

## MÔ HÌNH OAXACA – BLINDER TRONG PHÂN TÍCH KINH TẾ

Ngày nhận bài: 02/03/2015

Ngày nhận lại: 22/03/2015

Ngày duyệt đăng: 26/03/2015

**Lê Bảo Lâm<sup>1</sup>**  
**Nguyễn Minh Hà<sup>2</sup>**  
**Lê Văn Hưởng<sup>3</sup>**

### TÓM TẮT

*Oaxaca – Blinder được sử dụng rộng rãi trên thế giới để nghiên cứu sự khác nhau (khoảng cách) trong giá trị trung bình của hai nhóm. Chẳng hạn kỹ thuật này thường được sử dụng để phân tích sự khác biệt về lương theo giới tính, dân tộc, khu vực kinh tế,... Nhằm giúp các nhà nghiên cứu ở Việt Nam có thêm phương pháp phân tích trong lĩnh vực kinh tế, mục tiêu của nghiên cứu này là đề cập đến tổng quan về phương pháp phân rã Oaxaca – Blinder (1973) trong mô hình hồi quy tuyến tính và trong mô hình hồi quy xác suất và khả năng ứng dụng phương pháp này trong phân tích nhiều vấn đề kinh tế ở Việt Nam.*

**Từ khóa:** *Blinder, Oaxaca, sự khác biệt, khoảng cách.*

### ABSTRACT

*Oaxaca – Blinder model is popularly used in the world to study differences (gap) in average values of two groups. For instance, this technique is often applied to analyse the wage gap by gender, race, economic sector, etc. In order to provide the economists in Vietnam with an alternative method in economic analysis, the objectives of this research are to present Oaxaca – Blinder (1973) decomposition technique in linear regression models and non-linear regression models and to show the possibilities of applying this model in economic analysis in Vietnam.*

**Keywords:** *Blinder, Oaxaca, difference, gap.*

### 1. Giới thiệu

Một phương pháp được sử dụng rộng rãi trên thế giới trong phân tích kinh tế về sự khác biệt giữa các nhóm là phải phân rã sự khác nhau về giá trị trung bình trong các mô hình hồi quy. Về lý thuyết, quá trình phân rã này được biết đến từ quá trình phân rã của Oaxaca – Blinder (Oaxaca, 1973 và Blinder, 1973), và sự khác biệt giữa 2 nhóm là do 2 thành phần tạo ra: Một là, sự khác biệt giữa 2 nhóm mà có thể giải thích được (explained). Đây là thành phần thứ nhất trong mô hình, thành phần còn lại không thể tính vào sự khác biệt giữa 2 nhóm, tức là thành phần không giải thích (unexplained). Thành phần này thường được

sử dụng như là 1 phương pháp để nói về sự phân biệt đối xử (discrimination) giữa hai nhóm trong thị trường lao động và trong lý thuyết về phân biệt đối xử. Đồng thời, phương pháp này cũng được sử dụng rộng rãi trong những lĩnh vực kinh tế khác để so sánh sự khác biệt giữa hai nhóm. Ở Việt Nam, phương pháp phân rã này chưa được sử dụng rộng rãi trong nghiên cứu, do đó, mục tiêu của nghiên cứu này là giới thiệu về phương pháp phân rã Oaxaca – Blinder (1973) được áp dụng trong mô hình hồi quy tuyến tính và mô hình hồi quy xác suất và khả năng ứng dụng phương pháp này trong phân tích kinh tế.

<sup>1</sup> PGS.TS, Trường Đại học Mở TP.HCM.

<sup>2</sup> PGS.TS, Trường Đại học Mở TP.HCM. Email: ha.nm@ou.edu.vn

<sup>3</sup> ThS, Trường Đại học Mở TP.HCM.

## 2. Phương pháp phân rã Oaxaca – Blinder

### 2.1. Kỹ thuật phân rã Oaxaca – Blinder cho mô hình tuyến tính.

#### a. Mô hình lý thuyết

Có hai nhóm S và O và biến kết quả (biến phụ thuộc) là Y và tập hợp các biến giải thích X. Ví dụ: có 2 nhóm lao động nam và nữ, thu nhập là biến kết quả (biến phụ thuộc và các biến giải thích (biến độc lập) là giáo dục và kinh nghiệm. Hoặc, có 2 nhóm lao động trong khu vực công nghiệp và trong khu vực dịch vụ, thu nhập là biến phụ thuộc và các biến giải thích là tuổi, trình độ, và kinh nghiệm. Câu hỏi là sự khác nhau của biến phụ thuộc về giá trị trung bình giữa 2 nhóm là bao nhiêu.

Theo mô hình phân rã Oaxaca-Blinder (1973), biến phụ thuộc (Y) được biểu diễn theo phương trình sau:

$$Y = X\beta + \varepsilon \quad (1)$$

Trong đó X đại diện cho các vector của các biến giải thích,  $\beta$  là các vector hệ số hồi quy, và  $\varepsilon$  là các sai số phần dư.

