The Gender Wage Differentials among Rural-Urban Migrants in China*

Rong Zhu†

September 1, 2010

Abstract

This paper analyzes the gender wage disparities among rural-urban migrants in urban China using a nationally representative data set. On average, female migrants earn only 66% of their male counterparts' average hourly wage. And the gender wage gap is not uniform across migrants' wage distribution with differentials much higher at the top end than at the bottom and the middle. The mean decomposition method by Fortin (2008) and the unconditional quantile regression by Firpo et al. (2009) are combined to decompose the distributional gender wage differentials into the contribution of each covariate. We find that the discrimination effect contributes more to the wage gap than the endowment effect throughout the wage distribution. Although the gender wage differential is the largest at the higher end of rural migrants' wage distribution, our decomposition results show that gender wage discrimination effect attributable to unequal returns to observable characteristics is most serious among low wage earners. How selected labor market characteristics contribute to the discrimination effect across the whole wage distribution is also investigated.

JEL classification: J3; J71

Keywords: Rural-urban migrants; gender wage gap; decomposition

1 Introduction

Rural-urban migration has been an increasingly important social phenomenon since China launched its economic reform in 1978. With the loosening of administrative controls over

*I would like to thank Elisabetta Magnani, Alan Woodland, Chris Carter, Dennis Doiron, Denzil Fiebig and Jianfei Shen for valuable comments on an earlier draft. I am also grateful to Thomas Lemieux for advice on the decomposition method. All remaining errors are my own.

†School of Economics, University of New South Wales, Sydney 2052 Australia. E-mail address: rongzhu0101@gmail.com.

population movements between rural and urban areas and the increasing disparity in rural-urban income gaps, the number of rural-urban migrants has increased significantly from an estimate of 34 millions in 1990 (Cai, 1996) to 80 millions in 1995 (Meng and Zhang, 2001), and it further rose to nearly 132 millions in 2006 (Demurger et al., 2009). According to China's census data in 2005, migrant workers account for over 20% of the labor force in the urban labor market. As the most important form of internal migration in China, rural-urban migration has exerted profound effects on the labor market and attracted much attention from both economists and policy-makers (Zhao, 2005).

The unique Household Registration System (*Hukou* System) implemented in China makes its rural-urban migration different from the internal migration in other developing economies.[1] The *Hukou* system, which was introduced in the 1950s, was a civil status registration system through which the government allocated housing and jobs, rationed food and necessities. At that time, the close relationship between place of residence and access to consumer goods and job opportunities made it almost impossible for people with rural *Hukou* to live in urban areas (Zhao, 2005; Poncet, 2006). Since in the late 1970s, the implementation of the Household Responsibility System in rural China has led to the abandonment of the rationing system and brought greater freedom to rural workers for choosing their occupations, which made rural-urban migration possible. In the meantime, the development of labor-intensive industries created more job opportunities in urban areas and this constituted an important 'pull' factor for rural-urban migration. Furthermore, the rising disparities in earnings between rural and urban areas 'pushed' rural workers to leave the countryside and seek better employment opportunities in cities (Poncet, 2006). Constraints on migration has been eased thereafter. Since the mid and late 1980s, rural to urban migration has become a constant social phenomenon.

Although many reforms has been implemented, the Household Registration System still imposes many constraints on rural migrants and makes the urban labor market segregated between urban residents and rural migrants (Meng and Zhang, 2001). In the urban labor market, only an individual holding an urban *Hukou* is eligible to take up certain types of jobs.[2] It is usually very difficult for rural migrants to enter formal sector (Meng ,2001). Without urban *Hukou*, they are usually not covered in urban social security system and they are not entitled to other social benefits such as health insurance. In addition, rural migrants are generally less educated and low skilled. Deliberate discrimination against migrants in cities remained legal until very recently and they are usually treated as second-class workers by employers and government departments (Gagnon et al., 2009; Demurger et al., 2009). Migrant workers usually work for longer hours than urban residents, but

[1]See Zhao (2005) for the origin and evolution of the *Hukou* system and a survey of related literature on rural-urban migration in China.

[2]Using a nationally representative data, Demurger et al. (2009) found that 52% of urban workers were professionals, technicians or official workers. However, the percentage was only 7% for migrant workers.

the earnings they receive are often far below those of urban workers. The situation is exacerbated by the fact that the prejudiced city policies make it very difficult for rural migrants to change their rural *Hukou* to urban *Hukou*, despite the substantial amount of time they have spent working in urban areas (Wang and Zhuo, 1999).

The different treatments to workers in the same labor market intrigue some scholars to investigate what contributes to the hourly wage differentials between urban residents and rural migrants. Using data sets from Shanghai, Meng and Zhang (2001) found significant differences in occupational attainments and wage differentials between urban residents and rural migrants. They also found that productivity-related differences between the two groups cannot explain most of the average wage differentials, which is likely to be attributable to discrimination. With the data from 2002 China Household Income Project, Demurger et al. (2009) found a much contrasted sectoral allocation between the two groups. Their decomposition analysis based on simulation showed that a large share of earnings difference is due to the substantial differences in population characteristics. Other studies have been devoted to the understanding of the determination of rural-urban migration decisions (see, for example, Hare, 1999). Special attention has been paid to the effects of factors like education (Zhao, 1997), earnings differences (Zhao, 1999; Zhu, 2002) and human capital externalties (Liu, 2008) on migration decisions. In addition, some other studies focus on investigating whether and how the influx of rural migrants into urban areas contributes to China's economic growth (Liang, 2001; Zhang and Song, 2003).

Given the abundance of literature studying the wellbeing of rural migrants in China, it is also of interest to know whether female migrants are treated equally to male migrants in the urban labor market. This paper contributes to existing literature by examining rural migrants' wage determination for both genders and the wage differentials between them. While there are some previous literature documenting the wage determination of this special group (see Meng, 1998; Meng, 2001; Hare, 2002; Lu and Song, 2006), they generally suffer from three major drawbacks. First, the data sets they used for analysis are usually surveyed at one or several cities in China, which are not nationally representative. Second, they mainly focus on how average wage is affected by labor market characteristics. However, it can also be of interest and significance to evaluate wage determination at different points of wage distribution. The rate of return to factors like education may not be identical at all earnings levels. Furthermore, none of previous literature has decomposed migrants' gender wage gap across wage distributions.[3] It is important to know whether and how labor market characteristics contribute to distributional wage differentials.

In this paper, using a nationally representative rural migrants data set from 2002 China

[3]The gender wage gap problem for urban residents has been very documented. Using decomposition methods, some studies investigate the average gender gap gap among urban residents (Gustafsson and Li, 2000; Hughes and ,aurer-Faazio, 2002; Wang and Cai, 2008) and the distributional gender wage differential among them (Bishop et al., 2005; Chi and Li, 2008; Zhang et al., 2008).

Household Income Project, we analyze the wage determination for both genders by using OLS estimation and the unconditional quantile regressions developed by Firpo et al. (2009). To decompose mean gender wage gap, we apply a recently-developed regression compatible procedure by Fortin (2008), which can solve two invariance problems in the widely-used mean decomposition method by Oaxaca (1973) and Blinder (1973). Methodologically, to decompose the wage differentials at different quantiles between both genders, we propose to extend the mean decomposition method by Fortin (2008) to quantile decomposition with the unconditional quantile regression by Firpo et al. (2009). It is much computationally simpler than other quantile decomposition methods available (see, for example, Machado and Mata, 2005 and Melly, 2005) and it can divide up both the endowment effect and the discrimination effect into the contribution of each explanatory variable.

On average, female migrants earn only 66% of their male counterparts' average hourly wage. We find that the gender wage gap is not even across migrants' wage distribution. In contrast to the "sticky floor" effect found among China's urban workers in previous literature (Chi and Li, 2008), we find that gender wage differentials are the highest at the top end of rural-urban migrants' wage distribution. By using OLS and unconditional quantile regressions, we find that the returns to observable characteristics differ by gender and these differences change as we move up the wage distributions. Our decomposition results show that the discrimination effect attributable to unequal returns to labor market characteristics contributes more to the wage gap than the endowment effect across the wage distribution. Although the wage differential is the largest at the higher end of wage distributions, the decomposition results show that the gender wage discrimination problem is most serious among low wage migrants. We also investigate how selected observable characteristics contribute to the discrimination effect against female migrants throughout the wage distribution. We find a changing pattern that cannot be revealed by mean decomposition.

The remainder of this paper is organized as follows. Section 2 is a description of the data set and raw gender wage differentials among rural migrants. In section 3, we present the mean and quantile decomposition method. Section 4 presents the regression and decomposition results. Section 5 concludes.

2 Data and Descriptive Statistics

2.1 Data and Variables

The data used for this study is the rural-urban migrants survey from the 2002 China Household Income Project (CHIP 2002).[4] The survey was conducted by economists from

[4]The data set was obtained from the Inter-University Consortium for Political and Social Research (ICPSR). Website: http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/21741.

the Chinese Academy of Social Sciences in 2003. To ensure its representativeness, the sample was drawn by the National Bureau of Statistics and the survey was conducted in 10 provinces and 2 province-level municipalities in China.[5] Rural-urban migrants are defined to be those living in urban areas but with rural *Hukou* status. However, the migrant workers who live on construction sites or in factory dormitories were not covered in the survey. Only those living in residential neighborhoods and communities in cities were surveyed (Khan and Riskin, 2005; Demurger et al., 2009). As discussed in Demurger et al. (2009), the sample we use is considered to be nationally representative of those long-terms migrants who can settle in China's urban areas.[6]

The survey covers 5,327 individuals. We restrict our attention to the migrants aged between 18 and 60 with positive earnings in 2002, then the sample is reduced to 3,319 observations.[7] Following the standard approach in the literature, any observation with missing information on variables for this analysis is dropped, and this result in our final sample consisting of 3,285 rural migrants, of which 1,865 are males and 1,420 are females.

Table 1: Summary Statistics by Gender

	Male		Female		Normalized
	Mean	SD	Mean	SD	Difference
Monthly income	903.3	1144	635.4	462.5	0.217
Hourly wage	3.720	5.481	2.454	1.982	0.217
Log(Hourly wage)	0.994	0.715	0.692	0.621	0.319
Age	35.27	8.406	33.46	7.877	0.157
Education	8.295	2.624	7.469	2.873	0.212
Experience	7.672	5.361	6.219	4.389	0.210
Tenure	5.412	4.510	4.728	3.790	0.116
Married	0.904	0.295	0.897	0.304	0.016
Minority	0.091	0.287	0.077	0.267	0.034
Permwork	0.060	0.237	0.044	0.206	0.048
Interprov	0.337	0.473	0.313	0.464	0.036
Observations	1,865		1,420		3,285

Note: Normalized difference is calculated as $\frac{\overline{X}_m - \overline{X}_f}{\sqrt{S_m^2 + S_f^2}}$, where S_m^2 and S_f^2 are the within-group variances of X.

The descriptive statistics of the continuous and binary variables for this study are reported separately by gender in Table 1. The monthly income is the sum of monthly earnings in all forms received from the current job in 2002 (including regular wages, bonuses, sub-

[5]Among the 10 provinces and 2 municipalities, Beijing, Guangdong, Jiangsu and Liaoning are from China's coastal areas; Anhui, Henan, Hubei and Shanxi come from the central part; Chongqing, Gansu, Sichuan and Yunnan are from west China.

[6]Although investigating wage distributions of those short-term migrant workers are of interest itself, we cannot do such analysis due to the unavailability of suitable data set. For the sample scheme and other aspects of the rural-urban migrant data, we refer to Demurger et al. (2009).

[7]By restricting age between 18 and 60, we lose 1415 observations. We delete 530 observations for missing information on earnings. We also drop 63 observations who reported to have no income in 2002.

sidies and all other types of income). The average monthly income of male migrants is 903.3 *yuan*, around 42% higher than the 635.4 *yuan* earned by women. The hourly wage is calculated based on the average monthly income, the number of working days per week and the number of working hours per day. Female migrants' average hourly wage is only 66% of the wage of male migrants. This ratio is smaller than any other gender earnings ratios in urban China documented in existing literature (see, for example, Gustafsson and Li, 2000; Bishop et al., 2005; Chi and Li, 2008; Zhang et al, 2008).

Among the rural-migrants, males are about 2 years older and have about 0.8 more years of schooling than females. On average, a migrant is undereducated without finishing junior high school, which generally takes 9 years in China.[8] The variable *Experience* measures the number of years living and working in urban areas.[9] *Tenure* measures the years on the current job. The mean of *Experience* and *Tenure* are both greater for male migrants than their female counterparts. *Married* is a binary variable equal to 1 if the respondent is currently married and 0 otherwise. Negligible differences in marriage rates are found between the two genders. *Minority* is a dummy variable showing whether the respondent belongs to China's ethnic minority groups or not. Ethnic minorities make up 9.1% of male migrants and 7.7% of female migrants. *Permwork* is a binary variable equal to 1 if the respondent is a long-term contract worker or a permanent worker. The proportion of male migrants having permanent jobs is slightly higher than that of female migrants. *Interprov* is a binary variable that equals to 1 if it is an interprovincial migration and 0 otherwise. 33.7% of males and 31.3% of females have migrated to and worked in the urban area in another province.

