Human capital investments and gender earnings gap: Evidence from China's economic reforms

Haoming Liu*
Department of Economics
National University of Singapore
ecsliuhm@nus.edu.sg
+65 6516 4876

June 8, 2009

Abstract

This paper examines the contributions of gender differences in post school investments (PSI) in human capital to the gender earnings gap in China. First, by exploring the exogenous variations in the length of working life caused by differences in mandatory retirement age, we find that the gender earnings gap is mainly driven by the difference in the slope of age–earnings profiles. Namely, a shorter working life is associated with a flatter age–earnings profile. Second, by examining the relationship between the decline in employment rate and the gender earnings gap, we find that a one percentage point decrease in women's employment rate is associated with a 0.851 percentage points increase in gender earnings gap. If women's employment rate had been fixed at its 1988 level, China's gender earnings gap would have declined by 4.7 percentage points rather than increased by 8.7 percentage points between 1988 to 2002. Because both the length of working life and employment prospect affect people's incentive to invest in human capital, these results suggest that the gender difference in PSI is a significant contributing factor to gender earnings gap in China.

Keywords: Economic reform, gender inequality, discrimination
JEL classification: J3, J7, P25

---

*Corresponding author: Haoming Liu, Department of Economics, National University of Singapore, 1 Arts Link, Singapore 117570. Email: ecsliuhm@nus.edu.sg, Phone: +65 6516 4876, Fax: +65 6775 2646

1 Introduction

Trying to put some economic meanings into the *unexplained* part of the gender earnings gap is always a challenging task, and it is rightly to be so because it is *unobserved* to the researchers by definition. Economic theory suggests that the gap is attributable to at least two factors: gender discrimination and human capital. While it is almost impossible to gauge the contribution of discrimination, several researchers have tried to identify the magnitude of the contribution of human capital. For examples, O'Neill and Polachek (1993) find changes in the potential experience coefficient account for a substantial component of the rise in women's relative wages. Mulligan and Rubinstein (2008) show that changes in the unobserved skill of employed women is the primarily driving force for the narrowing gender earnings gap. Similarly, Olivetti and Petrongolo (2008) claim that selection into employment accounts for a considerable proportion of the cross country differences in gender earnings gap.

In this paper, we use a Chinese data set to address the contribution of unobserved human capital to gender earnings gap. While understanding the gender earnings gap in China is a topic that is worth pursuing by its own right, some unique features of Chinese labor market can also shed some lights on this issue for other countries. First, unlikely other countries where retirement is primarily an endogenous decision, the mandatory retirement age is almost universally observed to employees in the State Own Units. Men's retirement age is set at 60 regardless their occupations. Women's retirement age is 55 for office–workers and 50 for laborers. This exogenous variation in retirement age, hence in the length of working life, gives women less incentive to invest in their human capital compared with men. Second, because of the low fertility rate and easy access to affordable day care centers in urban areas, Chinese female workers have little interruptions in the middle of their careers. This makes potential years of experience a good proxy for actual experience even for women. If men and women invest the same amount in their human capital, they should follow similar age–earning profiles. Third, while women's employment rate increased considerably in the U.S. and many other countries, it decreased in China. At the mean time, both women's and men's annual earnings increased by more than 180% in real term. This suggests that changes in wages are unlikely to be the driving force for the decline in women's employment. In contrast, it is very difficult to tell whether the increase in women's relative earnings in the U.S. is the result that higher wages attracted more women to work or the increased labor market attachment attracted employed women to invest more in their human capital. Finally, as more and more workers are employed in the service sector that traditionally favor women, the decline in women's relative wages is unlikely to be driven by unfavorable sector shifts.

Chinese women had enjoyed a high relative earnings in the late 1980s and early 1990s compared with their counterparts in most OECD countries. According to Gustafsson and Li (2000), women's average earnings were 15.6% lower than that of men in 1988 and 17.5% lower in 1995. The earnings gaps are comparable with the corresponding figures in Finland and Germany and lower than that in the U.K. and the U.S. (Olivetti and Petrongolo 2008). Studies using more recent data find that China's gender earnings gap has widened over time. For example, Zhang, Han, Liu, and Zhao (2007) find that women's relative earnings declined from 86.3% in 1988 to 76.2% in 2004. To understand the contributions of various factors to China's gender earnings gap, existing studies mostly use the Blinder–Oaxaca (Blinder 1973 and Oaxaca 1973) or the Junh, Murphy, and Pierce (1991) method to decompose the observed gap into explained and unexplained components. To the best of our knowledge, none of the existing studies have tried to link the gender earnings gap to the gender difference in incentives to invest in human capital. Failing to relate gender earnings gap to gender difference in investment incentives makes it harder to explain why the rate of returns to education is higher for women while the rate of returns to experience is higher for men (Gustafsson and Li 2000 and Liu, Meng, and Zhang 2000).

Using data extracted from three waves of China Household Project (CHIP), we find that gender earnings gap increased from 10% in 1988 to 20% in 2002 after controlling for observed characteristics. Much of the increase is related to the gender differences in post school investments (PSI) in human capital. We use two approaches to examine the impact of PSI on China's gender earnings gap.

In the first approach, we exploit the impact of the exogenous variation in retirement age. Because the difference in the rate of return to the PSI between men and women increases with age, the gender earnings gap should increase with age as well, particular for laborers. Using a difference in difference method, we find that the age–earnings profile of female office–workers is steeper than that of female laborers, but flatter than that of both male office–workers and male laborers. Actually, we cannot detect any significant difference between the slope of the age–earnings profile of male office–workers and male laborers. This suggests that the difference between female office–workers and laborers is likely due to the difference in PSI. Moreover, the coefficient on the gender dummy is not significantly different from zero both economically and statistically in 1988 and 2002, suggesting that women do not earn significantly less than men at labor market entry.

It is women's slower earnings growth that accounts for their lower relative earnings.

In the second approach, we examine the effect of women's employment rate on the gender earnings gap. Because worsening employment prospect discourages human capital investment, a declining women's employment rate should be associated with a widening gap. We first group our sample into 5 age groups with each group contains individuals born within a 7–year interval, which is chosen to match the 7 years interval between the CHIP surveys so that we can pool different years of data to construct pseudo cohorts. We then estimate the gender earnings gap for each age group. To capture the potential difference between the age–earnings profile of female office–workers and laborers, separate regressions are run for less educated workers (without graduating from high school) and better educated workers (high school and above). We use the maximum likelihood method proposed by Heckman (1976) to control for the sample selection. The reason for grouping the sample by education rather than occupation is that we cannot observe the occupation of individuals who do not work, hence cannot control for the sample selection based on occupation. We use better educated workers as a proxy for office–workers as most office–jobs require at least a high school diploma. The estimated earnings gaps from difference years are then pooled together to construct cohort specific gender earnings gap.

By regressing the constructed gender earnings gap series on women's employment rate and cohort and age fixed effects, we find that a one percentage point drop in women's employment rate is associated with a one percentage point increase in gender earnings gap. Moreover, our counterfactual analyses show that the gender earnings gap would have decreased by 4.7 percentage points rather than increased by 8.7 percentage points if women's employment rate was fixed at its 1988 level. These results provide further support for the argument that gender difference in human capital plays a significant role in determining the gender earnings gap in China.

The remainder of the paper is organized as follows. Section 2 lays out the basic framework that will be used in our empirical analysis. Section 3 provides some background information and describes the data source. Section 4 reports our empirical results, and a short conclusion is provided in Section 5.

