International Journal of China Studies
Vol. 8, No. 1, April 2017, pp. 61-91

Behind the Invisible Wall: What Determine Wage Differentials between Urban and Migrant Workers in China

Yih-chyi ***Chuang**** *and Eric* ***Yan*****
Dept. of Economics, National Chengchi University, Taipei, Taiwan, ROC
Institute of Economics, Academia Sinica, Taipei, Taiwan, ROC

Abstract

Using a wider scope of cities data from 2008 survey of Rural-Urban Migration in China, this study employs a comprehensive aspect of explanatory variables to empirically estimate wage determination and decomposes the wage differentials between urban and migrant workers in the Chinese labour market. We find that differences in endowments, such as personal traits, geography, cohort, firm characteristics and industry type explain 85-89% of the wage differentials; however, it drops significantly to 42-60% if group membership, a likely proxy for the Hukou system, is considered. Among those explanatory factors, human capital proxies of personal traits are the crucial factors for wage differentials; moreover, compared to the urban workers the education resource-poor migrants have higher rates of return on human capital variables of work experience, height and health. The significant age cohort effect reflects better job opportunity and labour quality of new generations of migrants. Policy implications for institutional change to close the wage gap are also discussed.

Keywords: *Wage differentials, migrant workers, Hukou system, rural-urban income gap, decomposition method, human capital*

1. Introduction

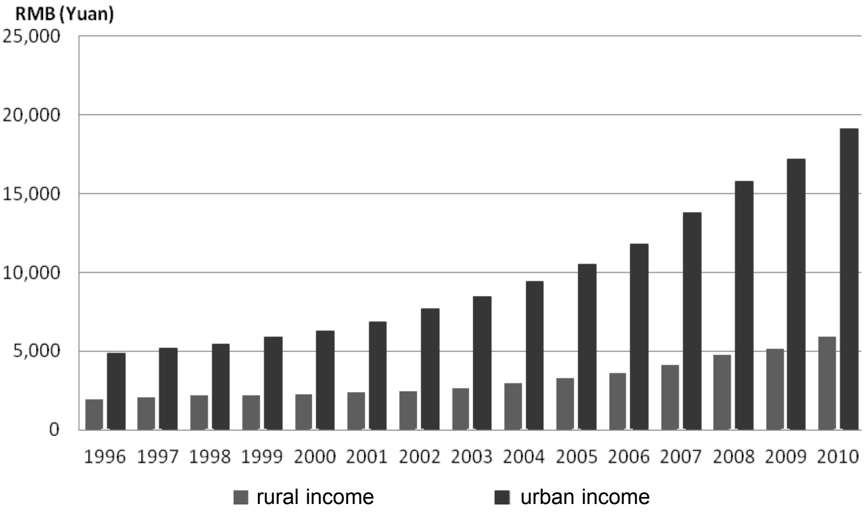
After 1978 economic reform, China experienced three decades of fast economic growth with an average annual growth rate of 9.7%. In this period, both the agriculture and industry sectors underwent rapid transformation. In 1958, in order to manage labour under the collective farm community arrangement, the implementation of a household registration system (Hukou) officially identifies a person as a resident of a city to control the movement of people between urban and rural areas. This Hukou system is considered as the major institutional arrangement that controls and discriminates migrants from urban workers in China, see for example, Wang (2005) and Chan and Buckingham (2008).

The economic reform in 1978 relaxed restrictions and regulations for rural and urban migration by allowing the transfer of surplus labour in agriculture sector to industry sector especially those located in the coastal area of China for speeding up the process of industrialization. According to statistics from the National Bureau of Statistics of China, in 1978 per capita output in primary industry was only RMB363 – about 14.44% of that in secondary industry. However, in 1990 it had increased to RMB1,301 – about 23.36% of secondary industry, while in 2010 it further jumped to RMB14,512 – a more than ten-fold increase in twenty years but its ratio with secondary industry dropped to 16.90%. Moreover, between 1980 and 2010, the share of non-agriculture employment in the agriculture sector amplified from 9.32% to 48.29%, implying more and more rural labour leave the low-productivity agriculture sector to high-productivity non-farm activities. This agriculture-industry transformation was also revealed in the employment share by sector, in 1978 the employment share of primary industry was 70.53% which then declined to 60.10% in 1990 after the opening up of special economic zones in Shenzhen, Zhuhai, Shantou and Xiamen along the southern coastal area since 1980. As a result, the employment share of secondary and tertiary sectors climbed to 21.40% and 18.50% in 1990, respectively. By 2010, the employment share of primary, secondary and tertiary industries reached 36.70%, 28.70%, and 34.60%, respectively. This shows that even GDP share of the primary sector had decreased from 28.19% in 1978 to 10.10% in 2010 under rapid industrialization, but the agriculture sector still maintained a high proportion of the labour force and population. Apparently, urbanization had not kept up with the process of industrialization because of different institutional and political arrangements between rural and urban areas in China (Chapter 5 in Naughton 2007).

