

ƯỚC LƯỢNG SUẤT SINH LỢI CỦA GIÁO DỤC Ở VIỆT NAM GIAI ĐOẠN 2010-2014

BẢNG THỦ TỤC HECKMAN HAI BƯỚC

Lê Đức Hoàng*, Nguyễn Thị Thùy Trang**, Hoàng Thanh Nghị***

Ngày nhận: 25/11/2015

Ngày nhận bản sửa: 20/12/2015

Ngày duyệt đăng: 25/12/2015

Tóm tắt:

Hàm thu nhập của Mincer (1974) được sử dụng rất rộng rãi trong các nghiên cứu về tiền công - tiền lương trên thế giới. Heckman (1979) đã chỉ ra việc hồi quy OLS nếu chỉ tiến hành với số liệu trên những người lao động có việc làm, được trả lương theo việc làm có thể gây ra tính chệch và không vững cho các ước lượng thu được. Một trong các kỹ thuật ước lượng mô hình để khắc phục tính chệch do chọn mẫu có thể gặp trong trường hợp này là sử dụng thủ tục Heckman hai bước. Bài viết này sử dụng thủ tục Heckman 2 bước để ước lượng hàm thu nhập của Mincer với số liệu từ Khảo sát mức sống hộ gia đình Việt Nam năm 2010, 2012 và 2014. Kết quả nghiên cứu cho thấy có sự giảm xuống trong suất sinh lợi của giáo dục ở Việt Nam những năm vừa qua. Bài viết chỉ ra một số nguyên nhân gây ra hiện tượng này và thảo luận về một vài hạn chế khi ước lượng hàm thu nhập của Mincer trong điều kiện số liệu của Việt Nam hiện nay.

Từ khóa: Hàm thu nhập Mincer, thủ tục Heckman 2 bước, suất sinh lợi của giáo dục.

Estimating the returns to education of Vietnam in the period 2010-2014 using 2-stage Heckman procedure

Abstract

Mincer Earnings function has been popularly used to study the relationship between income (salary/wages) and schooling. Heckman (1979) showed that the use of OLS regression only based on income data available from the people who are working could make the biasness and inconsistency to the estimators. One of the estimation methods can limit the biasness coming from the sampling problem is two-stage Heckman procedure. This paper utilizes two-stage Heckman procedure to estimate the Mincer earnings function based on Vietnam Household Living Standard Survey in 2010, 2012 and 2014. The results show that there has been a trend of decrease in the returns on education in Vietnam over the past few years. The paper also identifies some reasons for this situation as well as discusses the limitations when estimating Mincer earnings function in the condition of data available in Vietnam.

Keywords: Mincer earnings function, two-stage Heckman procedure, Returns to education.

1. Giới thiệu

Một trong những mô hình được sử dụng phổ biến trong các nghiên cứu về tiền công tiền lương là hàm thu nhập do Jacob Mincer đề xuất năm 1974. Sau hơn 40 năm được công bố, mặc dù hàm thu nhập do Mincer đề xuất và các dạng mở rộng của nó vẫn còn có một số hạn chế nhưng nó vẫn có vai trò quan

trọng trong việc xác định tác động của việc học tập đến tiền lương, và được sử dụng rất rộng rãi trong các nghiên cứu về tiền lương và thu nhập.

Trong các nghiên cứu sử dụng hàm thu nhập của Mincer, việc ước lượng thường chỉ được tiến hành với những người lao động có việc làm và được trả tiền công tiền lương theo việc làm đó, trong khi ở

trong mẫu tổng thể vẫn có những quan sát có những đặc điểm tương tự như những người có việc làm nhưng họ không đi làm, có thể vì mức lương được trả thấp hơn mức họ kỳ vọng, hoặc vì những lý do khác (tự làm cho công ty gia đình, tham gia hoạt động sản xuất nông nghiệp của gia đình...). Việc chọn mẫu như vậy có thể dẫn tới việc các hệ số ước lượng được bị chệch và không phải ước lượng vững, tương tự như khi mô hình bỏ sót biến hoặc bị định dạng sai. James J. Heckman (1979) đã đưa ra một kỹ thuật để ước lượng mô hình trong các trường hợp này nhằm khắc phục tính chệch của ước lượng do chọn mẫu không ngẫu nhiên, được gọi là thủ tục Heckman 2 bước. Từ nghiên cứu của ông đã hình thành hẳn một thủ tục mới trong thống kê, là một công cụ có ý nghĩa quan trọng đối với các chương trình ước lượng.

Bài viết này tập trung vào việc sử dụng thủ tục Heckman 2 bước để ước lượng suất sinh lợi của việc đi học ở Việt Nam giai đoạn 2010-2014 từ số liệu Khảo sát mức sống hộ gia đình Việt Nam (VHLSS), xem xét xu hướng biến đổi của suất sinh lợi qua các năm, giữa các nhóm đối tượng và thảo luận một số vấn đề liên quan.

2. Cơ sở lý thuyết

Với lập luận cho rằng số tiền công được trả của một người trong hiện tại phụ thuộc vào việc họ đã đầu tư vào vốn con người của bản thân họ trước đó như thế nào, Jacob Mincer (1974) đưa ra phương trình thu nhập (tiền lương/tiền công) trong đó logarit tiền lương là một hàm số của các yếu tố như: số năm đi học, số năm kinh nghiệm và bình phương số năm kinh nghiệm của người lao động. Phương trình của ông được xây dựng dựa trên rất nhiều các giả định và được diễn dịch từ các phương trình toán học khá phức tạp.

Phương trình tiền lương của Mincer có dạng:

$$\ln Y = \alpha + \beta.S + \gamma.T + \theta.T^2 \quad (1)$$

Trong đó Y: Thu nhập ròng của người lao động

S: Số năm đi học của người lao động

T: Số năm kinh nghiệm

Hệ số β chính là suất sinh lợi của việc đi học, cho ta biết phần trăm tăng thêm của thu nhập khi tăng thêm một năm đi học. Hệ số γ giải thích tỷ lệ tăng thêm của thu nhập khi người lao động có thêm một năm kinh nghiệm. Hai hệ số này kỳ vọng mang dấu dương, thể hiện tác động tích cực của học vấn và kinh nghiệm đến thu nhập của người lao động. Hệ

số θ kỳ vọng mang dấu âm, biểu thị mức độ suy giảm của thu nhập biên theo thời gian làm việc.