Phương trình nhóm S được ký hiệu (s) là

$$Y^s = X^s\beta^s + \varepsilon^s \quad (2)$$

và nhóm O được ký hiệu là (o) có phương trình  $Y^o = X^o\beta^o + \varepsilon^o$  (3)

Trong đó  $X^s$  and  $X^o$  đại diện cho các biến giải thích của mẫu nhóm S và của O.  $\beta^s$  and  $\beta^o$  là các thông số ước lượng cho mẫu của nhóm S và nhóm O.

Giá trị trung bình của các biến số và các thông số được ước lượng tại mỗi nhóm được biểu diễn lần lượt bằng hai phương trình:

$$\bar{Y}^s = \beta_o^s + \bar{X}^s \hat{\beta}^s + u^s \quad (4)$$

$$\text{và } \bar{Y}^o = \beta_o^o + \bar{X}^o \hat{\beta}^o + u^o \quad (5)$$

Khoảng cách đối với nhóm S và nhóm O trong giá trị trung bình của biến phụ thuộc, Y (hoặc ngược lại), có thể được viết như sau:

$$\bar{Y}^s - \bar{Y}^o = (\beta_o^s - \beta_o^o) + (\bar{X}^s \hat{\beta}^s - \bar{X}^o \hat{\beta}^o) \quad (6)$$

Sau khi cộng và trừ  $\bar{X}^o \hat{\beta}^s$ , phương trình khoảng cách được biến đổi thành:

$$\bar{Y}^s - \bar{Y}^o = (\beta_o^s - \beta_o^o) + [(\bar{X}^s - \bar{X}^o) \hat{\beta}^s - \bar{X}^o (\hat{\beta}^s - \hat{\beta}^o)] \quad (7)$$

Thành phần  $(\bar{X}^s - \bar{X}^o) \hat{\beta}^s$  được giải thích như là một phần của khoảng cách giữa các biến phụ thuộc do sự khác biệt trung bình ở những đặc tính có thể quan sát được (biến số giải thích) giữa nhóm S và nhóm O. Nếu các thành phần trong nhóm S và nhóm O có cấp độ giống nhau ở biến số X, thành phần này bằng 0. Thêm vào đó, thành phần khác,  $\bar{X}^o (\hat{\beta}^s - \hat{\beta}^o)$ , được coi là phần thể hiện sự khác biệt trong các hệ số hồi quy được ước lượng (phân biệt hoặc không thể giải thích được). Nếu những hệ số này là giống nhau giữa các thành phần nhóm S và nhóm O, thành phần này bằng 0 (khoảng cách hầu như phụ thuộc vào sự khác biệt trong các đặc điểm của từng thành phần trong nhóm).

Phương trình phân rã (7) xảy ra khi cộng và trừ  $\bar{X}^s \hat{\beta}^o$  vào phương trình (6); khoảng cách được viết như sau:

$$\bar{Y}^s - \bar{Y}^o = (\beta_o^s - \beta_o^o) + [(\bar{X}^s - \bar{X}^o) \hat{\beta}^s - \bar{X}^o (\hat{\beta}^s - \hat{\beta}^o)]$$

Như vậy, sự khác nhau được diễn tả:

i) Sự khác nhau tổng thể (R):

$$R = (\beta_o^s + \bar{X}^s \hat{\beta}^s) - (\beta_o^o + \bar{X}^o \hat{\beta}^o) = E + C + U \quad (8)$$

ii) Sự khác nhau do các đặc tính tạo ra

$$(E): E = (\bar{X}^s - \bar{X}^o) \hat{\beta}^s \quad (9)$$

iii) Sự khác nhau do sự khác nhau của các hệ số (C):  $C = \bar{X}^o (\hat{\beta}^s - \hat{\beta}^o)$  (10)

iv) Thành phần không thể lý giải được (U):  $U = (\beta_o^s - \beta_o^o)$  (11)

v) Thành phần khác biệt do sự phân biệt (D):  $D = C + U$  (12)

### **b. Các bước thực hiện phân rã**

Đối với mô hình hồi quy tuyến tính, các bước thực hiện phân rã như sau:

- i) Đầu tiên, chạy hồi quy cho từng mẫu S và O
- ii) Ước lượng giá trị của  $X^s$  và  $X^o$
- iii) Tính sự khác nhau theo các đặc tính và sự khác nhau theo các hệ số
- iv) Tính sự khác nhau do thành phần không thể lý giải
- v) Tính sự khác biệt do sự phân biệt

## **2.2. Mở rộng kỹ thuật phân rã Oaxaca – Blinder cho mô hình xác suất**

### **a. Mô hình lý thuyết**

Đối với mô hình xác suất, theo mô hình phân rã Oaxaca-Blinder (1973), biến phụ thuộc (Y) được biểu diễn theo phương trình sau:

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

trong đó, X đại diện cho các vector của các biến giải thích,  $\beta$  là các vector hệ số hồi quy, và  $\varepsilon$  là các sai số phần dư.