More importantly, the Hukou system still remains for workers' identification to keep wages of rural migrants in the city from rising so that a large cheap army of floating population can be used in the urban industry sector.[1] The population and labour policy reform in 1978 focused on three aspects: first, change from collective farm community to household responsibility system in agriculture production;[2] second, loosening labour mobility control to allow rural migrants to work in urban cities or manufacturing while maintaining the Hukou system; third, promote one-child policy in the urban area.[3] These labour policies have profound effects on the process of industrialization and demographic structure change in China. During the period 2001-2011, the rate of urbanization increased from 37.66% to 51.27%, while employment share in secondary and tertiary industries rose to over 60%, higher than the rate of urbanization. This decoupling effect between industrialization and urbanization was mainly due to the Hukou system that restricted labour mobility between rural and urban sectors. In 2012, China has a population of 1.37 billion people, and half of them lived in urban areas with a share of only 20% of permanent residents. With a large group of migrant workers living in cities, what happen to their wages relative to that of urban workers? What are the advantages and disadvantages determining migrant workers' wage compensation? Have migrants been discriminated while working in cities? These are important research questions for labour policy on further structural transformation in the Chinese economy as they affect the living standard and income of migrants and income disparity between rural and urban sectors.[4]

Over the past decades, there has been a problem of widening income gap in many economies in the world. In most literature, the cause of rising wage inequality may be related to trade that helps to spread technology, workers' level of human capital, workers' proficiency in applying technology for production, and discrimination towards workers with different background.[5] Undergoing three decades of fast growth since 1978, China has also encountered the problem of widening income gap, which can be observed from the diverging gap of per capita income between urban and rural residents in China. As shown in Figure 1, the income ratio of rural residents with respect to urban residents dropped significantly from 40% in 1996 to 31% in

Figure 1 Real Urban and Rural Per Capita Income, 1996-2010

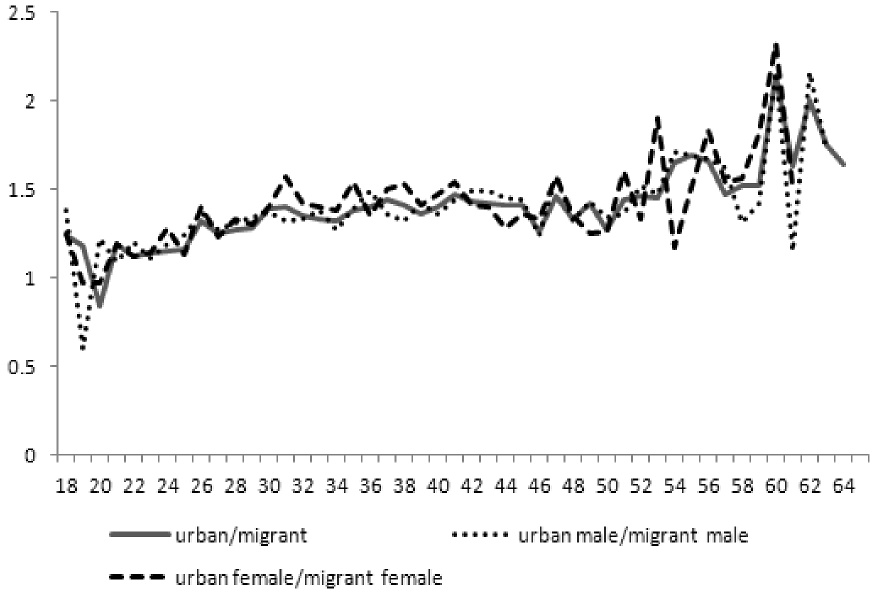


Source: National Bureau of Statistics of China (2011).

2010, implying that even with an increasing trend of rural residents' income the rural-urban income gap kept widening over time.

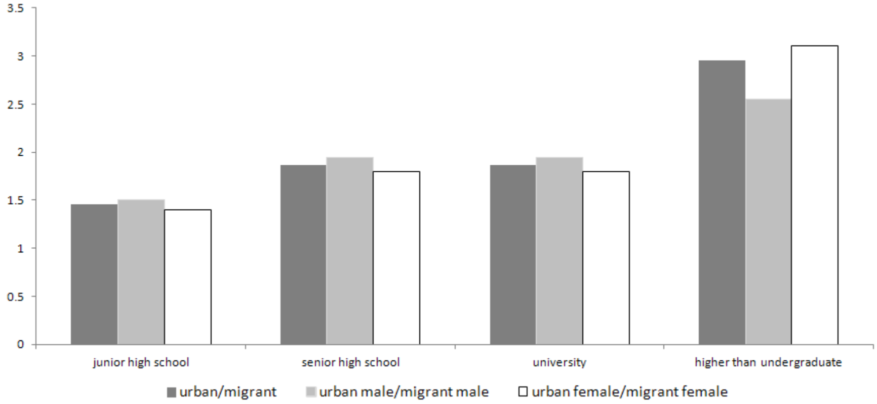
Among the aforementioned causes of income gap, discrimination has always been a rising focus to people who are concerned with the Chinese labour market. Some recent empirical studies on the Chinese labour market have found that women are paid lower than other groups (e.g. Rozelle *et al.*, 2002; *Liu et al.*, 2000), while others suggest that there is significant discrimination towards migrants in identity, occupation and industry segregation (e.g. Meng and Zhang, 2001; Lee, 2012). Using the 2005 China Urban Labor Survey data from five cities, Shanghai, Wuhan, Shenyang, Fuzhou, and Xi'an, Lee (2012) found 34% and 22% of wage and non-wage differences were unexplained for male and female migrants respectively. Zhang *et al.* (2016) used the China Household Income Project (CHIP) 2007 data and found that migrants only earned 49% of urban workers' income and 17% of the wage gap cannot be explained by observed factors. In detail, differences in educational attainment, work experience and distribution across industry, occupation and ownership of enterprises account for most of the explained wage gap. A coarse observation on wage differential between urban and migrant workers is shown in Figures 2 and 3. Accordingly, hourly

Figure 2 Ratio of Urban Log Hourly Wage to Migrant Log Hourly Wage, by Age



Source: National Survey Research Center at Renmin University of China (2008).

