

MỤC LỤC

KINH TẾ VÀ QUẢN LÝ

- 1. Nguyễn Hoàng Việt và Phan Đình Quyết** - Logistics đáp ứng nhu cầu xuất nhập khẩu qua các cửa khẩu tỉnh Cao Bằng: Thực trạng và giải pháp. **Mã số: 120.1IEM.11** 2
Logistics for Import and Export through Border Gates in Cao Bang Province: Current Situation and Recommendations
- 2. Tạ Huy Hùng** - Giải pháp hoàn thiện đánh giá cán bộ lãnh đạo sở, ngành tại tỉnh Hòa Bình. **Mã số: 120.1HRMg.11** 13
Solutions to Complete Appraisal of Department Leaders in Hoa Binh Province
- 3. Lê Văn Tuấn, Michel Simioni và Trịnh Thị Hương** - Ứng dụng phương pháp phân rã dựa trên COPULA để nghiên cứu bất bình đẳng thu nhập giữa nông thôn và thành thị của Việt Nam. **Mã số: 120.1MEIS.11** 23
Applying COPULA-based Factorization in Studying Income Inequality between Rural and Urban Areas in Vietnam

QUẢN TRỊ KINH DOANH

- 4. Lưu Thị Minh Ngọc, Vũ Tiến Hiệu và Nguyễn Tư Lương** - Nghiên cứu sự hài lòng của khách hàng khi sử dụng nhà ở của Tổng Công ty Đầu tư phát triển nhà và đô thị (HUD). **Mã số: 120.2BMkt.22** 32
A Study on Customer Satisfaction with Housing Use of Housing and Urban Development Corporation (HUD).
- 5. Lê Thị Tú Anh, Nguyễn Thu Hà và Phan Chí Anh** - Quan hệ giữa sự hợp tác với khách hàng trong quản trị chuỗi cung ứng và kết quả hoạt động của doanh nghiệp sản xuất chế tạo. **Mã số: 120.2BAadm.22** 42
Relationship between Customer Cooperation in Supply Chain Management and Business Performance of Manufacturing Enterprises
- 6. Tô Ngọc Thịnh** - Thực trạng và giải pháp hoàn thiện quản trị hoạt động lữ hành quốc tế theo tiếp cận liên kết đa ngành trên địa bàn Hà Nội. **Mã số: 120.2TRMg.21** 55
Reality and Solutions to Complete International Tour Operation Management in Hanoi under Multi-sector Link Approach

Ý KIẾN TRAO ĐỔI

- 7. Lê Thị Anh** - Sử dụng hồi quy tỷ lệ nghiên cứu các yếu tố tác động đến hành vi tiêu dùng hoa quả nhập khẩu của người dân thành thị Việt Nam trong điều kiện thông tin bất đối xứng - Trường hợp địa bàn Hà Nội. **Mã số: 120.3TrEM.31** 65
Applying Fractional Response Regression on Factors Affecting Behaviors of Imported Fruit Consumption of Viet Nam Urban Citizens in the Context of Asymmetric Information – A Case Study in Hanoi

ỨNG DỤNG PHƯƠNG PHÁP PHÂN RÃ DỰA TRÊN COPULA ĐỂ NGHIÊN CỨU BẤT BÌNH ĐẲNG THU NHẬP GIỮA NÔNG THÔN VÀ THÀNH THỊ CỦA VIỆT NAM

Lê Văn Tuấn

Đại học Thương mại

Email: tuanlevan@tmu.edu.vn

Michel Simioni

Toulouse School of Economics, France

Email: michel.simioni@inra.fr

Trịnh Thị Hoàng

Đại học Thương mại

Email: trinhthihuong@tmu.edu.vn

Ngày nhận: 09/07/2018

Ngày nhận lại: 01/08/2018

Ngày duyệt đăng: 06/08/2018

Bài viết sử dụng phương pháp phân rã dựa trên copula để nghiên cứu bất bình đẳng trong thu nhập giữa nông thôn và thành thị của Việt Nam (trên bộ dữ liệu VHLSS). Kết quả thực nghiệm cho thấy yếu tố giáo dục đóng vai trò quan trọng nhất trong việc giải thích sự chênh lệch thu nhập của dân cư ở hai khu vực này. Bên cạnh đó, kết quả còn cho thấy hiệu ứng phụ thuộc có vai trò đáng kể trong một số trường hợp (giải thích 1/6 chênh lệch thu nhập khi xét cho phân vị 90th ở năm 2014).

Từ khóa: Bất bình đẳng, thu nhập, nông thôn - thành thị, phân rã, copula.

1. Mở đầu¹

Bất bình đẳng thu nhập (giữa nông thôn và thành thị, giới tính, vùng miền, quốc gia hay giữa hai thời điểm) là một vấn đề trọng tâm trong nghiên cứu kinh tế, ở nhiều quốc gia phát triển và đang phát triển. Song song áp dụng vào thực nghiệm các phương pháp phân rã bất bình đẳng đã phổ biến trên thế giới, các nhà nghiên cứu vẫn tiếp tục phát triển về mặt lý thuyết và mô hình để tìm ra yếu tố/nguyên nhân dẫn đến bất bình đẳng thu nhập. Có hai xu hướng tiếp cận chính: tiếp cận từ các đặc điểm của dân số (giới tính, tuổi, vùng

miền); và tiếp cận theo cấu trúc của thu nhập (sự phân chia các nguồn thu nhập khác nhau).

Ở hướng tiếp cận thứ nhất, công trình mang tính khai sáng là của Oaxaca (1973) và Blinder (1973), về sau được gọi là phương pháp phân rã Oaxaca–Blinder. Nhóm các phương pháp này có hai bước:

- Bước thứ nhất (phân rã tổng hợp), phân chia bất bình đẳng thành hai phần: hiệu ứng đặc điểm (sinh ra do giá trị/đặc tính của biến giải thích), và hiệu ứng cấu trúc (sinh ra do sự khác biệt về tác động của các biến giải thích tới biến phụ thuộc).