Phương trình nhóm S được ký hiệu (s) là  $Y^s = X^s\beta^s + \varepsilon^s$

và nhóm O được ký hiệu là (o) có phương trình  $Y^o = X^o\beta^o + \varepsilon^o$

trong đó  $X^s$  and  $X^o$  đại diện cho các biến giải thích của mẫu nhóm S và của O.  $\beta^s$  and  $\beta^o$  là các thông số ước lượng cho mẫu của nhóm S và nhóm O.

Giá trị trung bình của các biến số và các thông số được ước lượng tại mỗi nhóm được biểu diễn lần lượt bằng hai phương trình:

$$\bar{Y}^s = \bar{X}^s \hat{\beta}^s \text{ và } \bar{Y}^o = \bar{X}^o \hat{\beta}^o.$$

Khoảng cách đối với nhóm S và nhóm O trong giá trị trung bình của biến phụ thuộc, Y, có thể được viết như sau:

$$\bar{Y}^s - \bar{Y}^o = \bar{X}^s \hat{\beta}^s - \bar{X}^o \hat{\beta}^o \quad (13)$$

Sau khi cộng và trừ  $\bar{X}^o \hat{\beta}^s$ , phương trình khoảng cách được biến đổi thành:

$$\bar{Y}^s - \bar{Y}^o = (\bar{X}^s - \bar{X}^o) \hat{\beta}^s - \bar{X}^o (\hat{\beta}^s - \hat{\beta}^o) \quad (14)$$

Thành phần  $(\bar{X}^s - \bar{X}^o) \hat{\beta}^s$  được giải thích như là một phần của khoảng cách giữa các biến phụ thuộc do sự khác biệt trung bình ở những đặc tính có thể quan sát được (biến số giải thích) giữa nhóm S và nhóm O. Nếu các thành phần trong nhóm S và nhóm O có cấp độ giống nhau ở biến số X, thành phần này bằng

0. Thêm vào đó, thành phần khác,  $\bar{X}^o (\hat{\beta}^s - \hat{\beta}^o)$ , được coi là phần thể hiện sự khác biệt trong các hệ số hồi quy được ước lượng (phân biệt hoặc không thể giải thích được). Nếu những hệ số này là giống nhau giữa các thành phần nhóm S và nhóm O, thành phần này bằng 0 (khoảng cách hầu như phụ thuộc vào sự khác biệt trong các đặc điểm của từng thành phần trong nhóm).

Phương trình phân rã (14) xảy ra khi cộng và trừ  $\bar{X}^s \hat{\beta}^o$  vào phương trình (13); khoảng cách được viết như sau:

$$\bar{Y}^s - \bar{Y}^o = (\bar{X}^s - \bar{X}^o) \hat{\beta}^o - \bar{X}^s (\hat{\beta}^s - \hat{\beta}^o) \quad (15)$$

Việc phân rã các biến số đầu ra tương tự như phương trình (14) là không thích hợp nếu kết quả đầu ra là nhị nguyên và các hệ số hồi quy được rút ra từ mô hình logit hoặc probit. Các hệ số ước lượng không thể được áp dụng một cách trực tiếp vào phương trình phân rã Oaxaca-Blinder chuẩn.

Fairlie (1999, 2003, và 2005) và Yun (2000 và 2004) đã mở rộng phương pháp phân rã Oaxaca-Blinder cho các mô hình phi tuyến. Phương pháp phân rã cho một phương trình phi tuyến,  $Y = F(X\hat{\beta})$ , được viết như sau:

$$\bar{Y}^s - \bar{Y}^o = \left[ \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^s)}{N^s} - \sum_{i=1}^{N^o} \frac{F(X_i^o \bar{\beta}^s)}{N^o} \right] + \left[ \sum_{i=1}^{N^o} \frac{F(X_i^o \bar{\beta}^s)}{N^o} - \sum_{i=1}^{N^o} \frac{F(X_i^o \bar{\beta}^o)}{N^o} \right] \quad (16)$$

Hoặc  $\bar{Y}^s - \bar{Y}^o = \text{Ảnh hưởng của các đặc điểm} + \text{Ảnh hưởng của các hệ số}$

$$\bar{Y}^s - \bar{Y}^o = \left[ \bar{P}(X^s \bar{\beta}^s) - \bar{P}(X^o \bar{\beta}^s) \right] + \left[ \bar{P}(X^o \bar{\beta}^s) - \bar{P}(X^o \bar{\beta}^o) \right] \quad (17)$$

trong đó biến số phụ thuộc Y là biến nhị nguyên nhận giá trị 1 nếu (là doanh nghiệp sống sót, hoặc là nam, hoặc là thành thị, ...) ,

bằng 0 nếu ngược lại.  $\bar{Y}^s = \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^s)}{N^s}$ , là xác suất dự đoán khả năng nhận giá trị 1 của

biến Y, và  $\bar{Y}^o = \sum_{i=1}^{N^o} \frac{F(X_i^o \bar{\beta}^o)}{N^o}$ , là xác suất dự

đoán khả năng nhận giá trị 0 của biến Y.  $N^s$  and  $N^o$  là số lượng mẫu của nhóm S và nhóm

O.  $\bar{P}$  đại diện cho xác suất trung bình dự đoán về kết quả nhị nguyên (Y là 1 hoặc Y là 0), và F là hàm phân phối cộng dồn từ phân phối logistic.

