
TÁC ĐỘNG CỦA GIÁO DỤC ĐẾN THU NHẬP TẠI VIỆT NAM GIAI ĐOẠN 2014-2020: KẾT QUẢ TỪ MÔ HÌNH HỒI QUY CỘNG TÍNH TỔNG QUÁT GAM

Trịnh Thị Hường

Bộ môn Toán, Trường Đại học Thương mại

Email: tringthi.huong@tmu.edu.vn

Nguyễn Minh Quang

Khoa Toán-Tin, Trường Đại học Sư phạm Hà Nội

Email: mquang12a1tqk@gmail.com

Nguyễn Thị Thanh Loan

Khoa Toán-Tin, Trường Đại học Sư phạm Hà Nội

Email: thanhloan21112000@gmail.com

Phan Văn Đức Nhật

Khoa Toán-Tin, Trường Đại học Sư phạm Hà Nội

Email: phanvanducnhat2000@gmail.com

Nguyễn Thị Ngọc Huyền

Khoa hệ thống thông tin kinh tế và Thương mại điện tử, Trường Đại học Thương mại

Email: ngn.huyen46@gmail.com

Ngô Hoàng Long

Khoa Toán-Tin, Trường Đại học Sư phạm Hà Nội

Email: ngolong@hnue.edu.vn

Mã bài: JED-408

Ngày nhận: 18/09/2021

Ngày nhận bản sửa: 12/01/2022

Ngày duyệt đăng: 11/02/2022

Tóm tắt:

Nghiên cứu này phân tích mối quan hệ giữa giáo dục và thu nhập cá nhân của người lao động tại Việt Nam trong các năm 2014, 2016, 2018 và 2020. Chúng tôi sử dụng dữ liệu cá nhân bao gồm thu nhập bình quân theo giờ, bằng cấp giáo dục cao nhất, số năm đào tạo và các thông tin nhân khẩu học từ các bộ số liệu Điều tra mức sống dân cư của Tổng cục Thống kê Việt Nam. Kết quả từ mô hình hồi quy cộng tính tổng quát (A Generalized Additive Model, GAM) thể hiện mối quan hệ phi tuyến và tích cực giữa số năm đi học và thu nhập theo giờ. Trong đó, lợi tức từ giáo dục của người lao động tăng 1 năm đào tạo ở trình độ cao là lớn hơn so với lợi tức từ tăng 1 năm đào tạo của các cá nhân ở trình độ thấp. Chúng tôi sử dụng biểu đồ xác suất q - q và tiêu chuẩn xác định chéo để kiểm chứng sự phù hợp của mô hình GAM so với mô hình hàm thu nhập Mincer. Kết quả nghiên cứu cho thấy vai trò quan trọng của việc đầu tư cho giáo dục, đặc biệt là đầu tư cho giáo dục ở trình độ cao.

Từ khóa: Giáo dục, thu nhập, hồi quy cộng tính tổng quát, tiêu chuẩn xác định chéo, điều tra mức sống dân cư.

Mã JEL: C30, J24, J31, I21.

Impact of education on income in Vietnam 2014-2020: New insights using GAM model

Abstract

This study analyzes the role of education in Vietnamese employees' wages in 2014, 2016, 2018 and 2020. We use individual hourly wage, their highest qualification as well as the number of years of training, and their demographic characteristics based on the Vietnam Living Standards Surveys, collected by the General Statistics Office. The results from a generalized additive model (GAM) show the nonlinear relationship and significantly positive between years of schooling and hourly earnings. In addition, the average return to education to one more year of schooling at a higher education level is greater than those of lower education levels. Statistical tests, including q - q plots and the cross-validation criterion, show that the GAM model is a significantly better fit than the widely used model in the literature, namely the Mincer earnings function. The results show the important role of education in promoting sustainable development, especially in higher education in Vietnam.

Keywords: Education; income; generalized additive model; cross-validation criterion; Vietnam Household Living Standards Survey.

JEL Codes: C30, J24, J31, I21.

1. Giới thiệu

Một trong 10 nhiệm vụ chủ yếu mà Chính phủ Việt Nam đặt ra trong Nghị quyết số 50/NQ-CP (2021) là “Phát triển nguồn nhân lực, giáo dục và đào tạo đáp ứng yêu cầu nhân lực chất lượng cao của cuộc cách mạng công nghiệp lần thứ tư và hội nhập quốc tế”. Điều đó cho thấy giáo dục và đào tạo được Đảng, Chính phủ xem là một trong những nhiệm vụ quan trọng trong phát triển bền vững, từ đó nâng cao các phúc lợi xã hội, thu nhập cá nhân và phát triển kinh tế. Về chính sách tiền lương, các Bộ ban ngành sửa đổi để phù hợp tình hình thực tế của xã hội như: Bộ luật Lao động 2019, Nghị định 90/2019/NĐ-CP mức lương tối thiểu vùng năm 2020 đối với người lao động.

Ở tầm vi mô, tức cấp độ hộ gia đình hoặc cấp độ cá nhân, nghiên cứu thực nghiệm đã được thực hiện để trả lời các câu hỏi như: Người có trình độ giáo dục cao hơn có mức thu nhập tốt hơn không? Suất sinh lợi từ đầu tư cho giáo dục có khác nhau giữa các cấp học? Lợi suất giáo dục đối với tăng 1 năm đào tạo là khác nhau giữa các cấp học (Doan, 2011; Doan & Tran, 2018; McGuinness & cộng sự (2021)). Trong đó, các tác giả sử dụng mô hình đánh giá lợi suất giáo dục được thực hiện đối với từng cấp học hoặc sử dụng bằng cấp giáo dục như các biến giả. Các mô hình hồi quy phi tuyến được quan tâm vì làm giảm các giả thuyết về ràng buộc giữa các biến số. Phương pháp phi tham số đã được sử dụng gần đây trong các nghiên cứu về kinh tế lao động (Henderson & cộng sự, 2018).

Trong bài báo này, chúng tôi sử dụng mô hình hồi quy cộng tính tổng quát GAM để phân tích bộ số liệu về người lao động làm công ăn lương và làm thuê trên các cuộc Điều tra mức sống Dân cư năm 2014, 2016, 2018 và 2020. Nghiên cứu này nghiên cứu 3 vấn đề sau:

- Phân tích mối quan hệ giữa tiền lương theo giờ của người lao động và giáo dục, theo bằng cấp hoặc số năm đào tạo;

- Phân tích mối quan hệ phi tuyến giữa tiền lương và số năm đào tạo, trong đó chúng tôi giả thuyết độ lớn số năm đào tạo có tác động khác nhau đến tiền lương;

- So sánh tác động của mối quan hệ giữa tiền lương theo giờ và trình độ giáo dục trong các năm gần đây.

Để đánh giá độ phù hợp của mô hình GAM, chúng tôi cũng phân tích bộ số liệu trên theo mô hình hàm thu nhập của Mincer và sau đó sử dụng Phương pháp kiểm định chéo với 10.000 vòng lặp để so sánh kết quả của hai mô hình.