1. Phần này tham khảo chính trong [Tuấn (2018)]

- Bước thứ hai (phân rã chi tiết), tiếp tục phân rã các hiệu ứng theo từng biến giải thích.

Nhóm các phương pháp này không chỉ dùng để nghiên cứu về bất bình đẳng trong thu nhập, mà còn được ứng dụng khi nghiên cứu về nhiều vấn đề bất bình đẳng khác như tiền lương, chi tiêu, cơ hội... và đặc biệt còn được sử dụng để lý giải về việc có hay không (hoặc mức độ) sự phân biệt đối xử trên thị trường lao động.

Phương pháp phân rã Oaxaca-Blinder áp dụng phân rã bất bình đẳng tại giá trị trung bình, biến phụ thuộc là biến liên tục và giả sử quan hệ giữa biến giải thích (biến độc lập, đồng biến) và biến phụ thuộc là tuyến tính. Phương pháp này cho phép cả phân rã tổng hợp và phân rã chi tiết. Nhiều mở rộng của phương pháp Oaxaca-Blinder nhằm phân rã không chỉ giá trị trung bình mà còn cho phương sai, các mức phân vị/hoặc phần phân vị, các chỉ số bất bình đẳng như Gini (gọi chung là các thống kê); hay áp dụng cho biến phụ thuộc là biến hữu hạn²; cũng như khắc phục giả thiết tuyến tính.

Hướng mở rộng phổ biến đầu tiên cần được nhắc tới là phương pháp dựa trên hồi quy phân vị (có điều kiện), tiêu biểu là công trình của [Machado-Mata (2005)]. Phương pháp Machado-Mata thực hiện phân rã tổng hợp và phân rã chi tiết cho hiệu ứng cấu trúc, nhưng không phân rã được hiệu ứng đặc điểm³.

Hướng mở rộng dựa trên hồi quy phân phối của [Chernozhukov (2013)] - bản tiên ấn phẩm năm 2009, và dựa trên hồi quy RIF (Recentered Influence Function)⁴ của [Firpo (2007)], đều có thể áp dụng đầy đủ cho phân rã tổng hợp và phân rã chi tiết. Tuy nhiên, mỗi phương pháp đều có những ưu nhược điểm riêng. Có thể xem tổng quan về các phương pháp phân rã cũng như đánh giá chi tiết cho từng phương pháp trong [Fortin (2011)].

Hướng mở rộng gần đây của [Rothe (2015)], dựa trên lý thuyết copula, cho phép phân rã tổng hợp và phân rã chi tiết cho hiệu ứng đặc điểm. Phương pháp của Rothe có thể xem là mở rộng tự nhiên của phương pháp Oaxaca-Blinder cho các thống kê và cho mô hình phi tuyến; trường hợp riêng của nó, phân rã cho giá trị trung bình và với giả thuyết tuyến tính, phương pháp của Rothe trùng với phương pháp Oaxaca-Blinder. Phương pháp này sẽ phân rã chi tiết hiệu ứng đặc điểm thành 3 thành phần:

(i) phân bổ trực tiếp cho từng biến - liên quan tới sự khác biệt ở các phân phối biên duyên của biến giải thích;

(ii) các hiệu ứng tương tác k-chiều ($k \geq 2$) - sinh ra do sự tương tác của các phân phối biên duyên;

(iii) hiệu ứng phụ thuộc - do sự khác biệt của cấu trúc phụ thuộc của các biến giải thích.

[Rothe (2015)] sử dụng phương pháp này để nghiên cứu về sự bất bình đẳng trong thu nhập từ lương của dân cư Mỹ giữa hai năm 1985 và 2005. Kết quả thực nghiệm cho thấy hiệu ứng cấu trúc chiếm 1/5 chênh lệch toàn phần về lương (được đo từ phân vị 10th đến 90th); và chiếm 1/4 hiệu ứng đặc điểm. Phương pháp phân rã của Rothe cho thấy hiệu ứng đặc điểm không chỉ bao gồm các đặc trưng của từng biến giải thích (như các công trình trước đó) mà còn được cấu tạo từ ảnh hưởng tương tác và cấu trúc phụ thuộc của các biến giải thích.

Vấn đề bất bình đẳng về thu nhập tại Việt Nam, trường hợp riêng là giữa nông thôn và thành thị, cũng thu hút được sự quan tâm của rất nhiều nhà khoa học trong và ngoài nước.

Công trình [Binh (2007)] đã áp dụng hướng nghiên cứu bất bình đẳng giữa nông thôn và thành thị ở các mức phân vị khác nhau tại Việt Nam, sử dụng bộ số liệu VHLSS giai đoạn 1993-

2. Các mở rộng cho biến phụ thuộc là biến hữu hạn xem trong [Fortin (2011)].

3. Đây là nhược điểm lớn của phương pháp này, vì hiệu ứng đặc điểm có ý nghĩa kinh tế hơn là hiệu ứng cấu trúc, hơn nữa việc phân rã hiệu ứng cấu trúc gặp vấn đề về "nhóm được bỏ". [Machado-Mata (2005)] có đưa ra gợi ý về kỹ thuật để phân rã hiệu ứng đặc điểm, tuy nhiên [Fortin (2011)] đã chỉ ra là giải pháp này không hợp lý.

4. Được xếp vào hồi quy phân vị không có điều kiện.

