Title	Estimation of discrimination coefficients in male-female wage differentials
Sub Title	
Author	ホリ, ハルヒコ(Hori, Haruhiko)
Publisher	Society of Business and Commerce, Keio University
Publication year	1996
Jtitle	Keio business review Vol.33, (1996.) ,p.53- 60
JaLC DOI	
Abstract	In this paper I discuss the theoretical framework of estimation of discrimination coefficients in male-female wage differentials, in order to estimate discrimination coefficients in Japan. For the theoretical framework of estimating discrimination coeffic
Notes	Festschrift for Prof. Yoko Sano
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AA00260481-19960001-00704503

KEIO BUSINESS REVIEW
No.33, 1996
pp.53-60.

Final version received on 26th February 1997

Estimation of Discrimination Coefficients in Male-Female Wage Differentials

By

Haruhiko Hori

Abstract

In this paper I discuss the theoretical framework of estimation of discrimination coefficients in male-female wage differentials, in order to estimate discrimination coefficients in Japan. For the theoretical framework of estimating discrimination coefficients, I expound on the Oaxaca decomposition which has been commonly utilized and then I introduce Neumark's theoretical framework which is a modified model of Oaxaca's. On the basis of Neumark's theoretical framework, I estimate Japan's discrimination coefficients. Male-female wage differentials as compared by average wages represent 62 percent, of which 22 percent, or about a third of the total wage disparity, account for wage differentials arising from discrimination.

Key Words

discrimination, wage differentials between the sexes, labor market, productivity differentials, wage function

Introduction

In this paper I will discuss the discrimination-based gap in wage differentials between the sexes. Looking at wages which are normally observable for males and females, we find wages for women amount to around 60 percent of those for men. The disparity in wages between men and women stands at about 40 percent. Comparison of wage differentials between males and females is made in terms of the mean value. If there are differentials of productivity between men and women, they are naturally incorporated in male-female wage differentials measured by the average value. Also, due to differences in the types of jobs in which males and females are engaged, wage differentials based on gaps in the distribution of job types are incorporated into the male-female wage disparity measured by the mean value. Any differences in size of the enterprise and industry in which men and women are engaged should be reflected in male-female wage differentials measured by the average value. In this paper I will define the residual observable, even after differences due to individual attributes or distribution of job types have been controlled, as wage differentials stemming from discrimination.

I will introduce the Oaxaca analysis, which is a theoretical framework for estimating discrimination coefficients, in Section 2 and a new analysis framework to modify the Oaxaca analysis in Section 3. Then in Section 4 I will estimate differential coefficients in Japan on the basis of models developed in Section 3. Finally, I will give a brief summary in Section 5.

Oaxaca Analysis of Male-Female Wage Differentials

There are numerous positive analyses of discrimination-based differentials in male-female wage differentials. Here I will present an analysis framework now in general use and will first take up Oaxaca's research work which laid a foundation for subsequent research. Oaxaca (1973) grouped wage differentials between the sexes into a portion stemming from the productivity gap between men and women and a portion other than that, i.e. a portion resulting from discrimination, to measure the portion due to discrimination.

If we let W denote wages, X the vector of an explanation variable affecting wages, β coefficient corresponding to the explanation variable and m and f subscripts for males and females, respectively, male and female wage functions can be expressed as follows.

$$\ln W_m = \beta X_m \qquad (1)$$
$$\ln W_f = \beta X_f \qquad (2)$$

Here if we consider a sample average of W and indicate the wage disparity measured in logarithm, male-female wage differentials are decomposed as follows:

$$\overline{\ln W_m} - \overline{\ln W_f} = (\overline{X_m} - \overline{X_f})'\beta_m + (\beta_m - \beta_f)'\ \overline{X_f} \qquad (3)$$

where $\overline{\ln W_m}$ is an average wage for men measured in logarithm and $\overline{\ln W_f}$ that for women. In addition, $\overline{X_m}$ and $\overline{X_f}$ represent the mean values of male and female explanation variants.

$(\overline{X_m} - \overline{X_f})'\beta_m$ is the value of a gap in individual attributes between the sexes evaluated in the coefficient of males, and may be considered as the portion of wage differentials based on the male-female productivity differential. $(\beta_m - \beta_f)'\ \overline{X_f}$ denotes a differential resulting from a coefficient gap in wage functions between the sexes. The male-female wage differentials should result in male-female productivity differentials if the difference between men and women in job types and industrial distribution are controlled, and if there is no discrimination in the labor market. Thus, the male-female differentials due to wage coefficients can be considered discrimination-based wage differentials.

