

# Composition effect matters: Decomposing the gender pay gap in Chinese university graduates

Qinying He, Lin Xu & Yao Men

To cite this article: Qinying He, Lin Xu & Yao Men (2020) Composition effect matters: Decomposing the gender pay gap in Chinese university graduates, Economic Research-Ekonomika Istraživanja, 33:1, 847-864, DOI: [10.1080/1331677X.2020.1734850](https://doi.org/10.1080/1331677X.2020.1734850)

To link to this article: <https://doi.org/10.1080/1331677X.2020.1734850>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 16 Mar 2020.



Submit your article to this journal [↗](#)



Article views: 945



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

# Composition effect matters: Decomposing the gender pay gap in Chinese university graduates

Qinying He<sup>a</sup>, Lin Xu<sup>b</sup> and Yao Men<sup>c</sup>

<sup>a</sup>College of Economics and Management, South China Agricultural University, Guangzhou, China;

<sup>b</sup>School of Data Sciences, Zhejiang University of Finance and Economics, Hangzhou, China; <sup>c</sup>Research and Development, China Asset Management Company, Beijing, China

## ABSTRACT

The purpose of this article is to decompose the composition effect and wage structure effect of the gender starting pay gap in Chinese university graduates at every quantile. The article aims to determine if the pay gap at every quantile is a result of gender characteristics difference, or the differences in returns to those characteristics. A 2007 Chinese university survey of new graduates employment and capacity conducted by an education research company MyCOS institute is used. This article exploits a counterfactual decomposition analysis using quantile regression to decompose the gender pay gap into one component that is based on differences in characteristics and one component that is based on differences in coefficients across the log wage distribution. We find that the majority of the gender pay differential is attributed to the gender difference in the endowment of human capital and the composition effect explains 30–60% of the pay difference at each quantile of the log wage distribution. It means that female graduates have almost the same rewards to characteristics as their male counterparts, especially at the bottom of the log wage distribution. We also find that female graduates have lower mean work capacity than male graduates and work capacity is positively related with wage. This article provides policy implications on how to reduce the gender pay gap after higher education reform in a transition economy.

## ARTICLE HISTORY

Received 26 July 2019

Accepted 20 February 2020

## KEYWORDS

gender pay gap; entry wage; quantile regression; counterfactual decomposition; higher education; China

## JEL CODES

J31; J40; I26

## 1. Introduction

Increased attention has been paid to examining the changes in the gender wage gap in urban China before and after introduction of market reforms (Gustafsson & Li, 2000; Hughes & Maurer-Fazio, 2002; Liu, Meng, & Zhang, 2000; Meng, 1998; Wang & Cai, 2008). The gender pay gap is attributable to the lower level of females' education. There is higher education expansion in China starting the early 2000s, then more and more high school students have the chance to enter college. The quality of

**CONTACT** Lin Xu  [xulin@zufe.edu.cn](mailto:xulin@zufe.edu.cn)

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

women's education and experience has improved with rising percentages of bachelor, master and Ph.D. degrees received by women since 2000. Before higher education reform, each college graduate in China was guaranteed a government-assigned job through a centralised placement system and the pay was also fixed. While the reform has offered graduates and their employers new opportunities, employers (including government recruiters) often openly discriminate against job applicants with certain characteristics, for examples, restrictions regarding gender, household registration status (Hukou), etc. then analysing the gender pay gap and the relative treatment of women with higher education by employers is of interest.

The existing evidence suggests that the gender pay gap is typically fairly small on entry to the labour market, but after a few years a significant wage gap emerges (e.g. Manning and Swaffield (2008)). There are several causes of the wider gender pay gap becoming noticeable after a few years. Bertrand, Goldin, and Katz (2010) pointed out that a large part of the gender gap in earnings can be attributed to differences in career interruptions between men and women due to traditional gender roles, in which women shoulder the responsibility of childcare and housework. As Babcock and Laschever (2003) stated, faced with competitive environments women are less willing to aggressively negotiate for promotion. What's more, Qu, Guo, and Wang (2019) found that returns to experience are higher for men than women, especially for married men and women, using data from China. When the gender pay gap between employees at different points of their career and with unequal years of labour market experience are examined, the gap may be caused by the aforementioned reasons. Restricting the analysis to entry wages can better identify how much the gap is attributed to gender discrimination. Concentrating on entry wages has several advantages, Bredtmann and Otten (2014) provided empirical evidence on the gender wage pay of labour market entrants and the determinants of their starting wages using a data set on economics graduates from a large German university.

This article will analyse the entry wage of Chinese university graduates to examine the gender pay gap for higher education groups. We disentangle the gender pay gap by counterfactual decomposition. We construct the counterfactual wage of female graduates if they had the same composition characteristics as male graduates. The difference between the counterfactual wage and the actual wage of females is then purely due to the differences in composition characteristics between male and female, and we call this part the composition effect (characteristics effect, endowment effect). The difference between the actual wage of males and the counterfactual wage is due to the difference in the returns to the covariates, which is called the wage structure effect (coefficient effect). The overall wage structure can also have an important effect on the gender pay because it describes the array of prices set for various labour market skills (measured and unmeasured) and the rents received for employment in particular sectors of the economy.

We use the counterfactual decomposition method of Machado and Mata ([MM] 2005)). The MM method has been widely used in wage distribution decomposition, and it generalises the traditional Oaxaca's (1973) decomposition of effects on mean wages to the entire wage distribution. The MM method uses conditional quantile regression and has previously been used by Fang and Sakellariou (2011) to decompose the gender pay gap in Thailand. An alternative decomposition method, the

unconditional distribution decomposition proposed by Firpo, Fortin, and Lemieux (2007, 2009), is used in Chi and Li (2008) and Xiu and Gunderson (2014) to make detailed sub-decompositions possible to see the contribution of each individual variable to the gender earnings gap.

This article contributes to the literature on the gender pay gap in higher education groups after higher education reform in a transition economy. This article aims to decompose the composition effect and wage structure effect of the gender starting pay gap in Chinese university graduates at every quantile. It means that we want to check if the pay gap at every quantile is mainly a result of gender characteristics difference, or the differences in returns to those characteristics. Specifically we calculate the contribution of composition effect to the gender pay gap over the entire wage distribution of Chinese university graduates, which reveals how much the gender pay gap in Chinese university graduates is explained by endowment in human capital. Using a 2007 Chinese university survey of new graduates' employment and capacity, conducted by education research company the MyCOS institute, our empirical analysis finds that the composition effect explains about 30–60% of the gender pay gap over the distribution. Our study gives a better picture of how much composition effect contributes to the gender pay gap in Chinese university graduates.

While traditional investigations into the gender gap tend to focus on gender discrimination, the latest studies drill into various gender differences as potential causes of the gap (Ahamed, Wen, & Gupta, 2019; Guo, Song, Sun, & Wang, 2016; Ichino & Moretti, 2009). This article also contributes to the literature on the causes of gender pay gap.

