg2h lnincome_m tuoi tuoiibp khuvuc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61
> & kn==1 & female==1,robust

. est store u_10

. qui:ivreg2h lnincome_m tuoi tuoiibp khuvuc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61
> & kn==2 & female==1,robust
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities
warning: -ranktest- error in calculating weak identification test statistics;
         may be caused by collinearities

. est store u_20

. qui:ivreg2h lnincome_m tuoi tuoiibp khuvuc dantoc honnhan female idnghe (edein1 - edein5= edf1-edf5) if tuoi>=20 & tuoi<=61
> & kn==3 & female==1,robust

. est store u_30

. est table u_10 u_20 u_30 , p stats(r2 N)
```

Variable	u_10	u_20	u_30
edein1	.71587439 0.0001	.55647533 0.0000	.79892187 0.0001
edein3	.22374195 0.5711	.2371254 0.0177	-.11018748 0.6038
edein4	.15330146 0.3800	-.01635571 0.9221	.18035155 0.3041
edein5	(omitted)	(omitted)	(omitted)
edein2	.27995744 0.1040	.43269767 0.0104	.37901383 0.0733
tuoi	.97665753 0.5834	-.20600263 0.8919	-.02202515 0.7782
tuoiibp	-.02051908 0.5870	.0037368 0.8905	.00004854 0.9573
khuvuc	-.02591915 0.7696	.19482391 0.0285	.22395991 0.0633
dantoc	-.3251178 0.0052	-.29662992 0.0534	.12776073 0.6288
honnhan	-.40072501 0.1535	(omitted)	(omitted)
female	(omitted)	(omitted)	(omitted)
idnghe	-.04869585 0.0986	-.06143797 0.0015	.09391149 0.4599
_cons	-3.8346781 0.8546	10.635338 0.6149	7.7875381 0.0000
r2	.20732984	.2474627	.31657287
N	142	123	174

Phụ lục 16: Phương pháp Kernel ước lượng hàm tiền lương năm 2010

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.4540944	2.223357
tuoiBP	20.89704	102.317
honnhan	.098996	.4847084
female	.2741596	1.342352
khuvuc	.2543519	1.245369
dantoc	.1678727	.8219454

Local-linear regression Number of obs = 609
Continuous kernel : epanechnikov E(Kernel obs) = 7
Discrete kernel : liracine R-squared = 0.2634
Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.232489
Effect	
tuoi	-.5150262
tuoiBP	.0116627
honnhan	-.942207
female	-.2139416
khuvuc	.2608918
dantoc	.4874075
educn	
(2 vs 1)	.1046752
(3 vs 1)	.0942646
(4 vs 1)	.1403553
(5 vs 1)	.1965678

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.8639174	5.906198
tuoi bp	47.43213	324.2712
honnhan	.0358388	.2450133
female	.2542602	1.738258
khuvuc	.2512655	1.717784
dantoc	.1573347	1.075623

Local-linear regression Number of obs = 1,195
Continuous kernel : epanechnikov E(Kernel obs) = 18
Discrete kernel : liracine R-squared = 0.2955
Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.438579
Effect	
tuoi	.1822787
tuoi bp	-.0030716
honnhan	0
female	-.33019
khuvuc	.2937963
dantoc	.5227042
educn	
(2 vs 1)	.1195086
(3 vs 1)	.1969675
(4 vs 1)	.2951539
(5 vs 1)	.3618922

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	4.200944	16.07508
tuoi bp	395.4004	1513.015
honnhan	.0120011	.0459228
female	.2292264	.8771439
khuvuc	.2315549	.886054
dantoc	.1267211	.4849032

Local-linear regression Number of obs = 2,922
Continuous kernel : epanechnikov E(Kernel obs) = 392
Discrete kernel : liracine R-squared = 0.3448
Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.31118
Effect	
tuoi	.0447548
tuoi bp	-.0007427
honnhan	0
female	-.4298009
khuvuc	.4532942
dantoc	.4545209
educn	
(2 vs 1)	.1066451
(3 vs 1)	.1599393
(4 vs 1)	.2338292
(5 vs 1)	.2641346

Phụ lục 17: Phương pháp Kernel ước lượng hàm tiền lương khu vực thành thị năm 2010

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.4888707	.6242447
tuoi bp	22.50747	28.74005
honnhan	.084909	.1084214
female	.2946901	.3762933
dantoc	.1114195	.1422728

Local-linear regression Number of obs = 178