2 The basic framework

Because workers' human capital investments varies over the life-cycle, the impact of China's economic reforms differs across birth cohorts. For example, a sudden reduction in the employment rate should have little impact on the human capital accumulation of women who are close to their retirement age because most of their human capital investments have already been committed. To incorporate these features into our model, we assume the log wage equation takes the following form:

$$w_{ict} = X_{it}\alpha + F_i g(e_{it}) + F_i \delta_c + \epsilon_{it}, \tag{1}$$

where w_{ict} represents person i of cohort $c's$ log wage in year t, X_i is a vector of personal characteristics, α is the price of these personal characteristics, F_i is a gender dummy (=1 for women), e denotes working experience, $g(\cdot)$ captures the difference in the rate of return to experience between gender due to difference in human capital accumulation, it could vary across cohorts due to changes in the length of working life, δ_c is the gender earnings gap at labor market entry and could differ among cohorts, and ϵ_{it} is the random error term. While it is reasonable to assume $E(\epsilon_i) = 0$ for the entire population, $E(\epsilon_i|i$ is employed) is unlikely to be zero if the selection into employment is not random.

One implicit assumption of equation (1) is that the impact of discrimination on earnings is fixed for each cohort. For workers who already entered the labor market, equation (1) will attribute the impacts of changes in discrimination to $g(\cdot)$, the difference in return to experience between genders, rather than to δ. This might not be an ideal assumption. It is made largely due to the inseparability among age effect, cohort effect and time effect. Nevertheless, this assumption can serve as a reasonable approximation as any changes in the taste for discrimination is likely to have a larger impact on new entrants than on existing workers. This is because changes in the degree of discrimination on currently employed workers will lead to a gender specific salary increase, which is unpopular to co-workers of the opposite gender.

The OLS estimate of the gender earnings gap without controlling for the potential gender difference in the return to experience, G_t, is equivalent to subtracting the average earnings of employed men from the average earnings of employed women after controlling for the common factor $X_{it}\alpha$:

$$G_t = \sum_c \omega_{ct}^f \overline{g(e_{ct}^f)} + \sum_c \omega_{ct}^f \delta_c + \left[\sum_c \omega_{ct}^f \bar{\epsilon}_{ct}^f - \sum_c \omega_{ct}^m \bar{\epsilon}_{ct}^m \right], \qquad (2)$$

where ω_c^f is the proportion of employed women who belong to cohort c and

ω_c^m is the proportion of employed men who belong to cohort c,

$$\overline{g(e_{ct}^f)} = \sum_{i=1}^{n_{ct}^f} g(e_{ict}^f)/n_{ct}^f,$$

where n_{ct}^f is the number of employed cohort c women in year t, and $\bar{\epsilon}_{ct}^f$ and $\bar{\epsilon}_{ct}^m$ are the average level of ϵ of employed women and men, respectively.

The first term of equation (2) shows that even if $g(\cdot)$ and δ are the same across cohorts, the OLS estimate, G_t, could still vary over time if ω_c varies. For example, the magnitude of G_t will increase with the employment share of older workers if women's relative earnings decreases with age. Obviously, G_t also changes over time if $g(\cdot)$ or δ varies across cohorts. The last term of equation (2) shows that variations in the sample selection affect the estimated gender earnings gap. For example, if the declining employment rate of younger cohort raises the average level of unobserved ability of employed women, then the magnitude of G_t will decrease as these younger cohort enter the labor market.

We use two strategies to address the impacts PSI on gender earnings gap. In the first approach, we exploit the difference in the mandatory retirement age between female office–workers and laborers. Given the shorter working life, female laborers have less incentive to invest in their human capital than female office–workers. In general, the effect of PSI on earnings is ambigous. The costs of PSI lower workers' earnings in their early career, but the returns raise earnings later on. Gronau (1988) argues that the second effect predominates. Because the gender difference in the rate of return to PSI increases with age in China, Gronau's claim is even more likely to be hold in China than in the U.S. This implies that female office–wrokers should have a steeper age–earnings profile than female laborers, but a flatter profile than male workers.

In the second approach, we examine the relationship between the cross cohorts variations in employment rate and gender earnings gap. If women invest less in their human capital because of the deterioration of their employment prospect, their relative wages should be negatively correlated with their employment rate.

To implement the first approach, we run the following regression

$$\begin{aligned} w_i = {} & X_i\alpha + \gamma_0 F_i + \gamma_1 C_i + \gamma_2 F_i * C_i + \gamma_3 F_i * A_i \\ & + \gamma_4 C_i * A_i + \gamma_5 F_i * C_i * A_i + \epsilon_i, \end{aligned} \tag{3}$$

where X_i is a vector of control variables, including province of residence, the ownership of the work unit, schooling, age and age squared, F_i is a gender dummy (=1 if i is a woman), C_i is an office–worker dummy (=1 if i is an office–worker and 0 otherwise), A_i is $i's$ age. In the above equation, we implicitly assume that $g(\cdot)$ is linear in age. In a cross-section regression, γ_3 captures both the cohort effect and the age effect. For example, if the degree of gender discrimination declines over time, γ_3 will be negative as women of younger cohort enjoy a higher starting relative earnings at labor market entry. As a result, γ_3 is not a consistent estimate of the difference between the slope of age–earning profile of female and male laborers unless the degree of gender discrimination is the same across cohorts. γ_4 captures the potential difference between the age–earnings profile of male office–workers and laborers. Because the cohort effect is captured by γ_3, and the occupational effect is captured by γ_4, γ_5 is a consistent estimate of the difference in the slope of age–earning profile between females office–workers and female laborers.

The human capital theory predicts that $\gamma_5 > 0$, i.e. female office–workers have a steeper age–earnings profile than female laborers.

The second approach is implement in two stages. In the first stage, we run the following regression for each year

$$w_i = X_i\alpha + \sum_g \phi_g A_{ig} + \sum_g \mu_g F_i * A_{ig} + \epsilon_i, \tag{4}$$

where X_i is a vector of control variables, including province of residence, the ownership of the work unit, schooling. A_{ig} is a vector of age group dummies (=1 if i belongs to age group g and 0 otherwise). The OLS estimate μ_g^{ols} consists of three terms

$$\hat{\mu}_g^{ols} = \sum_{i \in g} \frac{g(e_i^f)}{n_g^f} + \delta_g + \left(\sum_{i \in g} \frac{\epsilon_i^f}{n_g^f} - \sum_{i \in g} \frac{\epsilon_i^m}{n_g^m} \right), \tag{5}$$

where n_g^f is the number of employed women in age group g, and n_g^m the number of employed men.

To control for self-selection into employment, let us assume the employment equation as

$$Prob(E_i = 1) = \Phi(X_i\gamma_0 + \sum_g \gamma_g A_{ig} + Z_i\gamma_2 + \xi_i), \tag{6}$$

where E is a person's employment status, and Φ is the accumulative density function of standard normal distribution, Z is a vector of observed characteristics that affect $i's$ employment status, but not $i's$ wage. We can obtain consistent estimates of α and ϕ by jointly estimating the wage equation (4) and the employment equation (6) using the maximum likelihood method proposed by Heckman (1976). Because men's employment rate also declined considerably over time and γ differs between genders, we run men's and women's regressions separately. Given the consistent estimates of α and ϕ, the selection–corrected estimate μ_g^{hk} can be calculated using the following formula

$$\hat{\mu}_g^{hk} = (\bar{X}_g^f \hat{\alpha}_f + \hat{\phi}_g^f) - (\bar{X}_g^f \hat{\alpha}_m + \hat{\phi}_g^m), \tag{7}$$

where $\hat{\alpha}_j, \hat{\phi}_g^j,\ j = m, f$ are the estimates from men's and women's wage regressions after controlling for self-selection, and $\bar{X}_g^f$ is the average observed characteristics of women of age group g.