The rest of this article is organised as follows. Section 2 illustrates our decomposition method. Section 3 describes our data. Section 4 presents the empirical results. In Section 5 we do some robustness checks and we conclude in Section 6.

## 2. Decomposition method

In this section, we introduce the counterfactual MM (2005) decomposition method. The counterfactual decomposition requires an estimation of the log wage distribution that is conditional on the variables of interest. We accomplish the estimated coefficients of covariates firstly by means of quantile regressions (Koenker & Bassett, 1978), which provide a more general picture of the effect of covariates on log wage at different log wage levels. Then the analysis is followed by estimating the marginal density function of log wage that is consistent with the conditional distribution estimated in the previous quantile regression. The procedure enables provision of information on the relative contribution that differences in overall endowments of log wage determining characteristics and returns to those characteristics have on the pay gap in each decile of the earnings distribution.

The quantile regression model is given by:

$$Q(Y|\mathbf{X}) = \mathbf{X}\beta \quad (1)$$

where  $Y$  is the logarithm of monthly earnings in RMB Yuan, and  $\mathbf{X}$  is a  $K+1$  dimensional vector of exogenous control variables including university academic

performance, graduation time, work experience, and employers’ characteristics such as enterprise size and a constant term. We also include a vector of job location, for example, *region1* is a binary variable that takes value 1 if one’s job is in the developed east coast region of China and 0 otherwise. Dummy variable *Gender* equals 1 if graduate is male and 0 otherwise. The  $\tau$ th quantile regression solves the minimisation problem:

$$\hat{\beta} = \underset{\beta \in \mathbf{R}^{k+1}}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{X}_i \beta_{\tau}) \tag{2}$$

where  $\rho_{\tau}$  is the check function at  $\tau$ th quantile,  $\rho_{\tau} = (\tau - 1(u \leq 0))u$  with  $1(\cdot)$  being the indicator function,  $\mathbf{R}^{K+1}$  means a  $K + 1$  dimensional real set and  $n$  is the sample size.

We want to decompose the composition effect on the gender pay gap. To make the points clear, the calculation procedure is summarised as follows:

Step 1. Generate a random sample of size  $m$  from a uniform distribution  $U(0, 1)$ . Then we sort them into ascending order to obtain  $\tau_1, \dots, \tau_d, \dots, \tau_m$  with  $1 \leq d \leq m$ .

Step 2. For the data at stage  $s$  or stage  $t$ , we get  $m$  estimates of the quantile regression coefficients  $\hat{\beta}^s = \{\hat{\beta}^s(\tau_1), \dots, \hat{\beta}^s(\tau_d), \dots, \hat{\beta}^s(\tau_m)\}$  and  $\hat{\beta}^t = \{\hat{\beta}^t(\tau_1), \dots, \hat{\beta}^t(\tau_d), \dots, \hat{\beta}^t(\tau_m)\}$  where:

$$\hat{\beta}^s(\tau_d) = \{\hat{\beta}_0^s(\tau_d), \hat{\beta}_1^s(\tau_d), \dots, \hat{\beta}_k^s(\tau_d)\}^T$$

and

$$\hat{\beta}^t(\tau_d) = \{\hat{\beta}_0^t(\tau_d), \hat{\beta}_1^t(\tau_d), \dots, \hat{\beta}_k^t(\tau_d)\}^T.$$

Step 3. Generate  $m$  rows with replacement from an  $n \times (K + 1)$  data matrix  $\mathbf{X}(s)$  to obtain an  $m \times (K + 1)$  new design matrix  $\mathbf{X}^*(s)$  for stage  $s$ . Each row is denoted by  $\mathbf{X}_d^*(s)$ . Then:

$$\{y_d^*(s) = \mathbf{X}_d^*(s) \hat{\beta}^s(\tau_d) \quad d = 1, \dots, m\} \tag{3}$$

is a random sample of size  $m$  from the desire distribution for stage  $s$ . We use  $F^*(y(s))$  to denote the estimator of cumulative distribution function (C.D.F.) of  $y(s)$  for stage  $s$ , which is estimated from the data  $y_d^*(s) = \mathbf{X}_d^*(s) \hat{\beta}^s(\tau_d), d = 1, \dots, m$  where the coefficient estimates vector  $\hat{\beta}^s(\tau_d)$  is based on the observed sample  $\{y_i(s), \mathbf{X}_i(s) : i = 1, \dots, n\}$ . The same method could also be used for stage  $t$ . We obtain the estimation of marginal density function of log wage for male and female graduates.

Step 4. Let  $F^*(y(t); \mathbf{X}^*(s))$  denote the C.D.F. of the counterfactual log wage where  $\mathbf{X}^*(s)$  indicates the generated design matrix for stage  $s$ . For example, it is used to present the C.D.F. of the counterfactual log wage of female graduates if they had the same characteristics as male graduates (stage  $s$ ). We want to estimate the density function of log wages for female graduates, corresponding to the male graduates’ distribution of covariates. Just follow the third step of the algorithm

above but the counterfactual log wage is estimated from the data  $y_d^{**}(s) = \mathbf{X}_d^*(s)\hat{\beta}^t(\tau_d)$ ,  $d = 1, \dots, m$ .

Step 5. With the notations introduced above, we may measure the contribution of composition effect by looking at indicators such as:

$$Eff = \frac{\hat{q}_\theta(F^*(y(t); \mathbf{X}^*(s))) - \hat{q}_\theta(F^*(y(t)))}{\hat{q}_\theta(F^*(y(s))) - \hat{q}_\theta(F^*(y(t)))} \quad (4)$$

where  $\hat{q}_\theta(\cdot)$  is the estimate of the  $\theta$ th quantile.

The other details can be found in MM (2005).

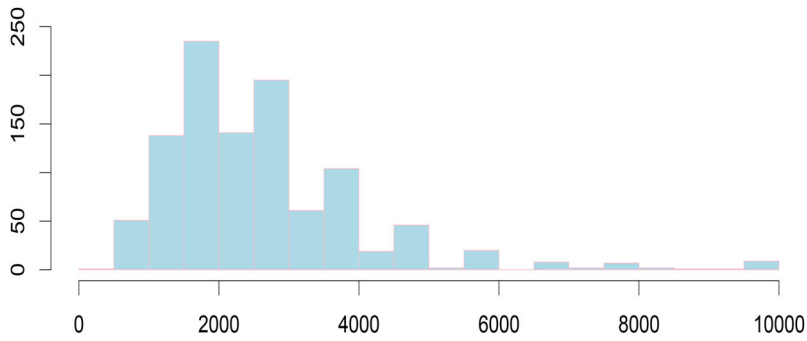
### 3. Data

The data used in this article is from the MyCOS institute<sup>1</sup> (<http://www.mycos.com/En/>). The survey was conducted among college students in China six months after their graduation, including vocational and professional school undergraduates from about 2,000 higher education institutions.<sup>2</sup> The purpose of the survey was to collect data on employment status, basic working ability and professional competence information of new university graduates, to establish an annual database. The main content of the questionnaire includes information about job-seeking, employment status and basic working ability, etc. About 167,000 questionnaires were distributed to the graduates by email at the beginning of 2007 and total of 85,000 valid questionnaires were received. Since Project 211 universities<sup>3</sup> take on the responsibility of training one-third of undergraduates, offer 85% of the state's key subjects, hold 96% of the state's key laboratories, and utilise 70% of scientific research funding, our sample in this article only includes the students who graduated from Project 211 universities. We excluded the respondents if the number of the respondents who are from the same major subject and same university is smaller than five. Finally, the sample size is 1,545 and 67.5% are male graduates, so our sample includes 1,043 male and 501 female graduates.