Nghiên cứu của chúng tôi có ba đóng góp: Kết quả nghiên cứu cung cấp bằng chứng về lợi suất giáo dục trong các năm gần nhất cho các nhà hoạch định chính sách. Đồng thời, người lao động cũng có thêm bằng chứng về mức thu nhập kỳ vọng khi quyết định đầu tư cho giáo dục (Patrinos, 2016). Cuối cùng, theo hiểu biết của nhóm tác giả, đây là lần đầu tiên mô hình GAM được ứng dụng trong nghiên cứu lợi suất giáo dục tại Việt Nam.

2. Cơ sở lý thuyết

2.1. Mô hình hồi quy cộng tính tổng quát

Mô hình hồi quy cộng tính tổng quát được giới thiệu lần đầu tiên bởi Hastie và Tibshirani (Hastie & Tibshirani, 1987) vào năm 1987. Mô hình này có dạng:

$$g(\mu_i) = \sum_k \alpha_k X_{ki} + \sum_j s_j(Z_{ji}) \quad (1)$$

trong đó:

Chỉ số i kí hiệu biến quan sát trong mẫu, chỉ số k, j kí hiệu biến độc lập.

$\mu_i \equiv E(Y_i)$ với Y_i là các biến phục thuộc có phân phối thuộc họ mũ (exponential family)

$EF(\mu, \varphi)$, μ_i là tham số trung bình và φ là tham số tỉ lệ.

X_k là các biến độc lập có tác động tuyến tính đến biến phục thuộc và các hệ số chưa biết α_k .

Z_j là các biến độc lập có tác động phi tuyến đến biến phụ thuộc và $s_j(.)$ là hàm trơn.

$g(.)$ là hàm liên kết (a link function), thường sử dụng là lôgarit hoặc hàm đồng nhất (an identity function).

Mô hình (1) được coi là trường hợp mở rộng của mô hình tuyến tính tổng quát (GLM):

$$g(\mu_i) = \sum_k \alpha_k X_{ki} + \sum_j \beta_j Z_{ji} \quad (2)$$

Trong (2), các biến độc lập được giả thuyết có tác động tuyến tính đến biến phụ thuộc Y_i . Biến phụ thuộc cũng được giả sử có phân phối thuộc họ phân phối mũ.

Mô hình (1) có thể được ước lượng bằng phương pháp hợp lý cực đại có phạt (penalized likelihood maximization), trong đó các hàm trơn được ước lượng trong thông qua các Spline và các nút. Các phương pháp ước lượng GAM được phát triển và cải thiện thông qua việc lựa chọn các nút và lựa chọn tham số trơn một cách hợp lý để tránh sự quá khớp (Overfitting) hoặc chưa khớp (Underfitting) (Wood, 2017). Về giải thích ý nghĩa tác động của từng biến độc lập lên biến quan sát được tiến hành tương tự như mô hình hồi quy tuyến tính cổ điển.

Trong phần mềm mã nguồn mở R, gói lệnh *mgvc* (Wood, 2001) và hàm *gam* thực hiện ước lượng mô hình GAM với các lựa chọn khác nhau về hàm phân phối của biến phụ thuộc, hàm liên kết, phương pháp ước lượng các hàm trơn và lựa chọn nút.

Mối quan hệ giữa giáo dục và thu nhập trong nghiên cứu kinh tế có thể có mối quan hệ phi tuyến (Henderson & cộng sự 2018), vì thế, chúng tôi giả thuyết mô hình GAM là phù hợp.

2.2. Tiêu chuẩn xác định chéo

Để so sánh độ phù hợp của hai mô hình ước lượng tuyến tính và phi tuyến, Racine & Parmeter (2012) đề xuất sử dụng tiêu chuẩn xác định chéo (Cross-validation criterion)¹. Phương pháp này được thực hiện như sau:

Chia bộ dữ liệu quan sát (cỡ mẫu n) thành 2 tập dữ liệu một cách ngẫu nhiên và không giao nhau: Tập dữ liệu thứ nhất, kí hiệu là S_T , gồm n_1 quan sát được dùng để ước lượng mô hình hồi quy (the training data) và tập dữ liệu còn lại, kí hiệu là S_E , được dùng để đánh giá mô hình (the evaluation data).

Khi đó, chỉ số thất thoát (the loss criterion) là:

$$L = \frac{1}{n - n_1} \sum_{i \in S_E} (Y_i - \hat{Y}_{i,S_T})^2$$

trong đó, $\hat{Y}_{i,S_T}$ là giá trị ước lượng Y được tính từ kết quả ước lượng mô hình trên mẫu S_T nhưng quan sát thuộc mẫu S_E . Chỉ số thất thoát là trung bình sai số giữa giá trị quan sát và giá trị ước lượng trên tập đánh giá S_E . Chỉ số thất thoát càng nhỏ thì mô hình càng phù hợp.

Trong trường hợp so sánh độ phù hợp của mô hình (1) và mô hình (2), lặp lại 10.000 lần quy trình trên cho mỗi mô hình. Kí hiệu $L_j^{(1)}, j=1,2,...,10.000$ là dãy giá trị thất thoát sau 10.000 lần lặp lại quy trình đối với mô hình (1). Tương tự, $L_j^{(2)}, j=1,2,...,10.000$ là dãy giá trị thất thoát sau 10.000 lần lặp đối với mô hình (2). Khi đó, so sánh độ phù hợp của hai mô hình (1) và mô hình (2) được thực hiện thông qua so sánh hai dãy giá trị thất thoát $L_j^{(1)}$ và $L_j^{(2)}$. Dãy chỉ số thất thoát của mô hình nào nhỏ hơn là phù hợp hơn. Trong bài báo này, chúng tôi sử dụng biểu đồ hình hộp (boxplot) và thống kê t -ghép cặp (t -paired test) để so sánh phân phối thực nghiệm của hai dãy số trung bình thất thoát.

2.3. Một số nghiên cứu thực nghiệm về ảnh hưởng của giáo dục tới thu nhập

Ảnh hưởng của giáo dục tới thu nhập tại Việt Nam được phân tích lần đầu trong (Glewwe & Patrinos, 1998) dựa trên bộ số liệu mức sống Dân cư. Năm 2004, Arcand & cộng sự 2005) đã tìm thấy tác động tích cực của tăng 1 năm đào tạo tới thu nhập thông qua mô hình hồi quy với biến công cụ. Doan (2011), Doan & Tran (2018) nghiên cứu trên bộ số liệu mức sống Dân cư 2008 cho thấy bậc đào tạo đại học trở lên giúp người lao động tăng trung bình 17% thu nhập. Nguyễn Hữu Dũng & Nguyễn Ngọc Thuyết (2015) chỉ ra suất sinh lợi trung bình từ đầu tư giáo dục là 5% và lên tới 12,1% đối với người lao động có bằng đại học và cao hơn. McGuinness & cộng sự (2021) tìm thấy mối quan hệ tuyến tính giữa giáo dục và thu nhập, dẫn tới suất sinh lợi dương của đầu tư giáo dục qua các năm 2002 - 2016. Một hướng tiếp cận khác là các tác giả sử dụng hồi quy phân vị để so sánh bất bình đẳng trong lợi suất sinh lợi từ giáo dục đối với các nhóm đối tượng có mức thu nhập khác nhau (xem Van Vu, 2020).