The reason for defining $\hat{\mu}_g^{hk}$ as the difference between predicted earnings rather than $\hat{\phi}_g^f - \hat{\phi}_g^m$ is that we are interested in the effect of $g(e_g^f)$ on women's relative earnings. Obviously, if $\hat{\alpha}_f = \hat{\alpha}_m$, equation (7) will be reduced to $\hat{\phi}_g^f - \hat{\phi}_g^m$. If $\hat{\alpha}_f \neq \hat{\alpha}_m$, then the variations in $\hat{\alpha}_f - \hat{\alpha}_m$ will be captured by variations in $\hat{\mu}_g^{hk}$ but not in $\hat{\phi}_g^f - \hat{\phi}_g^m$. Because we are interested in addressing the impact of human capital on gender earnings gap, the gender differences in the rate of returns to various forms of human capital are crucial parts of our analyses. For example, better educated women in China have a longer working life than less educated due to their higher probability of being office–workers, which gives them a stronger incentive to investment in their human capital. The positive correlation between schooling and post school investments in human capital biases the estimated rate of return to education upward. This could explain why previous studies (e.g. Gustafsson and Li 2000 and Liu, Meng, and Zhang 2000) find that Chinese women enjoyed a higher rate of return to schooling than Chinese men. In this example, the difference in the estimated rate of return to education is an integrated part of $g(e_g^f)$ and should not be excluded from our analyses.

In addition to $g(e_g^f)$, μ_g^{hk} is also affected by variations in δ_g. Under the assumption that δ_g is cohort specific and does not vary overtime, the year–to–year variations in $\hat{\mu}_g^{hk}$ for any given cohort will be the result of changes in $g(e_g^f)$, which should be positively correlated with women's employment rate if their human capital investments indeed depend on their employment prospect. Therefore, we can pool the cross section estimates $\hat{\mu}_g^{hk}$ and run the following fixed effects model

$$\hat{\mu}_{ct}^{hk} = \beta_0 + \sum_g \beta_g A_g + \beta_e ER_{ct}^f + \eta_c + \nu_{ct}, \tag{8}$$

where A_g is a vector of age group dummies, and ER_{ct}^f is the employment rate of cohort c women in year t, and η_c is a cohort fixed effect, and ν_{ct} is the random error term. The human capital theory predicts that $\beta_e > 0$.

3 Institutional Background and the Data

In the earlier reform period, both employment and compensation in China were controlled by the state. Employers had little control over who they employ and how much to pay. The wage system was centrally regulated into occupational based wage scales: administrative personnel were put into 24 salary grades before 1995 and into 15 grades after that, technicians into 17 grades and manual employees into 8 grades. As a result, men and women would have the same basic wage for any given grade. Beside the basic wage, workers' wages contain another two components: functional wage (relating to status and seniority), and the floating wage (including the bonus, determined at the enterprise level). The last two components were largely under the control of employers. The relative importance of basic wage declined gradually over time. According to Knight and Song (2003), the share of basic wage was 56% in 1988 and 47% in 1993. Moreover, employers have some degree of freedom on when to promote an employee to the next grade. Therefore, the gender earnings differentials could be the result of either the speed of promotion or the variation in functional or floating wage.

In addition to the basic wage, the retirement age is also regulated by the central government. The mandatory retirement age is 60 for men regardless of their occupations. It is 55 for female office–workers and 50 for female laborers. Given women's shorter working life due to their younger mandatory retirement age, the speed of promotion could be a major contributing factor to the gender earnings differentials. Presumably, employers have less incentive to promote workers who are close to their mandatory retirement age. For instance, a 54-year old woman will have a smaller chance to be promoted compared with a 54-year old man even if they have the same qualification. Given the lower probability of being promoted, women are discouraged to invest in their human capital. Consequently, a 54–year old women tend to accumulate less human capital over her working life than a 54–year old men, which further reduces her chance of being promoted.

Government control over the labor market loosened overtime as more and more reforms were implemented. Arguably, the most influential urban reform was commenced in the mid-1990s when the state owned sector started to lay-off workers, known as *xia gang*, in a large scale. *Xia gang* was first on trial in 1994 and finally launched in 1997 (Appleton et al. 2002). As a result of the mass layoff, the employment shares of SOEs and COEs have decreased considerably since 1997. Table 1 shows that both the total employment and the number of employees in State-Owned Units (SOUs) increased year by year between 1984–1995. The employment share of SOUs stayed at around

73% over this period with little year-to-year variation. The total employment and SOU employment started to fall in 1995, with the latter outpaced the former. As a result, SOU's employment share decreased from 73.5% in 1995 to only 60.9% in 2004. During the same period, the employment share of Other-Ownership Units increased from 5.9% to 31.1%. Because the wage system in the private sector is not subject to the restriction of the centralized wage system, the increase in private sector employment share moved Chinese labor market closer to a free market.

The mass layoff has a larger negative impact on women's employment. Appleton et al. (2002) find that the incident rate of layoff is 12% for men, and 22% for women. Female labor force participation rate fell alongside the decline in employment rate. Giles, Park, and Cai (2006, p67) show that labor force participation rate dropped from 74.4% in January 1996 to 63.1% in November 2001 for women and from 93.0% to 86.3% for men. Conditional on being laid-off, it was harder for women to find a new job. Women generally had to face a higher unemployment rate as well. In November 2001, the unemployment rate of the 40-50 year old was 10.3% for men and 17.1% for women. Consequently, they also had a lower reemployment probability. For example, Giles, Park, and Cai (2006) show that while 44.3% of 40-50 year old males were reemployed within 12 months of leaving their jobs, the corresponding figure for females was only 22.1%. The worsening employment prospect further reduced women's incentive to invest in their human capital, which could widen the gender earnings gap.

The data used in this paper are mainly extracted from 3 waves of the urban sample of China Household Income Project (CHIP). The first survey was conducted in 1989, the second in 1995 and the most recent one in 2002. The samples were drawn from a larger annual national household survey conducted by the National Bureau of Statistics (NBS). The geographic coverage of the sample varies slight across waves. To keep our sample comparable across different waves, we only use households from cities that were surveyed in every wave. We restrict our analysis to individuals aged 19–53, and exclude full time students, self-employed and retirees. The reason for excluding self-employed is that their earnings are not comparable with these of employees. The reason for excluding retirees is that they are more resemble to employed than to other non-employed individuals. Moreover, the retirees are paid by their former employers in 1988 and 1995, which makes it unreasonable to treat retirees as other non–employed individuals. Observations with missing value on schooling, age, gender or employment status are excluded as well. The lower bound of our age restriction is the most common high school graduation age. The upper bound of our age restriction is selected mainly for the convenience of cohort analyses. Because the survey is separated by a 7–year interval, we also grouped the sample population into 5 age groups: 19–25, 26–32, 33–40, 40–46, and 47–53.

The basic sample statistics are reported in Table 2. The sample female population was slightly younger than the male population. This is the result of excluding retirees from our sample. Because women tend to retire earlier than men, this restriction excludes more women who were around 50–year old. The average education level of women was 0.8 years lower than that of men in 1988, but the difference declined to 0.4 years in 2002. The narrowing education gap should improve women's relative earnings. The employment rate was extremely high with a value close to 100% for both men and women in 1988. It declined slightly in 1995 and considerably in 2002, particularly for women. Women's employment rate was 93.4% in 1995 and 81.4% in 2002. The steeper decline in women's employment should discourage their human capital investments. If individuals were not randomly selected into employment, the relatively low employment rate in 2002 suggests that the estimated gender earnings gap might be sensitive to the control of sample selection. Interestingly, the decline in employment rate was accompanied by a considerable increase in earnings. Both men's and women's real earnings (2000 is the base year) were more than doubled over the 14–year period. This evidence suggests that the decline in women's employment rate was not driven by changes in wages.