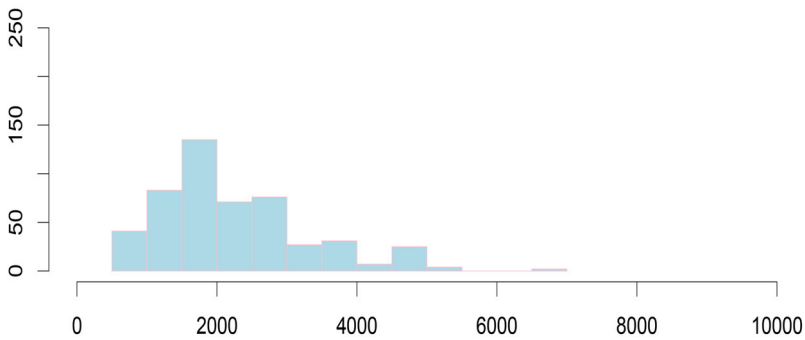
Based on data from the national census conducted in 2010, the average number of years of schooling for women was 1.1 years fewer than for men in China. while the gap was 0.93 years in urban China. Gustafsson and Li (2000) suggest that the most important source of gender wage difference is education. Here we consider the female and male samples with the same education levels to control the effect of gender education gap on gender pay gap.

Although the hourly wage rate would have been ideal to examine gender pay differentials, only monthly earnings were available in our data set. We will use the log of monthly wage<sup>4</sup> in the decomposition analysis. Figure 1 plots the histograms of the monthly wage for both sexes. The shapes of wage distribution are similar for both samples. It can be noted that the top of the wage for males is higher than that for women. The male wage is more dispersed than female wage.

The survey questionnaire contains 35 questions<sup>5</sup> to describe the different dimensions of job competency, which form an indicator of productivity affecting earnings. During the survey, each respondent first evaluates the importance of each dimension in his/her job position and then self-rates him/her upon graduation. Then the



(a) Male graduates

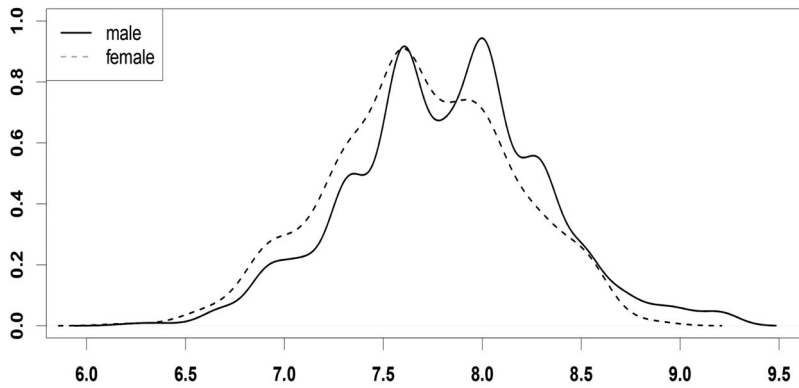


(b) Female graduates

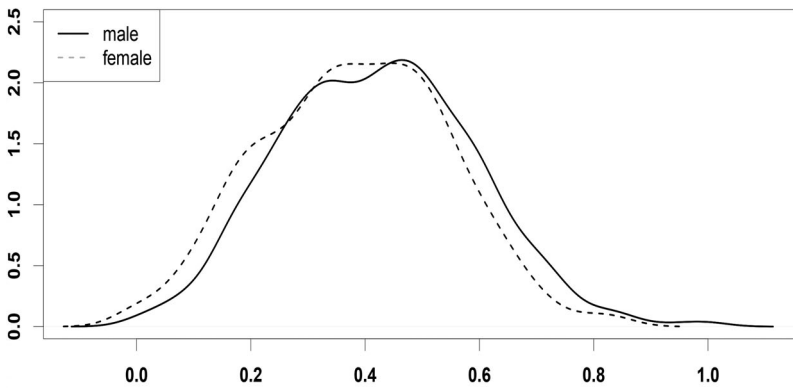
**Figure 1.** Histogram of monthly earnings for male and female university graduates.

capacity index (*indexac*) of each respondent is calculated as an importance-weighted average of individual ratings of these 35 variables. Then it is scaled to percentage with the maximum of 100%. This index represents the graduate's productivity. It considers the relative importance of capacity among different occupation and positions. The larger the value of capacity index is, the stronger job competency and productivity the graduate has. Figure 2 plots the density curves of logarithm of wage and capacity index for male and female samples. Both the distributions of logarithm of wage and the capacity index for males are clearly shifted to the right with respect to those of females.

Table 1 gives the sample mean or proportion of the key variables for samples of male and female graduates used in this empirical analysis. The questionnaire on college academic performance include four choices: poor, middle, good and excellent. We set three dummy variables (*Academic1-3*) to describe the middle, good, excellent levels of college academic performance, respectively and the base group is who has poor college academic performance. Table 1 shows that about 73% of graduates report they had middle or good college academic performance. Among our sample, 77% of graduates take the job related to their majors. In terms of job location, 41% of graduates work in sub-provincial cities<sup>6</sup> and 22% work in a municipality.<sup>7</sup> We also



(a) log wage

(b) *indexac*

**Figure 2.** Density curves of log wage and *indexac* for male and female university graduates.

set three dummy variables to denote the regions of job location based on their economic development and geographic location in China. And *region1-3* denotes the dummy variables of working in the east coast developed region, central west developing region, central west middle-developed region, respectively and working at east coast middle-developed region of China as the base group. Table 1 shows that 60% of graduates choose to work in the east coast developed region, and only 31% work at central west middle-developed region, only 3% work at central west developing region and 6% work at east coast middle-developed region of China.