Figure 3 Ratio of Urban Log Hourly Wage to Migrant Log Hourly Wage, by Education Level



Source: National Survey Research Center at Renmin University of China (2008).

wage ratios of urban to migrant workers in China are in general greater than 1 in 2008, whether male or female; and the ratios are quite close to a constant except for the widening dispersion of wage differential in the younger and elder groups and the group with education level higher than university.

Existing literature on research of wage differentials between migrants and urban workers finds that migrant workers work more hours and receive less pay than urban natives, see, for example, Meng and Zhang (2001), Knight and Song (2003), Demurger *et al.* (2009), Deng and Li (2010), Magnani and Zhu (2012) and Meng (2012). They indicated that wage gaps can only be partially explained by differences in the work-related characteristics and mostly be attributed to the divergent returns to endowments and institutional factors in China. These studies on rural-urban wage differentials covered samples from small groups of cities and were restricted to a small set of explained variables in wage determination. Moreover, only recent works by Lee (2012) and Zhang *et al.* (2016) adjusted for sample selection, which may arise due to employment and occupational choice of migrant and urban workers. However, Lee's (2012) estimated correction term was insignificant.

The aim of this study is to investigate the discrimination towards migrants in China with a wider scope of coverage of cities by using the data from the 2008 Rural-Urban Migration in China (RUMiC) Survey that includes more variables such as personal traits like gender, education, work experience, health, cohort, geography, firm characteristics and industry type. Our major contributions are that wage determination regression takes into account of a wider scope of cities and a comprehensive aspect of explanatory variables, and decomposition of wage differentials between migrants and urban workers confirms that differences in personal traits attributed to human capital variables account for a large proportion of explained part for the wage differentials. However, aside from their lack of education resources migrants incline to have higher return on work experience and health than the urban workers. The consideration of group membership, a proxy for the Hukou system, significantly reduces the explained part of wage differentials. Finally, we offer policy implications for future reform to improve wage inequality between migrants and urban workers.

2. Empirical Model

Our empirical estimation model consists of two parts. The first part uses Heckman two-stage regression model to estimate wage determination for migrants and urban workers, respectively. The second part uses the estimated coefficients obtained from wage regressions to decompose the wage differential between urban and migrant employees through a modified decomposition method of Oaxaca (1973) and Blinder (1973) approaches.

2.1. Wage Determination

Heckman's (1979, 1998) two-stage regression model is used for estimating the wage rate. The following briefly introduces the methodology of Heckman test.

Heckman test consists of two stages. The first stage model estimates the probability of an urban native or migrant being employed, whereby it gives the inverse Mills ratios ($\hat{\lambda}_i$) to correct the selection bias of sampling an employed urban (or migrant) in the second stage wage equation. Thus, the first stage Probit model for being employed or not can be expressed as:

$$\mathrm{P}(\mathrm{z}_\mathrm{i} = 1) = \mathrm{P}\,(\mathrm{z}_\mathrm{i}^{*} > 0) = X_i'{}^{2}{}_{\scriptscriptstyle 1} + \mathrm{v}_\mathrm{i}, \mathrm{i} = \mathrm{u}, \mathrm{m}, \tag{1}$$

where i represents individual, u for the urban native and m for the migrant, $\mathrm{z}_\mathrm{i}^{*}$ is the latent variable and z_i is the indicator satisfying $\mathrm{z}_\mathrm{i} = 1$ if $\mathrm{z}_\mathrm{i}^{*} > 0$ and $\mathrm{z}_\mathrm{i} = 0$ if $\mathrm{z}_\mathrm{i}^{*} \leq 0$, X_i represents the vector of explanatory variables for being employed, and v_i is the error term. The set of explanatory variables includes age, health, self-confidence, years of education, child rearing, gender, geography and age cohort.

We calculate the inverse Mills ratios ($\hat{\lambda}_i$) from equation (1) and then introduce it into the second stage wage regression denoted as:

$$\ln \mathrm{INC}_\mathrm{i} = Z_i'\alpha_\mathrm{i} + \hat{\lambda}_\mathrm{i}'\gamma_\mathrm{i} + \eta_\mathrm{i}, \mathrm{i} = \mathrm{u}, \mathrm{m}, \tag{2}$$

where ln INC_i is the hourly wage in the logarithmic form and Z_i is a vector of explanatory variables for wage determination. The explanatory variables include personal traits, gender, family background, *age cohort*, geography, firm characteristics and industry type.

It should be noted that inclusion restriction is required for solving the identification problem of equations (1) and (2). That is, equation (1) should contain at least one variable that is not in equation (2). We include extra individual's variables of age and self-confidence in equation (1) for inclusion restriction.

2.2. Wage Differential Decomposition

The mean value of log hourly wage for the urban and migrant workers is denoted as $\ln \overline{INC}_u$ and $\ln \overline{INC}_m$ respectively, and $\bar{Z}_m$ and $\bar{Z}_u$ are the vectors of respective explanatory variables that would influence wage level.

