The Impact of Gender on the Income of Floating Population in Beijing

Liyang Wang

Beijing Jiaotong University, Beijing, China
Email: wangly123@bjtu.edu.cn

ABSTRACT

Since the market-oriented reforms and the acceleration of urbanization, the number of floating population in Beijing has increased sharply. This has led to not only large-scale population migration, but also changes in gender relations and roles. Based on the 2017 Beijing population survey data, we construct multiple linear regression equations with Mincer income equation to analyse the impact of 11 variables on the income of male and female migrants. Further, we use the Oaxaca-Blinder decomposition method to calculate the contribution rate of each part. The results show that there are significant gender differences in the income of migrants which can be explained by individual characteristics and serious women sexism. Finally, this paper proposes two policy recommendations for enhancing women’s human capital investment and breaking institutional barriers that are not conducive to gender equity employment.

Keywords: Floating population, Gender difference in income, Mincer income equation, Oaxaca-Blinder decomposition method.

1. INTRODUCTION

With the acceleration of urbanization process, the relevant policies of labor market segmentation have been loosened and the free flow of labor force has been realized. The urban floating population in large cities with more job opportunities has exploded. By 2017, the total number of floating population in China had reached 244 million, which has become an inevitable development trend in the process of China's social and economic transformation and meanwhile, brought the changes in gender relations and roles [1]. Many studies have shown that the gender difference weighs heavily in income in the labor market [2]. Therefore, the study of the gender income difference of floating population is both a perspective of the status of gender relations in the process of market transformation, and an important dimension to study the income of floating population [3].

There is a certain "floor effect" in China, which deserves special attention when making relevant policies. Women are severely discriminated against in the labor market due to their low education level and lack of work experience. Increased investment in education and on-the-job training for them will increase their competitiveness and help narrow the gender income gap[4]. Beijing, as the capital city, has a huge economic scale and labor force, and its marketization degree and labor force endowment characteristics are not exactly the same as the national average level. So, it is of great practical significance to conduct a separate study.

This paper mainly applies the Oaxaca-Blinder decomposition method, proposed by Oaxaca and Blinder, which is the most classic academic method for analyzing income differences. This method is based on the Mincer income equation, which divides income differences into either can be explained by population characteristics or cannot be explained [5]. There are two main functions: firstly, the contribution degree of the explicable and unexplainable parts to the total income difference is determined respectively, to find whether there is group difference or discrimination between two groups and what is the main cause of the difference; secondly, finding out major variable in each part respectively to obtain the specific reasons leading to the difference[6].

1.1. Background Theory

Floating population refers to people who leave their domicile place and located in other places for the purpose of working or living. According to the
destination, they can be divided into inflow and outflow population. At present, the flow direction and trend in China are mainly from rural areas to cities, western areas to eastern coastal areas, and from economically underdeveloped areas to developed areas [7]. There, we define the floating population in Beijing as the incoming population aged above 15 and have lived in Beijing for at least one month with other domicile places.

**Human Capital.** In 1960s, Theodore W. Schultz proposed the concept of human capital as reflection in people's knowledge, skills, experience and proficiency [8]. Gary S. Becker established the basis of microeconomic analysis to determine the labor income distribution relationship with the human capital [9]. In the 1970s and 1980s, J. Mincer established the Mincer income equation to study the gender difference in wages. Paul Romer, Robert Lucas and M. Scott demonstrated that human capital is the fundamental driving force of modern economic growth.

**Discrimination.** There are three aspects [10]: (1) Personal preference discrimination was put forward by Becker in the Discriminatory Economy in 1971. It comes from employers, employees and customers. They have certain prejudice against certain groups of people, incurring their unwillingness to work with them. (2) Dual labor market discrimination theory divides the labor market into the senior and primary labor market. Women and other vulnerable working groups are mostly employed in the latter. Various social and institutional factors result discrimination against them continually, which increases the instability of their working experience. (3) Statistical discrimination, put forward by Edmund S. Phelps, is because of incomplete information and unscientific statistical methods. Employers usually take the common characteristics when hiring, hoping to find the best candidates, but in fact, they always unconsciously decide from subjective feelings.