Dạng tổng quát của phương trình thu nhập Mincer có dạng: $\ln Y_i = X_i \beta + \varepsilon_i$ (2)

Thông thường trong nghiên cứu, người ta tiến hành thu thập số liệu về tiền lương Y_i và các đặc điểm của người lao động X_i (bao gồm các yếu tố như số năm đi học, số năm kinh nghiệm và các biến độc lập khác có tác động đến tiền lương). Sau đó hồi quy Y_i theo X_i để tìm ra ước lượng của hệ số hồi quy β . Số liệu đưa vào hồi quy chỉ bao gồm những người lao động có việc làm và được nhận tiền lương theo việc làm đó.

Heckman trong nghiên cứu công bố năm 1979 đã chỉ ra việc ước lượng hàm tiền lương theo mô hình (2) dựa trên việc chọn mẫu chỉ lấy số liệu ở những người có việc làm và được nhận lương mà bỏ qua những người lao động không tham gia làm công ăn lương có thể làm cho ước lượng bình phương nhỏ nhất (OLS) thu được bị chệch và không vững. Heckman gọi đó là ước lượng chệch do chọn mẫu (Sample Selection Bias). Ông chỉ ra nguyên nhân gây ra ước lượng chệch đó như sau:

Giả sử xét hàm hồi quy: $E^* = Z_i \delta + u_i$ (*), trong đó $E^*_i = Y_i - Y'_i$

Y'_i là mức tiền lương tối thiểu mà người lao động đồng ý để nhận công việc. Người lao động sẽ không đi làm nếu mức lương thấp hơn Y'_i . Ta giả thiết $\varepsilon \approx N(0, \sigma^2_\varepsilon)$, $u \approx N(0, \sigma^2_u)$ với $\sigma^2_u = 1$. Hệ số tương quan giữa u và ε là ρ_{ue} , u , ε độc lập với X , Z .

$$\text{Đặt biến giả } E = \begin{cases} 1 & \text{ khi } E^*_i \geq 0 \\ 0 & \text{ khi } E^*_i < 0 \end{cases}$$

Trong các nghiên cứu, thông thường ta chỉ chọn mẫu ứng với $E^*_i \geq 0$ tức $E=1$

Lấy kỳ vọng có điều kiện ở cả hai vế của phương trình (2), ta có $E(\ln Y_i | E_i = 1) = X_i \beta + E(\varepsilon_i | E_i = 1)$

Do ta chỉ lấy các quan sát có $E_i = 1 \Leftrightarrow E^*_i \geq 0 \Rightarrow$

$$Z_i \delta + u_i \geq 0 \Rightarrow u_i \geq -Z_i \delta$$

suy ra $E(\ln Y_i | E_i = 1) = X_i \beta + E(\varepsilon_i | u_i > -Z_i \delta)$ (3)

Trong trường hợp u và ε không tương quan với nhau ($\rho_{ue} = 0$) thì $E(\varepsilon_i | u_i > -Z_i \delta) = 0$ do đó ước lượng thu được bằng OLS từ phương trình (3) vẫn là ước lượng vững và không chệch. Nếu u và ε có tương quan với nhau ($\rho_{ue} \neq 0$) thì $E(\varepsilon_i | u_i > -Z_i \delta) \neq 0$. Khi đó các ước lượng sẽ bị chệch và không là ước lượng vững nữa, tương tự như việc ta bỏ sót biến độc lập

trong mô hình (hay là chỉ định mô hình sai). Yếu tố $(\varepsilon_i|u_i > -Z_i \delta)$ trong phương trình (3) được xem như một biến tốt bị bỏ sót. Từ đó, Heckman đề xuất khắc phục bằng cách lượng hóa yếu tố bị bỏ sót này và đưa biến này vào mô hình. Theo Greene (2003), $E(\varepsilon_i|u_i > -Z_i \delta)$ có thể được ước lượng thông qua tỉ số Mill nghịch đảo (*Mill's ratio*). Tỷ số Mill nghịch đảo là tỉ lệ giữa hàm mật độ xác suất so với hàm phân phối tích lũy của một phân phối. Ở đây tỉ lệ này được xây dựng dựa trên một tính chất của hàm phân phối chuẩn, rút ra từ một định lý phát biểu trong công trình của Greene (2013): Nếu Y và X là hai biến ngẫu nhiên cùng có phân phối chuẩn với kỳ vọng là μ_Y và μ_X , phương sai là σ_Y^2 và σ_X^2 , hệ số tương quan là ρ thì $E(Y|X > a) = \mu_Y + \rho \cdot \sigma_Y \cdot \lambda(\alpha_X)$

$$\text{trong đó } \alpha_X = \frac{a - \mu_X}{\sigma_X} \text{ và } \lambda(\alpha_X) = \frac{\phi(\alpha_X)}{1 - \Phi(\alpha_X)}$$

Áp dụng công thức trên để tính $E(\varepsilon_i|u_i > -Z_i \delta)$, với $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ đóng vai trò của Y , $u_i \sim N(0, \sigma_u^2 = 1)$ đóng vai trò của X và $a = -Z_i \delta$

$$\text{Khi đó } E(\varepsilon_i | u_i > -Z_i \delta) = \underbrace{\rho_{\varepsilon u} \sigma_\varepsilon}_{\beta_\lambda} \lambda(-Z_i \delta)$$

$$\begin{aligned} &= \underbrace{\rho_{\varepsilon u} \sigma_\varepsilon}_{\beta_\lambda} \lambda(-Z_i \delta) = \beta_\lambda \lambda(-Z_i \delta) = \\ &= \beta_\lambda \underbrace{\frac{\phi(-Z_i \delta)}{1 - \Phi(-Z_i \delta)}}_{\text{Mill's ratio}} \end{aligned} \quad (4) \quad \text{Ta lại có:}$$

$$\rho_{\varepsilon u} = \frac{\text{Cov}(\varepsilon, u)}{\sigma_\varepsilon \sigma_u} \Rightarrow \sigma_\varepsilon \sigma_u \rho_{\varepsilon u} = \text{Cov}(\varepsilon, u) \quad (5)$$

Theo giả thiết ở trên, thay $\sigma_u = 1$ vào (5) suy ra $\text{Cov}(\varepsilon, u) = \sigma_\varepsilon \rho_{\varepsilon u}$

Phương trình hồi quy (3) trở thành $E(\text{Ln}Y_i | E_i = 1) = X_i \beta + \beta_\lambda \lambda(-Z_i \delta)$

$$\text{Hay } \text{Ln}Y_i = X_i \beta + \beta_\lambda \lambda(-Z_i \delta) + v_i \quad (6)$$

Trong đó, v_i là yếu tố ngẫu nhiên thoả mãn các giả thiết của hồi quy cổ điển.