The decomposition approach for wage differential depends on the choice of reference group. Conventionally, when the migrant is used as a reference group, the wage gap can be expressed as:

$$\ln \overline{INC}_u - \ln \overline{INC}_m = \bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m) + (\bar{X}_u' - \bar{X}_m')\hat{\beta}_m, \tag{3}$$

in which $(\bar{X}_u' - \bar{X}_m')\hat{\beta}_m$ is the explained part attributed to the difference of endowments between migrants and the urban using migrants' coefficients and $\bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m)$ is the unexplained part due to difference in the coefficients of the two groups using urban workers' endowments as the reference. Unexplained ratio in this case is defined as

$$U = [\bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m)]/[\bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m) + (\bar{X}_u' - \bar{X}_m')\hat{\beta}_m]. \tag{4}$$

When the urban worker is used as a reference group, the wage gap can be decomposed as:

$$\ln \overline{INC}_u - \ln \overline{INC}_m = \bar{X}_m'(\hat{\beta}_u - \hat{\beta}_m) + (\bar{X}_u' - \bar{X}_m')\hat{\beta}_u, \tag{5}$$

in which $(\bar{X}_u' - \bar{X}_m')\hat{\beta}_u$ is the explained part attributed to the difference of endowments between migrants and the urban using urban workers' coefficients and $\bar{X}_m'(\hat{\beta}_u - \hat{\beta}_m)$ is the unexplained part due to difference in the coefficients of the two groups using migrants' endowments as the reference. Unexplained ratio in this case is defined as:

$$U = [\bar{X}_m'(\hat{\beta}_u - \hat{\beta}_m)] / [\bar{X}_m'(\hat{\beta}_u - \hat{\beta}_m) + (\bar{X}_u' - \bar{X}_m')\hat{\beta}_u]. \tag{6}$$

These two approaches of decomposition provide a range for us to determine the explained and unexplained parts of wage differential.[6]

In many cases, the traditional decomposition method will generate an extraordinarily large unexplained ratio. This, however, does not necessarily mean the wage gap is largely unexplainable. Instead, it might be simply because the denominator of U is too close to zero. To solve the problem, we take the exponential values of $\ln \overline{INC}_u$ and $\ln \overline{INC}_m$ and obtain:[7]

$$\overline{INC}_u - \overline{INC}_m \propto e^{\bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m)} + e^{(\bar{X}_u' - \bar{X}_m')\hat{\beta}_m}, \tag{7}$$

where $e^{\bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}$ is the monotonic transformation of the unexplained part of the original decomposition mentioned above, while $e^{(\bar{X}_u' - \bar{X}_m')\hat{\beta}_m}$ is the monotonic transformation of the explained part of the original decomposition mentioned above. Unexplained ratio of $\overline{INC}_u - \overline{INC}_m$ under such a prerequisite is thus defined as:

$$U = e^{\bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m)} / [e^{(\bar{X}_u'(\hat{\beta}_u - \hat{\beta}_m)} + e^{(\bar{X}_u' - \bar{X}_m')\hat{\beta}_m}]. \tag{8}$$

Likewise, when the urban is used as a reference group,

$$\overline{INC}_u - \overline{INC}_m \propto e^{\bar{X}_m'(\hat{\beta}_u - \hat{\beta}_m)} + e^{(\bar{X}_u' - \bar{X}_m')\hat{\beta}_u}, \tag{9}$$

and its unexplained ratio is defined as

$$U = e^{\bar{X}_m'(\hat{\beta}_u - \hat{\beta}_m)} / [e^{(\bar{X}_m'(\hat{\beta}_u - \hat{\beta}_m)} + e^{(\bar{X}_u' - \bar{X}_m')\hat{\beta}_u}]. \tag{10}$$

Apart from the problem of extraordinary large unexplained ratio discussed above, another issue of the decomposition method is whether or not we should consider the "group membership" that differentiates income at the same bundle of productivity (Jones and Kelley, 1984). As has been observed previously, in the labour market of China, the classification of group membership is pronounced between urban and migrant workers due to the Hukou system. If it is considered in the decomposition, it might outweigh the effect of other explanatory variables being the cause of discrimination towards migrants leading the decomposition to drop its explanatory power. To manifest the influence of group membership on explained ratios, the empirics in the later part will simultaneously consider the cases with and without this group membership factor. That is, both the cases where constant terms of regression results are included or excluded in the decomposition of wage differentials are analyzed.

3. Data

Data used in this paper were compiled from the 2008 Urban-Rural Migration in China (RUMiC) Survey, a longitudinal survey consisting of three parts: the Urban Household Survey, the Rural Household Survey and the Migrant Household Survey. It was initiated by a group of scholars and researchers at the Australian National University, the University of Queensland and the

Beijing Normal University and was supported by the Institute for the Study of Labor (IZA). For urban data, the sample size is 14,683. Among them, 5,790 entries are valid for our empirical study. The sample size for migrant data is 8,446. Among them, 3,257 entries are used for the empirical study.

For labour market participation decision, a set of explanatory variables that would determine if an urban native or migrant is employed. These variables are age, self-evaluation of confidence, health condition, years of education, child rearing, gender, geography and cohort. The rationale to include two additional individual variables, age and self-evaluation of confidence, is for exclusion restrictions required in the Heckman selection equation. The older generation due to aging effect or being influenced by the prolonging socialist education movement could have a quite different value judgment and philosophy of life from those who received modern education system gradually adopted in China since the end of Mao Era in the early 1970s. Hence, age variable replicates not only differences in physical status but also in mental and work attitude. People with diverse degree of self-confidence may not only think but also behave differently in making their career decision. Thus, these two additional variables may properly explain people's employment decision in certain ways.