Additional summary statistics on the gender differences in distributions of occupations, employer ownerships, industries and provinces are displayed in Table 2. Over half of migrant workers reported to be self-employed or owners of private businesses and the differences between both genders are very small. Such a high proportion may be the result of occupational segregation in urban labor market (Meng, 1998 ; Meng and Zhang, 2001) and the discriminatory *Hukou* system (Liu, 2005). With lower education levels and rural *Hukou* status, compared with their urban counterparts, they are much less likely to work in formal sectors (Meng, 2001). Many of the migrants have chosen to start their own small businesses and be self-employed. Only 5.4% of male migrants and 2.1% of females hold professional or technical positions, which are far below the 32.7% reported for urban

[8]From the data we only know whether one migrant received any training before 1999. Training received between 1999 and 2002 was not reported so we cannot find a variable indicating training experience before 2002. In the existing literature, the effects of training on migrants' earnings are not clear-cut. Meng (1998) found that one more year of training can increase the wage level by 21% for rural-urban migrants working in Jinan city of Shangdong province, while Meng and Zhang (2001) found that job training is not an important factor determining rural migrants' earnings in Shanghai.

[9]Here we use a city experience measure because, for rural migrants, factors such as agricultural experience are believed to be less relevant to their earnings in urban areas (Meng and Zhang, 2001).

Table 2: Labor Market Characteristics by Gender

	Male		Female		Normalized
	Mean	SD	Mean	SD	Difference
Occupation:					
Professional, technician	0.054	0.227	0.021	0.143	0.123
Clerical staff	0.027	0.162	0.019	0.137	0.037
Service worker	0.169	0.375	0.230	0.421	-0.108
Self−employed or business owner	0.525	0.500	0.542	0.498	-0.025
Manufacturing worker	0.044	0.205	0.028	0.137	0.060
Commercial worker	0.059	0.236	0.092	0.288	-0.087
Construction worker	0.050	0.218	0.006	0.079	0.188
Other occupations	0.072	0.258	0.061	0.239	0.030
Employer Ownership:					
State−owned	0.076	0.264	0.054	0.227	0.061
Collective	0.033	0.179	0.042	0.201	-0.033
Private or self−employed	0.669	0.471	0.694	0.460	-0.039
Foreign or joint venture	0.006	0.080	0.005	0.070	0.014
Others	0.216	0.412	0.204	0.403	0.021
Industry:					
Manufacturing	0.101	0.301	0.091	0.287	0.024
Construction	0.072	0.258	0.014	0.118	0.203
Transportation, communication	0.042	0.201	0.007	0.084	0.162
Wholesale, retail, food services	0.427	0.495	0.537	0.499	-0.157
Finance, insurance, real estate	0.012	0.108	0.008	0.088	0.029
Social service	0.199	0.399	0.233	0.423	-0.059
Education, culture	0.014	0.119	0.013	0.115	0.007
Health, sports, social welfare	0.010	0.098	0.012	0.109	-0.016
Other industries	0.123	0.329	0.085	0.278	0.090
Province:					
Anhui	0.106	0.307	0.099	0.299	0.015
Beijing	0.055	0.227	0.059	0.236	-0.014
Chongqing	0.053	0.223	0.071	0.257	-0.055
Gansu	0.077	0.267	0.056	0.231	0.059
Guangdong	0.102	0.303	0.109	0.312	-0.015
Henan	0.099	0.299	0.098	0.297	0.003
Hubei	0.084	0.278	0.088	0.283	-0.010
Jiangsu	0.099	0.299	0.103	0.304	-0.008
Liaoning	0.101	0.301	0.113	0.316	-0.027
Shanxi	0.066	0.249	0.044	0.204	0.071
Sichuan	0.083	0.276	0.088	0.283	-0.012
Yunnan	0.075	0.263	0.072	0.258	0.007
Observations	1,865		1,420		3,285

Note: Normalized difference is calculated as $\frac{\overline{X}_m - \overline{X}_f}{\sqrt{S_m^2 + S_f^2}}$, where S_m^2 and S_f^2 are the within-group variances of X.

residents in Demurger et al. (2009).[10] 5% of male migrants are construction workers, and for females the proportion is lower, only 0.6%. This is consistent with our common belief that females are less likely to take jobs requiring much physical strength. Different from Meng (1998), the gender occupational segregation problem is not serious among the rural-urban migrants in our sample.

As for enterprize ownerships, the largest employers of migrants are private firms and self-employed businesses, hiring 67% of male migrants and 69% of females. In addition, around 10% of rural migrants are employed by state-owned enterprizes and collective firms. Rural migrants are least often hired by foreign firms and joint ventures. In terms of industry distribution, the commerce sector (including wholesale, retail, and food services) and social service sector are the two major employers of migrants. 42.7% of male migrants and 53.7% of females are working in the former industry, and the two percentages for the latter sector are 19.9% and 23.3%. While female migrants are more likely to be employed in the two sectors, they are less likely to work in almost all other industries. Finally, the gender differences in the distribution of regions are very small, as can be seen from the bottom part of Table 2.

2.2 Descriptive Gender Wage Differentials

To look at more detailed gender wage gap among rural migrants, we present preliminary statistics on hourly wage for each gender by occupation, ownership, industry and province in Table 3. We find that male migrants earn higher average wages in almost every category.[11] For occupation distribution, the gender wage gap is the smallest for commercial workers with females' average wage being 79% of that for males. The gender wage differential is the largest for manufacturing workers, both in absolute (hourly wage difference) and relative terms (female/male wage ratio). Among all employer ownerships, females earn less than male migrants and the female/male wage ratios are similar for different ownerships. This result is different from the findings for Chinese urban workers in Chi and Li (2008) that female/male wage ratios are higher for state-owned enterprizes than private, foreign and joint venture firms. In terms of industry distribution, the gap is the largest for those working in the finance, insurance and real estate industry. Female migrants are found to earn higher wages only in the education and culture industry. We also find interesting variations in wage differentials across different locations. The wage gap is likely to be smaller

[10] Using survey data from Shanghai, Meng and Zhang (2001) also documented that compared to urban residents, rural migrants were far less likely to be white-collar workers.

[11] Gustafsson and Li (2000) reported gender earnings ratios for China's urban workers using data from China Household Income Project 1988 and 1995. Using data from China Urban Household Survey 1987, 1996 and 2004, Chi and Li (2008) also reported gender earnings ratios in China's urban labor market. Both of them have found that female workers earned less in every occupation, ownership and industry. But the female/male wage ratios for our sample of rural migrants are considerably less than what they have reported. This might indicates that the gender wage gap problem is more serious among rural migrants.

in the more developed areas such as Beijing and Guangdong than in those less developed provinces like Shanxi and Yunnan.

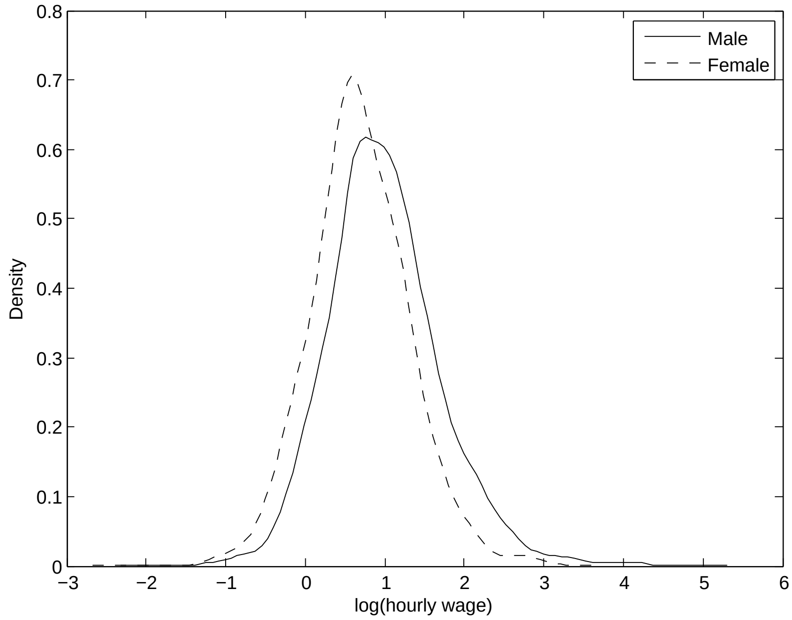

Figure 1: Kernel density estimates of log wage distributions by gender

To better describe the gender wage disparities among rural migrants, we present the kernel density estimates of the logarithmic hourly wages for both genders in Figure 1, from which we can see the contrasted wage distributions between the two groups. The two-sample Kolmogorov-Smirnov test rejects the null hypothesis that the logarithmic hourly wages for the two groups come from the same distribution (the p value is 0.000).

Following Albrecht et al. (2003), we also plot the raw gender log wage differential at each percentile in Figure 2. For example, at the 10th percentile, we see a gender wage gap of 0.288, which means that the log wage at the 10th percentile of male migrants' wage distribution is 28.8 log points higher than that at the same percentile of female migrants' wage distribution. Substantial gender wage differentials are found among the rural migrants and they are not even throughout the wage distribution. Because of the sample size we have, the graph is not smooth. To better describe the trend, we run an OLS regression of the raw percentile gender wage gap on the percentile variable and up to its fifth-order polynomials and plot the fitted wage differentials against the percentile in Figure 2. The gender log wage gap stays at relatively lower levels between the 10th and the 40th percentile. It becomes higher from the 40th percentile to the 70th. A sharp acceleration is observed from the 70th to upper percentiles. It is much higher at the top of the wage distribution than at the bottom, reaching as high as 0.426 at the 90th percentile. This is different to the findings in Bishop et al. (2005) and Chi and Li (2008) that gender wage differentials are higher at the bottom than at the middle and the top end of urban residents' wage distribution, although the two groups both work in China's urban labor

Table 3: Descriptive Average Gender Wage Gap

	Male	Female	M−F	F/M
Occupation:				
Professional, technician	4.364	3.318	1.226	71.91%
Clerical staff	4.296	2.882	1.414	67.09%
Service worker	2.892	2.156	0.736	74.53%
Self−employed or business owner	3.928	2.550	1.378	64.92%
Manufacturing worker	4.392	2.800	1.592	63.75%
Commercial worker	2.897	2.289	0.608	79.02%
Construction worker	4.156	3.461	0.695	83.28%
Other occupations	3.416	2.340	1.076	68.49%
Employer Ownership:				
State−owned	3.229	2.202	1.026	68.21%
Collective	3.299	2.302	0.997	69.77%
Private or self−employed	3.767	2.520	1.246	66.92%
Foreign or joint venture	4.955	3.381	1.574	68.23%
Others	3.775	2.304	1.471	61.02%
Industry:				
Manufacturing	4.488	2.975	1.513	66.28%
Construction	6.254	3.664	2.590	58.59%
Transportation, communication	5.142	3.478	1.664	67.64%
Wholesale, retail, food services	3.232	2.383	0.849	73.72%
Finance, insurance, real estate	4.817	2.183	2.634	45.33%
Social service	2.874	2.246	0.627	78.17%
Education, culture	3.093	3.977	-0.884	128.54%
Health, sports, social welfare	3.496	2.277	1.219	65.14%
Other industries	4.167	2.442	1.725	58.61%
Province:				
Anhui	2.886	1.901	0.985	65.88%
Beijing	4.646	3.361	1.284	72.35%
Chongqing	3.006	2.216	0.789	73.74%
Gansu	3.149	1.881	1.268	59.75%
Guangdong	4.464	3.121	1.343	69.91%
Henan	3.325	2.206	1.119	66.35%
Hubei	3.639	2.589	1.050	71.14%
Jiangsu	5.951	2.850	3.101	47.89%
Liaoning	3.587	2.580	1.007	71.93%
Shanxi	3.183	1.798	1.385	56.47%
Sichuan	2.708	2.266	0.442	83.68%
Yunnan	3.731	2.181	1.550	58.46%
Full Sample	3.720	2.454	1.266	65.97%

Note: M−F denotes the average wage gap between male and female migrants. F/M indicates the ratio of females' average wage over males'.