1998. Các tác giả sử dụng phương pháp tương tự Machado-Mata nhưng không đánh giá hàm mật độ thu nhập mà áp dụng trực tiếp vào các phân vị của giá trị thu nhập. Trong bài báo này, các tác giả đã sử dụng biến chỉ tiêu tiêu dùng làm thước đo cho thu nhập để nghiên cứu bất bình đẳng. Trong năm 1993, nguyên nhân của bất bình đẳng chủ yếu do hiệu ứng đặc điểm và tập trung ở các biến giáo dục, dân tộc, độ tuổi; kết quả đồng nhất trên tất cả các mức phân vị. Trong năm 1998, kết quả chỉ đúng cho mức phân vị thấp nhất (hộ nghèo); ở các mức phân vị còn lại, nguyên nhân của bất bình đẳng chủ yếu do hiệu ứng cấu trúc. Công trình [Huong (2014)] cũng thực hiện nghiên cứu tương tự trong giai đoạn 1993-2006, tuy nhiên, sử dụng phương pháp hồi quy RIF của [Firpo (2009)]. Kết quả cũng cho thấy giáo dục đóng vai trò quan trọng gây nên sự chênh lệch thu nhập. Phân rã hiệu ứng cấu trúc cho thấy một phần đáng kể thuộc về hệ số chặn - thể hiện vai trò của các biến không quan sát gây nên sự chênh lệch. Công trình của [Thanh (2017)] cũng sử dụng phương pháp tương tự như [Huong (2014)] cho giai đoạn 2008-2012.

Bên cạnh những công trình sử dụng biến chỉ tiêu, một số công trình khác sử dụng biến tiền lương để nghiên cứu bất bình đẳng thu nhập giữa nông thôn-thành thị. Công trình [Trần (2015)] đã sử dụng phương pháp Machado-Mata để phân rã chênh lệch tiền lương trên bộ số liệu VHLSS 2012. Kết quả cho thấy hiệu ứng đặc điểm chiếm hơn 50% chênh lệch tiền lương ở tất cả các mức phân vị được xét.

Bài viết này sẽ sử dụng phương pháp phân rã dựa trên copula của [Rothe (2015)] để nghiên cứu bất bình đẳng thu nhập giữa nông thôn và thành thị của Việt Nam. Các kết quả thực nghiệm dựa trên bộ dữ liệu VHLSS các năm 2004 và 2014.

2. Cơ sở lý thuyết của phương pháp phân rã dựa trên copula⁵

Ta xem xét hai tập con riêng rẽ ký hiệu bởi $g \in \{0, 1\}$. Với mỗi nhóm g , ta quan sát biến đầu ra Y^g

và vectơ d-chiều các biến giải thích X^g ; với hàm phân phối F_Y^g và F_X^g , giá Y^g và X^g , và hàm phân phối tích lũy (CDF) có điều kiện $F_{Y|X}^g$. Hơn nữa, với mỗi hàm phân phối tích lũy F , ta ký hiệu (F) là một đặc trưng phân phối, với $v: F \mapsto \mathbb{R}$. Chẳng hạn, giá trị trung bình $v: F \mapsto \int y dF(y)$ và τ -phân vị, $v: F \mapsto F^{-1}(\tau)$ các momen bậc cao, các thống kê liên quan tới phân vị như các khoảng giữa phân vị hay tỷ lệ phân vị, các thước đo bất bình đẳng như hệ số Gini. Mục đích chính của chúng ta là tìm hiểu sự chênh lệch toàn phần được quan sát giữa $v(F_Y^1)$ và $v(F_Y^0)$:

$$\Delta_O^v = v(F_Y^1) - v(F_Y^0)$$

liên hệ như thế nào với sự khác biệt giữa hai phân phối F_X^1 và F_X^0 . Để đạt mục đích này, ta định nghĩa phân phối phản thực của biến đầu ra $F_Y^{g|j}$: kết hợp phân phối điều kiện trong nhóm g với phân phối của các biến giải thích trong nhóm $j \neq g$

$$F_Y^{g|j}(y) = \int F_{Y|X}^g(y, x) dF_X^j(x)$$

Tích phân trên là định nghĩa-được miễn là $X_j \subset X_g$.

Khi đó, ta có thể viết: $\Delta_O^v = \Delta_S^v + \Delta_X^v$ với $\Delta_S^v = v(F_Y^1) - v(F_Y^{0|1})$ và $\Delta_X^v = v(F_Y^{0|1}) - v(F_Y^0)$

Ở đây, Δ_X^v là hiệu ứng đặc điểm: do sự khác biệt về phân phối của các biến giải thích giữa hai nhóm; và Δ_S^v là hiệu ứng cấu trúc: do sự khác biệt giữa các hàm phân phối tích lũy có điều kiện $F_{Y|X}^1$ và $F_{Y|X}^0$.

Nền tảng của phương pháp phân rã dựa trên copula là Định lý Sklar được phát biểu như sau:

Hàm phân phối tích lũy của X^g luôn có thể viết thành

$$F_X^g(x) = C^g(F_{X_1}^g(x_1), \dots, F_{X_d}^g(x_d)) \text{ với } g \in \{0, 1\},$$

ở đây C^g là hàm copula - hàm phân phối tích lũy nhiều biến với các biên duyên là phân phối đều chuẩn tắc, và $F_{X_k}^g$ là phân phối biên duyên của thành phần thứ k của X^g . Hàm copula mô tả phân phối chung của các thứ hạng riêng rẽ của các

5. Phần này thao khảo chính trong [Rothe (2015)]

thành phần trong X^g , và có thể được diễn giải là đối tượng xác định cấu trúc phụ thuộc.

Để đơn giản, ta sẽ giả định các cấu trúc phụ thuộc tuân theo copula Gaussian.