Hiệu ứng đặc điểm phụ thuộc vào sự khác biệt trong các xác suất dự đoán về khả năng nhận giá trị 1 hoặc 0 của biến Y khi các đặc tính của của từng thành phần trong nhóm

được sử dụng trong mô hình và vector tham số hồi quy được giữ không đổi. Nói cách khác, nó đại diện cho khoảng cách do những sự khác biệt nhóm về những phân phối của X.

Hiệu ứng hệ số hồi quy đo lường những khác biệt về xác suất dự đoán về khả năng nhận giá trị 1 hoặc 0 của biến Y khi các đặc điểm của từng thành phần trong nhóm được giữ không đổi nhưng những vector hệ số hồi quy của cả nhóm S và nhóm O được sử dụng. Nói cách khác, tác động của hệ số hồi quy đại diện cho phần khác biệt đối với nhóm mà xác định các cấp độ của Y, và nó là thành phần của khoảng cách được tạo thành do sự khác biệt nhóm về những yếu tố đầu vào không quan sát được.

Một ước lượng cân bằng hợp lý đối với phương trình (16) cho sự phân rã được viết như sau:

$$\bar{Y}^s - \bar{Y}^o = \left[ \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^o)}{N^s} - \sum_{i=1}^{N^o} \frac{F(X_i^o \bar{\beta}^o)}{N^o} \right] + \left[ \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^s)}{N^s} - \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^o)}{N^s} \right] \quad (18)$$

Trong phương trình (18), các hệ số hồi quy của nhóm S,  $\bar{\beta}^o$ , được dùng như là các trọng số cho các hiệu ứng đặc điểm trong phương trình phân rã, và phân phối của biến số độc lập,  $X^s$ , đối với nhóm O được dùng như là

các trọng số cho các hiệu ứng hệ số hồi quy.

Tương tự, thay đổi nhóm tham chiếu, phương trình thay thế cho việc phân rã được viết như sau:

$$\bar{Y}^o - \bar{Y}^s = \left[ \sum_{i=1}^{N^o} \frac{F(X_i^o \bar{\beta}^o)}{N^o} - \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^o)}{N^s} \right] + \left[ \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^o)}{N^s} - \sum_{i=1}^{N^s} \frac{F(X_i^s \bar{\beta}^s)}{N^s} \right] \quad (19)$$

Số hạng đầu tiên bên tay phải thể hiện sự khác biệt về xác suất khả năng biến Y nhận giá trị 1 hoặc 0, số hạng này thể hiện sự khác biệt của các hiệp biến X. Số hạng thứ hai là thành phần khác biệt về xác suất khả năng biến

nhận giá trị 1 hoặc 0 do sự khác biệt về hệ số hồi quy ( $\beta$ ) của cả nhóm S và O.

$$\bar{Y}^o - \bar{Y}^s = \left[ \sum_{i=1}^{N^o} \frac{F(X_i^o \beta^s)}{N^o} - \sum_{i=1}^{N^s} \frac{F(X_i^s \beta^s)}{N^s} \right]$$

Signature Valid:

Ký bởi(Signed By): HVCNBCVT

Ký ngày(Signing Date): 01/12/2024

Mô hình (16) giả thiết rằng sự phân biệt đối với nhóm O, và tỷ lệ của nhóm S là dựa vào sự so sánh giữa hai nhóm. Nói cách khác, nó phân rã sự khác biệt về tỷ lệ rời bỏ thị trường bằng cách so sánh sự khác nhau giữa tỷ lệ rời bỏ của doanh nghiệp ngoài quốc doanh (dựa vào cấu trúc rời bỏ của chúng) và tỷ lệ rời bỏ của doanh nghiệp ngoài quốc doanh (dựa vào cấu trúc rời bỏ của chúng). Hiệu ứng đặc điểm, thể hiện cho sự mở rộng những khác biệt về tỷ lệ rời bỏ giữa doanh nghiệp quốc doanh và ngoài quốc doanh, được được giải thích bởi những khác biệt không quan sát được về đặc điểm của doanh nghiệp. Hiệu ứng phần dư đo lường phần khác biệt về hệ số hồi quy và các nhân tố không quan sát được.