A two-sample *t* test is used to formally test whether there is statistically difference in characteristics between male and female samples and the corresponding results are given in the last two columns of Table 1. We found that the majority of *p*-values of the two sample *t* tests are less than the significance level of 5%. It is shown that there exists a significant difference in monthly earnings, work capacity, college academic performance, work experience and enterprise characteristics between female and male graduates. Female graduates have higher academic performance than male graduates on average. There is no significant difference in job location choice, especially in



**Table 1.** The descriptions of key variables of male and female university graduates.

Variable name	Descriptive statistics						<i>t</i> -test	
	Total mean (proportion)	SD	male mean (proportion)	SD	female mean (proportion)	SD	t-statistic	<i>p</i> -value
<b>Wage</b>								
<i>Inwage</i>	7.7769	0.4806	7.8193	0.4863	7.6889	0.4567	5.1468	0
<b>Work capacity</b>								
<i>indexac</i>	0.4088	0.1685	0.4215	0.1692	0.3826	0.1641	4.3184	0
<b>Academic performance</b>								
<i>Academic1</i>	0.3508	0.4774	0.3700	0.4836	0.3068	0.4616	2.5612	0.0175
<i>Academic2</i>	0.3832	0.4863	0.3605	0.4804	0.4303	0.4956	-2.6178	0.0089
<i>Academic3</i>	0.1877	0.3906	0.1601	0.3669	0.2450	0.4305	-3.8036	0.0002
<b>Work experience</b>								
<i>gradtime</i>	8.8777	2.1007	9.0336	1.9607	8.5538	2.3343	3.9787	0
<i>jtime</i>	9.2220	3.1030	9.4046	2.9742	8.8426	3.3261	3.2168	0.0013
<b>Employer's characteristics</b>								
<i>size</i>	7.2486	2.5303	7.5743	2.5099	6.5719	2.4393	7.4936	0
<i>major</i>	0.7761	0.4170	0.8082	0.3939	0.7091	0.4546	4.1854	0
<b>Job location</b>								
<i>Subpro</i>	0.4129	0.4925	0.4161	0.4931	0.4064	0.4916	0.3641	0.7159
<i>Munici</i>	0.2265	0.4162	0.2042	0.4033	0.2610	0.4396	-2.4395	0.0148
<i>region1</i>	0.6019	0.4897	0.5983	0.4905	0.6096	0.4883	-0.4249	0.6710
<i>region2</i>	0.0298	0.1700	0.0364	0.1875	0.0159	0.1254	2.5425	0.0111
<i>region3</i>	0.3113	0.4632	0.3068	0.4614	0.3207	0.4672	-0.5503	0.5822

Notes: Data are from Chinese university 2007 survey of fresh graduates' employment and capacity conducted by an education research company MyCOS institute. *Inwage* denotes the logarithm of monthly earnings and *indexac* denotes capacity index. And *Academic1-3* denote the dummy variables describing the middle, good, excellent levels of college academic performance, respectively and the base group is those with worst academic performance. Here *gradtime* means the time span from graduation to the survey time in months and *jtime* means the duration of holding current job position in months. *Subpro* and *Munici* are dummy variables denoting that the graduate is working at a sub-provincial city or municipality, respectively and the base group includes those who work at other cities. *Size* denotes the logarithm of the number of workers at the enterprise and *major* is a dummy variable that takes value 1 if one's job is related to their majors and 0 otherwise. And *region1-3* denotes the dummy variables of working at the east coast developed region, central west developing region, central west middle-developed region, respectively and working at east coast middle-developed region of China is the base group. If the *p*-value of two sample *t* test is zero, it means that the decimals are all zeros when we keep four decimals and the *p*-value is very small and close to zero.

**Table 2.** The coefficients estimates from quantile regression with quantile  $\tau = 0.5$ .

Variable name	Male sample	Female sample	Full sample
Intercept	6.9454 (6.6988, 7.1832)	6.5779 (6.3270, 7.0715)	6.7632 (6.5916, 7.9832)
Gender			0.1004 (0.0528, 0.1488)
<i>indexac</i>	0.2532 (0.1033, 0.4421)	0.3091 (0.0732, 0.5150)	0.2528 (0.1097, 0.3675)
<i>Academic1</i>	-0.0716 (-0.1657, 0.0175)	0.0358 (-0.4107, 0.2612)	-0.0486 (-0.1217, 0.0514)
<i>Academic2</i>	-0.0214 (-0.1076, 0.0697)	0.0347 (-0.3969, 0.2487)	-0.0220 (-0.0988, 0.0815)
<i>Academic3</i>	0.0272 (-0.0900, 0.1386)	0.0832 (-0.3484, 0.2977)	0.0430 (-0.0412, 0.1611)
<i>gradtime</i>	0.0088 (-0.0103, 0.0257)	0.0159 (-0.0423, 0.0020)	-0.0023 (-0.0176, 0.0100)
<i>jtime</i>	0.0136 (0.0027, 0.0236)	0.0328 (0.0200, 0.0471)	0.0236 (0.0159, 0.0316)
<i>size</i>	0.0202 (0.0074, 0.0299)	0.0478 (0.0349, 0.0636)	0.0300 (0.0186, 0.0384)
<i>majorm</i>	0.0819 (0.0099, 0.1611)	-0.0008 (-0.0823, 0.0863)	0.0266 (-0.0305, 0.0.0883)
<i>Subpro</i>	0.1456 (0.0750, 0.2180)	0.2243 (0.1389, 0.3074)	0.1962 (0.1407, 0.2561)
<i>Munici</i>	0.1915 (0.0884, 0.2723)	0.2374 (0.1331, 0.3517)	0.2302 (0.1636, 0.3043)
<i>region1</i>	0.4056 (0.3124, 0.5472)	0.4492 (0.2538, 0.5859)	0.4306 (0.3369, 0.5126)
<i>region2</i>	0.1740 (0.0008, 0.3947)	0.1987 (-0.0803, 0.4398)	0.1985 (0.0098, 0.3315)
<i>region3</i>	0.0478 (-0.0497, 0.1998)	0.1118 (-0.0807, 0.2593)	0.0700 (-0.0366, 0.1481)

Notes: The variables are described the same as in Table 1. The 95% bootstrap confidence intervals for the coefficient estimates are included in parentheses.

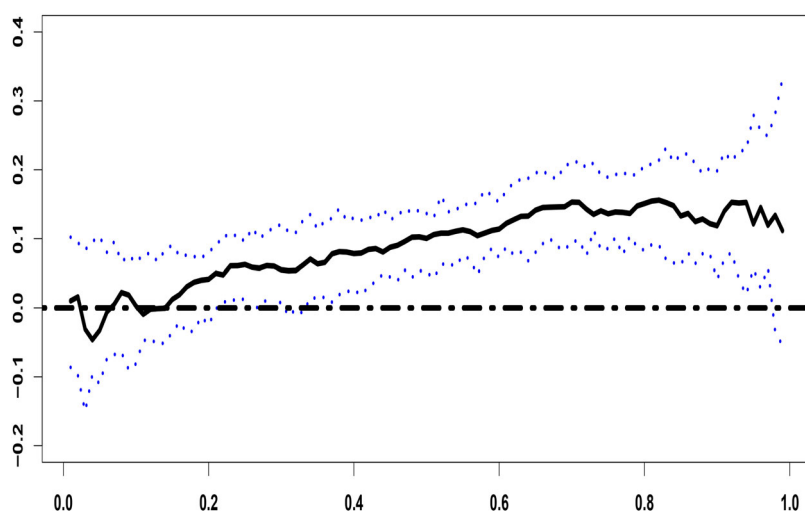
choosing to work at a sub-provincial city and east coast developed region, between female and male graduates. There is a higher percentage of female graduate working at municipalities and central west middle-developed regions than that of male graduates. The average graduation time is 8.8 months and average time for working experience is 9.2 months. It seems that some of the graduates have started working before graduation.