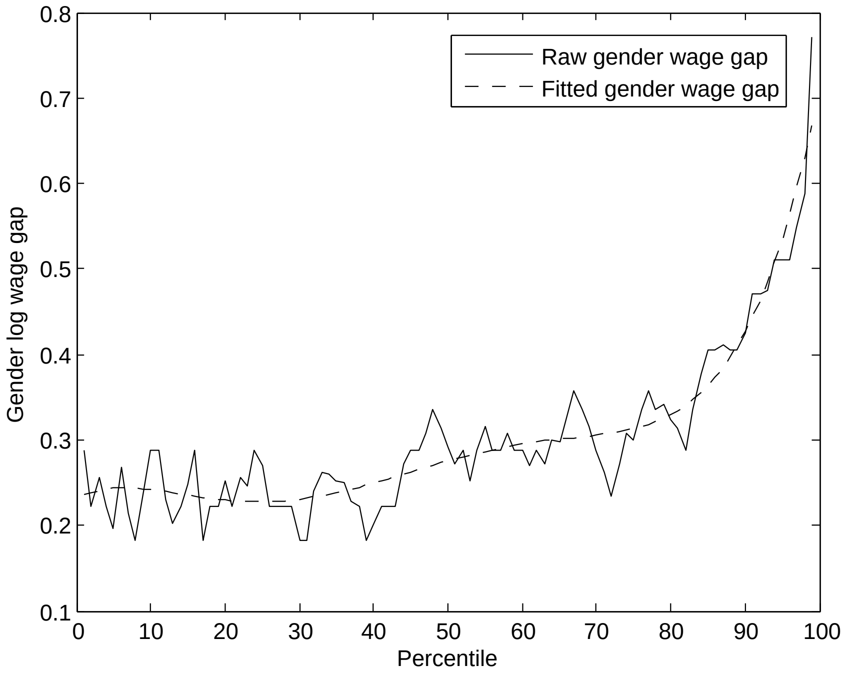

Figure 2: Raw gender log wage gap by percentile among rural migrants

market. This finding highlights the intrinsic differences between rural migrants and urban residents in gender wage distributions and provides a further evidence that gender wage differentials among rural migrants deserve a separate study.

3 Empirical Methodology

3.1 OLS Estimation and Mean Wage Gap Decomposition

3.1.1 OLS Regression

Suppose the mean log wage function for males (m) and females (f) is described by the following equation:

$$E[Y_g|X_g] = X_g\beta_g \tag{1}$$

where Y denotes the log hourly wages, X is the vector of labor market characteristics (including a constant term), β is the vector of coefficients and $g = m, f$ denotes the group. Then the OLS estimate of β_g indicates the impact of X on the conditional or unconditional average of Y for group g.

3.1.2 Mean Wage Gap Decomposition

The Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973) is widely used to decompose average earnings gap between two groups into an endowment effect explained by the differences in covariates and an unexplained discrimination effect due to the different returns to covariates. The mean gender log wage gap can be expressed as:

$$\overline{Y}_m - \overline{Y}_f = (\overline{X}_m - \overline{X}_f)\widehat{\beta}_m + \overline{X}_f(\widehat{\beta}_m - \widehat{\beta}_f) = (\overline{X}_m - \overline{X}_f)\widehat{\beta}_f + \overline{X}_m(\widehat{\beta}_m - \widehat{\beta}_f) \tag{2}$$

where $(\overline{X}_m - \overline{X}_f)\widehat{\beta}_g$ is the endowment effect and $\overline{X}_g(\widehat{\beta}_m - \widehat{\beta}_f)$ represents the discrimination effect. Both of them can then be decomposed into the contribution of each single variable.

However, it is easy to see that the decomposition results vary with the choice of base category (either male or female) (Oaxaca and Ransom, 1994). One proposition to solve this problem is to use the pooled wage structure ($\widehat{\beta}_p$) as the reference (Oaxaca and Ransom, 1994), then equation (2) can be written as:

$$\overline{Y}_m - \overline{Y}_f = (\overline{X}_m - \overline{X}_f)\widehat{\beta}_p + [\overline{X}_m(\widehat{\beta}_m - \widehat{\beta}_p) - \overline{X}_f(\widehat{\beta}_f - \widehat{\beta}_p)] \tag{3}$$

The first term in the bracket is interpreted as the wage advantage of men (the amount that men are overpaid relative to the reference wage structure) and the second term as the wage disadvantage of women (the amount that women are underpaid relative to the reference wage structure) (Oaxaca and Ransom, 1994). However, this approach does not include a gender intercept in the pooled regression, so it may inappropriately transfer some of unexplained part of the wage differential into the explained component. Furthermore, this reference wage structure ($\widehat{\beta}_p$) can not make men's advantage equal to women's disadvantage, so it is hardly nondiscriminatory (Fortin, 2008). Another problem is that once we have categorical predictors, the decomposition results for them depend on the choice of left-out category. For the unexplained part, changing the omitted base category not only alters the results for single dummy variables but also changes the whole contribution of that categorical variable because different part of its effect is hidden in the intercept (Jann, 2008; Fortin et al., 2010).

To obtain a nondiscriminatory wage structure, Fortin (2008) suggests estimating the γ in the following model using the male and female pooled sample:

$$Y_i = X_i\gamma + F_i\gamma_f + M_i\gamma_m + \varepsilon_i \tag{4}$$

where F_i is the female dummy and M_i is the male dummy. Equation (4) is estimated subject to the identification restriction that $\gamma_f + \gamma_m = 0$. It can be shown that

$$\overline{Y}_m - \overline{Y}_f = (\overline{X}_m - \overline{X}_f)\widehat{\gamma} + (\widehat{\gamma}_m - \widehat{\gamma}_f) = (\overline{X}_m - \overline{X}_f)\widehat{\gamma} + [\overline{X}_m(\widehat{\beta}_m - \widehat{\gamma}) - \overline{X}_f(\widehat{\beta}_f - \widehat{\gamma})] \tag{5}$$

where $\widehat{\gamma}_m$ equals to $\overline{X}_m(\widehat{\beta}_m - \widehat{\gamma})$ and $\widehat{\gamma}_f$ equals to $\overline{X}_f(\widehat{\beta}_f - \widehat{\gamma})$. Then the mean wage differential between males and females is decomposed into the endowment effect $((\overline{X}_m - \overline{X}_f)\widehat{\gamma})$ and the discrimination effect $(\widehat{\gamma}_m - \widehat{\gamma}_f)$. $\widehat{\gamma}_m$ is positive, representing men's advantage, and $\widehat{\gamma}_f$ is negative with $-\widehat{\gamma}_f$ representing women's disadvantage. The constraint $\gamma_f + \gamma_m = 0$

in the pooled regression ensures that the advantage of men is equal to the disadvantage of women. Thus the $\widehat{\gamma}$ obtained is a nondiscriminatory wage structure.

To address the noninvariance problem resulting from categorical predictors in the decomposition, Fortin (2008) adopts the solution by Gardeazabal and Ugidos (2004) by including all dummy variables in the regression but restricting the sum of estimated coefficients for each set of dummy variables to be 0. It is easy to see that the decomposition result is not contaminated by any left-out category. We will follow this approach in our decomposition exercise.

3.2 Unconditional Quantile Regression and Quantile Decomposition

The method described above only allows us to investigate wage determination and decompose gender wage gap at the mean. To analyze the wage differentials throughout the distribution, we resort to the unconditional quantile regression and the quantile decomposition based on it.

3.2.1 Unconditional Quantile Regression

The key concept in the unconditional quantile regression by Firpo et al. (2009) is the influence function (IF), which is a widely used tool in robust statistics that represents the influence of an individual observation on a distributional measure of interest such as quantile, variance or other statistics. By adding the influence function back to the statistic we care about, we can obtain the recentered influence function (RIF). For the quantile, the influence function IF(Y, q_τ) is known to be $(\tau - I(Y \leq q_\tau))/f_Y(q_\tau)$, where q_τ is the τth quantile of Y, I is an indicator function and f_Y is the density of the marginal distribution of Y. Then the recentered influence function RIF(Y, q_τ) is equal to q_τ+IF(Y, q_τ). We can model RIF(Y, q_τ) as the function of explanatory variables. Similar to OLS regression, a linear specification is typically assumed for RIF(Y, q_τ):

$$E[RIF(Y, q_\tau)|X] = X\beta_\tau \tag{6}$$

Unlike the conditional quantile regression by Koenker and Bassett (1978), the unconditional quantile regression by Firpo et al. (2009) can be directly used to evaluate the economic impact of a change of X on the corresponding quantiles of the unconditional distribution of Y, which is usually of real interest in economic applications. Since the RIF(Y, q_τ) is never observed in practice, following Firpo et al. (2009), we replace all unknown components with their sample estimators in our empirical application:

$$\widehat{RIF}(Y_i, \widehat{q}_\tau) = \widehat{q}_\tau + \frac{\tau - I(Y_i \leq \widehat{q}_\tau)}{\widehat{f}_Y(\widehat{q}_\tau)} \tag{7}$$

In equation (7), $\widehat{q}_\tau$, the τth sample quantile of Y, can be represented as $\widehat{q}_\tau$=arg $\min\limits_{q}$ $\sum_{i=1}^{N}\left(\tau - I(Y_i \leq q)\right)(Y_i - q)$. The nonparametric Rosenblatt's kernel density estimator $\widehat{f}_Y(\widehat{q}_\tau)$ is equal to $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{h_Y}K_Y(\frac{Y_i-\widehat{q}_\tau}{h_Y})$, where K_Y is the Gaussian kernel and h_Y is the scalar bandwidth for Y. The $\widehat{\beta}_\tau$ estimated as $(X'X)^{-1}X'\widehat{RIF}(Y,\widehat{q}_\tau)$ can be used to recover the marginal effect of X on the unconditional τ-quantile of Y. For technical details of the RIF-OLS regression, we refer to Firpo et al. (2009).

3.2.2 RIF-based Quantile Decomposition

Similar to the mean decomposition, it is also of interest to investigate what contributes to the gender wage gap at different percentiles. However, as pointed out by Firpo et al. (2007), the previous methods proposed share the same shortcoming that they cannot untangle both the endowment effect and discrimination effect into the contribution of each covariate.[12] Using the theoretical property that the mean of RIF(Y, q_τ) equals to q_τ, the mean differences in RIF(Y, q_τ) between two groups are equivalent to the differences in q_τ for them (Firpo et al., 2007). Assuming linear specifications, the Oaxaca-Blinder procedure can be easily generalized to decompose the difference in unconditional quantiles.[13]

One method, as applied in Nandi and Nicoletti (2009) and Fortin et al. (2010), is "... applying the Oaxaca-Blinder method using RIF-regression rather than the Y-regression, i.e., replacing the dependent variable Y with RIF(Y, q_τ)". However, this procedure shares all the problems of Oaxaca-Blinder decomposition we discussed in Section 3.1.2. Based on Firpo et al. (2009), an alternative two-stage approach is proposed by Firpo et al. (2007). In the first stage, the counterfactual wage for one group is obtained by using the reweighting method by DiNardo et al. (1996), and the overall differences in wages are then decomposed into a composition effect (endowment effect) and a wage structure effect (discrimination effect). In the second stage, the two effects are further decomposed into the contribution of every covariate.[14] However, the first-stage decomposition results also vary with the base group chosen (DiNardo et al., 1996; Fortin et al., 2010).[15] In the second stage, the contribution of each covariate to wage structure effect also has the noninvariance

[12] These methods include the "Plug-in"method of Juhn et al. (1993), the reweighting procedure of DiNardo et al. (1996) and the simulation-based methods of Machado and Mata (2005) and Melly (2005). While it is natural to ask how the differences in education between two groups contribute to the endowment effect at the τ-quantile, the question cannot be answered with these methods. See the survey of the literature in Fortin et al. (2010).

[13] The RIF of the mean is equal to Y itself, so RIF-regression with the mean statistic is identical to OLS estimation and the Oaxaca-Blinder decomposition is a special case of RIF-based decomposition.

[14] See Chi and Li (2008) for an application of this technique.

[15] Using our rural migrants data, among the deciles from the first to the ninth, the decomposition results are insensitive to the choice of base group only at the first and the fifth decile. For example, at τ=0.9, the raw gender log wage difference is 0.426. With males being the base group, the composition effect is estimated to be 0.223 and the wage structure effect is 0.203. However, if females are selected as the base group, the two effects are respectively 0.134 and 0.292. It is hard to conclude which effect contributes more the wage gap at the 90th percentile.

problem from using categorical predictors (Firpo et al., 2007).

The mean decomposition method by Fortin (2008) can be easily extend to quantile decomposition with unconditional quantile regression by Firpo et al. (2009). We substitute Y_i in section 3.1.2 for each gender group with corresponding RIF(Y_i, q_τ). In the pooled RIF-OLS regression to estimate the reference wage structure, the dependent variable we use is obtained by pooling the RIF$(Y_m, q_{m\tau})$ and RIF$(Y_f, q_{f\tau})$ for each gender. By using a similar approach to the method by Fortin (2008), we can always find a nondiscriminatory reference wage structure that can make men's advantage equal women's disadvantage at the τ-quantile, so the sensitivity problem from the choice of base group is avoided. Moreover, the problem resulting from categorical predictors can be easily resolved using the solution by Gardeazabal and Ugidos (2004) with restricted regressions. This RIF-based decomposition can divide up both the endowment effect and the discrimination effect into the contribution of each explanatory variable.