Công thức Sklar có thể sử dụng để xác định các phân phối đầu ra phân thực: kết hợp phân phối điều kiện trong nhóm g với phân phối giả thiết của các biến giải thích (chia sẻ tính chất của cả v và y). Ký hiệu mỗi phần tử d chiều thuộc tập $\{0, 1\}^d$ bằng chữ đậm, ta định nghĩa phân phối của biến đầu ra theo nghĩa phân thực: cấu trúc như trong nhóm g , phân phối của các biến giải thích có copula của nhóm j , và phân phối biên duyên của biến giải thích thứ l thì bằng phân phối biên duyên trong nhóm k_l

$$F_Y^{g|j,k}(y) = \int F_{Y|X}^g(y, x) dF_X^{j,k}(x)$$

với

$$F_X^{j,k}(x) = C^j(F_{X_1}^{k_1}(x_1), \dots, F_{X_d}^{k_d}(x_d))$$

Ta ký hiệu $1 = (1, 1, \dots, 1)$ và $0 = (0, 0, \dots, 0)$, ký hiệu e^l là vectơ đơn vị thứ l , và đặt $|k| = \sum_{l=1}^d k_l$. Tiếp theo, với bất kỳ đặc trưng phân phối v ta định nghĩa:

$$\beta^v(k) = v(F_Y^{0|0,k}) - v(F_Y^0)$$

có thể được diễn giải như là hiệu ứng của thực nghiệm phân thực trong nhóm 0 nhưng thay thế phân phối biên duyên của $|k|$ biến giải thích, trong đó $k_l = 1$ tương ứng với nhóm 1, trong khi các thành phần khác giữ nguyên (bao gồm cả cấu trúc phụ thuộc).

Cuối cùng ta định nghĩa:

$$\Delta_M^v(k) = \beta^v(k) + \sum_{1 \leq |m| \leq |k|-1} (-1)^{|k|-|m|} \beta^v(m)$$

với tổng trên tập rỗng bằng 0 (do đó, $\Delta_M^v(e^l) = \beta^v(e^l)$).

Quá trình phân rã hiệu ứng đặc điểm được thực hiện như sau:

- Ở bước thứ nhất, Δ_X^v phân rã thành hiệu ứng phụ thuộc Δ_D^v và hiệu ứng phân phối biên duyên toàn phần Δ_M^v :

$$\Delta_X^v = \Delta_D^v + \Delta_M^v$$

với $\Delta_D^v = v(F_Y^{0|1,1}) - v(F_Y^{0|0,1})$ và $\Delta_M^v = v(F_Y^{0|0,1}) - v(F_Y^{0|0,0})$

- Ở bước thứ hai, ta phân rã Δ_M^v thành các hiệu ứng phân phối biên duyên từng phần

$$\Delta_M^v = \sum_{1 \leq |k| \leq d} \Delta_M^v(k)$$

Như vậy, ta có:

$$\Delta_X^v = \sum_{1 \leq |k| \leq d} \Delta_M^v(k) + \Delta_D^v$$

Trong trường hợp $|k| = 1$, nghĩa là $k = e^l$ là vectơ đơn vị thứ l , được diễn giải là phân bố trực tiếp của sự khác biệt giữa hai nhóm theo phân phối biên duyên của biến giải thích thứ l . Với $|k| > 1$, số hạng Δ_M^v thể hiện hiệu ứng tương tác $|k|$ -chiều của các phân phối biên duyên mà có thành phần tương ứng trong K bằng 1.

3. Kết quả thực nghiệm tại Việt Nam

Báo cáo này sử dụng bộ dữ liệu Vietnam Household Living Standards Survey (VHLSS) của Tổng cục Thống kê, được khảo sát hai năm một (từ 2002) với sự hỗ trợ của Ngân hàng Thế giới (World Bank). Bộ dữ liệu bao gồm rất nhiều thông tin về các hộ gia đình ở Việt Nam: thu nhập, chi tiêu gia đình (lương thực, bảo hiểm, giáo dục...), và đặc điểm nhân khẩu của các hộ. Mỗi đợt khảo sát sẽ thực hiện trên tất cả 64 tỉnh thành của Việt Nam, với khoảng 9000 hộ. Trong bài viết này, chúng tôi sử dụng 2 bộ dữ liệu VHLSS của năm 2004 và 2014. Biến nghiên cứu là loga của chi tiêu tiêu dùng thực tế tính theo đầu người (lrpcc). Việc sử dụng chi tiêu tiêu dùng là hợp lý hơn thu nhập vì ta có thể giảm thiểu được những sai lệch do tự kê khai trong quá trình khảo sát cũng như sự biến động quá lớn trong thu nhập của các hộ - vì tính chất thời vụ hoặc tỷ lệ tiết kiệm⁶. Các tính toán

6. Việc sử dụng thay thế này xuất phát từ giả thiết PIH (Permanent income hypothesis) được khởi xướng bởi Milton Friedman. Trong các bộ dữ liệu khảo sát thì biến thu nhập thường bị sai lệch do vấn đề tự khai báo, và các kết quả thực nghiệm cho thấy biến này có sai số lớn về đo lường, mặc dù nhiều bộ khảo sát vẫn có biến thu nhập nhưng các tác giả thường bỏ qua biến này, xem [Gallup (2002)].

thống kê được dựa trên bộ dữ liệu đã được chọn mẫu có trọng số của VHLSS nên kết quả có thể xem là đại diện cho toàn dân số Việt Nam trong mỗi năm cũng như mỗi khu vực.

3.1. Thống kê mô tả

Đề tài nghiên cứu sự ảnh hưởng của 8 biến giải thích lên biến thu nhập: nhóm biến đặc trưng cho nhân khẩu gồm có Ethnic, Hsize, Gender, Ageh và Agehsq; biến đại diện cho vốn con người là Yedu; biến đại diện cho ngành nghề là Incom.Crop, biến đại diện cho vùng miền là South⁷. Biến kiểm soát là Urban. Cụ thể:

- + Urban: Biến giả để chỉ đặc trưng khu vực. Urban = 1 nếu là thành thị, Urban = 0 nếu là nông thôn.
- + Ethnic: Biến giả để chỉ đặc trưng sắc tộc của

chủ hộ. Ethnic = 0 nếu là dân tộc thiểu số, Ethnic = 1 nếu là người Kinh.

- + Hsize: Biến liên tục chỉ số người trong hộ.
- + Gender: Biến giả để chỉ giới tính của chủ hộ. Gender = 1 nếu là nam, Gender = 0 nếu là nữ.
- + Ageh và Agehsq: Biến liên tục chỉ độ tuổi (và bình phương độ tuổi) của chủ hộ.
- + Yedu: Biến liên tục chỉ số năm được đào tạo của chủ hộ.
- + Incom.Crop: Biến giả chỉ đặc trưng ngành nghề. Incom.Crop = 1 nếu có thu nhập trong lĩnh vực nông nghiệp, Incom.Crop = 0 nếu ngược lại.
- + South: Biến giả để chỉ đặc trưng vùng miền. South = 1 nếu là miền Nam, South = 0 nếu ngược lại.