### 1.2. Hypotheses

Income difference has rich theoretical and research roots. Sanborn[11] applied the Oaxaca-Blinder decomposition method to examined income difference in the United States in 1950, and found nearly 90% of the was determined by education and the type of work, rather than gender discrimination. Bielby [12] pointed out that the income difference was due to the different given job. Jacob[13] discussed the changes of caste and gender wage gap, and the results showed that about 55% of the wage gap could not be explained by production characteristics and endowments. Bobbitt Zeher[14] concluded that the improvement of women's education level reduced the income gap to a certain extent, while the difference was still significant due to the gender professional segmentation. Xie Sicheng[15] found that 45.6% of the gender wage difference in China's cities was caused by gender discrimination and attested that the income difference (93.1%) mainly came from the difference within the industry. Li Chunling[16] analyzed data at three time points, discovered the gender income gap was due to gender inequality, not the market competition mechanism in the past. Yuan Xiaoyan[17] supported the gender wage difference was caused by gender discrimination (62.13%). Qing Shisong and Zheng Jiamei[18] studied the impact of vertical occupational segregation on gender wage gap. Through comparative studies, He [19] found that the proportion attributable to gender discrimination in the gender wage gap decreased, with endowment difference increasing.

However, at present, the academic research on gender income difference is still unable to fully explain the gender difference in income between men and women in terms of empirical estimation. In the application of the Oaxaca-Blinder decomposition method, the effect of endogenous human capital has not been applied in empirical studies. In addition, most of the domestic researches stay in the qualitative description, which means the analysis of the causes is too macroscopic and relies more on subjective judgment. Also, there is no study on it in Beijing. Therefore, under the framework of economics and the micro theory, this paper attempts to decompose the gender difference in the income of floating people in Beijing by using research methods that are in line with China's actual situation, in an attempt to analyze the degree of discrimination against women.

### 2. RESEARCH METHOD

#### 2.1. Data Sources

The data comes from the 2017 National Floating population Dynamic Monitoring survey data released by the National Health Commission, which covers areas with concentrated floating population sampled from 31 provinces and the Xinjiang Production and Construction Corps. The main contents include basic demographic characteristics of the floating population, employment, income, basic public services, etc. The survey objects are the incoming population with characteristics that we defined as floating population. The study in this paper is based on data from 15 municipal districts in Beijing, with a total sample size of 7,000 people. After eliminating the missing values and outliers, the effective sample size obtained was 5281 people, among which the male sample accounted for 54% and the female sample accounted for 46%.

#### 2.2. Variable Declaration

In this article, 11 independent income influencing variables including gender, years of education, work experience, square of work experience, nationality,
nature of household registration, marital status, work intensity, major occupation, nature of unit, and employment status.

In general, the most vital non-discriminatory factor affecting individual income is individual endowment. Considering the unreachable of innate endowment, we use the maximum number of education-year to represent educational background, making it a fixed distance variable. Work experience and square are measured by years of experience. According to the previous researches, there is an inverted "U" pattern between work experience and income, especially for physically dependent jobs. Therefore, the square term of work experience is added to fit the effect to control nonlinear effect. Besides, the phenomenon of the "marriage premium", in which on average, married men earn more than unmarried men, has been recognized by academicians. However, whether it is true for female labor force has not been determined, so we added this variable to investigate its impact. About how to measure work intensity, there are two main variables: the number of working hours per week and for per month. The former is more universal and suitable for more mobile and stable jobs that fits this paper's purpose, thus the number of working hours per week is adopted to measure work intensity, which is a continuous variable.

The dependent variable should be the income of floating population through employment in the labor market, provided by the monthly income from survey data. Considering that the income difference after taking the logarithm of monthly income has the meaning of percentage and is closer to the normal distribution, we take the logarithm of monthly income as the dependent variable. The observation of missing monthly income, unemployment, unemployment and other non-employment status is deleted. A multiple linear regression model was constructed based on Mincer income equation.

<table>
<thead>
<tr>
<th>Table1. Description of variables</th>
<th>Variables</th>
<th>AVG</th>
<th>Std.</th>
<th>Assignment and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income Logarithm</td>
<td>8.51</td>
<td>0.67</td>
<td></td>
<td>The log of last month net income</td>
</tr>
<tr>
<td>Gender</td>
<td>0.46</td>
<td>0.50</td>
<td></td>
<td>female=1, male=0</td>
</tr>
<tr>
<td>Education Year</td>
<td>12.14</td>
<td>3.50</td>
<td></td>
<td>no=0, primary=1, junior=9, senior=12</td>
</tr>
<tr>
<td>Work Experience</td>
<td>18.03</td>
<td>10.57</td>
<td></td>
<td>work experience = age - years of education -6</td>
</tr>
<tr>
<td>Experience Square</td>
<td>486.7</td>
<td>484.51</td>
<td></td>
<td>work experience*work experience</td>
</tr>
<tr>
<td>Nationality</td>
<td>0.96</td>
<td>0.20</td>
<td></td>
<td>Han=1, other=0</td>
</tr>
<tr>
<td>Registered Residence</td>
<td>0.37</td>
<td>0.48</td>
<td></td>
<td>non-agriculture=1, agriculture=0</td>
</tr>
<tr>
<td>Marital Status</td>
<td>0.82</td>
<td>0.38</td>
<td></td>
<td>married=1, single=0</td>
</tr>
<tr>
<td>Work Strength</td>
<td>45.54</td>
<td>13.76</td>
<td></td>
<td>working-hour per week</td>
</tr>
<tr>
<td>Professional</td>
<td>3.70</td>
<td>1.07</td>
<td></td>
<td>principal=1, technical=2, clerks=3,</td>
</tr>
<tr>
<td>Nature of Unit</td>
<td>2.92</td>
<td>0.57</td>
<td></td>
<td>public unit=1, state-owned =2, private=3, other=0</td>
</tr>
<tr>
<td>Employment Status</td>
<td>1.48</td>
<td>0.82</td>
<td></td>
<td>employer=1, proprietary =2, other=3, employee=0</td>
</tr>
</tbody>
</table>