To see whether the decline in employment rate differs across education levels and birth cohorts, Table 3 reports the employment rate by gender, education level and birth cohort. The figures of each horizontal line trace changes in the employment rate as a birth cohort ages while the numbers of each vertical line reveal year-to-year changes in the employment rate of a particular age group. The figures at the top of each column are calculated using the 2002 data and the bottom the 1988 data. The statistics at the diagonal of the table show the variation in employment across age groups in a given year. Men's employment rates are reported in columns (1)–(5) and women's employment rates are reported in columns (6)–(10). The employment rates of people without a high school diploma are reported in Panel A and those with at least a high school diploma are reported in Panel B.

Employment rate declined over time for any given age and education group, particularly for less educated young women. Among less educated women, the employment rate of 19–25 year olds declined by about 40 percentage points, from 98.1% in 1988 to 58.8% in 2002, the employment rate of 26–32 year olds decreased by about 30 percentage points over the same period. In comparison, the employment rate of 19–25 year old less educated men dropped by 45.4 percentage points, which is even bigger than the decline experienced by women with the same age and education. However, the employment rate of 26–32 year old less educated men dropped 20 percentage points, which is much smaller than that of women of the same age and education level. People with at least a high school degree fared better than the less educated. But even for them, their employment rates still experienced 8 to 20 percentage points drop depending on gender and age. Except for people aged 19–25, women's employment rates were consistently lower and declined more than men's employment rates. These statistics suggest that younger cohorts faced a tougher labor market conditions than older cohorts in terms of finding employment.

Reading the values horizontally reveals that women's employment rates decreased monotonically as they age. For example, for women born in 1963–1969, the employment rate of less educated group declined by 8 percentage points between 1988 and 1995 and by another 16 percentage points in 2002. The employment rate of better educated group declined from 98.3% in 1988 to 92.5% in 2002. In contrast, the employment rate of better educated men of the same birth cohort was very stable with a value of 97.9% in 1988 and 97.3% in 2002. A comparison between men's and women's employment rate reveals that while women of the youngest age group always have a higher employment rate than men regardless of birth cohorts and education levels, women's employment rate is always lower than that of their male counterparts for other age groups. This suggests that gender discrimination at the labor market entry is unlikely the main reason for women's lower employment rate.

Reading the values diagonally down and to the right suggests that employment rates increased with age at first and then declined after 46 for all age and education groups except for better educated men. The employment rate of better educated men did not vary much between age 26 and 53. The difference between information revealed from reading horizontally and diagonally suggests that an age–employment profile constructed from cross–section data is biased by cohort effects.

To see whether the steeper decline in women's employment rate is mainly driven by more married women withdrawing from the labor market, Table 4 reports the employment rates of married individuals.[1] Except for less educated young women, the employment rates of married individuals are comparable to these of the entire population, suggesting the decline in women's employment is not attributable to married women withdrawing from the labor market. Moreover, the employment rate of married women (aged 26–32) who have a lagger probability of having young kids is actually higher than that of those who already past their reproductive age (aged 40–46). This suggests that changes in the employment rate of married women with young kids is unlikely to be the reason for the decline in women's employment either.

Overall, the evidence documented so far suggests that women's employment rate declined considerably from 1988 to 2002, and the decline was neither due to marriage nor to raising children. Because women's employment rate declined monotonically with age and fell to 70% in 2002 even before they reached age 55, women of later born cohorts should have less incentive to invest in their human capital.

[1]The reason that we do not report the employment rate for singles is that the small sample size makes it impossible to accurately estimate employment rate by age and gender.

4 Estimation results

4.1 The length of working life and gender gap

Table 5 reports the estimation results of our first approach. The reference group of our estimation is male laborers. The results show that the age–earnings profile for male laborers was very stable over this 14–year period. For example, a one year increase in age was associated with a 5.3% increase in earnings in 1988, 6.2% in 1995 and 5% in 2002. This is consistent with previous studies that generally found little changes in the rate of return to experience. Unlike the stable rate of return to experience, the rate of return to education increased considerably over time even though it was still lower than that in most market economies. The rate of returns to education was 1.4% in 1988, 2.6% in 1995, and 4.8% in 2002. The lower rate of returns to education is partially due to the use of age as a proxy for experience. Workers in state own units earned a considerable premium and the size of the premium increased slightly over time.

The coefficient on gender dummy is only significant in 1995, suggesting that the gender earnings gap in China is largely driven by the difference in the slope of age–earnings profile rather than the difference in starting salaries. The 1995 result is mostly driven by the sudden increase in the earnings gap of the 19–25 year olds. Because the sudden increase only applies to one particular age group, it is unlikely to be the result of a jump in the degree of discrimination against women or any factors that affect the earnings of the entire female work force.

The coefficient on the interaction between gender dummy and age is negative and statistically significant for each year. Because of the high employment rate and little institutional changes prior to 1988, selection into employment and cohort specific differences in the degree of discrimination are likely to have minute effects on the 1988 estimation results. Hence, the difference between men's and women's age–earnings profile is likely responsible for the negative coefficient on $F * A$ in 1988. Therefore, the 1988 regression result suggests that the gender gap increases by 0.8 percentage points for a one year increase in age. The coefficient on the interaction between age and office–workers is positive and statistically significant at the 5% level in 1988, suggesting office–workers had a steeper age–earnings profile than laborers. Because the average skill level of office–workers is higher than that of laborers, this difference might be driven by the fact that skilled workers generally have a steeper age–earnings profile than unskilled workers. The coefficient on the triple interaction between gender dummy, age and office–workers is also positive and significant in 1988, suggesting the gender earnings gap widened at a slower pace among office–workers than among laborers. We suggest that this difference is due to female office–workers' stronger incentive to invest in their human capital, which is the result of their relatively longer working lifes compared with female laborers.

Beside the human capital interpretation, one might argue that the negative coefficient on the interaction between gender dummy and age is the result of a positive correlation between gender discrimination and age, i.e. older women have to face a higher degree of discrimination than younger women. However, this interpretation cannot explain why the association between age and discrimination is significantly lower for office–workers.

The 1995 regression results suggest that while the gender earnings gap still increase with age among laborers, the slope of the age–earning profile of female office–workers does not significantly differ from that of female laborers. However, the average earnings of office–workers are significantly higher than laborers even after controlling for education and age. For example, the average earnings of male office–workers is 0.077 log points higher than male laborers, and the average earnings of female office–workers is 0.182 log points higher than female laborers. We suggest the difference between the 1988 and 1995 results is due to the shocks introduced by the profound urban reforms in 1995. Given the size of these shocks, the gradual changes in the gender earnings gap over life–cycle might be dominated by these one time changes. To test this hypophysis, we re-run the earnings regression using the retrospective 1993 earnings contained in the 1995 CHIP survey. The regression results are comparable to the 1988 results, suggesting 1995 is a abnormal year.

The 2002 regression results reveal similar pattern as the 1988 results do. Namely the coefficient on the interaction between gender and age is negative and the coefficient on the interaction between gender, age and office–worker is positive. However, some caveats need to be raised before we can conclude that the difference between the slope of age–earnings of office–workers and laborers is the result of the differences in human capital investment. If younger women faced less discrimination than older women at labor market entry, then the cross section estimate of the coefficient on the interaction between gender and age would be downward biased. We believe this is unlikely the case. Given the fact that many existing studies find that the gender earnings gap has increased considerably even after controlling for many observed characteristics, it is highly unlikely that women of younger cohort faced less discrimination. In addition, even if younger women indeed faced less discrimination, the impact of variations in the discrimination across birth cohorts on earnings should be captured by the coefficient on $F * A$.