Bảng 1 và bảng 2 thể hiện các thống kê mô tả

Bảng 1: Thống kê mô tả cho năm 2004

	Thành thị					Nông thôn				
	Mean	SD	Q10	Q50	Q90	Mean	SD	Q10	Q50	Q90
lrpcc	9.82	0.62	9.01	9.81	10.64	9.11	0.53	8.47	9.08	9.79
Gender	0.62	0.49	0.00	1.00	1.00	0.79	0.40	0.00	1.00	1.00
Yedu	8.34	5.45	0.00	9.00	16.00	5.63	4.57	0.00	5.00	12.00
Ageh	50.84	13.30	35.00	49.00	71.00	49.01	14.17	32.00	46.00	70.00
Ethnic	0.95	0.23	1.00	1.00	1.00	0.86	0.34	0.00	1.00	1.00
Hhsize	4.14	1.47	2.00	4.00	6.00	4.32	1.51	2.00	4.00	6.00
Incom.Crop	0.19	0.39	0.00	0.00	1.00	0.71	0.45	0.00	1.00	1.00
South	0.46	0.50	0.00	0.00	1.00	0.32	0.47	0.00	0.00	1.00

Bảng 2: Thống kê mô tả cho năm 2014

	Thành thị					Nông thôn				
	Mean	SD	Q10	Q50	Q90	Mean	SD	Q10	Q50	Q90
lrpcc	10.01	0.58	9.29	10.01	10.72	9.89	0.59	9.14	9.88	10.64
Gender	0.71	0.45	0.00	1.00	1.00	0.74	0.44	0.00	1.00	1.00
Yedu	7.62	5.35	0.00	9.00	14.00	7.20	5.15	0.00	9.00	14.00
Ageh	51.73	13.80	34.00	51.00	71.00	51.14	14.32	33.00	50.00	72.00
Ethnic	0.90	0.31	0.00	1.00	1.00	0.85	0.36	0.00	1.00	1.00
Hhsize	3.87	1.50	2.00	4.00	6.00	3.75	1.47	2.00	4.00	6.00
Incom.Crop	0.52	0.50	0.00	1.00	1.00	0.61	0.49	0.00	1.00	1.00
South	0.31	0.46	0.00	0.00	1.00	0.33	0.47	0.00	0.00	1.00

7. Việc chọn biến giải thích trong [Binh (2007)], [Huong (2014)] và [Thanh (2017)] là khác nhau và đều dựa trên lập luận định tính. Theo quan điểm của chúng tôi, cần phải đưa tất cả các biến liên quan tới hộ gia đình trong VHLSS (trừ biến thu nhập và các biến cấu thành nên thu nhập) vào mô hình.

của các biến trong các năm 2004 và 2014. Rõ ràng, thu nhập ở khu vực thành thị cao hơn ở nông thôn khi xét riêng cho từng năm, và thu nhập ở năm 2014 cao hơn năm 2004. Xét riêng từng năm, số năm trung bình được đào tạo của chủ hộ ở khu vực thành thị đều cao hơn ở nông thôn. Tuy nhiên, chênh lệch về số năm được đào tạo giữa nông thôn và thành thị giảm (2.71 năm vào năm 2004 so với 0.42 năm vào năm 2014). Một đặc điểm nổi bật khác là trong năm 2004, tỷ lệ có thu nhập từ nông nghiệp ở khu vực nông thôn là 71% trong khi ở thành thị là 19%, trong khi đó các kết quả của năm 2014 không có sự chênh lệch lớn, tương ứng là 61% và 52%.

3.2. Kết quả nghiên cứu thực nghiệm

Bảng 3 và bảng 4 mô tả các kết quả phân rã cho hai năm 2004 và 2014, bao gồm 5 thống kê: trung bình, phân vị 10th, trung vị, phân vị 90th và hệ số Gini. Để đơn giản, các bảng chỉ thể hiện các hiệu ứng tương tác 2 chiều, tuy nhiên các giá trị cũng không đáng kể.

Trước hết, ta xem xét ước lượng của hiệu ứng cấu trúc và hiệu ứng đặc điểm, những hiệu ứng cũng được ước lượng qua các phương pháp khác. Nếu chỉ xét cho ước lượng trung bình của thu nhập và các mức phân vị, hiệu ứng đặc điểm - được sinh do bản thân các đặc điểm về lao động - có thể giải thích 1/3 chênh lệch toàn phần ở năm 2004; kết quả chỉ tương đồng với năm 2014 ở trung bình và trung vị, ở năm này phân vị 10% có hiệu ứng đặc điểm khá nhỏ nhưng ở phân vị 90% hiệu ứng này lại chiếm tới hơn một nửa chênh lệch toàn phần. Với hệ số Gini, năm 2004 hiệu ứng đặc điểm chiếm gần như hoàn toàn chênh lệch toàn phần và có giá trị còn lớn hơn chênh lệch tại năm 2014. Như vậy, trong hầu hết các trường hợp, các đặc trưng về giáo dục, giới tính, dân tộc, độ tuổi, số nhân khẩu, nghề nghiệp hay vùng miền có thể giải thích ít nhất là khoảng 1/3 sự bất bình đẳng thu nhập của dân cư nông thôn - thành thị. Kết quả này tương đồng với kết

quả phân rã dựa trên hồi quy phân vị cho năm 1998 của [Binh (2007)].

Tiếp theo, ta sẽ đề cập tới hiệu ứng phụ thuộc, đây là hiệu ứng mới chỉ có trong phương pháp phân rã của [Rothe (2005)]. Trong năm 2004 hiệu ứng này có giá trị không đáng kể; tuy nhiên trong năm 2014, ở phân vị 90% hiệu ứng này chiếm tới 1/3 hiệu ứng đặc điểm, và ở hệ số Gini hiệu ứng này có giá trị còn lớn hơn hiệu ứng đặc điểm.