## 4. Empirical result

In this section, we first show the results of the quantile regression, and provide evidences for male–female pay gap. Then we decompose the gender log wage gap and find the contribution of composition effect.

### 4.1. Quantile regression analysis

We estimate our models separately for three sub-samples of data: first a full sample comprising both female and male university graduates, and then two disaggregated sub-samples, female and male graduates, respectively. Table 2 reports the estimation results of model (1) with  $\tau = 0.5$ . The 95% bootstrap confidence intervals for the

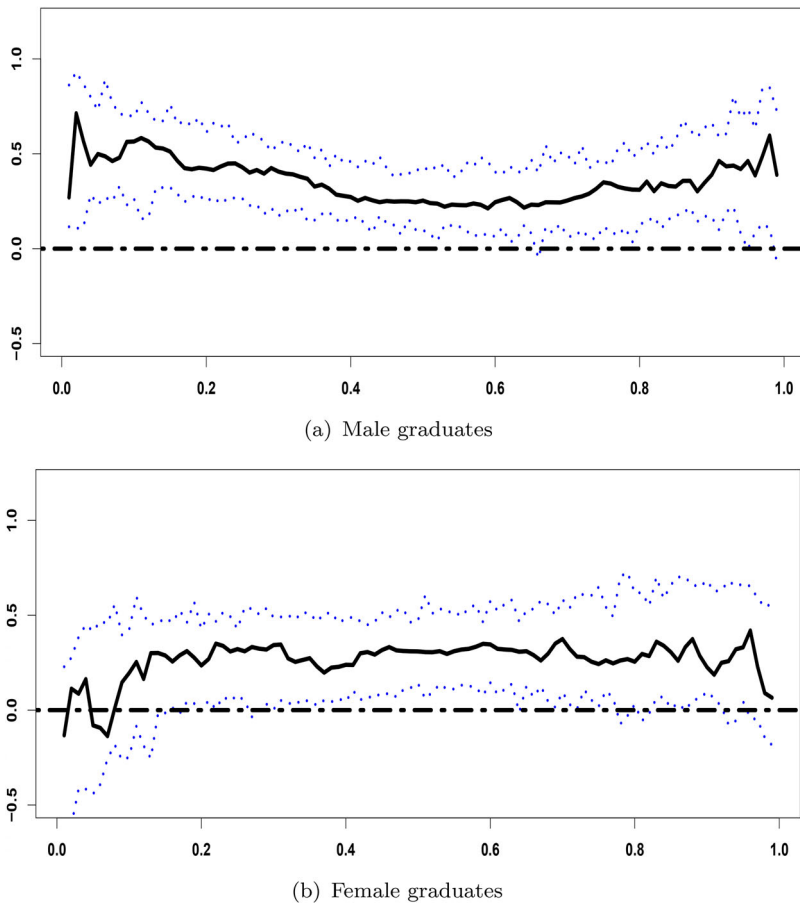


**Figure 3.** Coefficient estimates of Gender at different quantiles.

coefficient estimates are included in parentheses. If the confidence interval does not include zero, it indicates that the variable is significant at 5% significance level.

The coefficient of *Gender* is significantly positive in quantile regression based on full sample, which shows there exists gender pay gap and male has higher pay than the counterpart female. Figure 3 further presents the estimates of the coefficient of *Gender* on full sample for  $0.01 \leq \tau \leq 0.99$ , shedding some light on the magnitude of the relationship between wage and gender over the log wage distribution. The area between the two dot lines is the 95% bootstrap confidence band for the corresponding coefficient estimates. If the zero line is outside the coefficient interval, it indicates that the relationship between wage and gender is significant at 5% significance level. The association is insignificant at the lower tail of the log wage distribution. The empirical evidence reveals that wage and gender are significantly and positively related above the 20% quantile of the log wage distribution, which indicates that females earns significantly less than males. The positive relationship between wage and gender increases over almost all the log wage distribution except the top 10% quantiles.

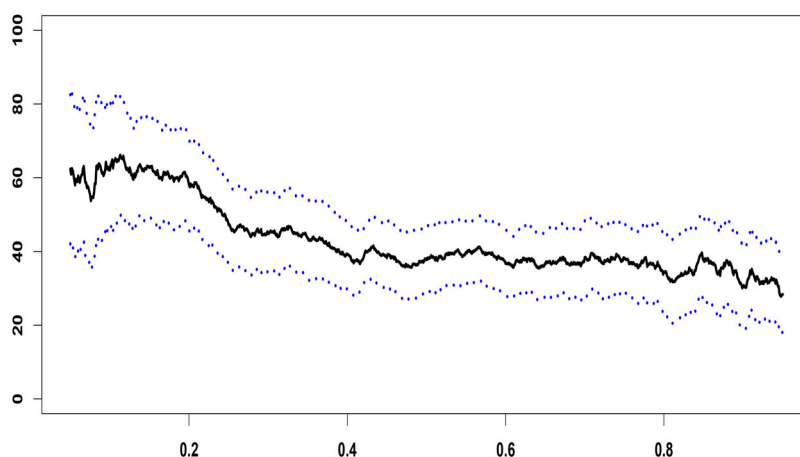
Other things being equal, graduates with higher *indexac* were more likely to have higher earnings. There is some evidence that returns to work capacity are higher for women from the median regression results shown in Table 2. In particular, for male and female graduates, respectively, the coefficients of *indexac* are significantly positive at different quantiles. Figure 4 further presents the estimates of the coefficient of *indexac* for male and female graduates for  $0.01 \leq \tau \leq 0.99$ , shedding some light on the magnitude of the relationship between wage and work capacity index over the log wage distribution. The area between the two dotted lines is the 95% bootstrap confidence band for the corresponding coefficient estimates. If the zero line is outside confidence interval, it indicates that the relationship between wage and work capacity index is significant at the 5% significance level. It shows that wage and work capacity index are significantly and positively related throughout the log wage distribution for



**Figure 4.** Coefficients estimates of *indexac* of quantile regression based on male and female graduates samples, respectively.

male graduates. However, the association is insignificant at the top of the log wage distribution for the female sample. The positive relationship between wage and work capacity index decreases and then increases over the log wage distribution in male sample, while the association seems to keep constant except at the bottom and top of the distribution for female sample.