Việc tìm $\lambda(-Z_i \delta)$ và ước lượng không chệch cho β trong phương trình (6) được thực hiện bằng thủ tục hồi quy theo hai bước do Heckman đề xuất như sau:

Bước 1: Ước lượng hồi quy Probit với biến phụ thuộc E được định nghĩa ở phương trình (*) để thu được ước lượng $\hat{\delta}$ của δ . Với mỗi quan sát

$$\text{trong mẫu, tính tỉ lệ } \hat{\lambda}_i = \frac{\phi(Z_i \hat{\delta})}{\Phi(Z_i \hat{\delta})}$$

Bước 2: Hồi quy $\text{Ln}Y_i$ theo X_i và $\hat{\lambda}$ theo phương trình (6) bằng OLS để thu được ước lượng của β cần tìm.

3. Kết quả nghiên cứu và một vài thảo luận

3.1. Số liệu

Trong nghiên cứu này, nhóm tác giả sử dụng số liệu từ cuộc Khảo sát mức sống hộ gia đình Việt Nam do Tổng cục Thống kê thực hiện các năm 2010, 2012, 2014. Đối tượng nghiên cứu là thành viên của các hộ gia đình được điều tra nằm trong độ tuổi lao động, bao gồm 23.762 quan sát năm 2010, 23.322 quan sát năm 2012 và 21.999 năm 2014. Ngoài thông tin cá nhân, các thông tin chính được sử dụng trong phân tích là thông tin về giáo dục, việc làm và thu nhập của các thành viên được quan sát.

3.2. Kết quả ước lượng mô hình

3.2.1. Các biến số trong mô hình

Các biến số sử dụng trong việc ước lượng hàm tiền lương của Mincer được ký hiệu và xác định như sau:

Y: Thu nhập bình quân theo giờ của người lao động, đơn vị tính là 1000 đồng/giờ.

Time: số năm đi học của người lao động. Giá trị của biến Time được tính gần đúng từ thông tin về việc người lao động đã học hết lớp mấy, đã đạt được bằng cấp cao nhất là gì, hiện có đang đi học không và đang học cấp bậc nào.

Exp: số năm kinh nghiệm tiềm năng, nó được tính bằng độ tuổi hiện tại trừ đi tuổi lúc không còn đi học.

Exp2: bình phương của Exp

Gender: Biến giả, nhận giá trị bằng 1 nếu giới tính là nam, bằng 0 nếu là nữ.

Urban: Biến giả, nhận giá trị bằng 1 với quan sát ở khu vực thành thị và bằng 0 nếu quan sát ở nông thôn.

Attend: Biến giả, nhận giá trị bằng 1 nếu quan sát hiện nay có đang đi học, bằng 0 nếu ngược lại.

Married: Biến giả, nhận giá trị bằng 1 nếu đối tượng quan sát hiện tại đang có gia đình, bằng 0 trong các trường hợp ngược lại.

3.2.2. Kết quả ước lượng mô hình Mincer cơ sở bằng OLS

Ta sử dụng hàm thu nhập Mincer cơ sở có dạng sau:

$$\text{Ln}Y = \alpha_1 + \alpha_2 \cdot \text{Time} + \alpha_3 \cdot \text{Exp} + \alpha_4 \cdot \text{Exp}^2 + \alpha_5 \cdot \text{Gender} + \alpha_6 \cdot \text{Urban} + U$$

Mô hình được ước lượng bằng phương pháp OLS với phần mềm Stata 13. Kỹ thuật Robust được sử dụng để khắc phục hiện tượng phương sai của sai số ngẫu nhiên thay đổi. Các quan sát được sử dụng ước lượng là những người trong độ tuổi lao động, có thu nhập từ công việc làm công ăn lương trong 12 tháng qua. Quyền sở cho các quan sát thành viên hộ được sử dụng trong quá trình ước lượng.

độc lập cơ bản của mô hình Mincer ban đầu, nhóm tác giả sử dụng thêm hai biến số đưa vào ước lượng phương trình Selection là *Attend* và *Married* (đã định nghĩa ở trên). Việc một người hiện có đang tham gia học tập ở một loại hình đào tạo nào đó hay không sẽ có ảnh hưởng tới khả năng để người đó tham gia các công việc làm công ăn lương. Tình trạng hôn nhân cũng là yếu tố quan trọng ảnh hưởng

Bảng 1: Kết quả ước lượng mô hình Mincer cơ sở bằng OLS

| Biến độc lập | Năm 2010 | | Năm 2012 | | Năm 2014 | |
|--------------------|----------------|--------|----------------|--------|----------------|--------|
| | Coef | T-Stat | Coef | T-Stat | Coef | T-Stat |
| <i>Time</i> | .096895(***) | 35.27 | .0955541(***) | 39.79 | .1044933(***) | 41.81 |
| <i>Exp</i> | .0351786(***) | 10.14 | .035924(***) | 11.35 | .0523334(***) | 14.89 |
| <i>Exp2</i> | -.0007593(***) | -9.31 | -.0006976(***) | -9.39 | -.0010256(***) | -12.27 |
| <i>Gender</i> | .2179794(***) | 10.05 | .1860804(***) | 9.43 | .1753767(***) | 8.56 |
| <i>Urban</i> | .6496658(***) | 28.52 | .4879556(***) | 24.02 | .4522843(***) | 22.33 |
| <i>C</i> | .367566 (***) | 7.86 | .8420038(***) | 20.15 | .7573467(***) | 16.17 |
| <i>Observation</i> | 9432 | | 9489 | | 9419 | |
| <i>F-statistic</i> | 715.56 | | 687.18 | | 731.17 | |
| <i>Prob>F</i> | 0.0000 | | 0.0000 | | 0.0000 | |
| <i>R-squared</i> | 0.3170 | | 0.2900 | | 0.3081 | |

Nguồn: Tính toán của nhóm tác giả từ số liệu VLSS 2010-2014

Tất cả các hệ số ước lượng của mô hình đều thực sự có ý nghĩa thống kê ở mức 1%. Kết quả ước lượng cho thấy suất sinh lợi của giáo dục trong giai đoạn 2010-2014 tương đối ổn định, trong khoảng xấp xỉ 10%. Tỷ lệ này có giảm nhẹ ở năm 2012 khi so với năm 2010, nhưng lại tăng lên đến 10,44% ở năm 2014, tăng thêm gần 1% so với năm 2012.