Cuối cùng, ta sẽ xem xét phân bổ trực tiếp lên các biến giải thích. Trừ trường hợp ngoại lệ là kết quả phân rã của hệ số Gini năm 2014, các kết quả cho thấy biến giáo dục chiếm phần lớn nhất trong hiệu ứng đặc điểm (từ 1/3 đến 1/2). Nói một cách khác, chênh lệch trong thu nhập giữa nông thôn-thành thị có một phần đáng kể do cư dân ở thành thị được học hành nhiều hơn nông thôn.

4. Các kết luận và phát hiện qua nghiên cứu

Như vậy, có một sự tăng trưởng ấn tượng về thu nhập ở cả khu vực nông thôn và thành thị (được đại diện bởi biến chi tiêu tiêu dùng) của dân cư ở Việt Nam từ năm 2004 tới 2014. Tuy nhiên, trong mỗi năm, có sự chênh lệch đáng kể về thu nhập của dân cư ở hai khu vực này. Sự bất bình đẳng giữa các khu vực phần lớn là do hiệu ứng cấu trúc, tuy nhiên, hiệu ứng đặc điểm cũng đóng một vai trò quan trọng. Trong một số trường hợp, hiệu ứng phụ thuộc đóng vai trò đáng kể trong việc giải thích chênh lệch toàn phần - đây là đóng góp mới của phương pháp phân rã dựa trên copula. Số năm được đào tạo của chủ hộ đóng vai trò lớn nhất trong việc giải thích sự bất bình đẳng trong thu nhập của cư dân nông thôn - thành thị. Kết quả này cũng tương tự với các nghiên cứu trước; một lần nữa khẳng định triết lý “đưa việc thu nạp kiến thức vào trung tâm phân tích kinh tế”, hay luận điểm phát triển “Giáo dục và đào tạo là quốc sách hàng đầu”. ♦

Bảng 3: Kết quả phân rã cho năm 2004 [thu nhập: $100 \cdot \log(rpcc)$]

	Mean		Q10		Median		Q90		Gini	
Chênh lệch toàn phần	66.66	(1.68)	44.40	(3.39)	76.02	(1.56)	72.55	(3.42)	0.34	(0.1)
Hiệu ứng cấu trúc	43.43	(1.5)	30.49	(2.59)	49.49	(1.57)	49.63	(3.96)	0.05	(0.09)
Hiệu ứng đặc điểm	23.22	(1.09)	13.91	(2.11)	26.54	(1.35)	22.93	(3.31)	0.30	(0.08)
Hiệu ứng phụ thuộc	0.67	(0.33)	2.85	(1.56)	0.82	(0.17)	-0.74	(2.39)	-0.05	(0.07)
Hiệu ứng biên duyên	22.55	(1.05)	11.06	(1.51)	25.72	(1.33)	23.66	(3.37)	0.35	(0.05)
<i>Phân bố trực tiếp:</i>										
Yedu	9.63	(0.62)	4.83	(0.87)	10.51	(0.79)	11.03	(1.97)	0.12	(0.03)
Gender	-0.14	(0.27)	-0.62	(0.48)	-0.04	(0.42)	0.37	(0.57)	0.04	(0.01)
Ageh	5.83	(0.08)	4.19	(0.16)	7.41	(0.13)	4.19	(0.17)	-0.01	(0.01)
Agehsq	-4.07	(0.21)	-2.63	(0.57)	-5.05	(0.34)	-7.83	(0.43)	-0.03	(0.01)
Ethnic	1.84	(0.1)	0.99	(0.19)	2.02	(0.13)	0.76	(0.23)	0.00	(0.01)
Hhsize	1.42	(0.3)	0.99	(0.56)	1.30	(0.37)	1.03	(0.27)	0.00	(0.01)
South	4.93	(0.37)	2.43	(0.77)	5.42	(0.48)	4.77	(0.7)	0.04	(0.01)
Incom.Crop	1.79	(0.44)	-3.05	(0.61)	2.44	(0.59)	5.00	(0.94)	0.21	(0.01)
<i>Hiệu ứng tương tác 2-chiều:</i>										
Yedu:Gender	0.08	(0.05)	-0.37	(0.28)	0.16	(0.28)	0.33	(0.75)	-0.00	(0)
Yedu:Ageh	0.14	(0.02)	1.76	(0.09)	0.21	(0.09)	1.60	(0.21)	-0.01	(0)
Yedu:Agehsq	-0.07	(0.03)	-0.19	(0.52)	0.10	(0.33)	2.63	(0.55)	0.01	(0)
Yedu:Ethnic	-0.03	(0.02)	0.52	(0.13)	-0.19	(0.07)	0.28	(0.26)	0.00	(0)
Yedu:Hhsize	0.01	(0.01)	0.56	(0.38)	0.06	(0.24)	0.60	(0.35)	-0.00	(0)
Yedu:South	0.05	(0.05)	1.58	(0.64)	-0.04	(0.48)	2.00	(1.03)	-0.00	(0)
Yedu:Incom.Crop	0.53	(0.05)	0.69	(0.7)	0.91	(0.52)	1.05	(1.35)	-0.01	(0)
Gender:Ageh	0.05	(0)	-0.29	(0.03)	0.31	(0.05)	0.48	(0.07)	-0.00	(0)
Gender:Agehsq	-0.03	(0.01)	-0.26	(0.19)	-0.18	(0.23)	0.41	(0.24)	0.00	(0)
Gender:Ethnic	0.02	(0)	-0.01	(0.03)	0.01	(0.03)	0.00	(0.13)	-0.00	(0)
Gender:Hhsize	0.01	(0.01)	-0.00	(0.15)	-0.03	(0.2)	0.02	(0.16)	-0.00	(0)
Gender:South	0.04	(0.02)	-0.01	(0.23)	0.21	(0.23)	0.55	(0.44)	-0.00	(0)
Gender:Incom.Crop	-0.01	(0.02)	-0.36	(0.22)	0.04	(0.23)	0.59	(0.63)	0.00	(0)
Ageh:Agehsq	-0.11	(0)	-0.10	(0.08)	0.22	(0.06)	3.84	(0.07)	0.00	(0)
Ageh:Ethnic	-0.02	(0)	0.53	(0.02)	0.11	(0.01)	0.57	(0.04)	-0.00	(0)
Ageh:Hhsize	0.03	(0)	0.50	(0.06)	0.25	(0.06)	0.81	(0.04)	-0.00	(0)
Ageh:South	0.06	(0.01)	1.38	(0.08)	-0.29	(0.07)	1.58	(0.11)	-0.00	(0)
Ageh:Incom.Crop	0.33	(0.01)	0.66	(0.09)	1.20	(0.07)	0.82	(0.15)	-0.01	(0)
Agehsq:Ethnic	0.02	(0)	0.43	(0.1)	0.20	(0.04)	1.87	(0.12)	0.00	(0)
Agehsq:Hhsize	-0.01	(0)	0.42	(0.38)	0.15	(0.23)	1.47	(0.15)	0.00	(0)
Agehsq:South	-0.03	(0.01)	0.47	(0.53)	-0.15	(0.38)	3.47	(0.36)	0.00	(0)
Agehsq:Incom.Crop	-0.21	(0.01)	-0.38	(0.53)	-0.11	(0.29)	3.28	(0.42)	0.01	(0)
Ethnic:Hhsize	-0.01	(0)	0.03	(0.08)	0.23	(0.03)	-0.07	(0.08)	-0.00	(0)
Ethnic:South	-0.03	(0.01)	0.14	(0.11)	-0.37	(0.05)	0.66	(0.17)	0.00	(0)
Ethnic:Incom.Crop	0.11	(0.01)	0.58	(0.11)	0.76	(0.05)	0.71	(0.24)	-0.00	(0)
Hhsize:South	0.01	(0.01)	0.13	(0.36)	-0.11	(0.24)	0.92	(0.22)	-0.00	(0)
Hhsize:Incom.Crop	0.08	(0.01)	0.36	(0.39)	0.94	(0.19)	0.98	(0.3)	-0.00	(0)
South:Incom.Crop	0.30	(0.02)	0.99	(0.55)	0.82	(0.42)	0.99	(0.9)	-0.00	(0)