Similar to equation (5), theoretically we have

$$\begin{aligned}\widehat{q}_{m\tau} - \widehat{q}_{f\tau} &= \overline{RIF}(Y_m, \widehat{q}_{m\tau}) - \overline{RIF}(Y_f, \widehat{q}_{f\tau}) \\ &= (\overline{X}_m - \overline{X}_f)\widehat{\gamma}_\tau + [\overline{X}_m(\widehat{\beta}_{m\tau} - \widehat{\gamma}_\tau) - \overline{X}_f(\widehat{\beta}_{f\tau} - \widehat{\gamma}_\tau)]\end{aligned} \tag{8}$$

$\widehat{q}_{m\tau}$ and $\widehat{q}_{f\tau}$ are the τth quantiles of the marginal distributions of Y_m and Y_f. $\widehat{\beta}_{m\tau}$ and $\widehat{\beta}_{f\tau}$ are the coefficient estimates from RIF-OLS regression for each gender. The first equality uses the fact that the mean of RIF(Y, q_τ) is q_τ. We obtain the second equality by replacing Y in equation (5) with RIF(Y, q_τ). Note that we consider $\widehat{\gamma}_\tau$ from the pooled RIF-regression as the counterfactual wage structure that would prevail at the τ-quantile in the absence of discrimination and in the pooled RIF-OLS regression the dependent variable we use is obtained by pooling the RIF$(Y_m, q_{m\tau})$ and RIF$(Y_f, q_{f\tau})$ for each gender.[16] $(\overline{X}_m - \overline{X}_f)\widehat{\gamma}_\tau$ is the endowment effect and $\overline{X}_m(\widehat{\beta}_{m\tau} - \widehat{\gamma}_\tau) - \overline{X}_f(\widehat{\beta}_{f\tau} - \widehat{\gamma}_\tau)$ as a whole represents the discrimination effect at the τ-quantile. The discrimination effect can be further decomposed into two equal components: men's advantage (amount that men are overpaid relative to the reference wage structure: $\overline{X}_m(\widehat{\beta}_{m\tau} - \widehat{\gamma}_\tau)$) and women's disadvantage (amount that women are underpaid relative to the reference wage structure: $-\overline{X}_f(\widehat{\beta}_{f\tau} - \widehat{\gamma}_\tau)$). Compared with the procedures by Firpo et al. (2007) using the reweighting method to obtain the counterfactual wage distribution for the base group, the approach we use relies on the

[16] We do not use the RIF(Y, q_τ) for the full sample in the regression. One reason is that the γ_τ obtained can not make men's advantage equal to women's disadvantage, so the reference wage structure is discriminatory. Another reason is that unlike OLS estimation, the coefficient estimates obtained in that way does not necessarily lie between the two sets of estimates for both genders. If the returns to some labor market characteristics using the pooled sample are larger (smaller) than those for both genders, then discrimination is (not) considered to be directed towards both males and females, which is quite unreasonable. We also obtain an alternative reference wage structure by weighting coefficient estimates obtained from RIF-OLS regressions at the τ-quantile for both genders by group sizes. The decomposition results are very similar. See Section 4.4.3.

linear structure and within-sample predictions.

As shown in Firpo et al. (2007) and Firpo et al. (2009), the recentered influence function RIF(Y, q_τ) is the leading term of a von Mises (1947) expansion of q_τ, through which RIF(Y, q_τ) approximates the nonlinear function of q_τ by expectation. The property that RIF(Y, q_τ) has a mean value of q_τ ensures that theoretically the decomposition in equation (8) is an exact one. However, in finite sample settings, $\overline{RIF}(Y, \widehat{q}_\tau)$ obtained may not be exactly equal to the sample quantile $\widehat{q}_\tau$, which is mainly caused by the presence of heaping displayed in wage distribution (Fortin et al., 2010).[17] Consequently, the predicted wage gap at the τ-quantile ($\overline{RIF}(Y_m, \widehat{q}_{m\tau}) - \overline{RIF}(Y_f, \widehat{q}_{f\tau})$) may be different from the raw wage gap ($\widehat{q}_{m\tau} - \widehat{q}_{f\tau}$). We use $\widehat{R}_\tau$ to denote the approximation error in quantile decomposition. Then the raw gender log wage gap at the τ-quantile can be decomposed as:

$$\widehat{q}_{m\tau} - \widehat{q}_{f\tau} = \overline{RIF}(Y_m, \widehat{q}_{m\tau}) - \overline{RIF}(Y_f, \widehat{q}_{f\tau}) + \widehat{R}_\tau \tag{9}$$

Since $\overline{RIF}(Y_m, \widehat{q}_{m\tau}) - \overline{RIF}(Y_f, \widehat{q}_{f\tau})$ equals to $(\overline{X}_m - \overline{X}_f)\widehat{\gamma}_\tau + [\overline{X}_m(\widehat{\beta}_{m\tau} - \widehat{\gamma}_\tau) - \overline{X}_f(\widehat{\beta}_{f\tau} - \widehat{\gamma}_\tau)]$, equation (9) now becomes:

$$\widehat{q}_{m\tau} - \widehat{q}_{f\tau} = (\overline{X}_m - \overline{X}_f)\widehat{\gamma}_\tau + [\overline{X}_m(\widehat{\beta}_{m\tau} - \widehat{\gamma}_\tau) - \overline{X}_f(\widehat{\beta}_{f\tau} - \widehat{\gamma}_\tau)] + \widehat{R}_\tau \tag{10}$$

The magnitude of $\widehat{R}_\tau$ shows how good the RIF-based quantile decomposition approximates the raw wage differentials. The standard errors for any components in equation (10) can be obtained with bootstrapping or using the method described by Jann (2008).

Using the decomposition methods, we can compute both the endowment effect and discrimination effect throughout the wage distribution, and the contribution of each covariate to the two effects. However, as pointed by Fortin et al. (2010) which finds the close link between the decomposition methods and the program evaluation literature (Imbens and Wooldridge, 2009), these methods have the limitation that the economic mechanisms behind the various elements of decomposition (or behind the treatment effect) are not clear since structural modeling is not required to perform a decomposition or estimate a treatment effect.

[17]The $\overline{RIF}(Y, \widehat{q}_\tau)$ is estimated as $\frac{1}{N}\sum_{i=1}^{N}\widehat{RIF}(Y_i, \widehat{q}_\tau)$, where N is the sample size. Since $\widehat{RIF}(Y_i, \widehat{q}_\tau)$ is defined as $\widehat{q}_\tau + \frac{\tau - I(Y_i \le \widehat{q}_\tau)}{\widehat{f}_Y(\widehat{q}_\tau)}$, $\overline{RIF}(Y, \widehat{q}_\tau)$ is equivalent to $\widehat{q}_\tau + \frac{1}{\widehat{f}_Y(\widehat{q}_\tau)}[\tau - \frac{1}{N}\sum_{i=1}^{N} I(Y_i \le \widehat{q}_\tau)]$, where $\frac{1}{N}\sum_{i=1}^{N} I(Y_i \le \widehat{q}_\tau)$ is the proportion of observations with Y not greater than the τ-percentile. However, when there are multiple $Y_i s$ equal to the sample quantile $\widehat{q}_\tau$, τ will be smaller than $\frac{1}{N}\sum_{i=1}^{N} I(Y_i \le \widehat{q}_\tau)$. And this makes the $\overline{RIF}(Y, \widehat{q}_\tau)$ different from the sample quantile $\widehat{q}_\tau$. But this problem is not serious if the number of $Y_i s$ equal to $\widehat{q}_\tau$ is small. This problem can also exist in the first-stage of the decomposition method by Firpo et al. (2007). For example, in **?**) which used the method of Firpo et al. (2007), their equation (4) can be written as $\widehat{q}_{m\tau} - \widehat{q}_{f\tau} = \overline{RIF}(Y_m, \widehat{q}_{m\tau}) - \overline{RIF}(Y_f, \widehat{q}_{f\tau}) + (\widehat{R}^s_\tau + \widehat{R}^c_\tau)$. If the predicted $\overline{RIF}(Y, \widehat{q}_\tau)$ always equals to the sample percentile $\widehat{q}_\tau$, then $\widehat{R}^s_\tau + \widehat{R}^c_\tau$ should be 0. However, their decomposition results in Table 5a, 5b and 5c show this does not hold all the time.

4 Results

4.1 OLS Estimation Results by Gender

As analyzed by Meng (1998), rural migrants' wages are determined by market forces, so we can safely rely on the human capital model to do the estimation. We first run an OLS regression with the male and female pooled sample, allowing the effect of each covariate varying with gender. Then we test the joint significance of all the interaction terms. We reject the null hypothesis that the corresponding OLS coefficient estimates are the same for both genders (the p value is 0.000). As a result, we examine the mean wage determination for each gender separately.[18]

Table 4 displays the OLS regression result for each gender with Huber-White standard errors to correct heteroscedasticity of unknown form. Note that we include those arguably endogenous variables such as occupation, ownership and industry dummies in the regressions. Albrecht et al. (2003) pointed out that as an accounting exercise they are useful in explaining wage differentials.[19] For other examples including these variables in wage determination and wage gap decomposition, see, for example, Gustafsson and Li (2000) and Chi and Li (2008).

The relation of rural migrants' logarithmic hourly wages to their ages has the widely documented inverted U-shape. The profile peaks at age 37 for male migrants, and at 31 for females. The returns to education are the same for males and females, with an additional year of schooling leading to 4.1% increase in hourly wage. This rate for rural migrants in our sample is lower than the 4.8% reported in Meng and Zhang (2001) using migrants data from Shanghai and it is also lower than the 5.5% for urban workers in Li (2003).[20] The returns to urban experience and tenure in the current job are both very small for males and females. This is expected because most migrants engage in low-skilled jobs and they can be easily substituted by competitors. Consequently, longer urban experience and tenure do not guarantee a big pay rise. Unlike their male counterparts, female migrants are rewarded for marriage, although it is not statistically significant. Contrary to the previous findings that being a ethnic minority put urban workers at a disadvantage in earning higher wages (Gustafsson and Li, 2000; Li, 2003; Yang, 2005), the coefficient on $Minority$ is 5.6% for male migrants. The premium to having a permanent job or being a long-term contract worker is very high, being 14.2% and 23.1% respectively for both genders. Furthermore,

[18]In Section 4.2, when we analyze the wage determination for rural migrants at different percentiles, we have done similar tests. The null hypotheses that the RIF-OLS coefficients at the 10th, 50th and 90th percentile are the same for both genders are all rejected at the conventional significance levels. So we choose to show the estimation results for each gender separately.

[19]In Section 4.4.3, as a sensitivity analysis, we do the decomposition excluding these variables.

[20]Zhang et al. (2005) documented that the rate of return to education in China's urban labor market has risen in a nonlinear way from 4.0% in 1988 to 10.2% in 2001. Compared with their 2001 urban counterparts, our rural migrants sample surveyed in 2002 have a much smaller wage premium from having more education.

Table 4: OLS Estimation Results by Gender

	Male		Female	
	Coef.	SD	Coef.	SD
Constant	-0.362	0.283	-0.313	0.279
Age	0.047	0.015	0.040	0.015
Agesq/100	-0.063	0.020	-0.064	0.021
Education	0.041	0.006	0.041	0.006
Experience	0.011	0.004	0.005	0.005
Tenure	0.008	0.005	0.018	0.006
Married	-0.045	0.065	0.013	0.060
Minority	0.056	0.062	-0.026	0.057
Permwork	0.142	0.064	0.231	0.070
Interprov	0.053	0.037	0.009	0.038
Occupation:				
Professional, technician	—	—	—	—
Clerical staff	0.003	0.117	-0.055	0.130
Service worker	-0.169	0.079	-0.239	0.099
Self−employed or business owner	-0.038	0.077	-0.064	0.098
Manufacturing worker	0.039	0.102	-0.220	0.122
Commercial worker	-0.120	0.093	-0.214	0.107
Construction worker	-0.337	0.131	-0.205	0.241
Other occupations	-0.200	0.091	-0.300	0.119
Employer Ownership:				
State−owned	—	—	—	—
Collective	0.121	0.080	0.019	0.081
Private or self−employed	0.163	0.061	0.071	0.061
Foreign or joint venture	0.350	0.128	0.282	0.121
Others	0.103	0.063	0.062	0.064
Industry:				
Manufacturing	—	—	—	—
Construction	0.445	0.127	0.369	0.135
Transportation, communication	0.184	0.093	0.309	0.177
Wholesale, retail, food services	-0.237	0.064	-0.154	0.066
Finance, insurance, real estate	0.093	0.190	-0.054	0.158
Social service	-0.190	0.065	-0.140	0.068
Education, culture	-0.192	0.140	0.021	0.145
Health, sports, social welfare	-0.091	0.174	-0.106	0.127
Other industries	-0.025	0.076	-0.057	0.087
Province dummies	Yes		Yes	
R-squared	0.190		0.203	
Observations	1,865		1,420	

male migrants benefit more from interprovincial migration that nonetheless has almost no impact on females' average earnings.