### b. Các bước thực hiện phân rã

Đối với mô hình xác suất, trong bài viết này đề cập đến mô hình logistic. Các bước phân rã được cho như sau:

Đầu tiên, ước lượng hồi quy logit sử dụng cho từng mẫu S và O. Sau đó, những kết quả rút ra từ các hệ số hồi quy này được sử dụng để tính toán các xác suất dự đoán mà được dùng để tính toán giá trị trung bình của các biến số độc lập:

$$X_1: F(\alpha^s + \bar{X}_1^o \beta_1^s + \bar{X}_2^s \beta_2^s + \bar{X}_3^s \beta_3^s) - F(\alpha^s + \bar{X}_1^s \beta_1^s + \bar{X}_2^s \beta_2^s + \bar{X}_3^s \beta_3^s)$$

Tương tự, các biến số  $X_2$  và  $X_3$  được viết như sau:

$$X_2: F(\alpha^s + \bar{X}_1^o \beta_1^s + \bar{X}_2^o \beta_2^s + \bar{X}_3^s \beta_3^s) - F(\alpha^s + \bar{X}_1^o \beta_1^s + \bar{X}_2^s \beta_2^s + \bar{X}_3^s \beta_3^s)$$

$$\bar{P} = \frac{1}{1 + \exp(-\bar{X} \beta)} \quad (21)$$

Tiếp theo, khoảng cách trong biến phụ thuộc giữa nhóm S và nhóm O được tính bằng sự khác biệt giữa các xác suất được dự đoán của hai nhóm, dựa vào phương trình (17).

Cuối cùng, các giá trị đối chứng được tính toán để nhận diện mức độ đóng góp của mỗi tập hợp biến tới khoảng cách của nhóm S và nhóm O đối với biến phụ thuộc Y. Mức độ đóng góp của mỗi biến số độc lập đối với khoảng cách của biến phụ thuộc giữa 2 nhóm (S và O) bằng với mức độ thay đổi trong xác suất dự đoán trung bình thông qua việc thay thế các phân phối của các biến số thuộc nhóm O với phân phối của các biến số thuộc nhóm S trong khi giữ cho phân phối của các biến độc lập khác không đổi<sup>4</sup>. Nó được mô tả như sau:

i) Các tác động do đặc điểm: giả sử có ba biến số ( $X_1$ ,  $X_2$ , and  $X_3$ ).

Ảnh hưởng của biến số độc lập  $X_1$  tới khoảng cách của biến phụ thuộc giữa 2 nhóm (S và O) được mô tả như sau:

<sup>4</sup> Trường hợp phi tuyến khác với trường hợp tuyến tính. Đó là: đóng góp độc lập của các biến  $X_1$ ,  $X_2$ , và  $X_3$  phụ thuộc vào giá trị của biến khác. Điều này ám chỉ rằng sự chọn lựa của 1 biến như  $X_1$ ,  $X_2$ , hoặc  $X_3$  (hoặc thứ tự của chuyển đổi các đóng góp) là quan trọng tiềm năng trong việc tính khoảng cách của biến phụ thuộc giữa 2 nhóm.

$$_3: F(\alpha^s + \bar{X}_1^o \beta_1^s + \bar{X}_2^o \beta_2^s + \bar{X}_3^o \beta_3^s) - F(\alpha^s + \bar{X}_1^s \beta_1^s + \bar{X}_2^s \beta_2^s + \bar{X}_3^s \beta_3^s)$$

Tổng của  $X_1$ ,  $X_2$ , and  $X_3$ , tác động do đặc điểm, được viết như sau:

$$F(\alpha^s + \bar{X}_1^o \beta_1^s + \bar{X}_2^o \beta_2^s + \bar{X}_3^o \beta_3^s) - F(\alpha^s + \bar{X}_1^s \beta_1^s + \bar{X}_2^s \beta_2^s + \bar{X}_3^s \beta_3^s) \quad (22)$$

ii) Tác động do hệ số: cho ba biến số ( $X_1$ ,  $X_2$ , và  $X_3$ )

Ảnh hưởng của biến số độc lập  $X_1$  tới khoảng cách rời bỏ được mô tả như sau:

$$X_1: F(\alpha^o + \bar{X}_1^o \beta_1^s + \bar{X}_2^o \beta_2^o + \bar{X}_3^o \beta_3^o) - F(\alpha^o + \bar{X}_1^s \beta_1^o + \bar{X}_2^s \beta_2^o + \bar{X}_3^s \beta_3^o)$$

Tương tự, biến số  $X_2$  và  $X_3$  được cho như sau:

$$X_2: F(\alpha^o + \bar{X}_1^o \beta_1^s + \bar{X}_2^o \beta_2^s + \bar{X}_3^o \beta_3^o) - F(\alpha^o + \bar{X}_1^s \beta_1^s + \bar{X}_2^s \beta_2^o + \bar{X}_3^s \beta_3^o)$$

$$X_3: F(\alpha^o + \bar{X}_1^o \beta_1^s + \bar{X}_2^o \beta_2^s + \bar{X}_3^o \beta_3^s) - F(\alpha^o + \bar{X}_1^s \beta_1^s + \bar{X}_2^s \beta_2^s + \bar{X}_3^s \beta_3^o)$$