3.2.3. Kết quả ước lượng mô hình Mincer cơ sở bằng thủ tục Heckman 2 bước

Trong thủ tục Heckman 2 bước ta sẽ đưa tất cả các quan sát trong độ tuổi lao động vào ước lượng, để xem xét ảnh hưởng của các nhân tố tới khả năng để một người lao động tham gia làm công ăn lương (phương trình Selection- mô hình Probit), từ đó tạo ra biến độc lập đại diện cho phần tác động này đưa thêm vào mô hình Mincer (phương trình Wage) để ước lượng nhằm thu được các ước lượng không chệch cho các hệ số.

Một số biến số có thể xem xét để đưa vào mô hình Probit: số người trong hộ gia đình, số người phụ thuộc trong hộ, thu nhập ngoài tiền lương tiền công của hộ, tình trạng sức khỏe của đối tượng quan sát.... Với điều kiện thực tế của số liệu khảo sát, dựa trên kết quả ước lượng thử mô hình, ngoài các biến

tới khả năng đi làm công làm thuê của người lao động. Những người lao động đã có gia đình thì một mặt, có thể bị hạn chế hơn trong khả năng tham gia làm công ăn lương do các ràng buộc gia đình, nhưng ở góc độ khác, có thể tăng khả năng tham gia hoạt động này, do các áp lực về kinh tế, nuôi sống gia đình.

Tất cả hệ số của các biến độc lập trong phương trình Selection và phương trình Wage đều thực sự khác 0 ở mức 1%. Các hệ số ước lượng bằng thủ tục Heckman hai bước có sự khác biệt so với sử dụng phương pháp OLS cả về độ lớn và cả về dấu.

Suất sinh lợi ước lượng được có xu hướng giảm từ năm 2010 đến năm 2014, tuy nhiên mức chênh lệch giữa năm 2012 với 2014 là không đáng kể. Các lợi suất này đều thấp hơn nhiều so với các ước lượng OLS, đặc biệt ở năm 2012 và 2014. Năm 2014 suất sinh lợi của giáo dục ước lượng được ở đây chỉ là 7,2% so với 10,4% khi ước lượng bằng OLS. Như vậy có thể nói phương pháp ước lượng OLS thông thường như ở trên đã ước lượng cao quá cho các mức lợi suất của giáo dục trong mô hình Mincer giai đoạn này (chệch lên).

Để thấy được rõ nét xu thế biến động của suất

Bảng 2: Kết quả ước lượng mô hình Mincer cơ sở bằng thủ tục Heckman 2 bước

| Biến độc lập | Năm 2010 | | Năm 2012 | | Năm 2014 | |
|---------------------------------------------|--------------------|----------------------------|--------------------|--------------------------|--------------------|---------------------------------------|
| | <i>Wage</i> | <i>Selection</i> | <i>Wage</i> | <i>Selection</i> | <i>Wage</i> | <i>Selection</i> |
| <i>Time</i> | .0946812 (***) | .0452258 (***) | .0723177 (***) | .0350813 (***) | .0720321 (***) | .0335117 (***) |
| <i>Exp</i> | .034616 (***) | -.0166844 (***) | .03001 (***) | -.0164589 (***) | .0383091 (***) | -.0148076 (***) |
| <i>Exp2</i> | -.0007295 (***) | | -.0003957 (***) | | -.000551 (***) | |
| <i>Gender</i> | .1928581 (***) | .4457851 (***) | -.0694236 (***) | .4312686 (***) | -.1171717 (***) | .3958971 (***) |
| <i>Urban</i> | .6431858 (***) | .1579701 (***) | .3642056 (***) | .1747722 (***) | .3676336 (***) | .0315257 (***) |
| <i>Attend</i> | | -1.593475 (***) | | -1.502215 (***) | | -1.198493 (***) |
| <i>Married</i> | | .1051583 (***) | | .0729105 (***) | | .1748839 (***) |
| <i>C</i> | .4740613 (***) | -.492599 (***) | 1.93932 (***) | -.3567362 (***) | 2.133118 (***) | -.3914571 (***) |
| <i>Lambda (λ)</i> | | -.084605 | | -.8776005 | | -1.024506 |
| <i>Prob> χ² (all coefs=0)</i> | 0.0000 | | 0.0000 | | 0.0000 | |
| <i>Selectivity Test (ρ = 0)</i> | | Chi2(1)=46744.46 (0.00) | | Chi2(1)=207.98 (0.00) | | Chi2(1)=3.2*10 ⁶ (0.00) |
| <i>Observations</i> | 9432 | 23762 | 9489 | 23322 | 9419 | 21999 |

Nguồn: Tính toán của nhóm tác giả từ số liệu VLSS 2010-2014

sinh lợi của giáo dục, chúng ta cần có khoảng thời gian nghiên cứu dài hơn, nhưng kết quả ước lượng ở đây cho phép nhận định thực sự có sự giảm xuống trong suất sinh lợi ở năm 2012 so với năm 2010 và đã giảm sút vẫn còn tiếp diễn sang năm 2014 tuy đã có dấu hiệu chững lại.

Kinh nghiệm là yếu tố thực sự ảnh hưởng tới tiền lương, thêm 1 năm kinh nghiệm làm mức lương tăng lên trên 3%, mức tăng tương đối ổn định giữa các năm khảo sát. Dấu hệ số của *Exp2* đều âm, thể hiện sự suy giảm lợi suất biên của tiền lương theo kinh nghiệm tiềm năng. Tiền lương thực sự có sự khác biệt giữa các lao động ở khu vực nông thôn và thành thị, khi các thuộc tính khác đề cập trong mô hình là như nhau. Năm 2012 và 2014, người lao động khu vực thành thị hưởng lương cao hơn những người tương tự nhưng ở khu vực nông thôn trung bình 36%, mức chênh lệch này được ước lượng là 64,31% ở năm 2010. Như vậy vẫn có sự khác biệt rõ ràng về mặt bằng tiền công tiền lương và các cơ hội về công việc - thu nhập giữa khu vực nông thôn và thành thị.