Another factor that can cause the gender earnings gap to be negatively correlated with age is the sample selection. If the sample selection is positive for the younger cohort and negative for the older cohort, the cross cohort differences in employed women's unobserved skills could also introduce a spurious negative correlation between gender earnings gap and age. If this is the case, then it is difficult to argue why we cannot observe a similar pattern among office–workers. While it is harder for these alternative explanations to reconcile the difference in the coefficients on $F * A$ and on $F * C * A$, the human capital theory can provide a consistent story. In addition, the theory can also explain why the difference between the age–earnings profile of female office–workers and female laborers should be bigger in 2002 than in 1988. This is because the higher probability of being laid off gives female laborers an even weaker incentive to invest in their human capital on top of the difference in the length of working life caused by mandatory retirement age.

4.2 Changes in employment rate and earnings gap

To jointly estimate the earnings equation and employment equation, we need variables that affect a person's employment decision but not her earnings. For women, we use a person's marital status, the presence of at least a young child in the household, and the total earnings of other household members. For men, we only use the latter two variables. The reason for not including men's marital status in their employment equation is that martial status affects both men's earnings and employment (e.g. Korenman and Neumark 1991). To capture the potential nonlinearity between other household members' income and employment, the income is included as three income quartile dummies. The reference group consists of people whose household income (excluding her own income) is at the first quartile of the income distribution. Because women's retirement age depends on her occupation, we would like to separate office–workers from laborers in our regressions. Unfortunately, we cannot observe a person's occupation if she is not employed. To capture the potential difference in the age–earnings profile across occupations, we split the sample into two groups—without a high school diploma, with at least a high school diploma. The rationale of doing this is that a high school diploma is almost the minimum requirement for an office–job, particularly for the younger cohorts.

The first stage regression results are reported in Tables 6 and 7 for women and men respectively. The coefficient on the inverse of Mill's ratio is also reported at the bottom of these tables. Because men's employment rate is close to 100% in 1988, we only jointly estimate men's earnings equation and employment equation in 1995 and 2002. Similar argument applies to better educated women.

The results in Tables 6 show that education is one of the most important determinants for women's employment status. However, the coefficient on marital status is never significant even at the 10% level. The coefficient on having a young child in the household is only significant in the 2002 regression. A young child has a negative impact on the employment rate of less educated women, but a positive impact on the employment rate of better educated women. The weak impact of marital status and having young child on women's employment rate is likely due to the small family size, and relatively cheap and readily available child care centers in China.

Income of other household members affect negatively the employment rate of less educated women, but positively the employment rate of better educated women. The coefficient on the inverse of Mill's ratio shows that less educated women are positively selected in 2002, but better educated women are negatively selected in both 1995 and 2002. While the increase in earnings inequality could explain the positive selection of less educated women's employment, the negatively selection of better educated women is a bit harder to explain. We suggest the negative selection is because better educated women with higher unobserved ability have a higher chance to marry wealthy men. As a result, withdrawing from the labor market is a viable option for them. Although we have controlled other household members' income in the employment equation, it is quite possible that annual income flow is not a good measure for a household's wealth.

The results reported in Table 7 show that sample selection has no significant effect on men's earnings equation except for the less educated in 1995. The negative selection of less educated men in 1995 is due to the mass layoff of state owned enterprises. The statistics reported in Table 2 shows that the probability of working for SOE is higher for men than for women. As a result, an unemployed man is also more likely to be a former SOE employee. Because SOE workers, on average, earn more than non-SOE workers even after controlling for observed characteristics, the over-representation of former SOE employees in the unemployment pool could lead to a negative selection.

Table 8 reports the estimated earnings gap for various age groups. Panel A reports the estimates after controlling for potential sample selection while Panel B reports the estimates without controlling for sample selection. The differences between Panel A and Panel B suggest that while the OLS estimates understate the gender earnings gap for less educated workers, they overstate the gap for better educated workers. For workers without a high school diploma, the results reported in both panels suggest that gender gap widened slightly from 1988 to 1995 and considerably from 1995 to 2002. For workers with at least a high school diploma, the gender gap widened for some age groups and narrowed for others. Hence, we can conclude that the widening gender earnings gap in the period of 1988–2002 is primarily driven by the decline in the relative earnings of less educated women.

Table 9 reports the estimation results where the estimated gender earnings gap reported in Table 8 are used as the dependent variable. The sample selection corrected gender earnings gap is used as the dependent variable in columns (1)–(3) and the OLS estimate is used as the dependent variable in column (4). The coefficient on women's employment rate is 0.851 (SD=0.180), suggesting that a one percentage point increase in women's employment rate reduces the gender earnings gap by 0.851 percentage points. We attribute the strong positive relationship between women's employment rate and relative earnings to their PSI decisions.

Nevertheless, the documented relationship is also consistent with the argument that variations in gender earnings gap are primarily driven by the demand for female workers. A lower demand for female workers will reduce both women's employment rate and the price of their labor services. If this is the case, then women's relative earnings should have an even stronger positive correlation with the relative demand for female labor. However, the coefficient on the gender employment gap, a measure of relative demand for women's labor services, is 0.594 (SD=0.230), which does not support the above argument. We also check whether the impact of women's employment rate on the gender earnings gap reflects the effect of overall labor market conditions. Presumably, men's employment rate should be a better indicator for overall labor market conditions. The coefficient on men's employment rate is 0.673 (SD=0.474) with a R^2 of 0.89. Both the coefficient and R^2 are smaller than what have been reported in column (1). Overall, the results from these robust check imply that it is the women's employment rate that has the strongest positive correlation with the gender earnings gap. This relationship provides further support for human capital interpretation of the gender earnings gap.

To see whether our estimations results are sensitive to the controlling for sample selection, column (4) of Table 9 reports the estimation results where the OLS estimate of the gender earnings gap is used as the dependent variable. The coefficient on women's employment is 0.490 and statistically significant at the 1% level, suggesting the gender earnings gap is strongly positively correlated with women's employment rate even without controlling for sample selection.

4.3 Counterfactual analyses

Once we have consistent estimates of gender earnings gap for different age and education groups, we can conduct counterfactual analyses to address the contributions of various factors to the changes in the gender earnings gap.

First, we would likely to know how much of the increase in gender earnings gap is due to changes in the age composition of the labor force. The results reported in Table 5 show that gender earnings gap increases with age and the sample statistics reported in Table 2 show the average age of women increased by 2.6 years between 1988 and 2002. The increase in women's age would have widened the gender earnings gap even if the wage structure did not change at all. To examine the contribution of changes in the age composition of the female work force to the gender earnings gap, we construct two type of aggregate earning gap measures. One is called varying weighted gap, G_t^{vs}, and is defined as

$$G_t^{vs} = \sum_g \omega_{gt}^s \hat{\mu}_{gt}^s, \tag{9}$$

where the superscripts v means varying weight and $s = l, h$ denote less educated and better educated respectively, ω_{gt}^s is the employment share of group g women with education s in year t, and $\hat{\mu}_{gt}^s$ is the estimated gender earnings gap for group g in year t. Another one is called fixed weighted gap, G_t^{fs}, and defined as

$$G_t^{fs} = \sum_g \omega_g^s \hat{\mu}_{gt}^s, \tag{10}$$

where $\omega_g^s = \sum_t n_{gt}^s / \sum_t n_t^s$, where n_{gt}^s is the number of employed group g women with education s and n_t^s is the number of employed women with education s in year t.