We find that the returns to occupational dummies differ by gender. Among all occupations, being a manufacturing worker pays most for male migrants, but work as a technician is the best occupational choice for female migrants, *ceteris paribus*. Both males and females are penalized most when working in state-owned enterprizes. Although foreign firms and joint ventures are the least often to hire rural migrants in China, the wages they offer for both genders are the highest among all ownerships. Furthermore, wages are the highest in construction industry and the lowest in wholesale, retail and food services industry for both male and female migrants. We can also find substantial heterogeneity in the returns to different industries. The differences in the coefficient estimates of industry dummies can reach as high as 0.68 for males and 0.62 for females, showing industry choices can substantially affect the wage inequality within each gender group.

4.2 Unconditional Quantile Regression Results by Gender

We report the unconditional quantile regression estimates separately by gender at the 10th, 50th and 90th percentiles in Table 5 and Table 6 with asymptotic standard errors. Similar to the OLS estimates that can be interpreted as the marginal effects of control variables on the mean of dependent variable, the coefficient estimates from RIF-OLS regressions are explained as the marginal effects of those covariates on the corresponding unconditional quantiles of dependent variable. The regression results suggest that the returns to observable characteristics differ by gender and these differences change as we move up the wage distributions.

4.2.1 RIF-OLS Regression Results for Male Migrants

For male migrants, education plays an increasingly important role in wage determination from the 10th to 90th percentile. At the 10th percentile of males' log wage distribution, one more year of schooling can increase the wage by around 1.5%. It goes up to 4.4% at the median and further increases to 10% at the 90th percentile. This shows that the OLS coefficient estimate of education can not fully describe the role that education plays in wage determination. The effects of working experience in the city are also monotonic, but they are much smaller in magnitude. Current job tenure exerts a diminishing effect on wage from the 10th to the 90th percentile. The effects of being married on earnings differ widely at the three percentiles. At the 10th percentile, it almost has no impact. It then turns negative at the median, although the magnitude is small. The penalty is the largest at the 90th percentile, indicating a 26% loss in hourly wages. The impact of being an ethnic minority is non-monotonic. We see a large positive impact at the higher quantile but a negative effect at the median. A permanent job benefits those at the median and the

90th percentile of males' wage distribution most. The effects of interprovincial migration on wages are similar at the three percentiles, ranging from 8% to 11%.

In terms of occupations, for males at the 10th and 50th percentiles of the wage distribution, working as a manufacturing worker earn more than all other occupations. But at the 90th percentile, it is the clerical staff who earns the most. The maximum difference in those coefficient estimates at the 10th percentile is 0.15 and it rises to 0.33 at the median. It further increases to 1.66 at the 90th percentile, indicating increasing differences in the returns to occupation dummies as percentile increases. In light of ownerships, foreign firms and joint ventures offer highest wages at the 10th and 90th percentile, but private firms or self-employed businesses pay most at the median. At both the 10th and 50th percentiles, the wage received from state-owned enterprizes is the least. At the 90th percentile, collective firms employ male migrants at the lowest wage levels. The industry that offers the highest and lowest wages also vary with the percentile. The maximum difference in the coefficients on industrial dummies rises from 0.29 at the 10th percentile to 0.55 at the median before reaching 2.20 at the 90th percentile of male migrants' wage distribution. We find that as the percentile increases, there is more heterogeneity in the coefficient estimates of both occupation and industry dummies. By comparing the OLS estimation result for male migrants and the results in Table 5, we find that the OLS regression cannot provide an adequate description of men's wage determination through the wage distribution.

4.2.2 RIF-OLS Regression Results for Female Migrants

The unconditional quantile regression results for women are displayed in Table 6. The differences in the returns to education at the three percentiles for women are not as large as those for men, ranging from 3% to 5%. The coefficients on $Experience$ and $Tenure$ are all very small in magnitude. Being married seems to increase women's wages only at the median. The premiums to having a permanent work or being a long-term contract worker are substantial at the three percentiles and the three coefficient estimates are all larger than those estimates for male migrants. As shown in Table 5, interprovincial migration has no negative impact on males' wages. But this does not hold for female migrants. Working outside one's home province increases women's earnings only at the 10th percentile. Around 4% loss in wages resulting from interprovincial migration is found at the median and 90th percentile.

Other things being the same, working as a technician earns most at the 10th and 90th percentiles of women's wage distribution. At the median, it is the clerical staff that earns the most. Unlike the estimates for males, the differences in the coefficient estimates of occupational dummies for females are the largest at the lower percentile. Similar to the coefficient estimates of ownership dummies for male migrants, the wages received from state-owned enterprizes are the least at both the 10th percentile and the median, while

Table 5: Unconditional Quantile Regression Results for Men

	10th		50th		90th	
	Coef.	SD	Coef.	SD	Coef.	SD
Constant	-1.112	0.411	-0.130	0.337	-0.487	0.834
Age	0.063	0.022	0.034	0.018	0.067	0.046
Agesq/100	-0.095	0.030	-0.048	0.023	-0.074	0.058
Education	0.015	0.008	0.044	0.007	0.100	0.020
Experience	0.008	0.006	0.013	0.005	0.021	0.013
Tenure	0.015	0.006	0.006	0.006	0.003	0.016
Married	0.002	0.086	-0.022	0.082	-0.259	0.220
Minority	0.042	0.070	-0.020	0.067	0.195	0.182
Permwork	0.019	0.084	0.193	0.079	0.191	0.238
Interprov	0.098	0.046	0.105	0.047	0.084	0.118
Occupation:						
Professional, technician	—	—	—	—	—	—
Clerical staff	-0.100	0.138	-0.126	0.126	0.507	0.477
Service worker	-0.129	0.084	-0.278	0.094	-0.079	0.275
Self−employed or business owner	-0.066	0.074	-0.270	0.089	0.323	0.277
Manufacturing worker	0.022	0.100	0.005	0.123	-0.203	0.352
Commercial worker	-0.134	0.116	-0.306	0.113	0.053	0.319
Construction worker	0.004	0.100	-0.334	0.141	-1.152	0.515
Other occupations	-0.199	0.105	-0.278	0.107	-0.116	0.304
Employer Ownership:						
State−owned	—	—	—	—	—	—
Collective	0.193	0.101	0.151	0.123	-0.252	0.245
Private or self−employed	0.036	0.086	0.203	0.074	0.329	0.191
Foreign or joint venture	0.258	0.096	0.135	0.188	0.468	0.677
Others	0.049	0.091	0.048	0.077	0.219	0.201
Industry:						
Manufacturing	—	—	—	—	—	—
Construction	0.058	0.097	0.335	0.112	1.441	0.476
Transportation, communication	0.139	0.079	0.155	0.105	0.049	0.375
Wholesale, retail, food services	-0.075	0.067	-0.211	0.073	-0.585	0.245
Finance, insurance, real estate	-0.150	0.221	0.125	0.169	0.151	0.596
Social service	-0.005	0.071	-0.137	0.078	-0.751	0.236
Education, culture	0.082	0.183	-0.266	0.146	-0.085	0.466
Health, sports, social welfare	-0.158	0.267	0.165	0.189	-0.833	0.431
Other industries	-0.055	0.079	-0.011	0.081	-0.283	0.268
Province dummies	Yes		Yes		Yes	
Observations			1,865			

Table 6: Unconditional Quantile Regression Results for Women

	10th		50th		90th	
	Coef.	SD	Coef.	SD	Coef.	SD
Constant	-2.892	0.793	-0.196	0.331	0.689	0.420
Age	0.167	0.044	0.031	0.017	0.029	0.021
Agesq/100	-0.235	0.062	-0.054	0.022	-0.038	0.027
Education	0.050	0.015	0.031	0.007	0.048	0.009
Experience	0.003	0.014	0.012	0.006	-0.010	0.007
Tenure	0.013	0.015	0.007	0.008	0.027	0.009
Married	-0.024	0.148	0.018	0.075	-0.050	0.100
Minority	0.119	0.143	-0.164	0.072	-0.064	0.090
Permwork	0.169	0.183	0.287	0.096	0.195	0.124
Interprov	0.044	0.092	-0.038	0.048	-0.042	0.061
Occupation:						
Professional, technician	—	—	—	—	—	—
Clerical staff	-0.115	0.241	0.100	0.181	-0.100	0.284
Service worker	-0.464	0.145	-0.208	0.145	-0.229	0.197
Self−employed or business owner	-0.256	0.123	-0.089	0.145	-0.000	0.198
Manufacturing worker	-0.177	0.130	-0.218	0.181	-0.372	0.251
Commercial worker	-0.539	0.179	-0.157	0.153	-0.152	0.210
Construction worker	-0.466	0.192	-0.230	0.322	-0.271	0.524
Other occupations	-0.678	0.213	-0.195	0.166	-0.232	0.210
Employer Ownership:						
State−owned	—	—	—	—	—	—
Collective	0.295	0.223	0.040	0.122	-0.194	0.113
Private or self−employed	0.057	0.200	0.070	0.086	0.068	0.095
Foreign or joint venture	0.330	0.230	0.345	0.171	0.015	0.459
Others	-0.001	0.202	0.037	0.090	0.076	0.102
Industry:						
Manufacturing	—	—	—	—	—	—
Construction	0.406	0.155	0.472	0.152	0.283	0.344
Transportation, communication	0.378	0.207	0.410	0.177	0.087	0.418
Wholesale, retail, food services	-0.088	0.129	-0.155	0.079	-0.277	0.122
Finance, insurance, real estate	-0.626	0.623	-0.017	0.232	-0.406	0.141
Social service	-0.118	0.147	-0.099	0.085	-0.245	0.124
Education, culture	0.136	0.152	-0.157	0.162	0.043	0.270
Health, sports, social welfare	0.130	0.351	-0.077	0.165	-0.337	0.211
Other industries	-0.131	0.187	-0.124	0.108	-0.213	0.146
Province dummies	Yes		Yes		Yes	
Observations			1,420			

collective firms pay the least at the 90th percentile. Foreign firms and joint ventures, which are estimated to offer the highest wage levels in the OLS regressions for both genders, pay the highest wages at the 10th and 50th percentiles. The construction industry are estimated to pay female migrants highest wages at those three percentiles, but the industry that pays the least vary with the percentile. Substantial differences in the returns to industry dummies are also found. The maximum differences in the coefficient estimates of industry dummies is 1.032 at the 10th quantile, and it is reduced to 0.629 at the median and 0.689 at the 90th percentile. Unlike the findings for male migrants that heterogeneity in the returns to industry dummies increases with the percentile, the heterogeneity is the largest at the bottom decile for women. By comparing the estimation results in Table 5 and 6, we find contrasted wage structures between the two groups.

4.3 Decomposition Results

From Table 1 and 2, we observe different labor market characteristics between male and female migrants. We also find different coefficient estimates for the two groups from Table 4, 5 and 6. Thus, we use the decomposition techniques described in Section 3 to untangle the distributional wage differentials into a part explained by differences in observed characteristics and a part explained by different returns to covariates. We present the endowment and discrimination effects at the mean and each decile in Table 7. The magnitude of approximation errors reported are all very small, showing RIF-based decompositions provide a very good approximation to the true gender wage differentials in our sample.

Consistent with Figure 2, the size of raw log wage gap is larger at higher deciles. However, the gender wage discrimination problem seems to be most serious at the lower end of rural migrants' wage distribution, which is evident from the ratios of discrimination effects to their corresponding wage gaps.

The detailed decomposition results at the mean and selected quantiles are displayed in Table 8 and 9. For discrete variables, the effect of each category is obtained by summing up the contribution of the dummy variables generated from the category. For example, the effect of "Industry" is obtained by summing up the contribution of all the 9 industry dummies. The effect of "Married" summarizes the effect of marital status, which is the sum of the contributions of a binary variable indicating currently being married and a binary variable indicating not being married. A positive sign suggests that the relevant variable contributes positively to the corresponding effect (endowment or discrimination effect).