The multiple linear regression model is based on the Mincer income equation, so its basic form is:

$$\ln Y = \beta_0 + \beta_1 Edu + \beta_2 Exp + \beta_3 Exp^2 + \varepsilon$$  \hspace{1cm} (1)

Where $Y$ represents personal monthly income; $\ln Y$ is the logarithm of monthly income; $Edu$ means years of schooling; $Exp$ for work experience; $Exp^2$ for work experience, indicating the work experience and the nonlinear relationship between personal income; $\varepsilon$ as random error; Coefficient $\beta_1$ said human capital of individual yields without considering the cost of education, referred to as return on education; $\beta_2$ and $\beta_3$ are personal returns from the human capital working experience.

At present, China's labor force is hard to move freely between occupations and work units, thus personal income may also be affected by factors other than years of education and work experience. Hence, combined with the characteristics of labor market and data in Beijing, this paper added other 8 independent variables into the Mincer income equation to explore the extent of their impacts on personal income, also solved the problem that the standard Mincer equation is prone to omission of variables. After adding the above variables, the multiple linear regression equation becomes as follow:

2.3 Descriptive Statistics

Statistics show that there are significant gaps in monthly income between men and women. The average monthly income of female floating population is about 79 percent of the average monthly income of men, which is 1,522 yuan less. As for the standard deviation, male income (6813) is still greater than that of female income (4962), identical to previous studies. Generally speaking, the higher the average income level, the greater the internal income difference.

Education is a significant variable affecting income, development of education is an effective way to solve income inequality in various regions. The average length of schooling for males is 11.98 years, while that for females is 12.40 years, indicating that nine-year compulsory education has been popularized in China. In terms of education policy, there is no significant gender difference. As for the nature of hukou, among the floating population in Beijing, the proportion of male (64.79%), female (60.33%) or overall (62.73) is higher than that of non-agricultural hukou. The rapid development of cities requires the support of a large number of labor forces than rural areas. Therefore, a large number of migrant workers have migrated to cities, and the proportion of agricultural hukou is much higher than that of non-agricultural hukou, which accords with the reality of China's society.

3. INSTRUMENT DEVELOPMENT

3.1. Mincer Income Equation

The multiple linear regression model is based on the Mincer income equation, so its basic form is:

$$\ln Y = \beta_0 + \beta_1 Edu + \beta_2 Exp + \beta_3 Exp^2 + \varepsilon$$  \hspace{1cm} (1)

Where $Y$ represents personal monthly income; $\ln Y$ is the logarithm of monthly income; $Edu$ means years of schooling; $Exp$ for work experience; $Exp^2$ for work experience, indicating the work experience and the nonlinear relationship between personal income; $\varepsilon$ as random error; Coefficient $\beta_1$ said human capital of individual yields without considering the cost of education, referred to as return on education; $\beta_2$ and $\beta_3$ are personal returns from the human capital working experience.
\[
\ln Y = \beta_0 + \beta_1 Edu + \beta_2 Exp + \beta_3 Exp^2 + \beta_4 Gen + \\
\beta_5 Nat + \beta_6 Reg + \beta_7 Mar + \beta_8 Hours + \sum_{i=1}^{5} \theta_i Job + \\
\sum_{j=1}^{2} \gamma_j Unit + \sum_{k=1}^{4} \phi_k Emp + \varepsilon
\]  
(2)

Hours, as a continuous variable, means work intensity. Gen, Nat, Reg, Mar, Job, Unit and Emp respectively represent the dummy variables of gender, nationality, hukou nature, marital status, main occupation, employer nature and employment status. We selected male, ethnic minority, agricultural hukou, single status, other occupations, other units and employees as reference categories. The above supplementary variables all enter the regression equation as explanatory variables. From geometric perspective, the regression coefficient of each dummy variable represents the difference of intercept terms, that is, the individual income difference under diversity categories when other variables remain unchanged.