Giới tính của người lao động cũng là yếu tố quyết định đến tiền lương, nhưng chiều hướng của tác động này là không thực sự nhất quán trong giai đoạn 2010-2014. Hệ số của biến *Gender* dương ở năm 2010, thể hiện thu nhập của lao động nam cao hơn

lao động nữ, khi các điều kiện khác tương đồng. Tuy nhiên hệ số này lại âm một cách có ý nghĩa ở năm 2012 và 2014. Như vậy có thể nhận định xét trong toàn bộ giai đoạn 2010-2014 thì ảnh hưởng của giới tính tới tiền lương là không nhất quán và sự thay đổi cấu trúc này có thể bắt nguồn từ những nguyên nhân gây ra sự biến động của suất sinh lợi của giáo dục mà chưa được kiểm soát trong mô hình.

Các kiểm định đối với hệ số tương quan đều cho thấy nó thực sự khác 0, như vậy việc sử dụng thủ tục Heckman 2 bước để ước lượng mô hình Mincer ở đây là thực sự cần thiết.

Hệ số của biến *Time* trong các mô hình Probit (*phương trình Selection*) đều thực sự dương, cho thấy số năm đi học tăng lên thì làm tăng khả năng để tham gia làm công ăn lương.

Để thấy rõ hơn ảnh hưởng của các biến độc lập trong phương trình chọn mẫu tới xác suất tham gia làm công ăn lương, tức là xác suất quan sát được chọn vào ước lượng trong phương trình tiền lương, ta tính tác động biên của các nhân tố này, dựa trên kết quả ước lượng mô hình Probit.

Theo kết quả tính toán trên, kết quả ước lượng mô hình Probit, trong năm 2010, thêm một năm đi học làm tăng khả năng để quan sát tham gia làm công ăn lương thêm hơn 2%, ở năm 2012 và 2014

Bảng 3: Tác động biên của các nhân tố tới xác suất tham gia làm công ăn lương

| Biến độc lập | 2010 | 2012 | 2014 |
|-----------------------------------------------------|-------------|-------------|-------------|
| <i>Time</i> | 0.02077 | 0.014492 | 0.013655 |
| <i>Exp</i> | -0.00766 | -0.0068 | -0.00603 |
| <i>Gender</i> | 0.204725 | 0.178162 | 0.161318 |
| <i>Urban</i> | 0.072547 | 0.0722 | 0.012846 |
| <i>Attend</i> | -0.7318 | -0.62058 | -0.48835 |
| <i>Married</i> | 0.048294 | 0.03012 | 0.071261 |
| <i>Wald χ^2 (6)</i> | 2644.91 | 2935.07 | 2385.85 |
| <i>Prob > χ^2 (all coeffs=0)</i> | 0.0000 | 0.0000 | 0.0000 |
| <i>Prediction of being wage-earners</i> | 0.297846381 | 0.39581612 | 0.418504937 |
| <i>Observations</i> | 23762 | 23322 | 21999 |

Nguồn: Tính toán của nhóm tác giả từ kết quả ước lượng mô hình và số liệu VHLSS 2010-2014

là xấp xỉ 1,4%. Như vậy, đây là một tác động kép của việc đi học, nó làm tăng khả năng để người lao động tham gia làm công ăn lương đồng thời làm tăng mức tiền công tiền lương khi người đó đi làm.

Nếu việc tăng số năm kinh nghiệm làm tăng mức tiền lương thì nó lại làm giảm khả năng để người lao động tham gia vào việc làm công làm thuê, tuy nhiên mức độ giảm là khá nhỏ. Khi có thêm một năm kinh nghiệm tiềm năng, mà ở đây thực chất là người lao động tăng thêm một tuổi thì xác suất làm công ăn lương giảm đi khoảng 0,6-0,7%.

Người lao động nam có khả năng tham gia làm công, làm thuê cao hơn so với lao động nữ. Ở năm 2014 mức chênh lệch này là 16,13%, với các yếu tố khác như nhau. Người lao động ở thành thị dường như cũng có nhiều khả năng để làm công ăn lương hơn so với lao động ở khu vực nông thôn; có thể do có nhiều cơ hội công việc làm công làm thuê hơn, cũng có thể do áp lực mưu sinh phải tham gia các công việc này nhiều hơn.

Khi đối tượng quan sát đang tham gia một chương trình học tập đào tạo nào đó, thì khả năng tham gia các công việc làm công ăn lương giảm đi rõ rệt. Khả năng này giữa những người đang học và những người hiện tại không học chênh lệch nhau đến 73,18% ở năm 2010 và giảm xuống còn 48,835% ở năm 2014.

Theo kết quả ước lượng, những người lao động đang có gia đình thì khả năng tham gia làm công ăn lương cao hơn so với những người chưa có gia đình, từ 3% đến hơn 7%. Điều này có thể do những người đã có gia đình có áp lực nhiều hơn để phải tham gia

làm công làm thuê.

Thông qua ước lượng mô hình Probit, ta cũng dự báo được khả năng để những người lao động có các thuộc tính ở mức trung bình tham gia vào hoạt động làm công làm thuê hưởng lương. Xác suất này là 29,78% ở năm 2010, tăng lên 39,58% ở năm 2012 và đến 41,85% trong năm 2014. Như vậy có thể thấy cơ hội tham gia công việc làm công ăn lương của người lao động đang tăng lên qua các năm, tuy tốc độ tăng có chậm lại ở giai đoạn 2012-2014.

4. Thảo luận

Vấn đề thứ nhất mà nhóm tác giả muốn thảo luận là một số điểm cần quan tâm trong khi ước lượng hàm thu nhập của Mincer và so sánh các kết quả nghiên cứu về vấn đề này trong điều kiện số liệu của Việt Nam hiện nay.