Data chosen for wage regression are workers who are identified as those with monthly income no less than RMB500 and weekly working hours of more than 30 hours.[8] We also restrict the sample with age below 60 for male and below 55 for female, according to the official retirement age in China. Explanatory variables that would influence hourly income are characterized in five categories: personal traits, geography, age cohort, firm characteristics and industry types. Personal traits include height, years of education, years of work, health condition, child rearing and gender. We additionally include individual's height in wage equation as employer usually pay a height wage premium, see, e.g., Persico, Postlewaite, and Silverman (2004) and Hübler (2016). The reasons may be that tall people tend to have higher productivity as the average height of the population is an indicator of the biological prosperity and standard of living (Komlos and Baur, 2003) or just because short people are discriminated in the labour market due to cultural and social stigma (Galbraith (1985). Geography includes the East, Central, and Southwest of China. Age cohort includes four generations aged below 30, between 30 and 44, between 45 and 60, and above 60. Firm characteristics include size of firm classified by small enterprises (with employees less than 50 persons), medium-sized (with employees above 50 persons but less than 500 persons) or large company (with employees above 500 persons) and type of ownership by foreign-owned, private-owned and state-owned enterprises.

Table 1 lists the abbreviations and definitions of all variables. Table 2 provides the descriptive statistics of the variables. The tables illustrates the

Table 1 Abbreviations and Descriptions of Variables

Variable Group	*Variable Name*	*Explanation*	*Reference Group*
	Lhrincome	log of hour wage	
	Employed	being hired as a permanent worker, long term contract worker (one year and above) or a short term contract worker (less than one year)	
Personal traits	Yrofwork	years of work experience in present occupation	
	Yrofworksq	the square of years of work	
	Eduyear	years of education	
	Age	years of age	
	Male	gender dummy (male=1 and female=0)	
	Health	dummy of good health (the value of dummy is 1 if the score of self-evaluation on health is between 1 and 3 otherwise its value is 0)	
	Noconfid	dummy of "lacking self-confidence"	
	Childum	dummy of children (1 if children number is greater than 0 and 0 otherwise)	
Geography	Central	dummy of central China	Southwest
	East	dummy of eastern China	
	Southwest	dummy of southwestern China	
Cohort	Oldgen	dummy of old generation (aged above 60)	Oldgen
	Midgen2	dummy of second mid-generation (aged between 45 and 60)	
	Midgen1	dummy of first mid-generation (aged between 30 and 45)	
	Youngen	dummy of young generation (aged below 30)	
Firm characteristics	Smallcom	firm size dummy for small companies (employees less than 50 persons)	Smallcom
	Midcom	firm size dummy for medium-sized enterprises (employees between 50 and 500 persons)	
	Bigcom	firm size dummy for big company (employees more than 500 persons)	
	Foreignown	ownership dummy for foreign-owned	

Table 1 (continued)

Variable Group	*Variable Name*	*Explanation*	*Reference Group*
	Privateown	ownership dummy for private-owned	
	Stateown	ownership dummy for state-owned	
Industry type	Indprim	industry dummy for Agriculture, Forestry, Animal husbandry, Fishery	Indsalntra
	Indmin	industry dummy for Mining	
	Indmanu	industry dummy for Manufacturing	
	Indelewatgas	industry dummy for Production and Supply of Electricity, Gas and Water	
	Indconst	industry dummy for Construction Enterprise	
	Indtransp	industry dummy for Transport, Storage and Post Industry	
	Indict	industry dummy for Information Transmission, Computer Services and Software Industry	
	Indsalntra	industry dummy for Wholesale and Retail Trade	
	Indhotel	industry dummy for Hotel and Catering Services	
	Indfinanc	industry dummy for Financial Intermediation	
	Indestate	industry dummy for Real Estate Industry	
	Indleasing	industry dummy for Leasing and Business Services	
	Indscience	industry dummy for Scientific Research, Technical Service and Geologic Prospecting	
	Indenvtmange	industry dummy for Management of Water Conservancy, Environment and Public Facilities	
	Indservice	industry dummy for Services to Households and Other Services	
	Indedu	industry dummy for Education	
	Indwelfare	industry dummy for Health, Social Security and Social Welfare	
	Indculture	industry dummy for Culture, Sport and Entertainment	
	Inddomesorg	industry dummy for Public Management and Social Organization	
	Indintlorg	industry dummy for International Organizations	

Table 2 Descriptive Statistics

	Urban Workers			*Migrants*		
	Obs	*Mean*	*Std.Dev.*	*Obs*	*Mean*	*Std.Dev.*
lhrincome	5643	2.39	0.67	3198	1.78	0.49
central	5790	0.34	0.47	3257	0.38	0.48
east	5790	0.46	0.50	3257	0.46	0.50
southwest	5790	0.20	0.40	3257	0.16	0.37
height	5776	167.05	7.47	3256	166.85	6.98
yrofwork	5775	14.70	10.58	3251	4.94	4.33
yrofworksq	5775	327.84	400.85	3251	43.17	88.47
age	5770	40.89	9.77	3251	30.91	9.62
highedu	5790	0.17	0.37	3257	0.02	0.14
midedu	5790	0.76	0.43	3257	0.77	0.42
eduyear	5724	12.44	3.08	3223	9.62	2.33
lowedu	5790	0.08	0.27	3257	0.21	0.41
oldgen	5790	0.02	0.14	3257	0.01	0.11
midgen2	5790	0.37	0.48	3257	0.08	0.27
midgen1	5790	0.46	0.50	3257	0.34	0.47
youngen	5790	0.15	0.36	3257	0.57	0.50
male	5790	0.57	0.49	3257	0.62	0.49
health	5790	0.98	0.13	3257	0.99	0.08
noconfid	3423	1.44	0.53	2890	1.46	0.57
childum	5790	0.80	0.40	3257	0.50	0.50
smallcom	5790	0.31	0.46	3257	0.43	0.49
midcom	5790	0.44	0.50	3257	0.14	0.35
bigcom	5790	0.25	0.43	3257	0.43	0.50
indprim	5790	0.01	0.11	3257	0.00	0.00
indmin	5790	0.01	0.11	3257	0.00	0.02
indmanu	5790	0.20	0.40	3257	0.31	0.46
indelewatgas	5790	0.04	0.20	3257	0.00	0.04
indconst	5790	0.03	0.18	3257	0.10	0.30
indtransp	5790	0.10	0.29	3257	0.03	0.18
indict	5790	0.04	0.20	3257	0.01	0.09
indsalntra	5790	0.09	0.28	3257	0.14	0.35
indhotel	5790	0.03	0.16	3257	0.16	0.37
indfinanc	5790	0.04	0.20	3257	0.00	0.06
indestate	5790	0.02	0.14	3257	0.05	0.21
indleasing	5790	0.03	0.17	3257	0.01	0.11