3.2. Oaxaca-Blinder Decomposition Model

To further study the impact of various factors on the income contribution degree, as well as measure gender discrimination, we use of Oaxaca-Blinder income decomposition method to analyze gender difference in earnings of the floating population, it can effectively decompose factors from income disparity background conditions and identity difference (such as gender) quantitatively. The method consists of two steps:

The first step is to construct the income regression equations of men and women respectively, using the Mincer equation after adding some variables in (2):

\[
Y = \hat{\beta}_g X_g + e_{g,g}g=m,f
\]  
(3)

The dependent variable \(Y\) is the logarithm of monthly income, and independent variable vector \(X\) is the factors affecting income, including years of education, working experience, square of working experience, nationality, nature of household registration, marital status, work intensity, main occupation, nature of unit, and employment status. Subscript \(g\) said of the floating population of different gender (m-men and f-women), regression coefficient \(\hat{\beta}_g\) contains constant intercept. After removing the influence of these factors, the income difference that cannot be explained under the corresponding income decomposition can be understood as the gender discrimination of income. According to the sample data, least squares estimation of regression coefficient \(\beta_g\) is

\[
\hat{\beta}_g = (X_g'X_g)^{-1}X_g'Y_g
\]  
(4)

\[
\hat{Y}_g = \hat{\beta}_g X_g \quad \text{and} \quad \bar{Y}_g = \hat{\beta}_g \bar{X}_g
\]  
(5)

The second step, decompose male and female income gap like:

\[
\Delta Y = \bar{Y}_m - \bar{Y}_f = \beta_m\bar{X}_m - \beta_f\bar{X}_f
\]  
(6)

\(\beta^*\) represents no discrimination of various factors under the indiscriminate regression coefficient, reflecting the income decision mechanism without discrimination. The first item \((\beta_m - \beta^*)\bar{X}_m + (\beta^* - \beta_f)\bar{X}_f\) on the right of the equation is the explicable part, which represents the income difference caused by the difference in individual characteristics of different genders or the difference in the mean value of explanatory variables between the two, and is the reasonable part of the gender income difference. The last two terms \((\beta_m - \beta^*)\bar{X}_m + (\beta^* - \beta_f)\bar{X}_f\) on the right are unexplainable, implying the income difference caused purely by gender under the same individual characteristics. \(\beta_m - \beta^*\bar{X}_m\)measures the unfair gain of men’s income, while \((\beta^* - \beta_f)\bar{X}_f\)measures the unfair loss of women's income. The two are equal to the unfair difference of income, namely gender discrimination.

4. MODEL ANALYSIS

4.1. Regression Results:

In theory, \(\beta^*\)in equation (6) represents the regression coefficients without discrimination in perfectly competitive labor market. For \(\beta^*\)-estimate, this paper respectively adopts samples of male, female, all and all samples after adjusting (including gender virtual variable) to regress, get four estimation results \(\beta^*\) under regression coefficients without discrimination, as shown in table 5. Comparing the regression results in the male sample and the female sample, the universality is that the longer the years of education and the more work experience, the higher the income. Moreover, work experience has strong influence on the income of both men and women. All of these conform to the classic Mincer equation. The income of non-agricultural population is higher than that of agricultural population; single person is lower than married; heads of state-owned enterprises or government, as well as professional and technical personnel earn more than other occupations. The income of a non-state-owned enterprise is higher than that of any other unit; self-employed workers earn more than employees. The discrepancy is that with same education, men earn a higher return than women. For every additional year of education, men's income increases by 7.3% on average, while women increase only 6.1%. This partly reflects gender discrimination.

Under other control variables stay unchanged, we are able to notice from adjustment after all samples of gender virtual variable regression coefficient that logarithmic income of women is 0.206 smaller than men, which can calculate the women earn only 81.4% of men, representing that with fixed education, work experience, marital status, individual characteristics of a
woman and a man, the woman’s income is only about 81.4% of the man, this conclusion directly reflects the income difference between men and women in serious gender discrimination.

The generally consistent results of the four decomposition methods reflect the robustness of the results. The total logarithm income gap between men and women is 0.199, meaning that on average, women earn 82% of men’s income, which is significant by international standards. According to the World Economic Forum’s 2018 annual Report on Gender Equality, the global gender pay gap in 2018 is 51 percent.