Next, the discrimination coefficient, D_{mf}, in the labor market is defined as follows:

$$D_{mf} = (W_m / W_f - W_m{}^0 / W_f{}^0) \ / \ (W_m{}^0 / W_f{}^0) \qquad (4)$$

where W_m/W_f denotes the male-female wage ratio observed and $W_m{}^0/W_f{}^0$ the male-female wage ratio assuming that there is no discrimination in the labor market.

As defined in equation (4), on the basis of the male-female wage ratio supposing that there is no discrimination in the labor market the discrimination coefficient is an indicator for the percentage of the difference between the wage ratio observed and the wage ratio when there is no discrimination. It is also a proxy indicator for discrimination in the labor market.

If we assume the logarithm of equation (4),

$$\ln (D_{mf}+1)=\ln (W_m/ W_f)-\ln (W_m^0/ W_f^0) \qquad (5)$$

If we consider a sample average of wages,

$$\overline{\ln W_m}-\overline{\ln W_f}=\ln (W_m^0/ W_f^0)+\ln (D_{mf}+1) \qquad (6)$$

Comparing equations (3) and (6) we have the following corresponding relation.

$$\ln (W_m^0/ W_f^0) = (\overline{X_m}-\overline{X_f})' \beta_m \qquad (7)$$
$$\ln (D_{mf}+1) =(\beta_m-\beta_f)'\overline{X}_f \qquad (8)$$

By estimating wage functions for males and females and calculating $(\beta_m-\beta_f)' \overline{X_f}$, we can obtain the value of the discrimination coefficient D_{mf}. But this method involves one important problem. Equation (3) may be rewritten as follows.

$$\overline{\ln W_m}-\overline{\ln W_f}=(\overline{X_m}-\overline{X_f})'\beta_f+(\beta_m-\beta_f)' \overline{X_m} \qquad (9)$$

Here we will again write down equation (3).

$$\overline{\ln W_m}-\overline{\ln W_f}=(\overline{X_m}-\overline{X_f})'\beta_m+(\beta_m-\beta_f)' \overline{X_f} \qquad (3)$$

It is quite obvious from a comparison of equations (3) and (9) that the productivity differential can be evaluated by β_f or β_m. There is no guarantee that the values of β_f and β_m are the same. It is, rather, considered natural that β_f and β_m are different from each other according to the estimated results of wage functions and therefore, it is predicted that the discrimination coefficient will change depending on whether it is evaluated by β_f or by β_m. Neumark (1988) presented a solution to this problem. I will expound on this in the next section.

Framework of New Analysis

In this section, on the basis of Oaxaca and Ransom (1994) I will discuss the framework of estimating new discrimination coefficients presented by Neumark. Male-female wage differentials are defined as follows:

$$G_{mf}=W_m/W_f-1 \qquad (10)$$

where G_{mf} represents male-female wage differentials and W_m and W_f denote male and female wages, respectively. In addition, subscripts m and f represent male and female, respectively.

If we let Q_{mf} represent the male-female productivity disparity,

$$Q_{mf}=W_m^0 / W_f^0-1 \qquad (11)$$

where the superscript 0 describes the state where there is no discrimination in the labor market.

Next, as expressed by means of equation (4), the discrimination coefficient in the job market D_{mf} is defined as follows:

$$D_{mf}=(W_m/ W_f-W_m^0/ W_f^0) / (W_m^0 / W_f^0) \qquad (4)$$

Using equations (10), (11) and (4) and taking natural logarithms, we can obtain equation (12).

$$\ln (G_{mf}+1)=\ln (D_{mf}+1)+\ln (Q_{mf}+1) \qquad (12)$$

Equation (12) indicates that male-female wage differentials represented by logarithm can be decomposed into discrimination coefficients (expressed by logarithm) and productivity differentials (represented by logarithm). Also, assuming the natural logarithm of equation (4),

$$\begin{aligned}\ln (D_{mf}+1) &= \ln (W_m/W_f)-\ln (W_m{}^0/W_f{}^0)\\ &= \ln (W_m/W_m{}^0)+\ln (W_f{}^0/W_f)\\ &= \ln (\delta_{m0}+1) + \ln (\delta_{0f}+1) \qquad (13)\end{aligned}$$

Here, $\delta_{m0}=W_m / W_m{}^0-1$ which denotes a gap between the current male wages and those when there is no discrimination in the labor market. In other words, it describes the benefits men enjoy from discrimination in the labor market.