Cơ sở dữ liệu phổ biến nhất có đồng thời và tương đối đầy đủ các thông tin liên quan tới học tập, thu nhập và chi tiêu của các cá nhân để ước lượng hàm thu nhập Mincer là số liệu Khảo sát mức sống hộ gia đình Việt Nam các năm. Tuy nhiên, từ cơ sở dữ liệu này ta không thể xác định được chính xác giá trị của một số biến số quan trọng, chẳng hạn như *Số năm đi học* và *Số năm kinh nghiệm*.

Hệ thống giáo dục nước ta có sự khác biệt về số năm đi học ở các cấp của bậc giáo dục phổ thông, tùy thuộc vào miền địa lý và thời kỳ lịch sử nhưng số liệu từ các cuộc khảo sát mức sống hộ gia đình không có thông tin về sự thay đổi nơi cư trú của cá nhân giữa hai miền Nam - Bắc kể từ khi đất nước thống nhất. Điều này làm giảm tính chính xác khi xác định số năm đi học ở bậc giáo dục phổ thông.

Giáo dục nghề nghiệp của Việt Nam hiện nay có các hình thức *Dạy nghề*, *Trung cấp chuyên nghiệp*, *Cao đẳng nghề*. Các hình thức của giáo dục nghề nghiệp có thời gian đào tạo rất khác nhau giữa các ngành đào tạo, ví dụ *trung cấp chuyên nghiệp* có chương trình đào tạo từ 3 đến 4 năm đối với người có bằng tốt nghiệp trung học cơ sở (THCS), từ một đến hai năm đối với người có bằng tốt nghiệp trung học phổ thông (THPT); *Cao đẳng nghề* có chương trình đào tạo từ hai đến ba năm tùy theo ngành nghề đào tạo.

Giáo dục đại học, gồm các trình độ *Cao đẳng*, *Đại học*, *Thạc sĩ*, *Tiến sĩ*. Thời gian đào tạo ở mỗi trình độ cũng khác biệt rất nhiều ở các khối ngành, các cơ sở đào tạo. Ví dụ đào tạo trình độ *Đại học (ĐH)* được thực hiện từ 4 đến 6 năm học đối với người có bằng tốt nghiệp THPT, đào tạo trình độ *Thạc sĩ* được thực hiện từ 1 đến 2 năm học đối với người có bằng tốt nghiệp đại học, đào tạo trình độ *Tiến sĩ* thực hiện trong 2 đến 3 năm học đối với người có bằng *Thạc sĩ*.

Từ những đặc điểm đó của hệ thống giáo dục Việt Nam, với dữ liệu thu thập được từ các câu hỏi liên quan tới đã học hết lớp mấy, bằng cấp giáo dục phổ thông và bằng cấp giáo dục nghề nghiệp, trình độ chuyên môn kỹ thuật đạt được, ta không thể xác định được chính xác số năm đi học của đối tượng quan sát, do đó biến này sẽ được xác định với cách thức khác nhau ở các nghiên cứu. Điều đó dẫn đến việc ước lượng suất sinh lợi của giáo dục theo hàm Mincer có thể cho kết quả rất khác biệt, cần rất thận trọng trong việc so sánh kết quả của các nghiên cứu với nhau, và càng khập khiễng hơn khi so sánh với kết quả ước lượng ở các nước khác.

Tương tự như vậy đối với biến *Số năm kinh nghiệm*. Trong các năm khảo sát 2004, 2006 và 2008 có câu hỏi về số năm kinh nghiệm nhưng các năm tiếp theo lại không có câu hỏi này nên các tác giả thường tính toán dựa trên giả định của mình là một người sẽ bắt đầu đi làm khi không đi học nữa và gọi đó là biến *Số năm kinh nghiệm tiềm năng*. Biến số năm kinh nghiệm tiềm năng được tính bằng khoảng thời gian từ sau khi không còn đi học cho tới năm khảo sát, với giả định kinh nghiệm là liên tục. Ta thấy rằng cách xác định như vậy trong nhiều trường hợp không phản ánh chính xác về kinh nghiệm trong thực tế mà cá nhân có được, và điều này dẫn tới các ước lượng về vai trò của kinh nghiệm đối với tiền công tiền lương có độ chính xác thấp và khó so sánh kết quả giữa các năm nghiên cứu khác nhau với nhau.

Vấn đề thứ hai mà nhóm tác giả muốn xem xét ở đây, đó là sự sụt giảm suất sinh lợi của giáo dục trong giai đoạn 2010-2014, từ kết quả ước lượng ở trên. Kết quả nghiên cứu của Tinh Doan & John Gibson (2009), sử dụng cùng phương pháp như trong bài viết này với số liệu giai đoạn 1998-2008, chỉ ra một sự tăng lên nhanh chóng trong suất sinh lợi của giáo dục ở toàn bộ giai đoạn, phù hợp với xu thế tăng lên của các nền kinh tế đang trong thời kỳ quá độ: suất sinh lợi tăng từ 3,81% năm 1998 lên đến 10,45% ở năm 2008. Còn theo kết quả nghiên cứu ở đây, suất sinh lợi của giáo dục ước lượng là 9,46% năm 2010, giảm xuống 7,23% trong năm 2012 và 7,203% ở năm 2014.

Đây là một vấn đề cần được quan tâm lí giải để từ đó có những gợi ý về chính sách thỏa đáng. Theo các tác giả, có thể chỉ ra 2 nguyên nhân quan trọng gây ra hiện tượng này.

Nguyên nhân thứ nhất, ảnh hưởng của cuộc khủng hoảng tài chính thế giới năm 2008 cùng với những vấn đề kinh tế nội tại, đã làm kinh tế Việt Nam ngày càng khó khăn, dẫn tới tiền công tiền lương của người lao động bị giảm sút, công ăn việc làm bị cắt giảm.