Table 2 (continued)

	Urban Workers			*Migrants*		
	Obs	*Mean*	*Std.Dev.*	*Obs*	*Mean*	*Std.Dev.*
indscience	5790	0.03	0.17	3257	0.02	0.15
indenvtmange	5790	0.01	0.12	3257	0.01	0.08
indservice	5790	0.09	0.29	3257	0.10	0.30
indedu	5790	0.06	0.24	3257	0.01	0.09
indwelfare	5790	0.04	0.21	3257	0.02	0.15
indculture	5790	0.02	0.14	3257	0.01	0.12
inddomesorg	5790	0.09	0.29	3257	0.00	0.05
indintlorg	5790	0.00	0.02	3257	0.00	0.00
foreignown	5790	0.05	0.22	3257	0.10	0.30
privateown	5790	0.18	0.38	3257	0.52	0.50
stateown	5790	0.59	0.49	3257	0.12	0.32
othercomtp	5790	0.19	0.39	3257	0.26	0.44

basic differences between urban and migrant workers. On average, urban workers received an hourly wage of RMB10.91 with average age of 40.89 years old, 12.44 years of education, and 14.7 years of work experience; while migrants received an hourly wage of RMB5.93 with average age of 30.91 years, 9.62 years of education, and 4.94 years of work experience. Among them, 15% of urban workers are below 30 years while 57% of migrants are below 30 years; 31% of migrant workers work in the manufacturing sector while only 20% of urban workers have jobs in manufacturing. Most urban workers (59%) are employed in state-owned enterprises, while most migrants (52%) are employed in private-owned firms.

In sum, the data show a general tendency that in contrast to migrant workers, urban workers on average are older, more educated, more experienced and are higher wage earners. Urban workers worked more in state-own enterprises, while migrant workers are employed mostly in private enterprises.

4. Estimation Results

Tables 3 and 4 respectively provide the results of Heckit test for both urban and migrant workers. Column (1) of the wage determination in the two tables is the base model, which regresses log hourly income on personal traits of the employed. Columns (2) to (5) of the two tables additionally adds the factor of geography, age cohort, firm characteristics, and industry type, separately to the base model. Column (6) jointly adds all the factors to the base model.

Table 3 Wage Regression for Urban Workers

Dep var = lhrincome

	(1) Personal traits		*(2) Personal traits + Geography*		*(3) Personal traits + Cohort*		*(4) Personal traits + Firm characteristics*		*(5) Personal traits + Industry*		*(6) Personal traits + Geography + Cohort + Firm characteristics + Industry*	
Height	0.005^{**}	(0.002)	0.000	(0.002)	0.005^{**}	(0.002)	0.003	(0.002)	0.004^{*}	(0.002)	-0.001	(0.002)
Yrofwork	0.030^{***}	(0.004)	0.031^{***}	(0.004)	0.030^{***}	(0.004)	0.027^{***}	(0.004)	0.027^{***}	(0.004)	0.026^{***}	(0.004)
Yrofworksq	0.000^{***}	(0.000)	0.000^{***}	(0.000)	0.000^{***}	(0.000)	0.000^{***}	(0.000)	0.000^{***}	(0.000)	0.000^{***}	(0.000)
Eduyear	0.063^{***}	(0.005)	0.065^{***}	(0.005)	0.060^{***}	(0.006)	0.060^{***}	(0.005)	0.055^{***}	(0.005)	0.051^{***}	(0.007)
Health	0.169^{**}	(0.084)	0.162^{*}	(0.084)	0.161^{*}	(0.088)	0.186^{**}	(0.083)	0.146^{*}	(0.083)	0.134	(0.086)
Childum	-0.007	(0.034)	-0.020	(0.033)	0.007	(0.041)	0.010	(0.033)	0.008	(0.034)	-0.006	(0.039)
Male	0.066^{*}	(0.034)	0.123^{***}	(0.035)	0.061	(0.039)	0.080^{**}	(0.034)	0.092^{***}	(0.034)	0.140^{***}	(0.040)
East			0.251^{***}	(0.030)							0.238^{***}	(0.029)
Central			0.284^{***}	(0.031)							0.283^{***}	(0.030)
Midgen2					-0.146	(0.115)					-0.146	(0.118)
Midgen1					-0.100	(0.127)					-0.143	(0.131)
Youngen					-0.089	(0.126)					-0.163	(0.129)
Midcom							0.151^{***}	(0.025)			0.143^{***}	(0.025)
Bigcom							0.102^{***}	(0.030)			0.133^{***}	(0.031)
Foreignown							0.411^{***}	(0.055)			0.401^{***}	(0.054)
Privateown							0.076^{**}	(0.035)			0.080^{**}	(0.034)
Stateown							0.164^{***}	(0.028)			0.127^{***}	(0.028)
Indmin									-0.098	(0.110)	-0.083	(0.107)

Table 3 (continued)