Tổng của  $X_1$ ,  $X_2$ , và  $X_3$ , tác động do hệ số, được mô tả như sau:

$$F(\alpha^o + \bar{X}_1^o \beta_1^s + \bar{X}_2^o \beta_2^s + \bar{X}_3^o \beta_3^s) - F(\alpha^o + \bar{X}_1^s \beta_1^o + \bar{X}_2^s \beta_2^o + \bar{X}_3^s \beta_3^o) \quad (23)$$

Lấy phương trình (22) trừ phương trình (23); khoảng cách của biến phụ thuộc giữa 2 nhóm là

$$F(\alpha^s + \bar{X}_1^s \beta_1^s + \bar{X}_2^s \beta_2^s + \bar{X}_3^s \beta_3^s) - F(\alpha^o + \bar{X}_1^o \beta_1^o + \bar{X}_2^o \beta_2^o + \bar{X}_3^o \beta_3^o) \quad (24)$$

### 3. Khả năng ứng dụng phương pháp Oaxaca – Blinder trong kinh tế

Phương pháp phân rã Oaxaca – Blinder (1973) đã được phát triển và ứng dụng rộng rãi trong nhiều khía cạnh khác nhau trong nghiên cứu kinh tế. Một số ứng dụng trong lĩnh vực nghiên cứu kinh tế như:

- Ứng dụng trong nghiên cứu kinh tế lao động, đặc biệt là tiền lương hoặc thu nhập và sự khác biệt việc làm giữa 2 nhóm khác nhau. Nghiên cứu tiên phong là nghiên cứu về sự khác biệt thu nhập giữa người da đen và người da trắng (Blinder, 1973) và sự khác biệt thu nhập giữa nam và nữ (Oaxaca, 1973; Simon và ctg, 1999; Michelle, 2002; Solomon, 2004; Donna, 2004; Julie và Melinda, 2007; Luiz, 2007), sự khác biệt thu nhập giữa khu vực công và khu vực tư ở Ai Cập (Bank of Greece, 2003; Aysit, 1999), sự khác biệt thu nhập giữa người Trung quốc và người Ấn độ (Olivier và ctg, 2007), sự khác nhau về năng suất (Jessica, 2006); sự khác nhau thu nhập giữa khu vực

nông thôn và thành thị ở Trung Quốc (Jonathan và Terry, 2002); sự khác nhau thu nhập trong nội bộ khu vực công nghiệp (Myeong-Su Yun, 2007); Sự khác biệt việc làm theo giới tính ở Chile (David và ctg, 2005); giáo dục tạo nên sự khác biệt thu nhập theo giới tính (Sami, 2006), ...

- Ứng dụng trong kinh tế vi mô và tài chính: Sự khác nhau trong tỷ lệ sống sót của doanh nghiệp nhà nước và doanh nghiệp không nhà nước (Nguyễn Minh Hà, 2011); Sự khác nhau trong khoản vay tín dụng giá cao (Chau Do và Irina, 2007); sự khác biệt về sở hữu cổ phiếu theo sắc tộc (Cong Wang, 2007); sự khác nhau trong tỷ lệ chi trả cổ tức (Jesus và Candra), ...

- Ứng dụng trong phúc lợi xã hội: Nghiên cứu về sự khác nhau về phúc lợi xã hội giữa nông thôn và thành thị (Maureen và Sonya, 2003); Sự khác nhau về tài sản theo giới tính (Eva và ctg, 2008), phân phối thu nhập trong hộ gia đình (Francois và ctg, 2008), ...

- Ứng dụng trong vốn con người: Nghiên cứu về sự khác nhau về phát triển vốn con người (Sylvie và ctg, 2007); sự khác biệt về chăm sóc y tế theo sắc tộc (Kirby và ctg, 2006), ...

Còn nhiều nghiên cứu ứng dụng trong các khía cạnh của lĩnh vực kinh tế được áp dụng phương pháp phân rã này trên thế giới.

Một ví dụ về sự khác nhau về thu nhập theo giới tính bằng cách tính toán bằng phương pháp Oaxaca-Blinder (1997) cho mô hình tuyến tính như sau:

Giả sử thu nhập (TN) bị phụ thuộc tuyến tính vào các biến giáo dục (GD), kinh nghiệm (KN) và công nhân có tham gia vào hiệp hội hay không (HH). Đặt Nam có ký hiệu là “m” và Nữ có ký hiệu là “f”. Phương trình thu nhập theo giới tính nam và nữ như sau:

$$TN_m = a_m + b_m GD_m + c_m KN_m + f_m HH_m \quad (25)$$

$$TN_f = a_f + b_f GD_f + c_f KN_f + f_f HH_f \quad (26)$$

Trong đó,  $GD_m$ ,  $KN_m$ ,  $HH_m$  đo lường số năm giáo dục, số năm kinh nghiệm và tham gia hiệp hội của lao động nam.