Giai đoạn 2002-2007, Việt Nam luôn được coi một trong những điểm sáng trong bản đồ kinh tế toàn cầu với tốc độ tăng trưởng bình quân đạt 7,8%. Khủng hoảng tài chính thế giới nổ ra năm 2008 đã ảnh hưởng xấu tới nền kinh tế Việt Nam. Những năm tiếp theo, mức tăng GDP luôn thấp hơn 7% và ngày càng đi xuống, đến năm 2012 chỉ còn 5,25%. Mãi đến năm 2014, kinh tế Việt Nam mới bắt đầu có dấu hiệu thoát đáy khủng hoảng với tốc độ tăng trưởng GDP đạt 5,98%

Khủng hoảng kinh tế tài chính và tình trạng nợ công toàn cầu đã ảnh hưởng tiêu cực đến nền kinh tế Việt Nam, làm cho lạm phát và mặt bằng lãi suất ở mức cao, ảnh hưởng đến sản xuất kinh doanh, thu hút đầu tư giảm, thị trường chứng khoán và thị trường bất động sản giảm sút, tiền đồng giảm giá. Khủng hoảng kinh tế đã làm cho các doanh nghiệp vừa và nhỏ của Việt Nam gặp rất nhiều khó khăn, một số bị từ chối hợp đồng, sản phẩm tiêu thụ chậm, hàng tồn kho ngày càng nhiều. Một số doanh nghiệp khác thì chịu ảnh hưởng của chính sách thắt chặt tiền tệ, hạn chế tăng trưởng tín dụng ngân hàng làm lãi suất cho vay cao vượt xa khả năng của doanh nghiệp. Nợ xấu ngân hàng ngày càng có xu hướng gia tăng. Sự suy yếu của môi trường kinh doanh đi liền với sự đóng băng tín dụng đã buộc hàng vạn

doanh nghiệp rời khỏi thị trường, khiến tỷ lệ thất nghiệp tăng nhanh, tiền công tiền lương bị cắt giảm, một số lượng lớn lao động buộc phải nghỉ việc hoặc làm công việc không phù hợp với chuyên môn kỹ thuật với thu nhập thấp.

Bên cạnh đó, theo nhóm tác giả, còn có một nguyên nhân quan trọng, mang tính chủ quan, dẫn tới sự suy giảm của suất sinh lợi trong giáo dục, đó là sự phát triển mạnh mẽ quá mức của hệ thống giáo dục bậc cao ở Việt Nam trong giai đoạn vừa qua.

Vấn đề này đã từng được đề cập trong một nghiên cứu của Jinhua (2014) về suất sinh lợi của giáo dục ở khu vực thành thị Trung Quốc. Trong nghiên cứu này, khi lí giải cho sự liên tục sụt giảm suất sinh lợi của giáo dục trong giai đoạn 2002 – 2007, tác giả đã chỉ ra một nguyên nhân là do sự mở rộng của hệ thống giáo dục bậc cao. Năm 1999, chính phủ Trung Quốc ban hành một số chính sách cho phép các trường đại học cao đẳng được mở rộng quy mô đào tạo, làm cho ngày càng nhiều học sinh tốt nghiệp THPT có cơ hội được học tiếp lên bậc cao hơn. Tỷ lệ người tốt nghiệp đại học tăng từ 10,7% năm 2002 lên 18,2% năm 2007 (tăng trên 70% trong vòng 5 năm). Nguồn cung về lao động có bằng cấp cao tăng lên nhanh chóng, vượt xa nhu cầu của thị trường lao động, dẫn tới một tỷ lệ lớn sinh viên tốt nghiệp ra trường không có việc làm, hoặc phải làm những công việc không cần chuyên môn kỹ thuật, lương thấp. Từ đó làm suất sinh lợi của giáo dục giai đoạn này liên tục giảm xuống.

Ở Việt Nam, năm 2007, Thủ tướng Chính phủ ra Quyết định quy hoạch mạng lưới các trường đại học, cao đẳng đến năm 2020 nước ta sẽ có 573 trường. Đến tháng 6/2013, Thủ tướng Chính phủ ra Quyết định điều chỉnh quy hoạch mạng lưới các trường đại học, cao đẳng. Theo đó, đến năm 2020 nước ta sẽ có 460 trường đại học, cao đẳng; giảm 100 trường so với quy hoạch năm 2007. Thế nhưng, tính đến tháng 7/2014, cả nước đã có tới 472 trường đại học, cao đẳng. Từ năm 2007 - 2013, đã có 133 trường đại học, cao đẳng được thành lập. Trong đó, có tới 108 trường được nâng cấp từ trung cấp lên cao đẳng hoặc từ cao đẳng lên đại học.

Việc tăng lên nhanh chóng của các trường đại học, cao đẳng với rất nhiều hệ như chính quy, vừa học vừa làm, liên thông, từ xa làm cho học sinh tốt nghiệp phổ thông trung học có rất nhiều cơ hội để vào học cao đẳng đại học, với mức điểm tuyển sinh vào có thể rất thấp. Số lượng sinh viên tăng một cách nhanh chóng, dẫn đến một loạt các vấn đề bất

cập như: cơ sở vật chất, số lượng giảng viên không đủ đáp ứng cho nhu cầu giảng dạy, chất lượng giáo dục thấp, nội dung đào tạo nặng về truyền thụ lý thuyết một chiều, nhẹ về thực hành, chưa chú trọng rèn luyện các kỹ năng mềm (làm việc nhóm, ứng dụng công nghệ thông tin, nâng cao trình độ ngoại ngữ và hoạt động xã hội) cho người học, khả năng cạnh tranh thấp dẫn tới tình trạng sinh viên thất nghiệp sau khi tốt nghiệp ngày càng đông.

Theo tính toán của nhóm tác giả từ số liệu VHLSS 2010-2014 thì tỷ lệ thanh niên trong độ tuổi từ 22-27 tốt nghiệp cao đẳng, đại học và sau đại học tăng nhanh qua các năm: 11,24% năm 2010; 14,52% năm 2012 và 20,16% ở năm 2014; tăng lên gần gấp đôi sau 4 năm, một tốc độ quá nhanh chóng.

Số lao động trình độ đại học, cao đẳng trong độ tuổi lao động thất nghiệp trong thời gian qua tăng cao hơn so với số tốt nghiệp và số có việc làm; số lao động trình độ đại học, cao đẳng thất nghiệp năm 2014 so với 2010 tăng 103%. Theo số liệu của Tổng cục Thống kê, đến quý 3 năm 2015, cả nước có 225.500 thạc sĩ, cử nhân thất nghiệp, tăng thêm 26.000 người so với đầu năm. Tỷ lệ thất nghiệp chung có dấu hiệu giảm nhẹ, song nhóm người có trình độ cao đẳng, đại học trở lên vẫn có tỷ lệ thất nghiệp cao nhất và có xu hướng tăng. Cụ thể, nhóm có trình độ cao đẳng nghề tăng từ 4,8% lên gần 8%; cao đẳng chuyên nghiệp tăng từ 6,8% lên 7,9%; đại học trở lên tăng từ 4,6% lên 4,9%.