 Continuous kernel : epanechnikov E(Kernel obs) = 5
 Discrete kernel : liracine R-squared = 0.2091
 Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.493433
Effect	
tuoi	.7933347
tuoi bp	-.016868
honnhan	0
female	0
dantoc	0
educn	
(2 vs 1)	.1154006
(3 vs 1)	.0685617
(4 vs 1)	.1261848
(5 vs 1)	.2234018

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.8758419	4.354475
tuoibp	48.16538	239.4667
honnhan	.0435907	.2167225
female	.2680103	1.332483
dantoc	.1050119	.522094

Local-linear regression Number of obs = 423
Continuous kernel : epanechnikov E(Kernel obs) = 22
Discrete kernel : liracine R-squared = 0.2922
Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.718445
Effect	
tuoi	.0434337
tuoibp	.0001295
honnhan	0
female	-.2490146
dantoc	.3090393
educn	
(2 vs 1)	.185165
(3 vs 1)	.3404265
(4 vs 1)	.4467942
(5 vs 1)	.6582497

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	4.535159	5.221364
tuoiBP	434.196	499.8933
honnhan	0	0
female	.2395153	.2757558
dantoc	.1134905	.1306625

Local-linear regression Number of obs = 1,069
Continuous kernel : epanechnikov E(Kernel obs) = 0
Discrete kernel : liracine R-squared = .
Bandwidth : cross validation

lnincome_m	Estimate
Mean	
lnincome_m	0
Effect	
tuoi	0
tuoiBP	0
honnhan	0
female	0
dantoc	0
educn	
(2 vs 1)	0
(3 vs 1)	0
(4 vs 1)	0
(5 vs 1)	0

Phụ lục 18: Phương pháp Kernel ước lượng hàm tiền lương khu vực nông thôn năm 2010

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.4412904	4.283439
tuoiBP	20.30307	197.0742
honnhan	.1014262	.984506
female	.2666976	2.588733
dantoc	.1778679	1.726496

Local-linear regression Number of obs = 430
 Continuous kernel : epanechnikov E(Kernel obs) = 19
 Discrete kernel : liracine R-squared = 0.2147
 Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.158235
Effect	
tuoi	-2.447079
tuoiBP	.0539979
honnhan	-.9092806
female	-.2169247
dantoc	.4695333
educn	
(2 vs 1)	.1095043
(3 vs 1)	.1376741
(4 vs 1)	.1856314
(5 vs 1)	.2621372

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.8578849	14.17932
tuoiBP	47.0608	777.8315
honnhan	.0256088	.4232679
female	.2459644	4.065355
dantoc	.1745545	2.885077

Local-linear regression Number of obs = 774
Continuous kernel : epanechnikov E(Kernel obs) = 34
Discrete kernel : liracine R-squared = 0.1902
Bandwidth : cross validation

lnincome_m	Estimate
Mean	
lnincome_m	7.310926
Effect	
tuoi	.4406266
tuoiBP	-.0080444
honnhan	0
female	-.3531458
dantoc	.5153514
educn	
(2 vs 1)	.1373313
(3 vs 1)	.2332884
(4 vs 1)	.3700899
(5 vs 1)	.4168862

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	3.913098	15.27019
tuoi bp	362.3119	1413.86
honnhan	.0148195	.0578304
female	.2193935	.8561452
dantoc	.126815	.4948736

Local-linear regression Number of obs = 1,909
Continuous kernel : epanechnikov E(Kernel obs) = 1,116
Discrete kernel : liracine R-squared = 0.2270
Bandwidth : cross validation

lnincome_m	Estimate
Mean	
lnincome_m	7.159086
Effect	
tuoi	.0299656
tuoi bp	-.0005793
honnhan	0
female	-.4529405
dantoc	.4305497
educn	
(2 vs 1)	.0842974
(3 vs 1)	.1163233
(4 vs 1)	.1591631
(5 vs 1)	.1418366

Phụ lục 19: Phương pháp Kernel ước lượng hàm tiền lương cho lao động nam năm 2010

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.4501159	2.57307
tuoiBP	20.71803	118.4338
honnhan	.0893251	.5106231
khuvuc	.2495416	1.426494
dantoc	.1748301	.9994094

Local-linear regression Number of obs = 362
 Continuous kernel : epanechnikov E(Kernel obs) = 13
 Discrete kernel : liracine R-squared = 0.2302
 Bandwidth : cross validation

lnincome_m	Estimate
Mean	
lnincome_m	7.323008
Effect	
tuoi	-1.884596
tuoiBP	.041992