$b_m$ ,  $c_m$ ,  $f_m$ ,  $b_f$ ,  $c_f$  và  $f_f$  là hệ số hồi quy đặt được khi chạy hồi quy cho từng nhóm mẫu tách rời nhau.

Sự khác biệt thu nhập trung bình theo giới tính giữa nam và nữ như sau:

$$TN_m - TN_f = a_m + b_m GD_m + c_m KN_m + f_m HH_m - (a_f + b_f GD_f + c_f KN_f + f_f HH_f) \quad (27)$$

Thêm và trừ bớt  $b_m GD_f$ ,  $c_m KN_f$ ,  $f_m HH_f$ , ta có:

$$TN_m - TN_f = b_m(GD_m - GD_f) + c_m(KN_m - KN_f) + f_m(HH_m - HH_f) + (a_m - a_f) + (b_m - b_f)GD_f + (c_m - c_f)KN_f + (f_m - f_f)HH_f \quad (28)$$

Từ phương trình (28), 3 nội dung đầu của vế phải là thành phần của sự khác biệt thu nhập do sự khác biệt bởi các đặc tính trung bình giữa nam và nữ (gọi là thành phần có khả năng giải thích). Nếu lao động nam và nữ có cùng mức giáo dục, kinh nghiệm và cùng tham gia vào hiệp hội (hoặc cùng không tham gia vào hiệp hội) thì 3 nội dung này sẽ bằng 0. Nếu lao động nam có giáo dục cao hơn, có kinh nghiệm nhiều hơn và tỷ lệ tham gia hiệp hội nhiều hơn thì khoảng cách khác biệt này là dương ( $>0$ ) được giải thích bởi những yếu tố này.

Bốn nội dung cuối của phương trình (28) là thành phần của sự khác biệt thu nhập do sự phân biệt hoặc không được giải thích. Vì sự khác biệt về giới tính nên các đặc tính của giới tính được xem xét khi trả thu nhập trong thị trường lao động. Nếu  $a$ ,  $b$ ,  $c$  và  $f$  là giống nhau giữa nam và nữ thì 4 nội dung trong phương trình này bằng 0, thì khoảng cách khác biệt thu nhập hoàn toàn là do sự khác biệt bởi các đặc tính.

Cụ thể,

Kết quả phương trình hồi quy theo thu nhập của nam và nữ như sau:

$$TN_m = -1.34 + 0.078GD_m + 0.27KN_m + 3.76HH_m \quad (29)$$

$$TN_f = -1.80 + 0.70GD_f + 0.21KN_f + 4.62HH_f \quad (30)$$

Giá trị trung bình của mẫu như sau:

	Nam	Nữ	Khoảng cách (Gap)
TN	\$14.74	\$11.69	\$3.05
Năm GD	14.15	14.56	
Năm KN	13.86	9.32	
Tỷ lệ tham gia HH	0.32	0.28	

Phân rã như sau:

$$TN_m - TN_f = b_m(GD_m - GD_f) + c_m(KN_m - KN_f) + f_m(HH_m - HH_f) + (a_m - a_f) + (b_m - b_f)GD_f + (c_m - c_f)KN_f + (f_m - f_f)HH_f$$

$$TN_m - TN_f = 0.78(14.15 - 14.56) + 0.27(13.86 - 9.32) + 3.76(0.32 - 0.28) + (-1.34 + 1.8) + (0.78 - 0.70)14.56 + (0.27 - 0.21)9.32 + (3.76 - 4.62)0.28$$

$$3.05 = 1.07 + 1.99$$

Tổng khoảng cách (gap) = Phần có thể giải thích và phần không thể giải thích. Điều này có nghĩa là  $(1.07/3.05) = 35\%$  của khoảng cách thu nhập theo giới tính là được giải thích bởi sự khác biệt bởi giáo dục, kinh nghiệm và tỷ lệ tham gia hiệp hội giữa lao động nam và nữ.

Vì vậy,  $(1.99/3.05) = 65\%$  của khoảng cách thu nhập theo giới tính là không thể giải

thích. Điều này có thể là chúng ta cần thêm các biến giải thích khác trong các biến độc lập chúng ta đưa vào bên cạnh 3 biến GD, HN và HH hoặc/và chúng ta xem như phần này là thành phần phân biệt đối xử do thị trường lao động tạo ra

#### 4. Kết luận

Oaxaca – Blinder (1973) được sử dụng rộng rãi trên thế giới để nghiên cứu sự khác nhau trong giá trị trung bình của hai nhóm và có thể sử dụng với mô hình hồi quy tuyến tính hoặc mô hình hồi quy xác suất. Phương pháp này không chỉ cho thấy các yếu tố tác động đến biến phụ thuộc, mà còn cho thấy các yếu tố tạo nên sự khác biệt giữa hai nhóm. Phương pháp này được ứng dụng trong phân tích ở nhiều vấn đề kinh tế. Nghiên cứu này là giới thiệu phương pháp để làm cơ sở cho nghiên cứu những vấn đề kinh tế ở Việt Nam.