4.3.1 Detailed Mean Decomposition Results

In the mean wage gap decomposition, the raw gender log wage gap (0.302) is decomposed into the endowment effect (0.101) and discrimination effect (0.201). The differences in

Table 7: Gender Wage Gap Decomposition at the Mean and Selected Quantiles

Selected Quantiles	Raw Gender Log Wage Gap	Endowment Effects	Discrimination Effects	Approximation Errors
	0.288	0.054	0.234	0.000
10th	(0.029)	(0.016)	(0.041)	(0.037)
	[100%]	[18.8%]	[81.3%]	[0.00%]
	0.251	0.087	0.166	-0.002
20th	(0.040)	(0.015)	(0.035)	(0.026)
	[100%]	[34.7%]	[66.1%]	[-0.80%]
	0.182	0.083	0.107	-0.008
30th	(0.025)	(0.014)	(0.030)	(0.023)
	[100%]	[45.6%]	[58.8%]	[-4.40%]
	0.201	0.087	0.116	-0.002
40th	(0.027)	(0.013)	(0.028)	(0.020)
	[100%]	[43.3%]	[57.7%]	[-1.00%]
	0.292	0.096	0.195	0.001
50th	(0.046)	(0.014)	(0.028)	(0.023)
	[100%]	[32.9%]	[66.8%]	[0.34%]
	0.288	0.105	0.180	0.003
60th	(0.027)	(0.014)	(0.028)	(0.018)
	[100%]	[36.5%]	[62.5%]	[1.04%]
	0.288	0.119	0.173	-0.004
70th	(0.028)	(0.017)	(0.031)	(0.024)
	[100%]	[41.3%]	[60.1%]	[-1.39%]
	0.324	0.145	0.176	0.003
80th	(0.042)	(0.021)	(0.033)	(0.028)
	[100%]	[44.8%]	[54.3%]	[0.93%]
	0.426	0.185	0.245	-0.004
90th	(0.052)	(0.031)	(0.052)	(0.023)
	[100%]	[43.4%]	[57.5%]	[-0.94%]
	0.302	0.101	0.201	0.000
Mean	(0.022)	(0.014)	(0.021)	(0.000)
	[100%]	[33.4%]	[66.6%]	[0.00%]

Note: Bootstrapped standard errors are reported in round parentheses. The ratios of effects to corresponding log wage gaps are reported in square brackets. In mean decomposition, the standard error of the approximation error is 0.000 because the decomposition is exact. All decomposition results reported are rounded to three digits after decimal.

observable characteristics can explain 33% of the average wage gap. Among all variables, education, experience and industry are the three largest contributors to the endowment effect. The impact of province on endowment effect is negative, indicating that it helps to narrow the gender wage gap due to the differences in locations. Age, experience and permanent job status are major sources of discrimination in favor of male migrants and against females. All other variables have negative signs, showing that employers have prejudices against male migrants in terms of those variables, which help decrease the overall discrimination effect against females. The discrimination effects are further decomposed into men's advantage and women's disadvantage, which are equal to each other. Age and experience in city contribute most to both men's advantage and women's disadvantage at the mean, as shown in Table 9.

4.3.2 Detailed Quantile Decomposition Results

In the quantile decomposition, the raw gender wage gap is becoming larger as we move from τ=0.1 to τ=0.9. We consider the endowment effect first. The endowment effect account for only 18.8% of the raw gender log wage gap at the lower end of the wage distribution. It rises to 32.9% at the median and reaches 43.4% at the 90th percentile. This shows that the extent of discrimination decreases with the percentile. Education contributes monotonically to the increase in the endowment effects as we move up the wage distribution. The effects of occupation are on decrease from the 10th percentile to the 90th percentile, with the effects being negative at the higher end of wage distribution. Ownership's influence is very limited across the three percentiles. The contribution of industry becomes greater from lower to upper percentiles. At the 90th percentile, the gender wage gap due to endowment effect is 0.185 log points. The effect of industry alone is 0.136, accounting for about 74% of the endowment effect.

As for the discrimination effect, both age and education have a very important role in contributing to it. At the 10th percentile, the impacts of both age and education are negative, showing male migrants are actually unfavorably treated in terms of returns to age and education. They are both positive at the median and the 90th percentile, with an increasingly larger contribution to the discrimination against females at the upper end of rural migrants' wage distribution. Experience does not has a monotonic positive impact on discrimination effects. The maximum of its effects is reached at the 90th percentile. The effects of tenure, marital status and ethnic minority status share the same trend. They contribute to the overall discrimination effect against females at the 10th percentile but help reducing discrimination at both the median and the 90th percentile of wage distribution. Occupation reduces the extent of discrimination at the lower quantile and the median but contributes substantially to the discrimination against female migrants at the 90th percentile. The effect of ownership on discrimination effect is positive at the three percentiles. The trend of the unexplained effects of industry is similar to that of tenure,

Table 8: Detailed Wage Gap Decomposition at the Mean and Selected Quantiles

	Mean	10th	50th	90th
Gender Log Wage Gap:	0.302	0.288	0.292	0.426
Endowment Effects:	0.101 (33.4%)	0.054 (18.8%)	0.096 (32.9%)	0.185 (43.4%)
Constant	0.000	0.000	0.000	0.000
Age	-0.005	-0.010	-0.009	0.013
Education	0.034	0.024	0.031	0.063
Experience	0.014	0.009	0.018	0.016
Tenure	0.007	0.009	0.004	0.007
Married	0.000	0.000	0.000	-0.001
Minority	0.000	0.001	-0.001	0.001
Permwork	0.003	0.001	0.003	0.003
Interprov	0.001	0.002	0.001	0.001
Occupation	0.001	0.017	0.012	-0.043
Ownership	-0.002	-0.002	-0.004	-0.001
Industry	0.059	0.016	0.051	0.136
Province	-0.010	-0.013	-0.011	-0.010
Discrimination Effects:	0.201 (66.6%)	0.234 (81.3%)	0.195 (66.8%)	0.245 (57.5%)
Constant	0.055	2.188	-0.024	-0.835
Age	0.219	-1.865	0.168	0.857
Education	-0.004	-0.271	0.104	0.407
Experience	0.040	0.033	0.007	0.208
Tenure	-0.050	0.014	-0.003	-0.118
Married	-0.023	0.010	-0.016	-0.084
Minority	-0.034	0.032	-0.060	-0.108
Permwork	0.040	0.067	0.042	0.002
Interprov	-0.008	-0.010	-0.025	-0.022
Occupation	-0.011	-0.012	-0.057	0.150
Ownership	0.015	0.021	0.087	0.049
Industry	-0.027	0.059	-0.018	-0.265
Province	-0.013	-0.032	-0.010	0.005
Approximation Errors:	0.000 (0.00%)	0.000 (0.00%)	0.001 (0.34%)	-0.004 (-0.94%)

Note: The ratios of effects to their corresponding log wage gaps are reported in parentheses. All decomposition results reported are rounded to three digits after decimal.

Table 9: Detailed Decomposition of Discrimination Effects

Discrimination Effects:	Mean	10th	50th	90th
	0.201	0.234	0.195	0.245
Men's Advantage:	0.100 (50.0%)	0.117 (50.0%)	0.097 (50.0%)	0.122 (50.0%)
Constant	-0.041	0.892	-0.079	-0.466
Age	0.173	-0.737	0.154	0.440
Education	0.001	-0.117	0.056	0.192
Experience	0.012	0.011	0.001	0.081
Tenure	-0.015	0.010	0.001	-0.041
Married	-0.016	0.000	-0.014	-0.044
Minority	-0.012	0.019	-0.022	-0.042
Permwork	0.013	0.022	0.010	0.015
Interprov	-0.003	-0.002	-0.009	-0.010
Occupation	-0.006	-0.004	-0.014	0.026
Ownership	0.007	0.008	0.031	0.029
Industry	-0.010	0.019	-0.017	-0.061
Province	-0.002	-0.005	-0.001	0.003
Women's Disadvantage:	0.100 (50.0%)	0.117 (50.0%)	0.097 (50.0%)	0.122 (50.0%)
Constant	0.096	1.295	0.056	-0.370
Age	0.046	-1.128	0.014	0.418
Education	-0.005	-0.154	0.049	0.214
Experience	0.028	0.022	0.006	0.127
Tenure	-0.034	0.004	-0.004	-0.077
Married	-0.007	0.010	-0.002	-0.040
Minority	-0.023	0.012	-0.038	-0.065
Permwork	0.027	0.045	0.032	-0.014
Interprov	-0.005	-0.007	-0.017	-0.012
Occupation	-0.003	-0.008	-0.043	0.124
Ownership	0.008	0.013	0.056	0.020
Industry	-0.017	0.040	-0.001	-0.204
Province	-0.011	-0.027	-0.010	0.002

Note: The ratios of effects to their corresponding discrimination effects are reported in parentheses. All decomposition results reported are rounded to three digits after decimal.

with a positive sign at lower quantile but negative impacts at both the median and 90th percentile.

The discrimination effects can be further decomposed into men's advantage and women disadvantage, which are equal to each other.[21] The advantage of men and disadvantage of women by each covariate show that how men are overpaid and women are underpaid relative to the nondiscriminatory wage structure in terms of that explanatory variable. The results is shown in Table 9. At the 10th percentile, one's permanent job status contributes most to men's advantage and it is the different returns to industry dummies that contributes to women's disadvantage most. Age is the most important determinant of men's advantage at both the median and the 90th percentile. It also contributes most to women's disadvantage at τ=0.9. Moreover, the different returns to ownerships account for most of women's disadvantage at the median. Education also plays a very important role. It help reducing the discrimination effects against female migrants at the 10th percentile. However, as we move up the wage distribution, it contributes more and more to both men's advantage and women's disadvantage.

4.3.3 Discrimination Effects by Selected Labor Market Characteristics: Graphical Results

In this section, we illustrate how the gender wage discrimination effect contributed by selected observable characteristics (age, education, marital status, occupation, ownership and industry) changes with each percentile.[22] In Figure 3, we can see the changing nature that cannot be revealed by mean decomposition. How the advantage of men and disadvantage of women are affected by each covariate at each percentile is also displayed. Note that each covariate's contribution to the discrimination effect is equal to the sum of its impacts on men's advantage and women's disadvantage at each percentile.

We focus on the effects of age first. From the first percentile to the median, we can see that the discrimination effect by age is negative, indicating that male migrants rather than females are prejudiced against in terms of age. The extent of discrimination against male migrants by age decreases as we move from lower percentiles to the median. However, the picture completely changes after reaching the 50th percentile. It is female migrants who are discriminated against from the median to upper percentiles and the magnitude of discrimination effect becomes increasingly larger. From the mean decomposition, we only know the discrimination in favor of male migrants and against females. Mean decomposition

[21]Note that the contribution of *all* control variables to men's advantage and women's disadvantage at each percentile should be the same. However, for each labor market characteristics, its contribution to men's advantage and women's disadvantage can differ. For some covariate, its contribution to men's advantage can be positive. This actually shows how men are underpaid relative to the reference wage structure in terms of that variable

[22]Hughes and Maurer-Fazio (2002) investigates how marital status, education and occupation choices affect the mean gender wage gap among urban residents in China.

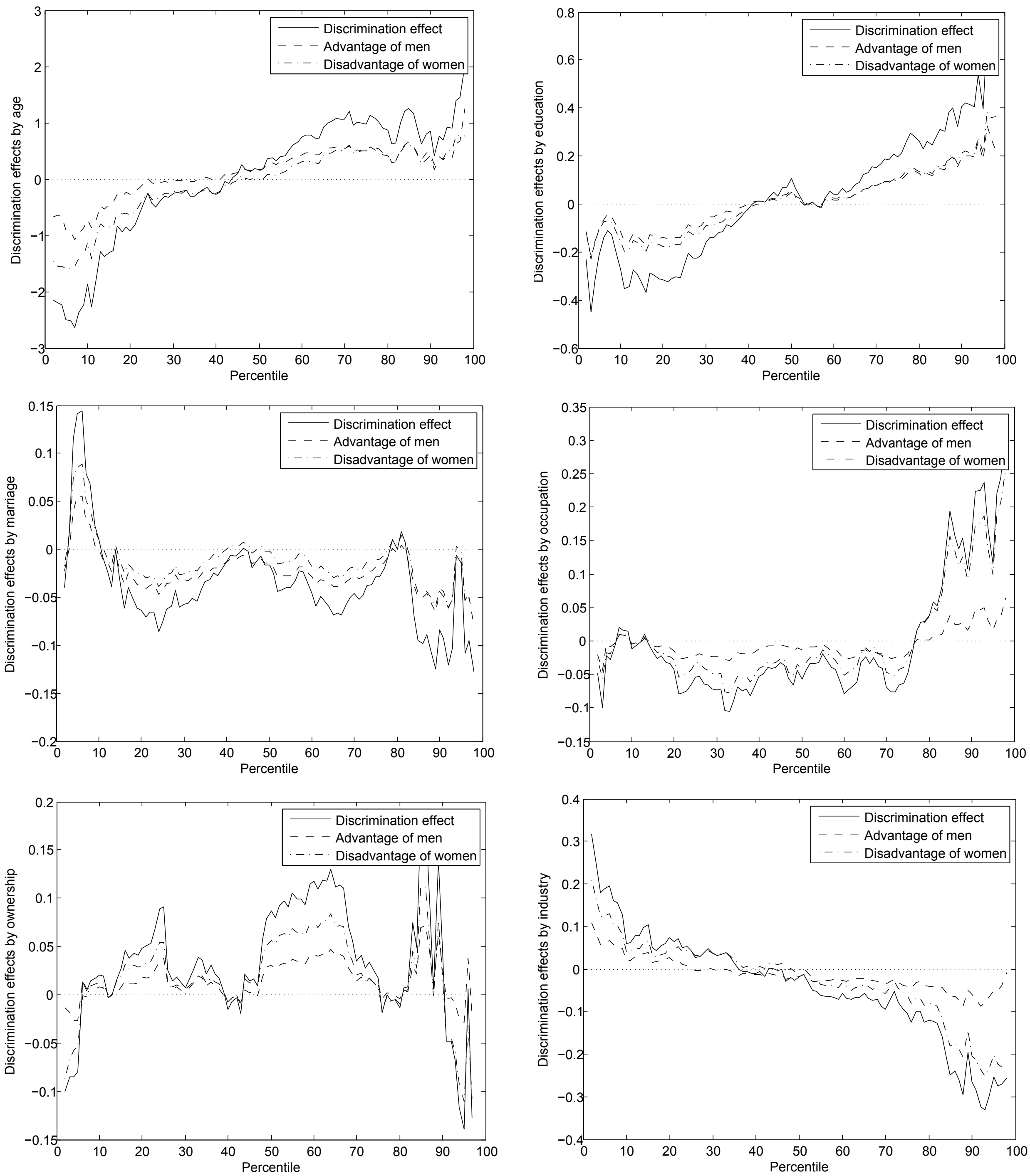

Figure 3: Discrimination effects by selected covariates throughout the wage distributions