Second, we would like to know the contribution of variations in women's employment rate to the gender earnings gap. To do this, we first predict gender age earning gap for each age group ($\tilde{\mu}_{gt}$) using the estimation results reported in Table 9. We then replace women's employment rate in year t with its corresponding 1988 value and re-run the prediction to get another predicted gender earnings gap $\tilde{\mu}_{gt}^{88}$. By aggregating $\tilde{\mu}_{gt}$ and $\tilde{\mu}_{gt}^{88}$ according to equations (9) and (10), we obtain two sets of predicted aggregate gender earnings gap. The difference between these two sets of predicted values reveals the contributions of variations in women's employment rate.

Tables 10 reports the aggregate earnings gap for less and better educated workers separately. The difference between the varying weighted and fixed weighted gap suggests that changes in the age composition indeed widened the gender earnings gap, but its contribution is very small. For instance, for less educated workers, results reported in panel A show that while G^v widened by 0.194 log points between 1988 and 2002, G^f widened by 0.178 log points. This suggests that changes in age composition widened the earnings gap by 0.016 log points, which accounts for 8.2% of the total change. A comparison between results in panels A and B shows that changes in sample selection mitigated the increase in gender earnings gap for less educated and exaggerated the increase for better educated. Results in panel C show that variations in $\tilde{\mu}_{gt}$ match the variations in $\hat{\mu}_{gt}^{hk}$ very well. However, results in panel D show $\tilde{\mu}_{gt}^{88}$ is a poor predictor of $\hat{\mu}_{gt}^{hk}$, suggesting most of the changes in gender earnings gap are due to changes in women's employment rate rather than cohort specific factors.

To have a complete picture on the variation in gender earnings gap, we further aggregate G_t^h and G_t^l into a population wide measure. The results are reported in Table 11. Interestingly, the contributions of age composition and sample selection to the changes in the gender earnings gap of the entire population are smaller than what have been documented in Table 10. For example, results in panel A show that while the varying weighted gender earnings gap widened by 0.087 log points between 1988 and 2002, the fixed weighted gender earnings gap widened by 0.09 log points. A comparison between the results reported in panel A and B suggests that controlling for sample selection also has little impact on the estimated gender earnings gap of the entire population. Finally, the difference between the results reported in panel C and D shows that variations in women's employment rate is the main driving force for the variations in gender earnings gap. If women's employment rate was fixed at its 1988 level, the gender earnings gap would have been narrowed by 4.7 percentage points rather than widened by 8.7 percentage points.

5 Conclusion

This paper examines the contribution of gender differences in post school investment in human capital to the gender earnings gap in China. We use two approaches to address this issue. First, we compare the age–earnings profile of office–workers and laborers. Because the mandatory retirement age of female laborers is 5 years younger than the retirement age of female office–workers and is 10 years younger than all male workers, female laborers have less incentive to invest than male workers and female office–workers, and female office–workers have less incentive to invest than male workers. If differences in human capital investments play any role in determining the size of the gender earnings gap, the theory predict female laborers have a flatter age–earnings profile than both female office–workers and male workers, and female office–workers have a flatter age–earnings profile than male workers. Our regression results support the prediction of theory.

In the second approach, we exploit the impact of the decline in women's employment rate caused by China's economic reforms on gender earnings gap. The data show that women's employment rate declined from 98.4% in 1988 to 81.4% in 2002. Because the decline in women's employment prospect shortens women's working life, it discourages them to invest in their human capital. By regressing the estimated cohort and age specific gender earnings gap on women's employment rate, we find that a one percentage point reduction in women's employment rate is associated with a 0.851 percentage points rise in gender earnings gap even after controlling for age and cohort fixed effects. We also find that gender earnings gap is only weakly positively correlated with men's employment rate and gender employment gap, suggesting the stronger correlation between gender earnings gap and women's employment rate is not the results of the responses of women's earnings to the overall labor market condition or to the relative demand for female workers. These findings provide further supports for the prediction of the human capital theory.

To gauge the contribution of various factors to the change in gender earnings gap in China, we conduct a counterfactual analysis by predicting the gender earnings gap if women's employment rate is fixed at its 1988 level and the age composition at the mean of the entire sample period. Our results suggests that while changes in the age composition exaggerates the increase in gender earnings gap, the size of the increase is small. The majority of the changes in earnings gap is associated with changes in women's employment rate. If women's employment rate is fixed at its 1988 level, the gender earnings gap would have been narrowed by 4.7 percentage points rather than widened by 8.7 points.

References

Appleton, S., J. Knight, L. Song, and Q. Xia (2002). Labor retrenchment in China: Determinants and consequences. *China Economic Review 13*(2-3), 252–275.

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources 8*(4), 436–455.

Giles, J., A. Park, and F. Cai (2006). How has economic restructuring affected China's urban workers? *China Quarterly 185*, 61–95.

Gronau, R. (1988). Sex-related wage differentials and women's interrupted labor careers-the chicken or the egg. *Journal of Labor Economics 6*(3), 277–301.

Gustafsson, B. and S. Li (2000). Economic transformation and the gender earnings gap in urban China. *Journal of Population Economics 13*(2), 305–329.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and social Measurement 5*(4), 475–492.

Junh, C., K. M. Murphy, and B. Pierce (1991). Accounting for the slowdown in Black-White wage convergence. In M. Koster (Ed.), *Workers and their wages*, pp. 107–143. Washington D.C.: AEI press.

Knight, J. and L. Song (2003). Increasing urban wage inequality in China. *Economics of Transition 11*(4), 597–619.

Korenman, S. and D. Neumark (1991). Does marriage really make men more productive? *Journal of Human Resources*, 282–307.

Liu, P.-W., X. Meng, and J. Zhang (2000). Sectoral gender wage differentials and discrimination in the transitional Chinese economy. *Journal of Population Economics 13*(2), 331–352.

Mulligan, C. B. and Y. Rubinstein (2008). Selection, investment, and women's relative wages over time. *Quarterly Journal of Economics 123*(3), 1061–1110.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*(3), 693–709.

Olivetti, C. and B. Petrongolo (2008). Unequal pay or unequal employment? a cross-country analysis of gender gaps. *Journal of Labor Economics 26*(4), 621–654.

O'Neill, J. and S. Polachek (1993). Why the gender gap in wages narrowed in the 1980s. *Journal of Labor Economics 11*(1), 205.

Zhang, J., J. Han, P.-W. Liu, and Y. Zhao (2007). Trends in the gender earnings differential in urban China, 1988-2004. *Industrial and Labor Relations Review 61*(2), 224–243.