Bên cạnh các tác động tích cực của giáo dục ở trên, lực lượng lao động của Việt Nam cũng còn tồn tại nhiều vấn đề bất cập như tỉ lệ lao động làm trái nghề và kĩ năng không phù hợp (ILSSA & ILO, 2018).

3. Phương pháp nghiên cứu

3.1. Dữ liệu nghiên cứu

Bộ dữ liệu sử dụng trong nghiên cứu được lấy từ cuộc khảo sát Điều tra mức sống Dân cư Việt Nam giai đoạn 2014 - 2020 được thực hiện bởi Tổng cục Thống kê Việt Nam. Đây là các cuộc khảo sát được tiến hành 02 năm một lần trên phạm vi cả nước bằng phương pháp phỏng vấn trực tiếp chủ hộ (trên 9000 hộ gia đình tham gia phiếu khảo sát về thu nhập, chi tiêu và các chủ đề khác). Kết quả điều tra phục vụ cho các đánh giá về phát triển kinh tế xã hội của đất nước (Tổng cục thống kê, 2019).

Nghiên cứu thực hiện ở mức độ cá nhân và nhóm lao động làm công ăn lương. Trong đó, biến phụ thuộc là trung bình thu nhập theo giờ của người lao động (nghìn đồng/giờ), được thu thập từ tiền công và phúc lợi của người lao động trong vòng 12 tháng trước thời điểm khảo sát.

$$\text{Wage} = \text{Thu nhập (nghìn đồng/giờ)} = \frac{\text{Tổng thu nhập từ lương và phúc lợi trong 1 năm}}{22 \times 12 \times (\text{Trung bình số giờ làm việc/ngày})}$$

Cách tính thu nhập trên tương tự như các nghiên cứu trước đã tiến hành (McGuinness & cộng sự, 2021). Đồng thời, chúng tôi sử dụng giá cố định năm 2020 trong nghiên cứu này (xem <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG>). Cụ thể, 1 đồng năm 2014 (2016 và 2018) tương ứng 1,086088 đồng (1,042789 đồng và 0,972907361 đồng) năm 2020. Từ đó, giá cố định cho phép so sánh mức thu nhập giữa các năm.

Trình độ giáo dục của từng cá nhân, kí hiệu $Yeduc$ được đo như sau:

- Thang đo định tính là bằng cấp giáo dục cao nhất cá nhân đạt được. Đối với hệ thống giáo dục phổ thông và cao đẳng-đại học trở lên bao gồm: không bằng cấp; tiểu học; trung học cơ sở; trung học phổ thông; cao đẳng, đại học và trên đại học. Đối với hệ thống giáo dục nghề nghiệp bao gồm: sơ cấp nghề, trung cấp nghề, trung học chuyên nghiệp và cao đẳng nghề.

- Thang đo định lượng là số năm đào tạo (năm) được quy đổi tương ứng từ các bằng cấp cao nhất ở trên ra số năm đào tạo.

Đối với các thông tin nhân khẩu học khác, chúng tôi trích xuất thông tin tương ứng từ phiếu điều tra cấp hộ và cá nhân. Cụ thể, chúng tôi quan tâm đặc điểm nhân khẩu học như: số thành viên hộ, chỉ số phụ thuộc hộ gia đình, tình trạng hôn nhân, dân tộc, khu vực làm việc, nơi sống. Các tính toán trong nghiên cứu được tiến hành trên phần mềm Rstudio, phiên bản 4.0.2. Gói lệnh được sử dụng là *mgvc*, *tidyverse*, *goft*, *fitdistrplus*, *tableone*, *flexsurv*. Các tính toán thống kê và kết quả hồi quy có sử dụng trọng số điều tra (sampling weights, biến wt9) để các kết quả nghiên cứu phù hợp với thiết kế điều tra của bộ số liệu.

3.2. Mô hình thực nghiệm và phương pháp ước lượng

Đầu tiên, chúng tôi sử dụng mô hình hàm thu nhập Mincer (Mincer, 1974):

$$\text{Log}(\text{Wage}_i) = \beta_0 + \beta_1 Yeduc_i + \sum_j B_j X_{ij} + \varepsilon_i, \quad (3)$$

Trong đó, i là chỉ số cá nhân, ε_i là phần dư và ε_i được giả sử tuân theo phân phối chuẩn $N(0, \sigma^2)$; Wage_i là thu nhập của người thứ i , $Yeduc_i$ là số năm đi học của người thứ i , các biến X_j bao gồm các thông tin nhân khẩu học khác, nơi sống và vùng sinh thái.

Tiếp theo, chúng tôi ước lượng mô hình GAM, trong đó số năm đào tạo $Yeduc$ được giả thuyết có tác động phi tuyến tới tiền lương, các biến độc lập khác được giả sử có tác động tuyến tính:

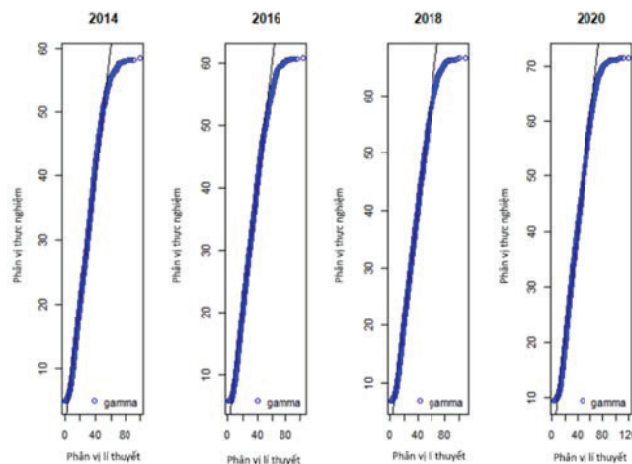
$$\text{Log}(\mu_i) = \beta_0 + s(Yeduc_i) + \sum_j B_j X_{ij} + \varepsilon_i, \quad (4)$$

Trong đó, $s(\cdot)$ là hàm trơn và $\mu_i \equiv E(\text{Wage}_i)$ và $g(\cdot) \equiv \log(\cdot)$ là hàm liên kết. Để xác định dạng phân phối thực nghiệm của biến Wage theo từng năm, chúng tôi tiến hành kiểm định Wage với các phân phối thuộc họ phân phối mũ và quan sát biểu đồ xác suất (q-q plot) của Wage . Kết quả thu được biến thu nhập năm 2014 tuân theo phân phối Gamma (xem biểu đồ q-q, Hình 1).