fails to reveal the changing nature of the effects. A similar pattern is also observed for education, though the magnitude of its impact is relatively smaller. In addition, education's contribution to men's advantage and women's disadvantage are very similar throughout the wage distribution.

Migrants' marital status also affect the discrimination effect. Women are discriminated against in terms of marital status only at very low percentiles. The picture is different for occupation. We see discrimination in favor of women from lower percentiles to the 80th percentile and the magnitude of discrimination effect stays relatively stable. From the 8th decile to upper percentiles, discrimination effect against females becomes positive and the magnitude is becoming increasingly larger, which also helps to explain why the wage gap and discrimination effect are much higher at top percentiles. We also find that at higher percentiles, the discrimination effect is largely contributed by women's disadvantage rather than men's advantage.

The impact of employer's ownership helps decrease the overall discrimination against female migrants only at very low and very high percentiles. It contributes to the discrimination against females most from the 50th to 70th percentile. The impact becomes more varied from the 7th decile to the higher end of wage distribution. In addition, industry and education have reversed roles in contributing to discrimination effect at each percentile. We see discrimination effect against women and in favor of men and the impact diminishes as we gradually move from lower percentiles to the median. From the median to upper percentiles, the discrimination becomes more and more negative, indicating increasing unfavorable treatment on males in returns to industry dummies. Throughout the wage distribution, the disadvantage of women accounts for the majority of the varied discrimination effect by industry.

4.4 Sensitivity Analysis

4.4.1 Model Specification and the Overlapping in Covariates Distributions

As indicated in previous sections, the approach we use rely heavily on linear specifications. To test the model specification of the OLS regressions in Section 5.1, we apply the widely-used Regression Error Specification Test (RESET) by Ramsey (1969). The p values for the male and female regressions are respectively 0.234 and 0.149, which provides no evidence of functional form misspecification. For the newly-developed RIF-OLS regressions, a linear form is typically assumed and we know of no test to test their specifications. As shown in Firpo et al. (2009), the coefficient estimates from RIF-OLS estimation are very similar when a nonlinear specification such as RIF-Logit is used.

In the following part of this section, we check the overlapping in covariates distributions for both genders. The significance of checking the common support is twofold: First, the decomposition using linear specification fails to recognize the gender differences in the supports by estimating wages equations for all working male and female migrants without restricting the comparison only to those individuals with comparable characteristics. If there is not considerable overlapping in individual characteristics between the two groups, the gap attributable to discrimination effect could be overestimated (Nopo, 2008). Second, if the covariates distributions for the two groups are very different, then the $\widehat{\gamma}$ and $\widehat{\gamma}_\tau$ estimated from pooled regressions can be sensitive to changes in the linear functional form we use (Imbens and Rubin, 2009; Imbens and Wooldridge, 2009).

One way to assess the overlapping is proposed in Imbens and Rubin (2009), which suggests calculating a scale-free measure of the difference in distributions for each covariate: the normalized difference. It is defined as $\triangle X = \frac{\overline{X}_m - \overline{X}_f}{\sqrt{S_m^2 + S_f^2}}$, where S_m^2 and S_f^2 are the sample variances of X for the two groups.[23] A rule suggested by them is that when the normalized differences for many of the control variables are greater than 0.25 in magnitude, linear regression results tend to be sensitive to changes in model specification. The normalized difference for each variable we use is reported in the last column of Table 1 and Table 2 and none of them exceeds one quarter. Following the program evaluation literature (Imbens and Wooldridge, 2009), we also estimate a logit model with gender as the dependent variable, using the explanatory variables in the OLS regressions as covariates. The overlap problem is assessed by inspecting the histograms of estimated propensity scores for both genders. Considerable regions of overlapping are found and the common support seems not a concern for us.[24]

4.4.2 Decomposition Results Excluding the Effects of Occupation, Ownership and Industry

Including the sets of occupation, ownership and industry dummies may bias our estimation results since these variables are likely to be jointly determined with wages, although as an accounting exercise they are useful in explaining wage differentials (Albrecht et al., 2003). In this section, we decompose the wage differentials across distribution using those basic controls (explanatory variables excluding occupation, ownership and industry dummies). The results are shown in Table 10.

Unlike the findings in Gustafsson and Li (2000) and Zhang et al. (2008) that excluding

[23]The normalized difference $\triangle X$ is different from the often reported t-statistic for the null hypothesis of equal means defined by $T = \frac{\overline{X}_m - \overline{X}_f}{\sqrt{S_m^2/N_1 + S_f^2/N_2}}$. The reason for focusing on $\triangle X$, instead of on T, as a measure of the degree of difficulty in adjusting for differences in covariates, comes from their relation to the sample size. Simply increasing the sample size does not make the problem of inference for the average treatment effect inherently more difficult. Increasing the sample size does not systematically affect the normalized difference $\triangle X$ but will lead to an increase in the t-statistic (Imbens and Rubin, 2009).

[24]This also explains why the discrimination effect contributes more to the wage gap than the endowment effect throughout the wage distribution. When the observable characteristics are similar for both genders, the wage gap attributable to the differences in covariates cannot be very large.

those arguably endogenous variables yields similar decomposition results for urban workers in China, the proportion of gender wage gap for rural migrants that can be explained by differences in covariates at each decile decreases dramatically. At the 10th percentile, less than 10% of the wage gap can be explained by endowment effects and over 90% can be attributed to discrimination. The highest explained proportion is at the 90th percentile, but it is still very small (23.7%). This strengthens our findings that the discrimination effect contributes more to the wage gap at every decile and the gender wage discrimination problem is most serious at the lower end of rural migrants' wage distribution. By comparing the results in Table 7 and 10, the occupation, ownership and industry dummies can explain almost the same percentage of wage gap as that can be explained by the differences in basic control variables. This confirms the findings in Albrecht et al. (2003) that these variables are useful in explaining wage differentials.

Table 10: Decomposition Results Using Basic Control Variables Only

Selected Quantiles	Raw Gender Log Wage Gap	Endowment Effects	Discrimination Effects	Approximation Errors
10th	0.288 (100%)	0.027 (9.38%)	0.261 (90.6%)	0.000 (0.00%)
20th	0.251 (100%)	0.041 (16.3%)	0.211 (84.1%)	-0.001 (-0.40%)
30th	0.182 (100%)	0.036 (19.8%)	0.153 (84.1%)	-0.008 (-4.40%)
40th	0.201 (100%)	0.041 (20.4%)	0.162 (80.6%)	-0.002 (-1.00%)
50th	0.292 (100%)	0.043 (14.7%)	0.247 (84.6%)	0.002 (0.68%)
60th	0.288 (100%)	0.053 (18.4%)	0.232 (80.6%)	0.003 (1.04%)
70th	0.288 (100%)	0.065 (22.6%)	0.227 (78.8%)	-0.004 (-1.39%)
80th	0.324 (100%)	0.072 (22.2%)	0.249 (76.9%)	0.003 (0.93%)
90th	0.426 (100%)	0.101 (23.7%)	0.328 (77.0%)	-0.003 (-0.70%)

Note: The ratios of effects to their corresponding log wage gaps are reported in parentheses. All decomposition results reported are rounded to three digits after decimal.

4.4.3 Decomposition Using An Alternative Reference Wage Structure

To address the sensitiveness of decomposition results to the choice of base group in Oaxaca-Blinder decomposition, Cotton (1988) suggested weighting $\widehat{\beta}_m$ and $\widehat{\beta}_f$ by the group sizes

N_m and N_f to obtain a reference wage structure, which is $\widehat{\beta}_r = \frac{N_m}{N_m+N_f}\widehat{\beta}_m + \frac{N_f}{N_m+N_f}\widehat{\beta}_f$.[25] Similarly, to check the robustness of our quantile decomposition results, we obtain an alternative reference wage structure by weighting coefficient estimates obtained from RIF-OLS regressions for both genders by group sizes at the τ-quantile. Then equation (10) is now changed to $\widehat{q}_{m\tau} - \widehat{q}_{f\tau} = \Delta\overline{X}\widehat{\beta}_{r\tau} + (\frac{N_f}{N_m+N_f}\overline{X}_m + \frac{N_m}{N_m+N_f}\overline{X}_f)(\widehat{\beta}_{m\tau} - \widehat{\beta}_{f\tau}) + \widehat{R}_\tau$.

Table 11: Decomposition Using An Alternative Reference Wage Structure

Selected Quantiles	Raw Gender Log Wage Gap	Endowment Effects	Discrimination Effects	Approximation Errors
10th	0.288 (100%)	0.059 (20.5%)	0.230 (79.9%)	-0.001 (-0.38%)
20th	0.251 (100%)	0.091 (36.3%)	0.161 (64.1%)	-0.001 (-0.40%)
30th	0.182 (100%)	0.090 (49.5%)	0.099 (54.4%)	-0.007 (-3.85%)
40th	0.201 (100%)	0.088 (43.8%)	0.115 (57.2%)	-0.002 (-1.00%)
50th	0.292 (100%)	0.096 (32.9%)	0.194 (66.4%)	0.002 (0.68%)
60th	0.288 (100%)	0.103 (35.8%)	0.182 (63.2%)	0.003 (1.04%)
70th	0.288 (100%)	0.112 (38.8%)	0.180 (62.5%)	-0.004 (-1.39%)
80th	0.324 (100%)	0.143 (44.1%)	0.177 (54.6%)	0.004 (1.23%)
90th	0.426 (100%)	0.175 (41.1%)	0.254 (59.6%)	-0.003 (-0.70%)

Note: The ratios of effects to their corresponding log wage gaps are reported in parentheses. All decomposition results reported are rounded to three digits after decimal.

Similar to the explanation in Section 4.2.2, on the right-hand side of the equation, the first term measures the endowment effect, the second term is the unexplained discrimination effect and the last term is the approximation error. The endowment and discrimination effects corresponding to each decile obtained using the above procedure are displayed in Table 11. The results are similar to those shown in Table 7. The decomposition results show that gender wage gap attributable to discrimination effect is most serious among low wage rural migrants.

[25] Then the mean gender wage gap becomes: $\overline{lnY}_m - \overline{lnY}_f = \Delta\overline{X}\widehat{\beta}_r + (\frac{N_f}{N_m+N_f}\overline{X}_m + \frac{N_m}{N_m+N_f}\overline{X}_f)(\widehat{\beta}_m - \widehat{\beta}_f)$. Note that with this reference wage structure, men's advantage and women's disadvantage are respectively $\overline{X}_m(\widehat{\beta}_m - \widehat{\beta}_r)$ and $-\overline{X}_f(\widehat{\beta}_f - \widehat{\beta}_r)$, but they are not equal to each other. Reimers (1983) proposed to use $\widehat{\beta}_r$=0.5$\widehat{\beta}_m$+0.5$\widehat{\beta}_f$ as the nondiscriminatory reference. Using this wage structure and the one proposed by Cotton (1988) will not lead to big differences in our decomposition results. The reason is that in our sample $\frac{N_m}{N_m+N_f}$ equals 0.568 and $\frac{N_f}{N_m+N_f}$ equals 0.432, which are not too different from 0.5.