Một khảo sát do Trung tâm Nghiên cứu và Phân tích chính sách thuộc Trường đại học Khoa học Xã hội và Nhân văn (Đại học Quốc gia Hà Nội) thực hiện, với quy mô gần 3.000 cựu sinh viên ra trường từ năm 2006-2010 của ba trường đại học: Đại học Quốc gia Hà Nội, Đại học Quốc gia Thành phố Hồ Chí Minh và Đại học Huế, đã cho thấy có đến 26,2% cử nhân cho biết chưa tìm được việc làm, dù khái niệm việc làm ở đây được hiểu rất rộng là bất cứ công việc gì tạo ra thu nhập, không nhất thiết phải đúng với trình độ, chuyên ngành đào tạo. Trong số này, 46,5% cho biết đã đi xin việc nhưng không thành công, 42% lựa chọn giải pháp là tiếp tục học lên hoặc học thêm một chuyên ngành khác; có 27% cử nhân được hỏi cho biết, họ rất khó kiếm việc làm do ngành học của mình không phù hợp với thị trường.

Bên cạnh đó, năng suất lao động ở nước ta còn rất thấp. Mặc dù năng suất lao động của Việt Nam đã liên tục tăng trong thời gian qua, bình quân đạt 3,7%/năm trong giai đoạn 2005 - 2014, góp phần

thu hẹp dần khoảng cách so với năng suất lao động của các nước trong khu vực nhưng hiện nay năng suất lao động Việt Nam chỉ bằng 1/18 năng suất lao động của Singapore; bằng 1/6 của Malaysia; bằng 1/3 của Thái Lan và Trung Quốc.

Các nguyên nhân xuất hiện trong cùng một giai đoạn dẫn tới mức tiền công tiền lương của người lao động nói chung và lao động có bằng cấp nói riêng ở nước ta giảm xuống đáng kể trong những năm qua. Điều đó giải thích cho sự sụt giảm của suất sinh lợi giáo dục trong giai đoạn 2010-2014, và cho thấy kết quả ước lượng được trong nghiên cứu này là hợp lý và đáng tin cậy.

5. Kết luận

Trong bài viết này, các tác giả đã sử dụng thủ tục Heckman hai bước để ước lượng hàm thu nhập của Mincer với số liệu từ Khảo sát mức sống hộ gia đình Việt Nam các năm 2010-2012-2014. Kết quả nghiên cứu cho thấy việc sử dụng kỹ thuật này nhằm khắc phục tính chệch của các ước lượng trong mô hình Mincer khi sử dụng phương pháp OLS với cách lấy mẫu thông thường là thực sự cần thiết. Kỹ thuật này cũng có thể áp dụng rộng rãi cho các mô hình khác

để hạn chế tính chệch của ước lượng khi mẫu được chọn không ngẫu nhiên.

Kết quả ước lượng đã chỉ ra xu thế giảm xuống của suất sinh lợi của giáo dục ở Việt Nam trong những năm gần đây. Đây là một nguy cơ và là thách thức lớn đối với ngành giáo dục đào tạo, đòi hỏi các nhà quản lý phải tìm ra các nguyên nhân và có các giải pháp, chính sách nhanh chóng và quyết liệt để khắc phục, nâng cao chất lượng và hiệu quả của việc giáo dục - đào tạo. Nhóm tác giả cũng chỉ ra hai nguyên nhân chính gây ra hiện tượng này. Một nguyên nhân khách quan đến từ môi trường kinh tế vĩ mô và một nguyên nhân mang tính chủ quan, xuất phát từ chính sách điều hành hệ thống giáo dục bậc cao.

Nhóm tác giả đề cập tới một số khó khăn khi ước lượng hàm thu nhập tiền lương của Mincer trong điều kiện số liệu của Việt Nam. Các khó khăn này làm giảm tính chính xác của kết quả ước lượng và hạn chế khả năng so sánh kết quả giữa các nghiên cứu với nhau. Vấn đề này có thể khắc phục khi chúng ta có được một khảo sát chuyên đề về Giáo dục - Việc làm - Thu nhập, hoặc đơn giản hơn là bổ sung thêm câu hỏi cần thiết khi thực hiện Khảo sát mức sống hộ gia đình Việt Nam ở những năm tiếp theo□.

Tài liệu tham khảo

- Doan, T., & Gibson, J. (2009), 'Do returns to schooling go up during transition? The not so contrary case of Vietnam', *Working papers in Economics No 09/08*, Retrieved from University of Waikato, Department of Economics, <ftp://mngt.waikato.ac.nz/RePEc/wai/econwp/0908.pdf>
- Greene, W.H. (2003), *Econometric Analysis*, 5th. Ed.. Upper Saddle River, NJ.
- Heckman, J.J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica: Journal of the Econometric Society*, 47(1), 153-161.
- JinHan (2014), 'An Empirical Study on Return to Education in Urban China', *Thesis Presented to the Graduate School of Clemson University*.
- Mincer, J. (1974), *Schooling Experience and Earnings*, New York: Columbia University Press.

Thông tin tác giả:

***Lê Đức Hoàng**, Thạc sỹ

- Tổ chức tác giả công tác: Khoa Toán kinh tế, Đại học Kinh tế quốc dân

- Lĩnh vực nghiên cứu chính: Kinh tế lượng ứng dụng, Thống kê xã hội

- Một số tạp chí đã đăng tải công trình nghiên cứu: Tạp chí Kinh tế và phát triển, Tạp chí Quản lý kinh tế, Tạp chí Khoa học và Đào tạo ngân hàng.

- Địa chỉ email: Duchoang08@yahoo.com

****Nguyễn Thị Thùy Trang**, Thạc sỹ

- Tổ chức tác giả công tác: Khoa Toán kinh tế, Đại học Kinh tế quốc dân

- Địa chỉ email: Trang.mfe.neu@gmail.com

*****Hoàng Thanh Nghị**, Thạc sỹ

- Tổ chức tác giả công tác: Trường Cao đẳng Kinh tế - Kỹ thuật Điện Biên

- Địa chỉ email: Hoangthanhnghi_cdb@yahoo.com.vn