The relationship between human capital accumulation within college (reflected in college academic achievement) and job-market outcomes in China is an open question. Li and Zhang (2010) has shown that female's advantage in college academic performance helped to produce a surprising gender employment gap favouring female graduates. However, in our analysis, three dummy variables describing college academic performance are insignificant. As we can see, the graduates with more experience have higher earnings for both the male and female samples. And whether job is related to their subject major or not is a significant factor affecting the wage of male graduates, but insignificant in the female graduate median wage determining function. The greater the enterprise size, the higher wage which female and male graduates can obtain, but the return to enterprise size is higher to male graduates. There is



**Figure 5.** Estimated log wage of males, females and the counterfactual log wage of females at different quantiles.

significant difference in wages between the east coast developed region and east coast middle-developed region for both female and male graduates. There exists significant differences in wages between the central west developing region of China and the east coast middle-developed region for male graduates, while the result does not hold for female graduates.

#### 4.2. Counterfactual decompositions

Following the procedure in Section 2, we then construct the counterfactual wage of female graduates if they had the same distributions of characteristics as male graduates. The difference between the counterfactual wage and the actual wage of females is then purely due to the differences in composition characteristics between males and females.

Figure 5 plots the approximated log wage of both male and female graduates and the counterfactual log wage of females at different quantiles. The log wage of male graduates is above that of female graduates at all quantiles. The counterfactual log wage of females is on top of the log wage of female graduates, and below the log wage of male graduates over all of the log wage distribution, i.e. if the female graduates had the same endowment as male graduates, their wage would definitely increase but would still be lower than the wage of male graduates in general.

Table 3 reports the detailed numerical results. Columns 2 and 3 give the observed and estimated log wage at the nine deciles for male graduates, and columns 4 and 5 report the corresponding log wage for female graduates. Again, the estimated log wage using our method is very close to the actual observed one, which proves the validity of our method. Column 6 is the counterfactual log wage of female graduates, and column 7 is the estimated log wage gap between male and female graduates. The last two columns report the estimated composition effect as well as their contributions to the gender log wage gap. The contribution of composition effect ranges about 30–60% over the log wage distribution. As we can see, the composition effect explains

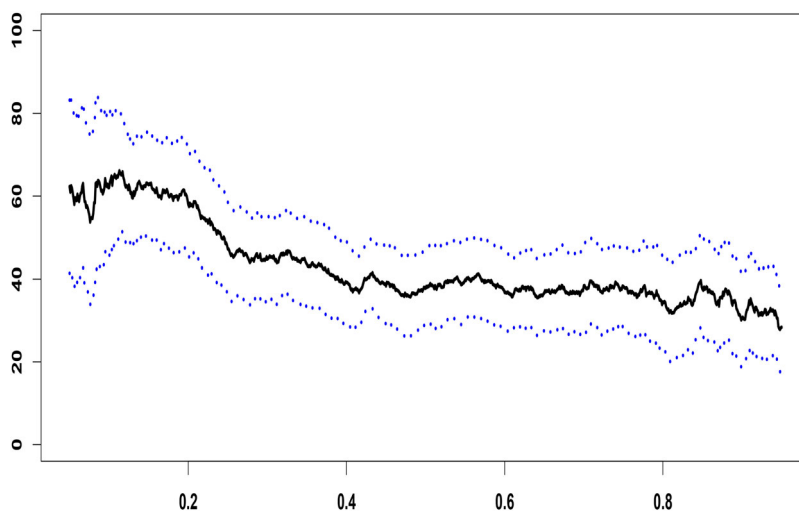
**Table 3.** Decomposition results.

Quantile	log wage					log wage Gap Decomposition		
	log wage of male		log wage of female		Counter-factual log wage (c)	Estimated log wage gap (d = a - b)	Composition effect	
	observed	Estimated (a)	observed	Estimated (b)			Gap (e = c - b)	Contribution (e/d)
0.1	7.2442	7.2075	7.0901	7.0913	7.1634	0.1162	0.0721	60.07%
0.2	7.3778	7.4195	7.3132	7.3044	7.3712	0.1152	0.0668	58.00%
0.3	7.6009	7.5785	7.4955	7.4550	7.5101	0.1235	0.0551	44.61%
0.4	7.6009	7.7142	7.6009	7.5798	7.6325	0.1345	0.0528	39.24%
0.5	7.8240	7.8360	7.6009	7.6914	7.7461	0.1446	0.0547	37.80%
0.6	8.0064	7.9504	7.8240	7.8110	7.8623	0.1393	0.0512	36.78%
0.7	8.0064	8.0633	8.0064	7.9410	7.9879	0.1224	0.0470	38.38%
0.8	8.2940	8.2129	8.0064	8.0860	8.1297	0.1270	0.0438	34.47%
0.9	8.4118	8.4261	8.2941	8.2827	8.3264	0.1434	0.0437	30.46%

Notes: Data are from Chinese university 2007 survey of fresh graduates employment and capacity conducted by an education research company MyCOS institute. The counterfactual log wage of female graduates is constructed if they had the same distributions of characteristics as male graduates.

45–60% of the gender pay gap below 30% quantile of the log wage distribution. Right after graduation, male and female graduates have almost identical labour income and most of the gap is attributed to the difference in the endowment. Our results indicate there is a higher explained portion from endowment in gender pay gap than previous studies since we use higher educated sample and entry wage. Napari (2008) found that only about 20–26% of the average early career gender wage gap in the Finnish private sector is explained by gender differences in experience, the field of education, employer characteristics using the Blinder-Oaxaca decomposition method.

Bishop, Luo, and Wang (2005) used data from the Chinese Household Income Projects for the years 1988 and 1995, a standard earnings equation, and quantile regressions to estimate and decompose the earnings gap, and then find that the unexplained share is 78.4, 67.2, 57.4, 55.1 and 59.9% in 1995, and 97.8, 57.2, 52.2, 58.6 and 60.3% in 1988, respectively, at the 10%, 25%, 50%, 75% and 90% quantiles. These numbers indicate that the unexplained or ‘discrimination’ portion is largest in the lower tail of the distribution, suggesting sticky floors. A similar pattern was documented in Zhang, Han, Liu, and Zhao (2008) based on Chinese Urban Household Survey Data. Fang and Sakellariou (2011) also found that there is a case of sticky floors in Thailand. However, Piazzalunga (2018) showed that the unexplained gap increases along the wage distribution, indicating a glass ceiling effect using Italian college graduates. Christofides, Polycarpou, and Vrachimis (2013) also found evidence of glass ceilings, suggesting more female disadvantage in better job. Figure 6 presents the confidence bound for the contribution of composition effect and the pattern of a glass ceiling effect is found in our analysis. Moreover, our decomposition results indicate that the contribution of difference in rewards is quantitatively less important than that of different covariates at the bottom of log wage distribution. More precisely, the characteristics effect accounts for 60% of the gender pay gap at the 10th percentile and for 58% at the 20th percentile. Figure 6 shows that the confidence band does not include zero over the log wage distribution, which reveals that the estimated composition effect on gender pay gap is always significantly different from zero, accounts about 30–60% of the gap on average. It means that the gender wage



**Figure 6.** The contribution percentage (%) of composition effect to the gender wage gap at different quantiles.

gap is mostly due to the difference in the endowment of human capital, which can result from gender difference in risk aversion as Guo et al. (2016) stated. Especially at the bottom quantiles, about 60% of the wage gap is almost due to the characteristics effect. The quality of female graduates' characteristics matters in wage determination and gender wage gap since higher education expansion in China.