Ghi chú: Các giá trị trong ngoặc là sai số chuẩn được tính qua Bootstrap

Bảng 4: Kết quả phân rã cho năm 2014 [thu nhập: $100 \cdot \log(rpcc)$]

	Mean		Q10		Median		Q90		Gini	
Chênh lệch toàn phần	50.21	(1.32)	44.87	(3.15)	51.07	(1.3)	60.05	(2.6)	0.23	(0.08)
Hiệu ứng cấu trúc	30.29	(1.16)	39.87	(2.35)	29.22	(1.39)	28.51	(2.66)	-0.25	(0.07)
Hiệu ứng đặc điểm	19.92	(0.92)	5.00	(2.66)	21.85	(1.11)	31.55	(1.92)	0.48	(0.07)
Hiệu ứng phụ thuộc	-1.24	(0.32)	-17.92	(1.67)	0.62	(0.19)	11.13	(0.93)	0.64	(0.06)
Hiệu ứng biên duyên	21.17	(0.82)	22.93	(1.86)	21.23	(1.12)	20.41	(1.69)	-0.17	(0.04)
<i>Phân bố trực tiếp:</i>										
Yedu	10.34	(0.55)	10.71	(1.17)	10.07	(0.72)	11.96	(1.32)	-0.02	(0.03)
Gender	-0.46	(0.2)	-0.97	(0.4)	-0.29	(0.26)	-0.10	(0.49)	0.02	(0.01)
Ageh	2.73	(0.14)	3.95	(0.2)	2.63	(0.14)	1.64	(0.23)	-0.03	(0)
Agehsq	-1.91	(0.36)	-1.91	(1.36)	-2.19	(0.42)	-1.33	(0.52)	0.00	(0.03)
Ethnic	4.39	(0.04)	10.23	(0.23)	3.64	(0.03)	1.60	(0.18)	-0.16	(0.01)
Hhsize	-0.29	(0.28)	-0.37	(0.51)	-0.22	(0.26)	-0.22	(0.25)	0.00	(0.01)
South	1.56	(0.2)	2.49	(0.45)	1.59	(0.24)	1.70	(0.27)	-0.02	(0.01)
Incom.Crop	4.49	(0.24)	-0.01	(0.45)	6.09	(0.34)	7.02	(0.59)	0.11	(0.01)
<i>Hiệu ứng tương tác 2-chiều:</i>										
Yedu:Gender	0.05	(0.03)	0.17	(0.32)	0.14	(0.21)	0.04	(0.4)	-0.00	(0)
Yedu:Ageh	-0.08	(0.01)	-0.43	(0.12)	-0.05	(0.08)	-0.41	(0.22)	-0.00	(0)
Yedu:Agehsq	0.07	(0.04)	0.24	(1.19)	0.48	(0.38)	0.52	(0.65)	0.00	(0)
Yedu:Ethnic	-0.18	(0.02)	-2.72	(0.12)	-0.49	(0.03)	-0.18	(0.12)	0.01	(0)
Yedu:Hhsize	-0.02	(0.01)	0.01	(0.3)	0.03	(0.1)	0.07	(0.13)	-0.00	(0)
Yedu:South	-0.06	(0.01)	-0.19	(0.47)	-0.17	(0.19)	-1.02	(0.31)	0.00	(0)
Yedu:Incom.Crop	0.29	(0.03)	0.98	(0.57)	0.47	(0.32)	-1.38	(0.69)	-0.04	(0)
Gender:Ageh	0.01	(0)	0.20	(0.03)	0.01	(0.04)	-0.04	(0.09)	-0.00	(0)
Gender:Agehsq	-0.01	(0.02)	-0.01	(0.31)	-0.07	(0.14)	-0.09	(0.22)	0.00	(0)
Gender:Ethnic	0.04	(0)	0.10	(0.03)	0.09	(0)	-0.05	(0.03)	-0.00	(0)
Gender:Hhsize	-0.00	(0)	-0.01	(0.06)	-0.01	(0.06)	-0.01	(0.05)	0.00	(0)
Gender:South	0.01	(0)	0.11	(0.11)	0.04	(0.1)	-0.03	(0.11)	-0.00	(0)
Gender:Incom.Crop	-0.01	(0.01)	-0.02	(0.12)	0.03	(0.14)	-0.02	(0.23)	0.00	(0)
Ageh:Agehsq	0.04	(0)	0.30	(0.13)	0.23	(0.07)	-0.22	(0.13)	-0.01	(0)
Ageh:Ethnic	-0.06	(0)	-0.53	(0.02)	-0.02	(0)	0.54	(0.03)	0.00	(0)
Ageh:Hhsize	-0.00	(0)	0.02	(0.04)	-0.06	(0.03)	-0.06	(0.04)	-0.00	(0)
Ageh:South	-0.02	(0)	-0.44	(0.06)	0.07	(0.05)	0.60	(0.09)	0.00	(0)
Ageh:Incom.Crop	0.09	(0)	0.06	(0.06)	-0.55	(0.09)	0.85	(0.13)	-0.01	(0)
Agehsq:Ethnic	0.05	(0.01)	0.02	(0.13)	0.07	(0.02)	-0.27	(0.07)	-0.00	(0)
Agehsq:Hhsize	0.00	(0.01)	-0.01	(0.31)	-0.04	(0.08)	-0.02	(0.1)	-0.00	(0)
Agehsq:South	0.02	(0.01)	0.04	(0.45)	0.20	(0.16)	-0.12	(0.21)	0.00	(0)
Agehsq:Incom.Crop	-0.05	(0.01)	0.01	(0.51)	-0.44	(0.29)	-0.44	(0.49)	0.01	(0)
Ethnic:Hhsize	-0.00	(0)	-0.05	(0.03)	-0.03	(0)	-0.09	(0.02)	-0.00	(0)
Ethnic:South	-0.05	(0)	-0.50	(0.05)	-0.10	(0.01)	0.80	(0.04)	0.00	(0)
Ethnic:Incom.Crop	0.23	(0)	0.91	(0.05)	-1.07	(0.02)	0.98	(0.08)	-0.02	(0)
Hhsize:South	-0.00	(0)	-0.00	(0.12)	-0.02	(0.09)	-0.05	(0.05)	-0.00	(0)
Hhsize:Incom.Crop	-0.02	(0)	-0.01	(0.13)	-0.06	(0.11)	-0.15	(0.1)	0.00	(0)
South:Incom.Crop	0.05	(0)	0.05	(0.25)	-0.46	(0.2)	0.99	(0.22)	-0.01	(0)