Table 1: Total number of “staff and workers”, by registration type

		Number of workers in 10,000			Employment share in %		
Year	Total	State owned	Collective owned	Other	State owned	Collective owned	Other
1984	11890	8637	3216	37	72.6	27.0	0.3
1985	12358	8990	3324	44	72.7	26.9	0.4
1986	12809	9333	3421	55	72.9	26.7	0.4
1987	13214	9654	3488	72	73.1	26.4	0.5
1988	13608	9984	3527	97	73.4	25.9	0.7
1989	13742	10108	3502	132	73.5	25.5	1.0
1990	14059	10346	3549	164	73.6	25.2	1.2
1991	14508	10664	3628	216	73.5	25.0	1.5
1992	14792	10889	3621	282	73.6	24.5	1.9
1993	14849	10920	3393	536	73.5	22.9	3.6
1994	14849	10890	3211	747	73.3	21.6	5.0
1995	14908	10955	3076	877	73.5	20.6	5.9
1996	14845	10949	2954	942	73.8	19.9	6.3
1997	14668	10766	2817	1085	73.4	19.2	7.4
1998	12337	8809	1900	1628	71.4	15.4	13.2
1999	11773	8336	1652	1785	70.8	14.0	15.2
2000	11259	7878	1447	1935	70.0	12.9	17.2
2001	10792	7409	1241	2142	68.7	11.5	19.8
2002	10558	6924	1071	2563	65.6	10.1	24.3
2003	10492	6621	951	2920	63.1	9.1	27.8
2004	10576	6438	851	3287	60.9	8.0	31.1

Data Source: The information is extracted from Table 1-14 of the 2005 China Labour Statistical Yearbook.

Table 2: Sample Statistics

	Females			Males		
	1988	1995	2002	1988	1995	2002
Age	35.5	36.5	38.1	36.8	37.6	39.5
Age 19–25	.16	.129	.097	.155	.132	.094
Age 26–32	.216	.198	.163	.182	.173	.135
Age 33–39	.285	.259	.262	.268	.231	.23
Age 40–46	.207	.301	.298	.199	.289	.274
Age 47–53	.132	.113	.18	.196	.175	.267
Schooling	10.2	10.4	11.1	11	11.1	11.5
Employed	.984	.936	.814	.997	.965	.908
Inc. of other mem.	4916	6705	14022	4088	5731	11340
Married	.789	.87	.855	.774	.84	.838
With young kids	.513	.395	.3	.512	.383	.278
Annual Earnings	3885	5788	10933	4487	6716	13418
SD of earnings	2289	3402	7674	2098	3823	9466
Sectors:(%)						
Primary	4.25	2.17	2.91	6.56	3.12	4.8
Manufacturing	45.3	40.9	25	42.8	43.4	28
Construction	2.64	2.3	2.24	3.93	3.02	4.06
TCP	5.4	4.12	5.58	8.88	5.99	11
TRW	17	16.3	14.8	10.2	11.6	10.2
HEA	14.1	14.2	16.2	9.42	9.27	11.7
Government	5	10.3	9.97	11	14.1	13
Services	6.17	8.85	20	7.14	8.41	14
Others	.137	.846	3.25	.119	1.04	3.22
Occupation:(%)						
Managers	2.41	6.83	9	9.67	16.4	18.7
Professionals	16.7	23.5	22.3	15.3	21.6	19.2
Clerks	20.8	21.7	24	25.1	19.8	18.1
Laborers	60	48	44.6	50	42.2	44
Ownership:(%)						
SOE	72	76.7	58.6	85	85.6	64.1
Non–SOE	28	23.3	41.4	15	14.4	35.9
No. of observations	5970	4157	4037	5928	4206	4195

Data Source: Full time students, self-employed workers and retirees are excluded from the sample.

Table 3: Employment rate by age, gender and birth cohort

Birth year	Males					Females				
	19-25	26-32	33-39	40-46	47-53	19-25	26-32	33-39	40-46	47-53
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	A: Lower than high school									
1977-83	53.3	–	–	–	–	58.8	–	–	–	–
1970-76	76	80	–	–	–	79.8	70	–	–	–
1963-69	98.7	94.7	91.2	–	–	98.1	89.4	73.5	–	–
1956-62	–	100	98.7	93.2	–	–	98.6	94.8	78	–
1949-55	–	–	100	98.4	89.5	–	–	98.7	96.6	71.2
1942-48	–	–	–	100	98.9	–	–	–	97.8	82.3
1935-41	–	–	–	–	99.6	–	–	–	–	92.5
	B: High school and above									
1977-83	71.9	–	–	–	–	79.7	–	–	–	–
1970-76	86.2	93.9	–	–	–	90.2	85.5	–	–	–
1963-69	97.9	98.8	97.3	–	–	98.3	97.2	92.5	–	–
1956-62	–	99.9	100	96.3	–	–	99.8	98.4	94.3	–
1949-55	–	–	100	99.8	97.1	–	–	100	98.9	94
1942-48	–	–	–	100	99.5	–	–	–	99.3	98.4
1935-41	–	–	–	–	100	–	–	–	–	99.4

Table 4: Employment rate of married individuals by age, gender and birth cohort

Birth year	Males					Females				
	19-25	26-32	33-39	40-46	47-53	19-25	26-32	33-39	40-46	47-53
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	A: Lower than high school									
1977-83	100	–	–	–	–	33.3	–	–	–	–
1970-76	87.1	88.8	–	–	–	78.8	70.6	–	–	–
1963-69	94.4	97.6	92.9	–	–	100	90.3	73.1	–	–
1956-62	–	100	99	93.3	–	–	98.8	95	78.2	–
1949-55	–	–	100	98.8	89.8	–	–	98.7	96.7	70.3
1942-48	–	–	–	100	99.1	–	–	–	97.7	82.3
1935-41	–	–	–	–	100	–	–	–	–	93
	B: High school and above									
1977-83	88.2	–	–	–	–	69.7	–	–	–	–
1970-76	91.7	98.5	–	–	–	95.7	87	–	–	–
1963-69	100	99.4	97.9	–	–	100	97	92.6	–	–
1956-62	–	100	100	96.4	–	–	99.7	98.4	94.3	–
1949-55	–	–	100	99.8	97.1	–	–	100	99.2	94.1
1942-48	–	–	–	100	99.5	–	–	–	99.3	98.4
1935-41	–	–	–	–	100	–	–	–	–	99.4

Table 5: The relationship between occupation and gender earnings gap

	1988	1995	2002
	(1)	(2)	(3)
Age-19	.053	.062	.050
	(.001)***	(.003)***	(.004)***
$(\text{Age-19})^2$	-.0009	-.001	-.0008
	(.00004)***	(.00007)***	(.00009)***
Schooling	.014	.026	.048
	(.001)***	(.002)***	(.003)***
Office worker	-.046	.077	.197
	(.020)**	(.038)**	(.048)***
State owned units	.156	.182	.183
	(.008)***	(.015)***	(.014)***
Women	.0005	-.115	-.022
	(.016)	(.038)***	(.050)
Women*age	-.008	-.004	-.011
	(.001)***	(.002)**	(.002)***
Age*office worker	.002	-.0002	-.0003
	(.001)**	(.002)	(.002)
Women*office worker	.011	.115	-.121
	(.028)	(.053)**	(.068)*
Women*age*office worker	.003	.001	.010
	(.001)**	(.003)	(.003)***
Province of residence	Yes	Yes	Yes
Obs.	11783	7948	7096

Note: Numbers in parenthesis are standard errors. *** means significant at the 1% level, ** means significant at the 5% level and * means significant at the 10% level.