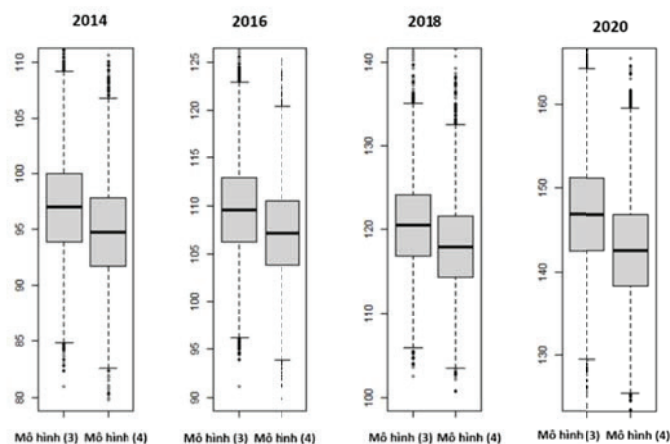
Chúng tôi ước lượng mô hình (3) và (4) cho từng năm. Sau đó, chúng tôi áp dụng tiêu chuẩn xác định chéo

với 10000 vòng lặp cho từng năm, thống kê t -ghép cặp và biểu đồ hình hộp được dùng để so sánh độ phù hợp của mô hình. Hình 2 là biểu đồ hộp của chuỗi trung bình thất thoát qua các năm: mô hình (4) phù hợp với câu hỏi nghiên cứu hơn mô hình (3) trong tất cả các năm. Cuối cùng, kiểm định thống kê t -ghép cặp của năm 2014 được giá trị trung bình của chuỗi thất thoát của mô hình (3) lớn hơn chuỗi thất thoát của mô hình (4) là 220,27 (p -value < 0,001). Kết quả kiểm định của các năm 2016, 2018 và 2020 tương ứng là 160,43 (p -value < 0,001), 141,85 (p -value < 0,001) và 254,8 (p -value < 0,001).

Hình 1: Biểu đồ xác suất (q-q plot) của tiền lương theo giờ



Hình 2: Tiêu chuẩn xác định chéo giữa hai mô hình (3) và mô hình (4)



Nguồn: Tác giả tính toán từ bộ số liệu mức sống Dân cư 2014, 2016, 2018 và 2020.

4. Kết quả nghiên cứu

4.1. Thông tin chung về đối tượng nghiên cứu

Tiền lương trung bình của người lao động làm công ăn lương tăng qua các năm (Bảng 1). Biểu đồ tần số của biến thu nhập (Hình 3) thể hiện sự dịch chuyển của các tần số khi thu nhập tăng thêm. Tần số của các mức thu nhập cao có xu hướng tăng qua các năm, như đối với các mức thu nhập lớn hơn 40.000 đồng/giờ. Đồng thời, độ lệch chuẩn cũng có xu hướng tăng qua các năm, tức là mức độ bất bình đẳng có xu hướng tăng. Độ tuổi trung bình của người lao động là khoảng 36 tuổi. Đa số người lao động sống trong gia đình có khoảng 4-5 thành viên. Chúng tôi thêm vào biến “Chỉ số phụ thuộc hộ gia đình”, được tính bằng tỷ lệ số người ngoài độ tuổi lao động trên số người trong độ tuổi lao động (quy ước từ 15 đến 60 tuổi) trong mỗi gia đình. Chỉ số phụ thuộc là khoảng trên 55% trong các kỳ điều tra, thể hiện mức độ phụ thuộc cao trong các gia

đình Việt Nam. Tỷ lệ người lao động đã kết hôn khoảng 72-73% trong các năm. Để đo lường ảnh hưởng của mức sống tại từng tỉnh, chúng tôi sử dụng biến mức thu nhập bình quân đầu người/tháng cấp tỉnh (<https://www.gso.gov.vn>). Mức thu nhập cấp tỉnh được điều chỉnh về giá của năm 2020 và cho thấy xu hướng tăng theo các năm. Tỷ lệ lao động làm việc cho cơ quan nhà nước giảm dần từ 27% số quan sát năm 2014 xuống 15,83% năm 2020. Trong khi đó số lượng và tỷ lệ lao động tư nhân hoặc làm trong công ty có vốn đầu tư nước ngoài có xu hướng tăng dần từ 9,53% năm 2014 lên 15,28% năm 2020, xu hướng này phù hợp với xu hướng mở rộng của các công ty nước ngoài và doanh nghiệp FDI trong các năm gần đây. Số năm đi học trung bình cả bốn kỳ điều tra đều rơi vào khoảng 9 - 10 năm, tỷ lệ lao động có trình độ từ trung học phổ thông trở xuống chiếm khoảng 80% số quan sát mỗi năm, số lượng và tỷ lệ người lao động có trình độ cao đẳng, đại học và trên đại học tăng dần. Tỷ lệ người lao động đã qua học nghề, đào tạo nghề chỉ chiếm 15,9% số quan sát năm 2014 và 2016, giảm xuống còn 13,7% năm 2018. Trong đó, bằng sơ cấp nghề là phổ biến nhất.

4.2. Thu nhập và bằng cấp giáo dục

Bảng 1: Thông tin chung

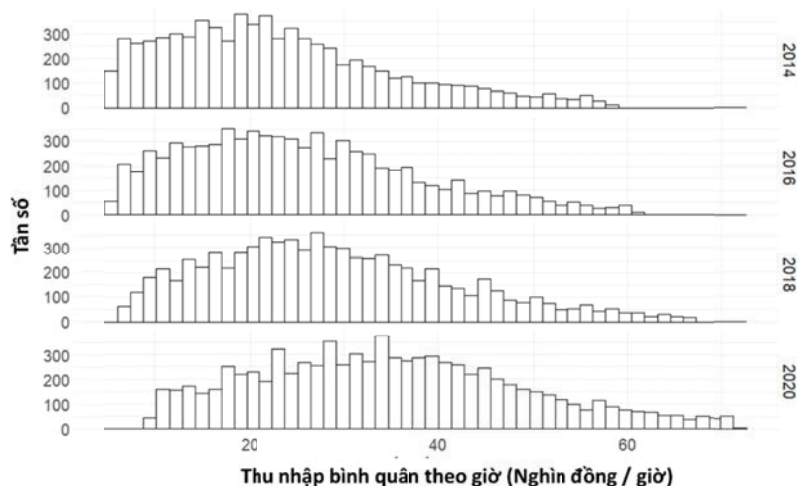
Biến quan sát		Năm 2014	Năm 2016	Năm 2018	Năm 2020
Số quan sát		7174	7548	7671	8266
Tiền lương (Nghìn đồng/giờ)		23.9 (12.2)	26.3 (12.4)	30.2 (13.3)	36.2 (14.1)
Tuổi (năm)		35.6 (11.1)	36.1 (11.3)	36.8 (11.2)	37 (10.9)
Số thành viên hộ gia đình		4.3 (1.5)	4.4 (1.5)	4.3 (1.5)	4.2 (1.6)
Tổng số năm đi học (năm)		9.6 (5)	9.8 (4.9)	9.8 (4.9)	9.2 (5.1)
Chỉ số phụ thuộc hộ gia đình		55.2 (58.3)	57.1 (59.1)	60.4 (64)	62.1 (63.6)
Thu nhập bình quân đầu người/tháng cấp tỉnh		3145.6 (1139.5)	3487.6 (1176)	4049.1 (1334.7)	4490.9 (1363.6)
Giới tính (%)	Nam	58%	57.80%	56.34%	55.20%
	Nữ	42%	42.20%	43.66%	44.80%
Tình trạng hôn nhân (%)	Đã kết hôn	72.50%	71.45%	72.22%	73.72%
	Khác	27.50%	28.55%	27.78%	26.28%
Dân tộc (%)	Kinh	91.67%	89.64%	89.75%	89.45%
	Dân tộc khác	8.33%	10.36%	10.25%	10.55%
Nơi làm việc (%)	Tư nhân	64.36%	66.19%	67.26%	68.89%
	Nhà nước	26.11%	22.84%	19.99%	15.83%
	Công ty có vốn đầu tư nước ngoài	9.53%	10.97%	12.74%	15.28%
Bằng cấp giáo dục cao nhất	Không bằng cấp	10.09%	9.79%	9.46%	13.71%
	Tiểu học	20.40%	19.54%	19.49%	17.42%
	Trung học cơ sở	25.60%	25.97%	26.62%	25.79%
	Trung học phổ thông	24.76%	24.79%	23.88%	28.85%
	Cao đẳng, đại học và trên đại học	19.14%	19.92%	20.55%	14.24%
	Sơ cấp nghề	5.28%	6.34%	5.97%	5.05%
	Trung cấp nghề	3.90%	3.37%	2.75%	2.45%
	Trung học chuyên nghiệp	5.44%	5.05%	4.10%	2.94%
	Cao đẳng nghề	0.99%	1.08%	0.95%	5.06%