Dep var = lhrincome

	(1) Personal traits		*(2) Personal traits + Geography*		*(3) Personal traits + Cohort*		*(4) Personal traits + Firm characteristics*		*(5) Personal traits + Industry*		*(6) Personal traits + Geography + Cohort + Firm characteristics + Industry*	
Indmanu									-0.060	(0.041)	-0.112***	(0.041)
Indelewatgas									0.135**	(0.064)	0.113*	(0.063)
Indconst									0.037	(0.068)	0.018	(0.067)
Indtransp									-0.003	(0.048)	-0.046	(0.048)
Indict									0.258***	(0.061)	0.197***	(0.060)
Indhotel									0.001	(0.075)	0.013	(0.073)
Indfinanc									0.298***	(0.061)	0.289***	(0.060)
Indestate									0.240***	(0.086)	0.285***	(0.083)
Indleasing									0.094	(0.073)	0.111	(0.071)
Indscience									0.154**	(0.071)	0.163**	(0.070)
Indenvtmange									0.046	(0.094)	-0.009	(0.092)
Indservice									-0.171***	(0.046)	-0.121***	(0.045)
Indedu									0.155***	(0.053)	0.131**	(0.053)
Indwelfare									0.187***	(0.059)	0.145**	(0.059)
Indculture									0.116	(0.079)	0.109	(0.077)
Inddomesorg									0.151***	(0.046)	0.155***	(0.048)
Lambda	-0.165***	(0.057)	-0.153***	(0.059)	-0.194**	(0.084)	-0.125**	(0.057)	-0.134**	(0.057)	-0.158*	(0.092)
cons	0.444	(0.385)	0.939**	(0.383)	0.611	(0.452)	0.464	(0.379)	0.592	(0.379)	1.343***	(0.449)

Table 3 (continued)

Dep. Var = Employed

	(1) Personal traits	*(2) Personal traits + Geography*	*(3) Personal traits + Cohort*	*(4) Personal traits + Firm characteristics*	*(5) Personal traits + Industry*	*(6) Personal traits + Geography + Cohort + Firm characteristics + Industry*
Age	-0.047^{***} (0.005)	-0.047^{***} (0.005)	-0.047^{***} (0.005)	-0.047^{***} (0.005)	-0.047^{***} (0.005)	-0.047^{***} (0.005)
Health	0.571^{***} (0.097)	0.571^{***} (0.097)	0.571^{***} (0.097)	0.571^{***} (0.097)	0.571^{***} (0.097)	0.571^{***} (0.097)
Noconfid	-0.135^{***} (0.032)	-0.135^{***} (0.032)	-0.135^{***} (0.032)	-0.135^{***} (0.032)	-0.135^{***} (0.032)	-0.135^{***} (0.032)
Eduyear	0.116^{***} (0.006)	0.116^{***} (0.006)	0.116^{***} (0.006)	0.116^{***} (0.006)	0.116^{***} (0.006)	0.116^{***} (0.006)
Childum	0.169^{**} (0.077)	0.169^{**} (0.077)	0.169^{**} (0.077)	0.169^{**} (0.077)	0.169^{**} (0.077)	0.169^{**} (0.077)
Male	0.586^{***} (0.037)	0.586^{***} (0.037)	0.586^{***} (0.037)	0.586^{***} (0.037)	0.586^{***} (0.037)	0.586^{***} (0.037)
East	0.152^{***} (0.048)	0.152^{***} (0.048)	0.152^{***} (0.048)	0.152^{***} (0.048)	0.152^{***} (0.048)	0.152^{***} (0.048)
Central	-0.107^{**} (0.050)	-0.107^{**} (0.050)	-0.107^{**} (0.050)	-0.107^{**} (0.050)	-0.107^{**} (0.050)	-0.107^{**} (0.050)
Midgen2	0.955^{***} (0.094)	0.955^{***} (0.094)	0.955^{***} (0.094)	0.955^{***} (0.094)	0.955^{***} (0.094)	0.955^{***} (0.094)
Midgen1	0.764^{***} (0.136)	0.764^{***} (0.136)	0.764^{***} (0.136)	0.764^{***} (0.136)	0.764^{***} (0.136)	0.764^{***} (0.136)
Youngen	0.076 (0.187)	0.076 (0.187)	0.076 (0.187)	0.076 (0.187)	0.076 (0.187)	0.076 (0.187)
Cons	-0.601^{*} (0.331)	-0.601^{*} (0.331)	-0.601^{*} (0.331)	-0.601^{*} (0.331)	-0.601^{*} (0.331)	-0.601^{*} (0.331)
obs.			6257	6257	6257	6257
censored_obs.			2957	2957	2957	2957
uncensored_obs.			3300	3300	3300	3300
Wald_chi^2			355.58	488.99	515.41	768.61

Note: Significance level: 1% = ***, 5% = ** and 10% = *. For explanations of abbreviations, see Table 1.