Likewise, $\delta_{0f}=W_f{}^0/W_f-1$, which denotes the difference between the wages women receive when there is no discrimination in the labor market and the current wages for females. To put it another way, it is the disadvantage women sustain because of discrimination in the labor market. Substituting equation (13) for equation (12),

$$\ln (G_{mf}+1) = \ln (\delta_{m0}+1) + \ln (\delta_{0f}+1) + \ln (Q_{mf}+1) \qquad (14)$$

Thus, it is found that depending on discrimination factors in the labor market, male-female wage differentials can be decomposed into three different parts, the benefits men enjoy, the disadvantage women sustain and male-female productivity differentials.

Next, we will examine decomposition of equation (14) in the framework of wage functions used in ordinary estimation. Respective wage functions for men and women estimated by using cross-section data are:

$$\overline{\ln W_m}=\overline{X_m}'\beta_m \qquad (15)$$
$$\overline{\ln W_f}=\overline{X_f}'\beta_f \qquad (16)$$

where W denotes the average wage, X the vector of the mean value for explanatory variables and β the vector of estimated coefficients. Considering their corresponding relation to equation (10),

$$\begin{aligned}\ln (G_{mf}+1) &= \ln (W_m/W_f)\\ &= \overline{\ln W_m}-\overline{\ln W_f}\\ &= \overline{X_m}'\beta_m-\overline{X_f}'\beta_f \qquad (17)\end{aligned}$$

Furthermore, considering their corresponding relation to equation (14),

$$\ln (G_{mf}+1) = \overline{X_m}'(\beta_m-\beta^*)+\overline{X_f}'(\beta^*-\beta_f)+(\overline{X_m}-\overline{X_f})'\beta^* \qquad (18)$$

where β^* denotes the estimated coefficient of wage functions when there is no discrimination in the labor market.

As is apparent from its corresponding relation with equation (14), the first, second and third terms of the right member describe the wage benefits males enjoy $\ln (\delta_{m0}+1)$, the wage disadvantage females sustain $\ln (\delta_{0f}+1)$ and male-female productivity differentials $\ln (Q_{mf}+1)$, respectively .

When there is no discrimination observable in the labor market, the market evaluation for men and women who possess the same attributes becomes equal and thus $\beta_m = \beta^* = \beta_f$ is predicted in relation to the wage coefficient. Consequently, male-female wage differentials all arise from male-female productivity differentials.

One important problem arises out of the estimation of equation (18). β^* represents the coefficient of the wage function when there is no discrimination in the labor market, and unless clear definition of the wage structure when there is no discrimination in the job market is assumed, it will be impossible to establish β^*.

Neumark presents one framework for estimating β^*. He theoretically shows that the wage coefficient β^* when there is no discrimination in the labor market agrees with the wage coefficient obtained by pooling male and female samples together and applying the ordinary least squares method. In this paper, we will estimate the discrimination differential in the male-female wage differentials on the basis of the method presented by Neumark.

Estimation of Discrimination Coefficient in Japan

Estimation Model and Data Used

We will use the Ministry of Labour's 1991 Basic Survey on the Wage Structure for this analysis, with the sample being limited to firms with a workforce of 100 employees or more. The explanatory variables will be specified to estimate the wage functions expressed in equations (15) and (16). In estimating wage functions, dummy variables for firm size, job type and industry type will be utilized, in addition to human capital variables normally used (years of education, years of tenure, years of experience in the labor market, etc). To be more specific, the following equation will be estimated:

$$\ln W = f\,(S, T, T^2, EXP, EXP^2, OCCEXP, SIZE, OCC, IND) \qquad (19)$$

For the dependent variable, the hourly wage (W) will be obtained by dividing scheduled earnings by scheduled hours worked. The natural logarithm will be utilized for estimation.

The explanatory variables are defined as follows.

Years of education (S) for those with middle-school diplomas, those with high-school diplomas, those with junior-college and technical-college diplomas, and those with college and university diplomas are set at 9, 12, 14 and 16, respectively, to estimate wage functions. Wages are predicted to increase with a rise in the years of education.