Nơi sống (%)	Nông thôn	60.98%	61.24%	59.81%	59.32%
	Thành thị	39.02%	38.76%	40.19%	40.68%
Khu vực (%)	Đồng bằng sông Hồng	27.72%	26.46%	24.10%	26.37%
	Đồng bằng sông Cửu Long	16.90%	16.83%	17.51%	16.54%
	Trung du và miền núi phía Bắc	8.78%	10.37%	9.92%	9.98%
	Bắc Trung Bộ và duyên hải miền Trung	20.48%	20.64%	21.21%	20.02%
	Tây Nguyên	3.96%	4.14%	4.09%	4.02%
	Đông Nam Bộ	22.16%	21.57%	23.17%	23.07%

Chú thích: Biểu liên tục thể hiện giá trị trung bình và độ lệch chuẩn trong ngoặc kép, biến rời rạc thể hiện tần số và tỉ lệ phần trăm của từng mức độ.

Nguồn: Tác giả tính toán từ bộ số liệu mức sống Dân cư 2014 - 2020.

Hình 3: Biểu đồ tần số của thu nhập bình quân theo giờ qua các năm



Nguồn: Tác giả tính toán từ bộ số liệu mức sống Dân cư 2014 - 2020

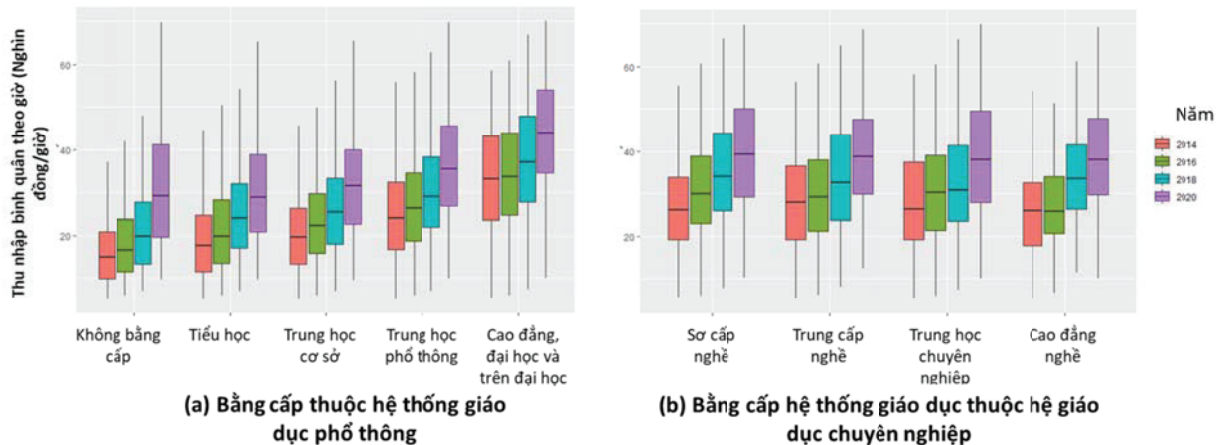
Hình 4 là thu nhập bình quân theo giờ đối với các loại bằng cấp và trong từng năm. Xu hướng tăng thu nhập qua các năm này cũng thể hiện ở tất cả các loại bằng cấp. Xu thế tăng thu nhập theo giờ có thể do tác động trực tiếp từ các chính sách của chính phủ Việt Nam về mức lương tối thiểu theo vùng, lộ trình tăng lương cơ bản (Nguyen Viet, 2017) hoặc tác động của các doanh nghiệp FDI (McGuinness & cộng sự, 2021). Đối với người lao động sở hữu bằng cấp thuộc hệ giáo dục chuyên nghiệp, Hình 4 cho thấy xu hướng thu nhập giữa các bằng cấp khác nhau là khá tương đồng. Đối với bằng cấp thuộc hệ thống giáo dục phổ thông và cao đẳng đại học: Các bằng cấp cao hơn thì có mức thu nhập theo xu hướng cao hơn. Đặc biệt, có sự tăng đáng kể giữa những cá nhân sở hữu bằng cấp cao đẳng và đại học so với các cấp học thấp hơn và có sự bất bình đẳng thu nhập gia tăng ở nhóm lao động này, như các nghiên cứu trước đó.

4.3. Kết quả mô hình hồi quy

Bảng 2 thể hiện kết quả hồi quy mô hình log-tuyến tính. Hệ số hồi quy của “Tổng số năm đi học” chính là lợi suất giáo dục khi tăng 1 năm đào tạo. Tăng 1 năm đi học mang lại mức thu nhập bình quân theo giờ tăng thêm tương ứng là 3,28% (tương tự, 2,9%, 2,9% và 2,2 %) trong năm 2014 (tương tự 2016, 2018 và 2020) và các hệ số đều có ý nghĩa thống kê. Kết quả chúng tôi thu được là thấp hơn các nghiên cứu đã thực hiện trước đó khi nghiên cứu ở cấp độ hộ gia đình (Anh & cộng sự, 2020).

Bảng 3 là kết quả hồi quy đối với các biến độc lập được giả thuyết tác động tuyến tính đến thu nhập theo giờ, kết quả ước lượng hàm tron được thể hiện ở Hình 5. Trong cả 4 kỳ điều tra, các biến quan sát đều có

Hình 4: Biểu đồ hộp của thu nhập theo giờ



Nguồn: Tác giả tính toán từ bộ số liệu mức sống Dân cư 2014 – 2020.