## 5. Robustness check

One might wish to compare the wage of male graduates with the counterfactual wage of male graduates if they had the same characteristics as female graduates. One must ensure, therefore, that, qualitatively, the conclusions drawn with one decomposition are resilient to changes in the order of that decomposition. Following the procedure in Section 2, we then construct the counterfactual log wage of male graduates if they had the same endowment as female graduates. The difference between the actual log wage of male and the counterfactual log wage is then purely due to the differences in the endowment, and it is also the composition effect. Table 4 reports the detailed numerical results. The first five columns are the same as those in Table 3. Column (f) is the counterfactual log wage. The last two columns report the estimated composition effect as well as their contributions to the gender pay gaps. The composition effect explains at most 60% of the gender pay gap, which is attained at the 10% quantile of log wage distribution. It also indicates that the wage structure effect is largest at the top tail of the distribution. And the contribution of difference in rewards is quantitatively less important than that of different covariates at the bottom of log wage distribution. It means that female graduates have almost the same rewards to characteristics as their male counterparts and there is no strong gender discrimination at the bottom of log wage distribution.

One major concern in the gender pay gap literature is self-selection in that the probability of participating in labour force or being employed may depend on

**Table 4.** Decomposition result.

Quantile	log wage					log wage Gap Decomposition		
	log wage of male		log wage of female		Counter-factual log wage ( <i>f</i> )	Estimated log wage gap ( <i>g</i> = <i>a</i> - <i>b</i> )	Composition effect	
	observed	Estimated ( <i>a</i> )	observed	Estimated ( <i>b</i> )			Gap ( <i>h</i> = <i>a</i> - <i>f</i> )	Contribution ( <i>h/g</i> )
0.1	7.2442	7.2075	7.0901	7.0913	7.1378	0.1162	0.0697	60.01%
0.2	7.3778	7.4195	7.3132	7.3044	7.3589	0.1152	0.0607	52.67%
0.3	7.6009	7.5785	7.4955	7.4550	7.5227	0.1235	0.0557	45.13%
0.4	7.6009	7.7142	7.6009	7.5798	7.6598	0.1345	0.0545	40.52%
0.5	7.8240	7.8360	7.6009	7.6914	7.7881	0.1446	0.0479	33.11%
0.6	8.0064	7.9504	7.8240	7.8110	7.9154	0.1393	0.0351	25.20%
0.7	8.0064	8.0633	8.0064	7.9410	8.0387	0.1224	0.0246	20.14%
0.8	8.2940	8.2129	8.0064	8.0860	8.1910	0.1270	0.0219	17.27%
0.9	8.4118	8.4261	8.2941	8.2827	8.4116	0.1434	0.0145	10.09%

Data are from Chinese university 2007 survey of new graduate's employment and capacity conducted by an education research company MyCOS institute. The counterfactual log wage of male graduates is constructed if they had the same composition characteristics as female graduates.

unobservable that are different for male and female graduates. As Li and Zhang (2010) discussed that though there are no official national statistics available, regional statistics also suggest that the female graduates unemployment rate is not in fact higher than that of corresponding males in Shanghai (2005), Beijing (2000–2002), or Hainan (2005–2006). We therefore proceed without a selection correction procedure, there is no selectivity bias without considering the unemployment rate difference between male and female graduates.

## 6. Conclusion

A major transformation of Chinese higher education has taken place over the past decades. China has reshaped its higher education sector from elite to mass education with a rapid growth in the number of its graduates from less than 1 million a year in 2000 to some 6.3 million a year by 2010. This article aims to decompose the composition effect and wage structure effect of gender starting pay gap in Chinese university graduates at every quantile. It means that the pay gap at every quantile is a result of differences in gender characteristics, or the differences in returns to those characteristics after the decomposition. Using Chinese university 2007 survey of new graduates employment and capacity conducted by an education research company MyCOS institute, we exploit a counterfactual decomposition analysis using quantile regression to decompose the gender pay gap into one component that is based on differences in characteristics and one component that is based on differences in coefficients across the log wage distribution.

The counterfactual decomposition analysis has displayed a significant gender pay gap over almost all the wage distribution. Although there are significant gender differences in the determinants of wages from the quantile regressions, we found that 30–60% of the gender pay gap is attributed to the gender difference in the endowment of human capital. We also find evidence of glass ceilings, suggesting more female graduates have disadvantages in better jobs. The similar results are obtained in robustness check. The two sample *t* tests show that female graduates' characteristics are significantly different from their male counterparts, which strengthens the theory



that composition effect may matter in the gender pay gap. Saygin (2019) supported our findings that graduates' characteristics, especially university attainment affect students lifetime earnings. What's more, gender pay gap may depend on the assessment methods of university attainment.

This article examines the gender entry wage gap in Chinese university graduates after higher education reform so it can contribute to the literature on gender pay gap in higher education groups after higher education reform in a transition economy. Our findings that the gap is mostly due to the difference in characteristics indicate that female graduates have almost same rewards to characteristics as their male counterparts and the quality of female graduates' characteristics matters in gender pay gap since higher education expansion in China. It is also shown the higher education expansion and other labour market reform bring the gender equality in reward to characteristics. In this article, we find that higher work capacity helps to produce higher wage in the wage determination. Our findings will provide policymakers with some implications on improving the quality of graduates' characteristics such as work capacity to reduce the gender pay gap.

There are several weaknesses of the study, which provide directions in which further research might develop. Since the hourly wage is not available, we use the monthly wage, which causes a little overestimation of the gender wage gap. In addition, we only used a cross-sectional data from 2007 due to the unavailability of the pooled cross-sectional data and more recent data. However, the labour market performance of university graduates in 2007 is representative of those affected by the policy of Chinese higher education expansion starting in 2000 and the latest data may not be suitable because the more recent data will be affected by other policies, such as the policy stimulating 2008 financial crisis. The further research can be conducted based on hourly wage using pooled cross-sectional data with consideration of sample selection correction.