### TÀI LIỆU THAM KHẢO

- Aysit, T. (1999). Public – Private Employment Choice, Wage Differentials and Gender in Turkey. Có thể download từ [http://www.econ.yale.edu/growth\\_pdf/cdp797.pdf](http://www.econ.yale.edu/growth_pdf/cdp797.pdf)
- Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, Vol. 8, No. 4, pp. 436-455
- Chau Do and Irina, P. (2007). Explaining the Growth of Higher – Priced Loans in HMDA: A Decomposition Approach. *The Journal of Real Estate Research*, 29, 4.
- Cong Wang (2007). Racial/Ethnic Disparities in Stock Ownership: A Decomposition Analysis. *Consumer Interests Annual*, Vol 53, 2007
- David, B., Cloudia, S and Sergio, U. (2005). Gender Discrimination and Economic Outcomes in Chile. Departamento De Economia Universidad De Chile.
- Donna, K. G (2004). Gender Differences in Salary and Promotion in Political Science” American Political Science Association. Có thể download từ <http://apsanet3b.inetu.net/imgtest/gintherAPSA.pdf>
- Dorothe, B. , and Michael. G. (2001). The unequal distribution of unequal pay – An empirical analysis of the gender wage gap in Switzerland. *Empirical Economics*. May 2001, Vol.26, Iss.2, pp 407-427
- Fairlie, R. W. (1999). The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment. *Journal of Labour Economics*, Vol. 17. No. 1, pp.80-108
- Fairlie, R. W. (2005). An Extension of the Blinder-Oaxaca decomposition technique to Logit and Probit Models. *Journal of Economic and Social Measurement* 30(2005), 305-316.



- François, B., Francisco, H. G. F, and Phillippe G. L. (2008). Beyond Oaxaca–Blinder: Accounting for differences in household income distributions. *The Journal of Economic Inequality*. June 2008, Volume 6, Iss. 2, pp 117-148
- Jessica, W. K. (2007), “Gender Gap in Income and Productivity of Obstetricians and Gynecologists”, *Obstetric and Gynecology*, Vol 109, No5, 2007
- Jonathan, M. and Terry, S. (2002). Rethinking Inequility Decomposition with Evidence from Rural China. *The Economic Journal*, 112, pp.93-106
- Julie, L. H. and Melinda, M. P. (2007). The Role of Labour Market Intermittency in Explaining Gender Wage Disffentials. *Working Paper* 2007, 1, Federal Reserve Bank of Altanta.
- Kirby, J. B., Taliaferro, G., Zuvekas, S. H. ( 2006). Explaining Racial and Ethnic Disparities in Health Care. *Medical Care*. May 2006, Vol. 44, Iss.5 - pp I-64-I-72
- Luiz, G. S. and Elaine, T. P. (2007). Using normalized equations to solve the indetermination problem in the Oaxaca-Blinder decomposition: an application to the gender wage gap in Brazil. *Rev. Bras. Econ.* vol.61, no.4
- Maureen, K. and Sonya, K. H (2003). Rural/Urban Welfare Program and Labour Force Participation. *American Journal of Agricultural Economic*, 85 (5), 914-927
- Michelle, A. G. (2002). The Gender Pay Gap in the New Zealand Public Service. *Working Papers*, The State Services Commission.
- Myeong-Su Yun (2007). Revisiting Inter-Industry Wage Differentials and the Gender Wage Gap: Another Identification Problem and an Identical Solution. Có thể download từ <http://ftp.iza.org/dp2427.pdf>
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, Vol. 14, No. 3, pp. 693-709
- Olivier, B., Sumon, R. B., Manisha, C., Zhong, Z. (2007). Earnings Diferences between Chinese and India Wage Eaners, 1987-2004: Trends and the Impact of Education. Có thể download từ [http://www.iza.org/conference\\_files/prizeconf2007/zhao\\_z1966.pdf](http://www.iza.org/conference_files/prizeconf2007/zhao_z1966.pdf)
- Sami, N. (2006). Type of Education and the Gender Wage Gap. *Discussion Paper* No. 128. Helsinki School of Economics.
- Simon, A., John, H. and Pramila, K. (1999). The Gender Wage Gap in Three African Countries. The University of Chicago, *Economic Development and Cultural Change*.
- Solomon, W. P. (2004). How the Human Capital Model Explains Why the Gender Wage Gap Narrowed. Có thể download từ <http://ftp.iza.org/dp1102.pdf>
- Yun, M. S. (2000). Decomposition Analysis for a Binary Choice Model. *IZA Discussion Paper No. 145*, April 2000.
- Yun, M. S. (2004). Decomposing Differences in the First Comment. *Economics Letters* 82 (2004) 275 – 280.