Bảng 2: Hệ số hồi quy mô hình tuyến tính

Biến quan sát	Hệ số hồi quy			
	Năm 2014	Năm 2016	Năm 2018	Năm 2020
Tổng số năm đi học (năm)	0,038 *** (0,001)	0,029 *** (0,001)	0,029 *** (0,001)	0,022 *** (0,001)
R^2	0,328	0,295	0,296	0,279

Chú thích: *, ** và *** lần lượt biểu diễn các mức ý nghĩa 10%, 5% và 1%.

Nguồn: Tác giả tính toán từ bộ số liệu mức sống Dân cư 2014 – 2020.

tác động có ý nghĩa thống kê đến thu nhập. Hệ số R^2 hiệu chỉnh cho thấy các biến phụ thuộc giải thích được khoảng 30-36% mức thu nhập. Mô hình 4 sử dụng hàm liên kết là logarit cơ sở tự nhiên, nên hệ số hồi quy của mô hình là phần trăm thay đổi của từng nhân tố tác động (sau khi nhân hệ số hồi quy với 100%) và giữa các yếu tố khác không đổi. Người lao động đang kết hôn có thu nhập cao hơn từ 11% so với lao động có các tình trạng hôn nhân khác. So với lao động dân tộc Kinh, lao động người dân tộc khác có thu nhập thấp hơn lên đến 13,9 % và 15,8 % trong năm 2018 và 2020. Tương tự, lao động nữ có thu nhập thấp hơn khoảng 11% so với lao động nam giới năm 2014 và có xu hướng thấp hơn trong các năm tiếp theo, lên đến 14,6% năm 2020. Tương tự, lao động ở khu vực nông thôn có thu nhập thấp hơn lao động ở khu vực thành thị là 11,9% năm 2014 và xu hướng bất bình đẳng có giảm vào năm 2020 (3,7%). Các nhân tố về dân tộc, giới tính và khu vực nông thôn thành thị là các yếu tố dẫn đến bất bình đẳng thu nhập tại Việt Nam và đã được nhiều nghiên cứu trước chỉ ra (Benjamin & cộng sự, 2017). Tuổi của người lao động có tác động dương và có ý nghĩa thống kê đến thu nhập theo giờ nhưng mức tác động rất nhỏ. So với người lao động làm việc trong lĩnh vực tư nhân, người lao động làm trong công ty nhà nước và các công ty có vốn đầu tư nước ngoài có mức lương cao hơn. Đặc biệt, khu vực có vốn đầu tư nước ngoài có mức thu nhập cao hơn lên tới khoảng 20% so với khu vực tư nhân. Khu vực có vốn đầu tư nước ngoài được kỳ vọng có mức thu nhập cao hơn, tuy nhiên, khu vực này cũng có yêu cầu cao hơn về trình độ và kỹ thuật của người lao động. Hiện nay, lao động Việt Nam làm công ăn lương trong các doanh nghiệp này vẫn thiếu và yếu về nhu cầu (ILSSA & ILO, 2018; McGuinness & cộng sự, 2021). So với lao động tại khu vực đồng bằng sông Hồng, các khu vực khác đều có mức thu nhập bình quân thấp hơn, 1% đến 10%, trừ khu vực Đông Nam Bộ. Đồng bằng sông Hồng là một trong các vùng kinh tế phát triển nhất cả nước, có lợi suất giáo dục cao trong nghiên cứu. Đồng thời, khu vực đồng bằng sông Hồng cũng có mức đầu tư có giáo dục khá cao trong năm 2018 (Ngô & cộng sự, 2021). Ngược lại, khu vực Đông Nam Bộ, nơi có nền kinh tế lớn nhất đất nước là thành phố Hồ Chí Minh, người lao động có mức thu nhập trung bình cao hơn người lao động tại Đồng bằng sông Hồng khoảng 10- 13%.

Hình 5 mô phỏng thu nhập qua các năm của các cá nhân có đặc điểm phổ biến nhất trong các năm quan

Bảng 3: Hệ số hồi quy mô hình GAM (4)

	Hệ số hồi quy			
Biến quan sát	Năm 2014	Năm 2016	Năm 2018	Năm 2020
Hệ số chặn	2,642*** (0,039)	2,947*** (0,039)	3,047*** (0,035)	3,262*** (0,033)
Tuổi (năm)	0,003*** (0,001)	0,001 (0,00)	0,001* (0,00)	0,00 (0,00)
Chỉ số phụ thuộc hộ gia đình (%)	0,001*** (0,00)	0,00** (0,00)	0,00*** (0,00)	0,00*** (0,00)
Thu nhập bình quân đầu người/tháng cấp tỉnh	0,00*** (0,00)	0,00*** (0,00)	0,00*** (0,00)	0,00*** (0,00)
Giới tính (mặc định: Nam)				
Nữ	-0,121*** (0,011)	-0,114*** (0,01)	-0,141*** (0,009)	-0,146*** (0,008)
Tình trạng hôn nhân (mặc định: Khác)				
Đã kết hôn	0,128*** (0,013)	0,11*** (0,012)	0,103*** (0,011)	0,105*** (0,01)
Dân tộc (mặc định: Kinh)				
Dân tộc khác	-0,082*** (0,019)	-0,141*** (0,017)	-0,139*** (0,015)	-0,158*** (0,013)
Khu vực làm việc (mặc định: Tự nhân)				
Cơ quan nhà nước	0,088*** (0,014)	0,066*** (0,014)	0,039** (0,013)	0,05*** (0,012)
Công ty có vốn đầu tư nước ngoài	0,217*** (0,018)	0,233*** (0,016)	0,211*** (0,014)	0,173*** (0,011)
Nơi sống (mặc định: Thành thị)				
Nông thôn	-0,119*** (0,012)	-0,133*** (0,011)	-0,092*** (0,01)	-0,037*** (0,008)
Khu vực (mặc định: đồng bằng sông Hồng)				
Đồng bằng sông Cửu Long	0,027 (0,018)	-0,029 (0,017)	-0,013 (0,015)	-0,01 (0,014)
Trung du và miền núi phía Bắc	0,105*** (0,024)	0,067** (0,022)	0,034 (0,02)	0,003 (0,018)
Bắc Trung Bộ và duyên hải miền Trung	0,003 (0,018)	-0,061*** (0,018)	-0,042** (0,015)	-0,051*** (0,014)
Tây Nguyên	0,043 (0,03)	-0,001596	-0,067** (0,025)	-0,101*** (0,023)
Đông Nam Bộ	0,134*** (0,017)	0,095*** (0,015)	0,12*** (0,014)	0,117*** (0,013)
Hệ số	0,36	0,309	0,312	0,29