Table 6: Coefficients from employment equation, women

	Without high school			High school and higher		
	1988	1995	2002	1988	1995	2002
Schooling	.211	.170	.141		.198	.167
	(.031)***	(.020)***	(.015)***		(.066)***	(.029)***
Age 26–32	-.196	.228	.622		.172	-.118
	(.322)	(.162)	(.191)***		(.279)	(.144)
Age 33–39	-.083	.653	.602		.410	.354
	(.303)	(.167)***	(.185)***		(.297)	(.159)**
Age 40–46	.014	1.059	.581		.443	.550
	(.307)	(.173)***	(.182)***		(.357)	(.175)***
Age 47–53	-.305	.315	.540		.238	.550
	(.304)	(.179)*	(.188)***		(.378)	(.190)***
HH inc. 2nd quartile	.205	.283	-.252		.289	-.033
	(.182)	(.108)***	(.080)***		(.234)	(.117)
HH inc. 3rd quartile	-.005	.109	-.510		.275	.206
	(.191)	(.118)	(.088)***		(.199)	(.121)*
HH inc. 4th quartile	-.354	-.079	-.651		.427	.252
	(.206)*	(.136)	(.109)***		(.195)**	(.127)**
Married	.247	.147	-.097		.311	.148
	(.225)	(.153)	(.140)		(.264)	(.137)
With young kid	.154	.165	-.277		.012	.182
	(.173)	(.111)	(.084)***		(.200)	(.110)*
Constant	.412	-.414	.148		-.520	-.948
	(.497)	(.353)	(.288)		(.973)	(.424)**
No. of obs.	3212	2575	2070		1582	1967
λ	-.014	.024	.166		-.331	-.430
	(.051)	(.061)	(.055)***		(.068)***	(.034)***

Note: Numbers in parenthesis are standard errors. *** means significant at the 1% level, ** means significant at the 5% level and * means significant at the 10% level.
HH inc. is the total income of other household members. People in the first quartile of the HH inc. distribution are used as the reference group.
λ is the inverse of Mill's ratio.

Table 7: Coefficients from employment equation, men

	Without high school			High school and higher		
	1988	1995	2002	1988	1995	2002
Schooling		.053	.035		.056	.136
		(.028)*	(.020)*		(.046)	(.031)***
Age 26–32		.755	.575		.960	.759
		(.163)***	(.172)***		(.242)***	(.142)***
Age 33–39		1.184	1.054		8.580	.961
		(.187)***	(.169)***		(3.51e+08)	(.164)***
Age 40–46		1.238	1.393		1.755	1.098
		(.161)***	(.156)***		(.356)***	(.140)***
Age 47–53		1.286	1.179		1.485	1.334
		(.218)***	(.147)***		(.273)***	(.155)***
HH inc. 2nd quartile		-.120	-.109		.237	-.043
		(.132)	(.099)		(.273)	(.144)
HH inc. 3rd quartile		-.095	-.185		-.058	-.329
		(.125)	(.122)		(.228)	(.141)**
HH inc. 4th quartile		.194	-.132		.066	-.141
		(.147)	(.155)		(.220)	(.166)
With young kid		.197	.543		.365	.500
		(.124)	(.118)***		(.262)	(.150)***
Constant		.628	.252		.316	-.961
		(.411)	(.304)		(.656)	(.438)**
No. of Obs.		2204	1949		2002	2246
λ		-.455	.005		-.007	-.037
		(.025)***	(.097)		(.047)	(.070)

Note: Numbers in parenthesis are standard errors. *** means significant at the 1% level, ** means significant at the 5% level and * means significant at the 10% level. HH inc. is the total income of other household members. People in the first quartile of the HH inc. distribution are used as the reference group. λ is the inverse of Mill's ratio.

Table 8: Earnings gap by age and year

	Less than high school			Graduated from high school		
	1988	1995	2002	1988	1995	2002
	(1)	(2)	(3)	(4)	(5)	(6)
			A: Two steps			
Age 19-25	-0.020	-0.381	-0.159	-0.037	-0.039	0.090
Age 26-32	-0.108	-0.149	-0.388	-0.057	-0.047	-0.032
Age 33-39	-0.129	-0.124	-0.293	-0.083	-0.097	-0.057
Age 40-46	-0.168	-0.208	-0.361	-0.080	-0.057	-0.157
Age 47-53	-0.239	-0.280	-0.317	-0.091	-0.041	-0.150
			B: OLS			
Age 19-25	-0.020	-0.181	-0.063	-0.037	-0.102	-0.041
Age 26-32	-0.109	-0.091	-0.311	-0.057	-0.065	-0.136
Age 33-39	-0.130	-0.106	-0.228	-0.083	-0.110	-0.114
Age 40-46	-0.168	-0.188	-0.304	-0.080	-0.064	-0.198
Age 47-53	-0.241	-0.260	-0.251	-0.091	-0.052	-0.195

Note: Earnings gap is the difference between the predicted earnings using the coefficients from men's

Table 9: Gender earnings gap and women's employment prospect

	Two steps			OLS
	(1)	(2)	(3)	(4)
Women's employment rate	.851			.490
	(.180)***			(.172)***
Employment rate gap		.594		
		(.230)***		
Men's employment rate			.673	
			(.474)	
Age 26-32	.018	-.020	-.082	-.037
	(.028)	(.032)	(.031)***	(.027)
Age 33-39	.002	-.064	-.132	-.071
	(.039)	(.042)	(.035)***	(.037)*
Age 40-46	-.033	-.127	-.194	-.124
	(.051)	(.051)**	(.046)***	(.048)**
Age 47-53	.004	-.132	-.225	-.122
	(.069)	(.068)*	(.058)***	(.066)*
Cohort fixed effect	Yes	Yes	Yes	Yes
Obs.	30	30	30	30
R^2	.963	.922	.894	.928

Note: Numbers in parenthesis are standard errors. *** means significant at the 1% level, ** means significant at the 5% level and * means significant at the 10% level. The dependent variable in columns (1)–(3) is the estimated gender earnings gap reported in panel A of Table 8 and the dependent variable in column (4) is the estimated gender earnings gap reported in panel B of Table 8.

Table 10: Predicted earnings gap by education level

	Less than high school			Graduated from high school		
	1988	1995	2002	1988	1995	2002
	(1)	(2)	(3)	(4)	(5)	(6)
A: Two steps						
Varying weight	-0.137	-0.196	-0.331	-0.066	-0.060	-0.071
Fixed weight	-0.143	-0.197	-0.321	-0.069	-0.059	-0.057
B: OLS						
Varying weight	-0.138	-0.155	-0.267	-0.066	-0.080	-0.141
Fixed weight	-0.144	-0.157	-0.254	-0.069	-0.080	-0.135
C: Predicted μ_g 2-step						
Varying weight	-0.142	-0.185	-0.327	-0.063	-0.063	-0.076
Fixed weight	-0.148	-0.189	-0.315	-0.064	-0.061	-0.063
D: Predicted μ_g 2-step, using 1988 women's ER						
Varying weight	-0.142	-0.137	-0.129	-0.063	-0.042	0.004
Fixed weight	-0.148	-0.138	-0.105	-0.064	-0.038	0.026

Note: The statistics reported in panel A and B are constructed using the estimated gender earnings gap reported in panel A and B of Table 8 respectively. The statistics reported in panel C are constructed using the predicted gender earnings gap. The statistics reported in panel D are constructed using the predicted gender earnings gap by fixing women's employment rate at its 1988 level.

Table 11: Predicted earnings gap

	1988	1995	2002
	A: Two steps		
Varying weight	-0.104	-0.142	-0.191
Fixed weight	-0.109	-0.133	-0.199
	B: OLS		
Varying weight	-0.104	-0.125	-0.199
Fixed weight	-0.109	-0.121	-0.199
	C: Predicted μ_g		
Varying weight	-0.105	-0.137	-0.192
Fixed weight	-0.109	-0.129	-0.198
	D: Predicted μ_g based on 1988 women's ER		
Varying weight	-0.105	-0.099	-0.058
Fixed weight	-0.109	-0.092	-0.045

Note: The statistics reported in panel A and B are constructed using the estimated gender earnings gap reported in panel A and B of Table 8 respectively. The statistics reported in panel C are constructed using the predicted gender earnings gap. The statistics reported in panel D are constructed using the predicted gender earnings gap by fixing women's employment rate at its 1988 level.