4.4.4 Discussion on the Sample Selection Problem

In our analysis, we drop those observations with no earnings in 2002. This can bias our estimates and decomposition results if the sample of employed migrants is systematically different to those unemployed ones. However, we choose to ignore this problem. The first reason is that only 63 migrants reported no earnings in 2002 (16 males and 47 females), who account for a very small proportion of the sample (1.9%). Second, given the data limitation, we cannot find at lease one valid instrumental variable that affects the job participation but exerts no impact on earnings. Thus, we cannot test the sample selection hypothesis or address the problem. Third, even if we find a valid instrument, it is still technically infeasible to *both* address the selection problem and decompose wage gap across the distribution into the contribution of *each* covariate using quantile regressions.[26] Finding how each covariate contributes to gender earnings gap is of more interest to us. As a result, we follow the previous literature (see, for example, Bishop et al., 2005; Chi and Li, 2008) and ignore this problem.

5 Conclusion

Using a nationally representative cross-sectional data set on Chinese rural migrants, we investigate whether female rural migrants are treated equally to male migrants in China's urban labor market. We analyze the wage determination of both genders using OLS and unconditional quantile regressions. We also untangle the gender wage differentials across the wage distributions by applying decomposition methods.

We find that OLS regressions cannot provide an adequate description of wage determination for both genders. We estimate unconditional quantile regressions for each gender and find substantial differences in the coefficients on the labor market characteristics at various quantiles of the wage distributions for both genders. For example, from OLS regression, the return to one more year of education is 4.1% for both male and female migrants. However, this return increases significantly from 1.5% at the 10th percentile to 4.4% at the median and 10% at the 90th percentile for male migrants, while it remains relatively stable for females (ranging from 3.1% to 5.0%). We also find that, as percentile increases with male migrants' wage distribution, there is more heterogeneity in the coefficient estimates of both occupation and industry dummies, indicating that the choices of occupation and industry can explain more variations in wage levels among high income migrants. Nonetheless, the differences in returns to industry dummies are the largest at the lower end of females' wage distribution.

[26]The only method we know of that can address the sample selection problem in quantile decomposition is developed by Albrecht et al. (2009). However, that method is computationally intensive and will disable us from finding the impact of each covariate on wage gap.

The gender wage gap is found to be substantial among rural migrants and it is not uniform across the wage distribution. It becomes increasingly higher at the higher end. The finding is different from the gender wage gap structure previously found for urban residents in Bishop et al. (2005) and Chi and Li (2008), although the two groups both work in urban China. We further decompose the gender wage gap across the distribution into a part explained by differences in productive characteristics (endowment effect) and another part explained by different returns to explanatory variables (discrimination effect). We find that the discrimination effect contributes more to the wage gap than the endowment effect across the wage distribution. While gender wage differentials are larger at higher percentiles, the gender wage discrimination problem seems most serious among low wage rural migrants, which is evident from the ratios of discrimination effects to their corresponding wage gaps.

Furthermore, the endowment effect and discrimination effect are decomposed into the contribution of each covariate. Education, urban experience and industry differences are the three most important factors contributing to the endowment effect at the 10th, 50th and 90th percentile. The effects of education and industry rise drastically from the 10th percentile to 90th percentile with their effects at the higher end more than doubled the corresponding effects at the 50th percentile. In terms of discrimination effect, female migrants are discriminated against in experience in city, whether having a permanent job or not and the ownership of employer at all the 10th, 50th and 90th percentiles of wage distribution. We find a more varied impact of other variables across the wage distribution and this shows that mean decomposition can not provide an adequate description of the picture. For example, at the lower decile, female migrants actually enjoy a advantage in returns to age over males. However, age becomes a factor contributing more and more discrimination against females as percentile increases. The effects of education share the same patterns as those of age, although the magnitude is relatively smaller. Women's occupation choices lower the discrimination effect at the 10th and 50th percentile before increasing it at the 90th percentile, although its mean impact on unexplained effects is negative. From the mean wage gap decomposition, we know that the different returns to industry dummies for both genders lower the discrimination effect. However, they contribute positively to the unexplained effect at the 10th percentile before lowering it at the median and the 90th percentile.

In conclusion, we find a large gender wage gap among the rural migrants in China's urban labor market. With the increase in percentiles, the wage differential is also on the rise with an acceleration observed from 70th to upper percentiles. Discrimination effect contributes to over 50% of the gap across the wage distribution and the extent of gender wage discrimination is the largest at the lower end. Labor market characteristics play a different role in contributing to the endowment effect and discrimination effect from lower percentile to the higher end. These are the findings that were not unveiled in previous literature. While migrant workers usually suffer from discrimination when compared to their urban resident counterparts (Meng and Zhang, 2001), special attention should be paid by the policymakers to the fact that female migrants are facing double whammy in China's urban labor market. With China being in the process of building a harmonious society and fostering a competitive labor market, more efforts should be directed to ensure equal pay for equal job. One limitation of this study is that we only have data at a single point in time. It is also of interest for us to provide a moving picture of the gender wage gap among rural migrants and show its changing pattern during the economic transition of China. A second limitation is that we fail to take the family issues into consideration. From the data set, we cannot identify whether the spouses of these migrants come to work in urban areas or not, and we cannot investigate the impact of their children on wage levels. We leave these things for future research.

References

Albrecht, J., A. Bjorklund, and S. Vroman (2003). "Is there a glass ceiling in Sweden?". *Journal of Labor Economics 21*, 145–177.

Albrecht, J., A. van Vuuren, and S. Vroman (2009). "Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands". *Labour Economics 16*, 383–396.

Bishop, J. A., F. Luo, and F. Wang (2005). "Economic transition, gender bias, and the distribution of earnings in China". *Economics of Transition 13*, 239–259.

Blinder, S. A. (1973). "Wage discrimination: Reduced form and structual estimates". *Journal of Human Resources 8*, 436–455.

Cai, F. (1996). "An economic analysis for labour migration and mobility". *Social Sciences in China 17*, 120–135.

Chi, W. and B. Li (2008). "Glass ceiling or sticky floor? Examining the gender earnings differential across the earnings distribution in urban China, 1987-2004". *Journal of Comparative Economics 36*, 243–263.

Cotton, J. (1988). "On the decomposition of wage differentials". *Review of Economics and Statistics 70*, 236–243.

Demurger, S., M. Gurgand, S. Li, and X. Yue (2009). "Migrants as second-class workers in urban China? A decomposition analysis". *Journal of Comparative Economics 37*, 610–628.

DiNardo, J., M. N. Fortin, and T. Lemieux (1996). "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach". *Econometrica 64*, 1001–1044.

Firpo, S., M. N. Fortin, and T. Lemieux (2007). "Decomposing wage distributions using recentered influence regressions". Working paper, Department of Economics, University of British Columbia.

Firpo, S., M. N. Fortin, and T. Lemieux (2009). "Unconditional quantile regressions". *Econometrica 77*, 953–973.

Fortin, M. N. (2008). "The gender wage gap among young adults in the United States: the importance of money versus people". *Journal of Human Resources 43*, 886–929.

Fortin, M. N., T. Lemieux, and S. Firpo (2010). "Decomposition methods in economics". In D. Card and O. Ashenfelter (Eds.), *Handbook of Labor Economics, Vol 4, forthcoming.*

Gagnon, J., T. Xenogiani, and C. Xing (2009). "Are all migrants really worse off in urban labour markets: New empirical evidence from China". *OECD Development*

Center Working Paper No. 278.

Gardeazabal, J. and A. Ugidos (2004). "More on identification in detailed wage decompositions". *Review of Economics and Statistics 86*, 1034–1036.

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

Hare, D. (1999). "'Push' versus 'Pull' factors in migration outflows and returns: Determinants of migration status and spell duration among China's rural population". *Journal of Development Studies 35*, 126–142.

Hare, D. (2002). "The determinants of job location and its effects on migrants' wage: Evidence from China". *Economic Development and Cultural Change 50*, 557–579.

Hughes, J. and M. Maurer-Fazio (2002). "Effects of marriage, education and occupation on the female/male wage gap in China". *Pacific Economic Review 7*, 137–156.

Imbens, G. W. and D. B. Rubin (2009). *"Causal Inference in Statistics, and in the Social and Biomedical Sciences"*. Cambridge University Press, New York. To appear.

Imbens, G. W. and J. M. Wooldridge (2009). "Recent developments in the econometrics of program evaluation". *Journal of Economic Literature 47*, 5–86.

Jann, B. (2008). "The Blinder-Oaxaca decomposition for linear regression models". *Stata Journal 8*, 453–479.

Juhn, C., K. Murphy, and B. Pierce (1993). "Wage inequality and the rise in returns to skill". *Journal of Political Economy 101*, 410–422.

Khan, A. R. and C. Riskin (2005). "China's household income and its distribution, 1995 and 2002". *The China Quarterly 182*, 356–384.

Koenker, R. and G. Bassett (1978). "Regression quantiles". *Econometrica 46*, 33–50.

Li, H. (2003). "Economic transition and returns to education in China". *Economics of Education Review 22*, 317–328.

Liang, Z. (2001). "The age of migration in China". *Population and Development Review 27*, 499–524.

Liu, Z. (2005). "Institution and inequality: the hukou system in China". *Journal of Comparative Economics 33*, 133–157.

Liu, Z. (2008). "Human capital externalities and rural-urban migration: Evidence from rural China". *China Economic Review 19*, 521–535.

Lu, Z. and S. Song (2006). "Rural-urban migrantion and wage determination: the case of Tianjin, China". *China Economic Review 17*, 337–345.

Machado, J. and J. Mata (2005). “Counterfactual decomposition of changes in wage distributions using quantile regression”. *Journal of Applied Econometrics 20*, 445–465.

Melly, B. (2005). “Decomposition of differences in distribution using quantile regression”. *Labour Economics 12*, 577–590.

Meng, X. (1998). “Gender occupational segregation and its impact on the gender wage differential among rural-urban migrants: A Chinese case study”. *Applied Economics 30*, 741–752.

Meng, X. (2001). “The informal sector and rural-urban migrantion: A Chinese case study ”. *Asian Economic Journal 15*, 71–89.

Meng, X. and J. Zhang (2001). “The two-tier labor market in urban China”. *Journal of Comparative Economics 29*, 485–504.

Nandi, A. and C. Nicoletti (2009). “Explaining personality pay gaps in the UK”. ISER Working Paper, University of Essex.

Nopo, H. (2008). “Matching as a tool to decompose wage gaps”. *Review of Economics and Statistics 90*, 290–299.

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

Oaxaca, L. R. and M. Ransom (1994). “On discrimination and the decomposition of wage differentials”. *Journal of Econometrics 61*, 5–21.

Poncet, S. (2006). “Provincial migration dynamics in China: Borders, costs and economic motivations”. *Regional Science and Urban Economics 36*, 385–398.

Ramsey, J. (1969). “Tests for specification errors in classical linear least squares regression analysis”. *Journal of the Royal Statistical Society, Series B 31*, 350–371.

Reimers, C. (1983). “Labour market discrimination against Hispanic and black men”. *Review of Economics and Statistics 65*, 570–579.

von Mises, R. (1947). “On the asymptotic distribution of differentiable statistical functions”. *The Annals of Mathematical Statistics 18*, 309–348.

Wang, F. and X. Zuo (1999). “Inside China’s cities: institutional barriers and opportunities for urban migrants”. *American Economic Review, Papers and Proceedings 89*, 276–280.

Wang, M. and F. Cai (2008). “Gender earnings differential in urban China”. *Review of Development Economics 12*, 442–454.

Yang, D. T. (2005). “Determinants of schooling returns during transition: Evidence from Chinese cities”. *Journal of Comparative Economics 33*, 244–264.

Zhang, H. and S. Song (2003). "Rural-urban migration and urbanization in China: Evidence from time-series and cross-section analysis". *China Economic Review 14*, 386–400.

Zhang, J., J. Han, P.-W. Liu, and Y. Zhao (2008). "Trends in the gender earnings differential in urban China, 1988-2004". *Industrial and Labour Relations Review 62*, 223–243.

Zhang, J., Y. Zhao, A. Park, and X. Song (2005). "Economic returns to schooling in urban China, 1988-2001". *Journal of Comparative Economics 33*, 730–752.

Zhao, Y. (1997). "Labour migration and returns to rural education in China". *American Journal of Agricultural Economics 79*, 1278–1287.

Zhao, Y. (1999). "Leaving the countryside: Rural-urban migration decision in China". *American Economic Review, Papers and Proceedings 89*, 281–286.

Zhao, Z. (2005). "Migration, labour market flexibility, and wage determination in China: a review". *The Developing Economies 44*, 285–312.

Zhu, N. (2002). "The impacts of income gaps on migration decisions in China". *China Economic Review 13*, 213–230.