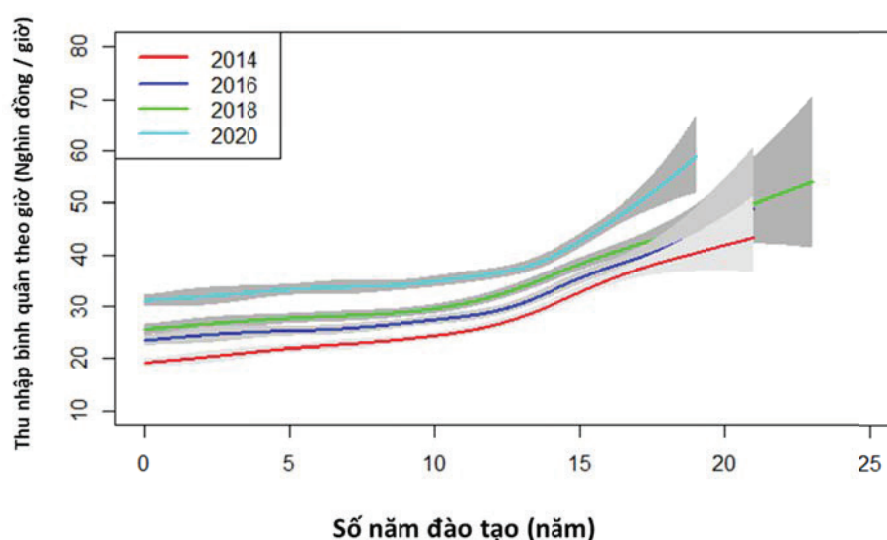
Chú thích: *, ** và *** lần lượt biểu diễn các mức ý nghĩa 10%, 5% và 1%.

Nguồn: Tác giả tính toán từ bộ số liệu Điều tra mức sống dân cư 2014, 2016 và 2018, 2020.

sát, cụ thể²: thuộc khu vực thành thị của Đồng Bằng sông Hồng, có kết hôn, dân tộc Kinh, là nam giới, tỉ lệ phụ thuộc của gia đình là 60% và làm việc trong công ty tư nhân. Đường màu là giá trị trung bình và dải màu xám là độ dao động (mức ý nghĩa 95%). Trên đồ thị, chúng tôi đã tính tiến hàm trơn $s()$ để đưa về giá trị của thu nhập (giá trị trên trục O_y).

Mô Hình 4 có kết quả hồi quy hàm trơn $s(Y_{educ})$ có mức ý nghĩa p-value < 0,001 trong tất cả các năm và xu hướng tác động của số năm đào tạo không thay đổi đáng kể. Trong khi mô hình (3), Bảng 2 là lợi suất của tăng 1 năm đào tạo đối với thu nhập, không phân biệt mức tăng này ở các trình độ khác nhau: Ví dụ, lợi suất tăng 1 năm đào tạo từ 3 năm lên 4 năm giống như tăng 1 năm đào tạo từ 12 năm lên 13 năm. Ưu điểm của mô hình GAM được thể hiện ở sự thay đổi giá của số năm đào tạo (trục O_x) đối với thu nhập. Đối với lao động có từ 1 đến khoảng 10 năm đào tạo, lợi suất của đầu tư giáo dục có xu hướng dương và tăng đều (có xu hướng tuyến tính). Sau khoảng 10 năm đào tạo, lợi suất của đầu tư giáo dục có tăng nhanh hơn so với giai đoạn trước (thể hiện ở rõ xu hướng phi tuyến và khác biệt so với giai đoạn trước). Đối với lao động có số năm đào tạo trên 18 năm, lợi suất thu nhập tiếp tục tăng nhưng mức thu nhập theo giờ có sự phân hóa khá lớn.

Hình 5: Mối quan hệ giữa thu nhập và số năm đào tạo từ mô hình GAM



Nguồn: Tác giả tính toán từ bộ số liệu mức sống Dân cư 2014 – 2020.

5. Kết luận

Nghiên cứu phân tích ảnh hưởng của giáo dục tới thu nhập theo giờ của người lao động làm công ăn lương tại Việt Nam trong các năm 2014, 2016, 2018 và 2020 trên bộ số liệu mức sống Dân cư. Sử dụng một số công cụ thống kê hiện đại, chúng tôi tìm thấy mối quan hệ phi tuyến giữa số năm đào tạo và thu nhập. Lợi suất của đầu tư giáo dục dương và tương đối ổn định đối với lao động có dưới 10 năm đào tạo, sau đó lợi suất tăng cao hơn. Đặc biệt, đối với lao động có số năm đào tạo cao (trên 18 năm) thì thu nhập có xu hướng bất bình đẳng nhiều hơn. Các yếu tố khác có tác động có ý nghĩa thống kê đến thu nhập như giới tính, nơi sống, dân tộc. Nghiên cứu của chúng tôi chỉ ra lao động làm công ăn lương được đào tạo có xu hướng tăng trong các năm gần đây, đặc biệt là lao động có trình độ cao đẳng và đại học trở lên. Tuy nhiên, hiện trạng khó tìm việc làm sau quá trình đào tạo, thiếu kỹ năng đáp ứng yêu cầu công việc trong bối cảnh cách mạng công nghiệp 4.0 cũng đưa đến nhiều rủi ro cho người đầu tư giáo dục (Hà & cộng sự, 2018).

Nghiên cứu của chúng tôi ủng hộ và cung cấp thêm các bằng chứng khoa học phục vụ các chính sách phát triển giáo dục, nguồn nhân lực và quá trình hội nhập quốc tế của Việt Nam. Giáo dục có tác động tích cực đến thu nhập người lao động. Đặc biệt, giáo dục trình độ cao cần nên được đầu tư vì lợi ích của cả nền kinh tế và của người lao động. Chính sách tiền lương đối với lao động có số năm đào tạo lớn và chất lượng cao cũng cần cải thiện để người lao động yên tâm công tác.

Nghiên cứu của chúng tôi có một số hạn chế do sử dụng số liệu có sẵn (số liệu thứ cấp). Hệ số R^2 hiệu chỉnh có xu hướng giảm qua các năm, cho thấy ngày càng thêm các nhân tố có thể tác động đến thu nhập mà nghiên cứu của chúng tôi không quan sát được. Về các biến ảnh hưởng đến thu nhập có thể bị thiếu do các biến về kinh nghiệm, tính chất công việc (bán thời gian, thời vụ, toàn thời gian) hay các kỹ năng của người lao động (như trình độ ngoại ngữ) chưa được đưa vào mô hình. Chúng tôi chưa đánh giá được tác động của giáo dục tới lao động trong lĩnh vực nông nghiệp và tự sản xuất kinh doanh dịch vụ, hoặc đối với các nhóm thu nhập thấp. Đây là các đối tượng dễ bị tổn thương hơn trong bối cảnh hội nhập và đặc biệt là dịch bệnh COVID-19 trong các năm gần đây. Trong đó, các nghiên cứu về cực trị (như mô hình hồi quy phân vị) sẽ cung cấp thêm bằng chứng đối với các nhóm thu nhập khác nhau. Về giả thuyết mô hình, do sử dụng số liệu chéo nên kết quả nghiên cứu có thể bị chệch do ảnh hưởng của vấn đề nội sinh. Từ đó, các nghiên cứu tiếp theo sử dụng biến công cụ trên số liệu mảng có thể cải thiện được các ước lượng. Cuối cùng, nghiên cứu của chúng tôi có đóng góp về sử dụng công cụ thống kê mới, tức mô hình GAM, tuy nhiên theo hiểu biết của tác giả, một số kỹ thuật thống kê sử dụng đối với mô hình tuyến tính chưa được nghiên cứu đối với mô hình GAM, như lựa chọn biến thông qua phương pháp Heckman Selection (Puhani, 2000). Ước lượng mô hình GAM đã tích hợp trọng số điều tra (sampling weights) nhưng chưa nghiên cứu các đặc điểm phức tạp