honnhan	-.6579814
khuvuc	.2437621
dantoc	.5311247
educn	
(2 vs 1)	.0950924
(3 vs 1)	.0705534
(4 vs 1)	.1281725
(5 vs 1)	.2154142

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.8487307	14.01681
tuoi bp	46.57177	769.1343
honnhan	.0367228	.6064776
khuvuc	.2435012	4.02143
dantoc	.1585367	2.618239

Local-linear regression Number of obs = 744
Continuous kernel : epanechnikov E(Kernel obs) = 42
Discrete kernel : liracine R-squared = 0.2881
Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.55063
Effect	
tuoi	.5702072
tuoi bp	-.0099265
honnhan	-.392364
khuvuc	.2159754
dantoc	.624941
educn	
(2 vs 1)	.1340006
(3 vs 1)	.2379524
(4 vs 1)	.3365774
(5 vs 1)	.4605109

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	4.008945	21.02408
tuoi bp	375.9416	1971.547
honnhan	.0144569	.075816
khuvuc	.2224345	1.166511
dantoc	.1235684	.6480287

Local-linear regression Number of obs = 1,952
Continuous kernel : epanechnikov E(Kernel obs) = 1,169
Discrete kernel : liracine R-squared = 0.3018
Bandwidth : cross validation

lnincome_m	Estimate
Mean	
lnincome_m	7.457331
Effect	
tuoi	.0488202
tuoi bp	-.0007653
honnhan	0
khuvuc	.3837527
dantoc	.4829819
educn	
(2 vs 1)	.1358271
(3 vs 1)	.217631
(4 vs 1)	.323916
(5 vs 1)	.3832715

Phụ lục 20: Phương pháp Kernel ước lượng hàm tiền lương cho lao động nữ năm 2010

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.4704638	2.368971
tuoiBP	21.64471	108.9896
honnhan	.1138856	.5734589
khuvuc	.2669301	1.344099
dantoc	.1607723	.8095521

Local-linear regression Number of obs = 261
 Continuous kernel : epanechnikov E(Kernel obs) = 13
 Discrete kernel : liracine R-squared = 0.2647
 Bandwidth : cross validation

lnincome_m	Estimate
Mean lnincome_m	7.12923
Effect	
tuoi	1.711501
tuoiBP	-.0375232
honnhan	-1.26852
khuvuc	.2614104
dantoc	.4080031
educn	
(2 vs 1)	.1157278
(3 vs 1)	.1520055
(4 vs 1)	.1787636
(5 vs 1)	.2015228

Bandwidth

	Mean	Effect
educn	.5	.5
tuoi	.8932525	3.096979
tuoi bp	49.0868	170.1879
honnhan	.0343941	.1192472
khuvuc	.2640989	.9156524
dantoc	.1557767	.5400905

Local-linear regression Number of obs = 465
Continuous kernel : epanechnikov E(Kernel obs) = 29
Discrete kernel : liracine R-squared = 0.2552
Bandwidth : cross validation

lnincome_m	Estimate
Mean	
lnincome_m	7.25639
Effect	
tuoi	-.4209746
tuoi bp	.0077457
honnhan	0
khuvuc	.3992739
dantoc	.3386117
educn	
(2 vs 1)	.1180379
(3 vs 1)	.1973773
(4 vs 1)	.2822896
(5 vs 1)	.2638971

Phụ lục 21: Kiểm tra chất lượng lựa chọn năm 2010

Variable	Mean		%bias	t-test		V(T)/ V(C)
	Treated	Control		t	p> t	
honnhan	0	0*
female	.42269	.41935	0.7	0.14	0.886	1.00
dantoc	.93548	.94438	-2.8	-0.79	0.427	1.15*
khuvuc	.60734	.60289	1.0	0.19	0.847	1.00
suckhoe	.07008	.04783	5.0	2.00	0.045	1.43*
exp	22.769	27.451	-51.4	-11.19	0.000	1.35*
educn	2.2336	2.2336	0.0	0.00	1.000	0.97

* if variance ratio outside [0.88; 1.14]

Ps	R2	LR	chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.056	139.63	0.000	10.1	1.9	57.6*	2.01*	57		

* if B>25%, R outside [0.5; 2]

Phụ lục 22: Kiểm tra chất lượng lựa chọn năm 2014

Variable	Mean		%bias	t-test		V(T)/ V(C)
	Treated	Control		t	p> t	
honnhan	.06218	.04231	9.9	2.15	0.032	1.44*
female	.41537	.42832	-2.6	-0.63	0.528	0.99
dantoc	.92746	.94128	-4.1	-1.34	0.180	1.22*
khuvuc	.59845	.59499	0.8	0.17	0.866	1.00
suckhoe	.07599	.08117	-2.0	-0.46	0.643	0.94
exp	22.128	22.775	-8.1	-2.03	0.043	1.35*
educn	2.1088	2.114	-0.4	-0.08	0.938	0.98

* if variance ratio outside [0.89; 1.12]

Ps	R2	LR	chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.004	11.38	0.123	4.0	2.6	14.0	1.37	43		

* if B>25%, R outside [0.5; 2]