## Notes

1. MyCOS Data was established in 2006 and it uses complex and sophisticated data collection and mining technologies to develop graduate employment and education quality evaluation databases and consulting services for colleges and universities, governmental education bureaus, and research institutions. The resulting survey data form the foundation for China's first nationwide college graduate employment database. It has published Chinese College Graduates' Employment Annual Reports since 2007.
2. By April 2012, China had a total of 2,138 higher education institutions.
3. Project 211 is a project of National Key Universities and colleges initiated in 1995 by the Ministry of Education of China, with the intent of raising the research standards of high-level universities and cultivating strategies for socio-economic development. The name for the project comes from an abbreviation of the twenty-first century and 100 (approximate number of participating universities). Today China has more than 118 higher education institutions (about 6%) designated as Project 211 institutions that have met certain scientific, technical, and human resources standards and offered advanced degree programs.
4. Chi and Li, (2014) found that the gender pay gap was larger when annual earnings were used in the calculation, suggesting that the gender earnings gap is to some extent caused by difference in the number of working hours. Research has also demonstrated that women tend to invest less time in work because they shoulder a larger share of housework

at home than men. This is true in both developing and developed countries (Cao & Chai, 2007). So the gender pay gap may be overestimated by using monthly earning. However, the difference in the number of working hours between male and female graduates should be pretty small because nearly none of them are married at the age and female does not need to invest more time in housework.

5. The 35 dimensions of work ability include reading comprehension, active listening, writing, effective oral communication, math's solutions, scientific analysis, critical thinking, active learning, learning method, performance monitoring, understanding others, coordination arrangements, persuading others, negotiation skills, instructing others, service to others, solving complex problems, new product ideas, technical design, equipment selection, installed capacity, computer programming, quality control analysis, operation monitoring, operation and control, equipment maintenance, troubleshooting, repair of machinery and systems, systems analysis, system evaluation, judgement and decision-making, time management, financial management, materials management and human resource management.
6. A sub-provincial city in China is governed by a province, but is administered independently in regard to economy and law. There are currently 15 sub-provincial cities in China including Changchun, Chengdu, Dalian, Guangzhou, Hangzhou, Harbin, Jinan, Nanjing, Ningbo, Qingdao, Shenyang, Shenzhen, Wuhan, Xiamen and Xi'an.
7. A municipality is the highest level classification for cities used by China. These cities have the same rank as provinces, and form part of the first tier of administrative divisions of China. Current Municipalities in China include Beijing, Tianjin, Shanghai and Chongqing.

## Acknowledgements

We sincerely thank the anonymous reviewers for their thoughtful comments, but we take full responsibility for all errors in this article.

## Disclosure statement

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

This research was supported by Zhejiang Federation of Humanities and Social Sciences Circles [grant number 16NDJC154YB], National Natural Science Foundation of China [grant number 71573092] and First Class Discipline of Zhejiang-A (Zhejiang University of Finance and Economics-Statistics).

## References

- Ahamed, M. M., Wen, J., & Gupta, N. (2019). Does board composition affect the gender pay gap? *Economics Letters*, 184, 108624.
- Babcock, L., & Laschever, S. (2003). *Women don't ask: Negotiation and the gender divide*. Princeton, NJ: Princeton University Press.
- Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3), 228–255.
- Bishop, J. A., Luo, F., & Wang, F. (2005). Economic transition, gender bias, & the distribution of earnings in china. *The Economics of Transition*, 13(2), 239–259.

- Bredtmann, J., & Otten, S. (2014). Getting what (employers think) you're worth: Evidence on the gender gap in entry wages among university graduates. *International Journal of Manpower*, 35(3), 291–305.
- Cao, X., & Chai, Y. (2007). Gender role-based differences in time allocation: Case study of Shenzhen, China. *Transportation Research Record: Journal of the Transportation Research Board*, 1, 58–66.
- Chi, W., & Li, B. (2008). Glass ceiling or sticky floor? Examining the gender earnings differential across the earnings distribution in urban China. *Journal of Comparative Economics*, 36(2), 243–263.
- Chi, W., & Li, B. (2014). Trends in China's gender employment & pay gap: Estimating gender pay gaps with employment selection. *Journal of Comparative Economics*, 42(3), 708–725.
- Christofides, L., Polycarpou, A., & Vrachimis, K. (2013). Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe. *Labour Economics*, 21, 86–102.
- Fang, Z., & Sakellariou, C. (2011). A case of sticky floors: gender wage differentials in Thailand. *Asian Economic Journal*, 25(1), 35–54.
- Firpo, S., Fortin, N. M., & Lemieux, T. (2007). *Decomposing wage distributions using recentered influence functions regressions*. In Mimeo. Vancouver: University of British Columbia.
- Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77, 953–973.
- Guo, Q., Song, Y., Sun, W., & Wang, Y. (2016). Experience, tenure & gender wage difference: Evidence from China. *China Economic Review*, 41, 104–113.
- Gustafsson, B., & Li, S. (2000). Economic transformation & the gender earnings gap in urban China. *Journal of Population Economics*, 13(2), 305–329.
- Hughes, J., & Maurer-Fazio, M. (2002). Effects of marriage, education, & occupation on the female/male wage gap in China. *Pacific Economic Review*, 7, 137–156.
- Ichino, A., & Moretti, E. (2009). Biological gender differences, absenteeism, and the earnings gap. *American Economic Journal: Applied Economics*, 1, 183–218.
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
- Liu, W., Meng, X., & Zhang, J. (2000). Sectoral gender wage differential & discrimination in the transitional Chinese economy. *Journal of Population Economics*, 13, 331–352.
- Li, T., & Zhang, J. (2010). What determines employment opportunity for college graduates in China after higher education reform? *China Economic Review*, 21(1), 38–50.
- Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4), 445–465.
- Manning, A., & Swaffield, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, 118, 983–1024.
- Meng, X. (1998). Male-female wage determination & gender wage discrimination in China's rural industrial sector. *Labour Economics*, 29, 67–89.
- Napari, S. (2008). The early-career gender wage gap among university graduates in the Finnish private sector. *Labour*, 4, 697–733.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 3, 693–709.
- Piazzalunga, D. (2018). The gender wage gap among college graduates in Italy. *Italian Economic Journal*, 1, 33–90.
- Qu, D., Guo, S., & Wang, L. (2019). Experience, tenure & gender wage difference: evidence from China. *Economic Research-Ekonomska Istraživanja*, 32, 1169–1184.
- Saygin, P. (2019). Gender bias in standardized tests: evidence from a centralized college admissions system. *Empirical Economics*. Advance online publication.
- Wang, M., & Cai, F. (2008). Gender earnings differential in urban China. *Review of Development Economics*, 12(2), 442–454.
- Xiu, L., & Gunderson, M. (2014). Glass ceiling or sticky floor? Quantile regression decomposition of the gender pay gap in China. *International Journal of Manpower*, 35(3), 306–326.
- Zhang, J., Han, J., Liu, P., & Zhao, Y. (2008). Trends in the gender earnings differential in urban China. *Industrial & Labor Relations Review*, 61(2), 224–232.