của các điều tra phức tạp, như strata hay độ lệch tiêu chuẩn cụm (Cluster standard errors). Mô hình GAM có ưu điểm là đưa ra bằng chứng phi tuyến về mối quan hệ trong lợi suất giáo dục và chúng tôi tìm được đáng điều của hàm phi tuyến, tuy nhiên các phân tích tìm ra tác động tăng giáo dục thêm 1 năm tại từng số năm đào tạo cụ thể của người lao động thì không trực quan như mô hình log-tuyến tính và cần thêm các tính toán cụ thể. Phương pháp ước lượng của mô hình GAM và các hàm trơn được sử dụng cũng phức tạp và đòi hỏi nhiều kỹ thuật tính toán hơn so với ước lượng bằng phương pháp bình phương tối thiểu. Tiêu chuẩn chéo sử dụng để so sánh ước lượng của mô hình GAM và mô hình tuyến tính cũng tốn nhiều chi phí về thời gian tính toán do số lần lặp lớn. Các nghiên cứu bổ sung về tác động của giáo dục đối với các đối tượng trên là cần thiết để có các chính sách hợp lý về đào tạo và phát triển nguồn nhân lực, đảm bảo công bằng thu nhập và công bằng xã hội.

Ghi chú:

1. Tác giả sử dụng cụm từ trong bài báo là “revealed performance test”, tuy nhiên, tác giả cũng bình luận bản chất của phương pháp này là Cross-validation. Do đó, chúng tôi sử dụng cụm Cross-validation trong bài báo này vì thuật ngữ này quen thuộc hơn.

2. Phân tích tác động của mô hình GAM theo quy ước thông thường, tức là phải giữ các nhân tố khác không thay đổi. Trong trường hợp thay đổi đặc điểm cá thể, đồ thị hàm trơn sẽ tịnh tiến theo trục Oy nhưng hình dáng hàm sẽ không thay đổi.

Tài liệu tham khảo

- Anh, T., Tran Quang, T., Tran The, N., & Nguyen Thi, H. (2020), ‘The role of education in the livelihood of households in the Northwest region, Vietnam’, *Educational Research for Policy and Practice*, 19(1), 63–88. <https://doi.org/10.1007/s10671-018-9242-6>
- Arcand, J. L., D’Hombres, B., & Gyselinck, P. (2005), ‘Instrument choice and the returns to education : New evidence from Vietnam’, *Revue Economique*, 56(3), 563–572. <https://doi.org/10.3917/reco.563.0563>
- Benjamin, D., Brandt, L., & McCaig, B. (2017), ‘Growth with equity: income inequality in Vietnam, 2002–14’, *Journal of Economic Inequality*, 15(1), 25–46. <https://doi.org/10.1007/s10888-016-9341-7>
- Doan, T. (2011), ‘Labour Market Returns to Higher Education in Vietnam’, *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.1812533>.
- Doan, T., Le, Q., & Tran, T. Q. (2018), ‘Lost in transition? Declining returns to education in Vietnam’, *The European Journal of Development Research*, 30(2), 195–216.
- Glewwe, P., & Patrinos, H. A. (1998), *The role of the private sector in education in Vietnam*, <https://doi.org/10.1596/0-8213-4167-7>
- Hà, T. T. T., & Hương, N. T. L. (2018), ‘Giáo dục-đào tạo với thị trường lao động trong bối cảnh Cách mạng công nghiệp lần thứ tư’, *Tạp chí khoa học giáo dục Việt Nam*, Số 01, tháng 01/2018, http://vjes.vnies.edu.vn/sites/default/files/bai_so_7_-_so_1.thang_01_2018.pdf
- Hastie, T., & Tibshirani, R. (1987), ‘Generalized additive models: Some applications’, *Journal of the American Statistical Association*, 82(398), 371–386. <https://doi.org/10.1080/01621459.1987.10478440>
- Henderson, D. J., & Souto, A. C. (2018), ‘An introduction to nonparametric regression for labor economists’, *Journal of Labor Research*, 39(4), 355–382. <https://doi.org/10.1201/9781315370279>
- ILSSA & ILO. (2018), *Labour and Social Trends in Viet Nam 2012–2017*.
- McGuinness, S., Kelly, E., Pham, T. T. P., Ha, T. T. T., & Whelan, A. (2021), ‘Returns to education in Vietnam: A changing landscape’, *World Development*, 138, 105205.
- Mincer, J. (1974), *Schooling, Experience, and Earnings*, Human Behavior and Social Institutions. ERIC.
- Ngô, T. N., Đàm, T. T. T., Nguyễn, T. T. M., & Trịnh, T. H. (2021), ‘Nghiên cứu các yếu tố ảnh hưởng đến việc chi tiêu giáo dục của hộ gia đình tại các tỉnh đồng bằng sông Hồng’, *TNU Journal of Science and Technology*, 226(04),

- Nguyễn Hữu Dũng & Nguyễn Ngọc Thuyết (2015), ‘Suất sinh lợi từ đầu tư cho giáo dục tại Việt Nam. Tạp Chí Phát Triển Kinh Tế Châu Á’, 26(5), 60–75.
- Nguyen Viet, C. (2017), ‘Do minimum wages affect firms’ labor and capital? Evidence from Vietnam’, *Journal of the Asia Pacific Economy*, 2(22), 291–308.
- Patrinos, H. A. (2016), *Estimating the return to schooling using the Mincer equation*, IZA World of Labor.
- Puhani, P. (2000), ‘The Heckman correction for sample selection and its critique’, *Journal of economic surveys*, 14(1), 53–68.
- Racine, J. S., & Parmeter, C. F. (2012), *Data-driven model evaluation: a test for revealed performance*, Citeseer.
- Tổng cục thống kê (2019), *Kết quả Khảo sát mức sống dân cư Việt Nam năm 2018*, NXB thống kê.
- Van Vu, H. (2020), ‘The impact of education on household income in rural Vietnam’, *International Journal of Financial Studies*, 8(1), <https://doi.org/10.3390/ijfs8010011>
- Wood, S. N. (2001), ‘mgcv: GAMs and generalized ridge regression for R’, *R News*, 1(2), 20–25.
- Wood, S. N. (2017), ‘Generalized additive models: An introduction with R, second edition’, In *Generalized Additive Models: An Introduction with R, Second Edition*,