It can be found from Table 6 that only a small part of the income difference can be explained, less than 34%, among which the main role is the income difference caused by different working experience and employment status. Taking the adjusted full sample as an example, the explicable part of the wage difference between men and women is 0.048, accounting for only 24.1% of the total difference, that is, only 24.1% can be explained by the differences in individual characteristics. The explanation degree of work experience, with a contribution rate of 30.0% is the largest. Employment status, which explains 6.2% of the difference stands in second place. Since the year of the education and registered permanent residence of the nature of decomposition coefficient is negative, these two factors in the difference of income is the impact of men than women, while the remaining factors are all women have more advantages than men, with consistent conclusions as last section in the regression results. 75.9% income difference cannot be explained. Such a high proportion gives the main reason for women's lower income and reflects the serious gender discrimination against women. Women are still treated unfairly, including taking on more housework in family, facing employment discrimination in labor market, suffering "ceiling effect" in promotion.

### 4.2. Sex Differences in Income

Additionally, by posing regression coefficients without discrimination $\beta*$ four estimation results, using formula (6) to test the gender decomposition of income disparity, the results as shown in table 6.

The generally consistent results of the four decomposition methods reflect the robustness of the results. The total logarithm income gap between men and women is 0.199, meaning that on average, women earn 82% of men’s income, which is significant by international standards. According to the World Economic Forum’s 2018 annual Report on Gender Equality, the global gender pay gap in 2018 is 51 percent.

It can be found from Table 6 that only a small part of the income difference can be explained, less than 34%, among which the main role is the income difference caused by different working experience and employment status. Taking the adjusted full sample as an example, the explicable part of the wage difference between men and women is 0.048, accounting for only 24.1% of the total difference, that is, only 24.1% can be explained by the differences in individual characteristics. The explanation degree of work experience, with a contribution rate of 30.0% is the largest. Employment status, which explains 6.2% of the difference stands in second place. Since the year of the education and registered permanent residence of the nature of decomposition coefficient is negative, these two factors in the difference of income is the impact of men than women, while the remaining factors are all women have more advantages than men, with consistent conclusions as last section in the regression results. 75.9% income difference cannot be explained. Such a high proportion gives the main reason for women's lower income and reflects the serious gender discrimination against women. Women are still treated unfairly, including taking on more housework in family, facing employment discrimination in labor market, suffering "ceiling effect" in promotion.

### 4.3. Robustness Analysis

In order to test the robustness of the results of the model, we adopt the substitution variables, by replacing work experience and their squared terms with age and their squared terms. The results show that coefficients of all variables after the substitution have no obvious change, and the direction is consistent. Therefore, it is considered that the model estimation results of factors...
affecting the income of the floating population in Beijing are robust.

5. CONCLUSION

In this section, the focus is on the interpretation of the emergent findings. Additionally, several substantive suggestions are given.

5.1 Concluding Marks

Among the variables affecting income, the number of years of education, working experience and the nature of household registration are the three most significant variables. Controlling other human capital variables, there is an obvious gender difference in income. Specifically, the income of floating female population is 81.4% of that of males. With the Oaxaca-Blinder method, we learned that about 24% of the gender income difference can be explained by different characteristics. Among them, work experience and employment status played a major role, respectively accounting for 30.0% and 6.2% of the total differences. About 76% of the income gap is unexplainable, and such a high percentage reflects serious sexism against women.

5.2 Policy Suggestions

The study proposes 2 advice. (1) Increase investment in women's human capital. On the one hand, the improvement of female education level can reduce gender income difference. On the other hand, it helps eliminate employment discrimination. Investment in education is valued in increased years of schooling and improved on-the-job training. It is particularly important to note the investment in girls' human capital which is often reduced by factors such as the low expected return and the traditional marriage system. These factors may have disadvantages, such as women's lower initial human capital, less opportunities and so on. Therefore, the government must play its role in providing basic education and pay attention to the interruption of the accumulation due to childbirth.

(2) Break down institutional barriers that are not conducive to gender equality in employment. Society should, to the extent possible, create fair career choices environment and pay system for both sexes. Institutional barriers include labor protection laws, widely used permanent contracts, unemployment benefits, etc. China should learn from the effective laws of developed countries, improve relevant policies and regulations, adopt more detailed supervision methods, and ensure the enforcement, so as to effectively reduce discrimination in the labor market. Employers who violate the principle of gender discrimination will be punished more severely, which aims to create a fair and reasonable competitive environment in the industry.

REFERENCES