The increase in years of tenure with the current employer (T) is predicted to move the worker's productivity as well as wages upwards. The term T^2 will also be included. It is predicted that the effects of T^2 on ln W will be negative.

Years of experience in the labor market (EXP) means the total number of years in the labor market since graduation. More specifically, years of experience in the labor market is defined as age−years of education−6. In addition, the term EXP^2 is also incorporated.

Years of job experience (OCCEXP) means the number of years in which the employee has been engaged in the current occupation. Years of experience were grouped into categories of 'under one year,' '1 through 4 years,' '5 through 9 years,' '10 through 14 years,' and '15 years or more.' Thus, job experience will be dealt with by dummy variables. Those with under one year of experience are the reference category.

OCCEXP1	1-4 years of experience=1, others=0
OCCEXP2	5-9 years of experience=1, others=0
OCCEXP3	10-14 years of experience=1, others=0
OCCEXP4	15 years or more experience=1, others=0

Firm size (SIZE) is classified into five different categories, firms with 100-299 employees, firms with 300-499 employees, firms with 500-999 employees, firms with 1,000-4,999 employees and firms with 5,000 or more employees. The reference category is firms with 100-299 employees.

SIZE1	firms with 300-499 employees=1, others=0
SIZE2	firms with 500-999 employees=1, others=0
SIZE3	firms with 1,000-4,999 employees=1, others=0
SIZE4	firms with 5,000 or more employees=1, others=0

In the samples, occupations (OCC) are classified into small groups, and were reclassified into large groups on the basis of the Occupational Classification compiled by the Ministry of Labour. The reference category is production workers.

OCC1	professional and technical workers=1, others=0
OCC2	managers and officials=1, others=0
OCC3	clerical and related workers=1, others=0
OCC4	sales workers=1, others=0
OCC5	service workers=1, others=0
OCC6	transportation and telecommunications workers=1, others=0
OCC7	craftsmen=1, others=0

Industries (IND) are classified into eight categories, with mining as the referent.

IND1	construction=1, others=0
IND2	manufacturing=1, others=0
IND3	transportation and telecommunications=1, others=0
IND4	wholesale and retail trade and eating and drinking places=1, others=0
IND5	finance and insurance=1, others=0
IND6	real estate=1, others=0
IND7	services=1, others=0

Estimation Results

The estimation results of the wage functions in Table 1 largely agree with the theoretical results. As for the male wage function, all the coefficients are significant at a level of 1 percent. However, figures for the years-of-job experience dummies are all negative, and the longer the years of job experience, the larger the negative values. This means that wages are lower as years of job experience become longer, the reason for which is difficult to explain.

The female wage function coefficients are all significant at a level of 1 percent, excluding the occupation dummies for professional and technical workers and service workers, and the industry dummy for services. The female coefficients for education and length of employment (rate of return) are larger than those for males; however, the female coefficient for years of experience in the labor market (rate of return) is considerably smaller than that for males.

The coefficients of the pooled wage function of male-female samples are all significant, excluding the occupation dummies for service workers and craftsmen and the industry dummy for finance and insurance.

Let us next look at the results of the discrimination coefficient, which are shown in Table 2. The male-female wage differential expressed in natural logarithm is 48