Ghi chú: Các giá trị trong ngoặc là sai số chuẩn được tính qua Bootstrap

Tài liệu tham khảo:

1. Chernozhukov, V., I. Fern´andez-Val, and B. Melly(2013), *Inference on counterfactual distributions*, *Econometrica* 81 (6), 2205-2268.

2. Firpo, S., Fortin M. N., Lemieux T. (2007), *Decomposing Wage Distributions using Recentered Influence Functions Regressions*, mimeo, University of British Columbia.

3. Fortin, N., Lemieux, T., and Firpo, S. (2011), *Decomposition methods in economics*, *Handbook of labor economics*4: 1-102.

4. Machado, J. A., and Mata, J. (2005), *Counterfactual decomposition of changes in wage distributions using quantile regression*(2005), *Journal of applied Econometrics*20, no. 4: 445-465.

5. Rothe, C. (2015), *Decomposing the composition effect: the role of covariates in determining between-group differences in economic outcomes*. *Journal of Business & Economic Statistics*33, no. 3: 323-337.

[Nghiên cứu thực nghiệm tại Việt Nam]

6. Binh, T. N., James W. A., Susan B. V., and M. Daniel W. (2007), *A quantile regression decomposition of urban–rural inequality in Vietnam*, *Journal of Development Economics*83, no. 2: 466-490.

7. Gallup , L. J. (2002), *The Wage Labor Market and Inequality in Vietnam in the 1990s*, Policy, Research working paper ; no. WPS 2896.

8. Huong T. L., Booth L. A. (2014), *Inequality in Vietnamese Urban-Rural Living Standards, 1993-2006*. *Review of Income and Wealth*. Series 60, Number 4.

9. Thanh B. & Katsushi S. I. (2017), *Determinants of Rural-urban Inequality in Vietnam: Detailed Decomposition Analyses Based*

on Unconditional Quantile Regressions, *Discussion Paper Series DP2017-01*, Research Institute for Economics & Business Administration, Kobe University, revised Jun.

10. Trần T.T.A. (2015), *Phân rã chênh lệch tiền lương thành thị - nông thôn ở Việt Nam bằng phương pháp hồi quy phân vị*, *Tạp chí Kinh tế và Phát triển*, Trường Đại học Kinh tế Quốc dân Hà Nội, số 219, tháng 9-2015, 20-29.

11. Tuấn L. V., Hùng T. T (2018), *Sử dụng phương pháp phân rã dựa trên copula để nghiên cứu bất bình đẳng thu nhập giữa nông thôn và thành thị của Việt Nam*, *Đề tài NCKH cấp trường Đại học Thương mại*.

Summary

The authors used Copula-based factorization to study the income inequality between rural and urban areas in Vietnam (on VHLSS data). The empirical research showed that education played the most important role in explaining the difference in resident income in the two areas. Besides, the findings revealed that dependence effect is significant in some cases (explaining one-sixth of income difference for 90th percentile in 2014).