Table 1. Wage Regression

Variable	Male coefficients	Female coefficients	Pooled coefficients
S	.0549(2.09E−04)	.0603(5.22E−04)	.0662(2.11E−04)
T	.0157(1.88E−04)	.0273(3.42E−04)	.0260(1.76E−04)
T^2	−2.46E−05(4.81E−06)	−6.21E−05(1.04E−05)	−9.29E−05(4.65E−06)
EXP	.0393(1.85E−04)	.0087(2.62E−04)	.0268(1.62E−04)
EXP^2	−6.25E−04(3.97E−06)	−2.19E−04(6.08E−06)	−4.77E−04(3.59E−06)
OCCEXP1	−.1515(.0048)	.0444(.0137)	−.0319(.0051)
OCCEXP2	−.1592(.0049)	.1065(.0138)	−.0357(0051)
OCCEXP3	−.1738(.0050)	.1005(.0140)	−.0575(0052)
OCCEXP4	−.1860(.0048)	.0859(.0140)	−.0588(.0051)
SIZE 1	.0466(.0014)	.0607(.0139)	.0564(.0014)
SIZE 2	.0910(.0014)	.1031(.0024)	.1021(.0013)
SIZE 3	.1535(.0011)	.1441(.0021)	.1639(.0011)
SIZE 4	.2204(.0011)	.2706(.0022)	.2350(.0011)
OCC1	.2549(.0054)	.0264(.0139)	.0925(.0054)
OCC2	.1942(.0014)	.3659(.0103)	.2383(.0015)
OCC3	.0394(.0060)	−.0775(.0150)	−.0682(.0061)
OCC4	.1007(.0056)	.0716(.0138)	.0204(.0053)
OCC5	.0332(.0060)	.0275(.0142)	−2.35E−04(.0058)
OCC6	.1354(.0049)	−.0698(.0160)	.0986(.0052)
OCC7	.1151(.0048)	−.2560(.0138)	−.0034(.0050)
IND1	−.0565(.0025)	−.0800(.0077)	−.0582(.0026)
IND2	−.1254(.0017)	−.1551(.0054)	−.1637(.0018)
IND3	−.1704(.0018)	−.0276(.0060)	−.1559(.0020)
IND4	−.1090(.0021)	−.1649(.0058)	−.1510(.0022)
IND5	.0698(.0021)	.0590(.0055)	.0029(.0021)
IND6	.0384(.0048)	.0285(.0098)	.0181(.0047)
IND7	−.0394(.0020)	−.0092(.0056)	−.0787(.0021)
(CONSTANT)	6.0866(.0035)	5.9758(.0089)	5.9360(.0035)
SAMPLE SIZE	467,418	167,059	634,477
Adjusted R^2	.6435	.5267	.6224

Note: The figures in parentheses denote standard errors. All the coefficients, excluding OCC1, OCC5 and IND7 of the female wage function and OCC5, OCC7 and IND5 of the pooled wage function, are significant at a level of 1 percent. The above variables are not significant even at a level of 5 percent.

Table 2. Decomposition of the Male/Female Wage Differential

$\ln(G_{mf}+1) = 0.4804$	$G_{mf} = 0.6168$
$\ln(D_{mf}+1) = 0.1988$	$D_{mf} = 0.2199$
	$\delta_{mo} = 0.0509$
	$\delta_{of} = 0.1608$
$\ln(Q_{mf}+1) = 0.2817$	$Q_{mf} = 0.3253$

percent, of which approximately 20 percent is the wage disparity arising from discrimination and the remaining 28 percent is the wage disparity resulting from the male-female productivity differential.

The male-female wage differential as measured by average wages is 62 percent, of which 22 percent is caused by discrimination. The components of this discrimination-based wage differential are 5 percent for the benefits men enjoy and 16 percent for the disadvantage women sustain. The disparity arising from male-female productivity differences represents 33 percent. In comparison, the U.S. results obtained by Oaxaca and Ransom (1994), showed a male-female wage disparity as measured by average wages of 35 percent, or around half the size of Japan's. The discrimination coefficient, however, is 22 percent for both countries. This is due to the relatively smaller productivity disparity in the U.S. It is interesting to note that the value for discrimination coefficients is the same in both Japan and the U.S. in which the labor market system is different.

Conclusion

In this paper I have discussed the theoretical framework of estimation of discrimination coefficients in male-female wage differentials, in order to estimate discrimination coefficients in Japan. For the theoretical framework of estimating discrimination coefficients, I have expounded on the Oaxaca decomposition which has been commonly utilized and then have introduced Neumark's theoretical framework which is a modified model of Oaxaca's. On the basis of Neumark's theoretical framework, I have estimated Japan's discrimination coefficients. The male-female wage differential as measured by average wages is 62 percent, of which 22 percent, or about a third of the total wage disparity, accounts for wage differentials arising from discrimination. It is interesting to observe that the value of 22 percent for the discrimination coefficient has been found not only in Japan, but also in the U.S.

References

Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." *Journal of Human Resources* 23(3): 279-95.

Oaxaca, Ronald L. 1973. "Male-female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3): 693-709.

Oaxaca, Ronald L., and Michael R. Ransom. 1994. "On Discrimination and the Decomposition of Wage Differentials." *Journal of Econometrics* 61(1): 5-21.