Table 4 Wage Regression for Migrant Workers

Dep var = lhrincome

	(1) Personal traits		*(2) Personal traits + Geography*		*(3) Personal traits + Cohort*		*(4) Personal traits + Firm characteristics*		*(5) Personal traits + Industry*		*(6) Personal traits + Geography + Cohort + Firm characteristics + Industry*	
Height	0.010^{***}	(0.002)	0.006^{***}	(0.002)	0.010^{***}	(0.002)	0.009^{***}	(0.002)	0.010^{***}	(0.002)	0.006^{***}	(0.002)
Yrofwork	0.058^{***}	(0.005)	0.051^{***}	(0.005)	0.053^{***}	(0.005)	0.051^{***}	(0.005)	0.055^{***}	(0.005)	0.042^{***}	(0.005)
Yrofworksq	-0.002^{***}	(0.000)	-0.002^{***}	(0.000)	-0.002^{***}	(0.000)	-0.002^{***}	(0.000)	-0.002^{***}	(0.000)	-0.001^{***}	(0.000)
Eduyear	0.055^{***}	(0.004)	0.058^{***}	(0.003)	0.055^{***}	(0.004)	0.051^{***}	(0.004)	0.053^{***}	(0.004)	0.053^{***}	(0.003)
Health	0.191^{*}	(0.101)	0.185^{*}	(0.097)	0.208^{**}	(0.101)	0.186^{*}	(0.098)	0.154	(0.099)	0.163^{*}	(0.094)
Childum	0.003	(0.018)	0.020	(0.018)	-0.055^{**}	(0.025)	-0.001	(0.018)	-0.009	(0.018)	-0.015	(0.023)
Male	0.004	(0.026)	0.055^{**}	(0.025)	-0.114^{***}	(0.037)	0.018	(0.025)	-0.024	(0.026)	-0.049	(0.032)
East			0.361^{***}	(0.023)							0.312^{***}	(0.026)
Central			0.261^{***}	(0.024)							0.190^{***}	(0.026)
Midgen2					0.448^{***}	(0.112)					0.388^{***}	(0.098)
Midgen1					0.697^{***}	(0.121)					0.619^{***}	(0.106)
Youngen					0.459^{***}	(0.111)					0.464^{***}	(0.096)
Midcom							0.005	(0.025)			-0.015	(0.024)
Bigcom							0.116^{***}	(0.018)			0.077^{***}	(0.019)
Foreignown							0.160^{***}	(0.031)			0.116^{***}	(0.031)
Privateown							-0.052^{***}	(0.019)			-0.026	(0.018)
Stateown							0.008	(0.028)			0.012	(0.027)
Indmin									0.357	(0.442)	0.350	(0.427)

Table 4 (continued)

Dep var = lhrincome

	(1) Personal traits		*(2) Personal traits + Geography*		*(3) Personal traits + Cohort*		*(4) Personal traits + Firm characteristics*		*(5) Personal traits + Industry*		*(6) Personal traits + Geography + Cohort + Firm characteristics + Industry*	
Indmanu									0.144***	(0.025)	0.040	(0.026)
Indelewatgas									0.223	(0.219)	0.090	(0.209)
Indconst									0.144***	(0.034)	0.103***	(0.033)
Indtransp									0.060	(0.048)	0.009	(0.046)
Indict									0.185**	(0.090)	0.156*	(0.086)
Indhotel									-0.118***	(0.028)	-0.102***	(0.027)
Indfinanc									0.089	(0.123)	0.127	(0.118)
Indestate									0.029	(0.041)	0.046	(0.040)
Indleasing									0.245***	(0.072)	0.232***	(0.069)
Indscience									0.139**	(0.055)	0.104**	(0.052)
Indenvtmange									-0.025	(0.103)	-0.045	(0.099)
Indservice									-0.025	(0.032)	-0.054*	(0.031)
Indedu									0.115	(0.090)	0.111	(0.086)
Indwelfare									-0.090	(0.057)	-0.129**	(0.054)
Indculture									0.087	(0.068)	0.095	(0.064)
Inddomesorg									-0.264*	(0.147)	-0.355**	(0.140)
Lambda	0.009	(0.069)	0.120*	(0.067)	-0.755***	(0.146)	0.075	(0.068)	0.049	(0.068)	-0.411***	(0.132)
Cons	-0.779***	(0.292)	-0.518*	(0.283)	-0.482	(0.324)	-0.725**	(0.286)	-0.840***	(0.286)	-0.376	(0.298)

Table 4 (continued)

Dep var = Employed

	(1) Personal traits		*(2) Personal traits + Geography*		*(3) Personal traits + Cohort*		*(4) Personal traits + Firm characteristics*		*(5) Personal traits + Industry*		*(6) Personal traits + Geography + Cohort + Firm characteristics + Industry*	
Age	-0.004	(0.005)	-0.004	(0.005)	-0.004	(0.005)	-0.004	(0.005)	-0.004	(0.005)	-0.004	(0.005)
Health	0.622^{***}	(0.162)	0.622^{***}	(0.162)	0.622^{***}	(0.162)	0.622^{***}	(0.162)	0.622^{***}	(0.162)	0.622^{***}	(0.162)
Noconfid	0.008	(0.029)	0.008	(0.029)	0.008	(0.029)	0.008	(0.029)	0.008	(0.029)	0.008	(0.029)
Eduyear	0.100^{***}	(0.008)	0.100^{***}	(0.008)	0.100^{***}	(0.008)	0.100^{***}	(0.008)	0.100^{***}	(0.008)	0.100^{***}	(0.008)
Childum	-0.295^{***}	(0.053)	-0.295^{***}	(0.053)	-0.295^{***}	(0.053)	-0.295^{***}	(0.053)	-0.295^{***}	(0.053)	-0.295^{***}	(0.053)
Male	0.214^{***}	(0.029)	0.214^{***}	(0.029)	0.214^{***}	(0.029)	0.214^{***}	(0.029)	0.214^{***}	(0.029)	0.214^{***}	(0.029)
East	0.022	(0.041)	0.022	(0.041)	0.022	(0.041)	0.022	(0.041)	0.022	(0.041)	0.022	(0.041)
Central	0.023	(0.042)	0.023	(0.042)	0.023	(0.042)	0.023	(0.042)	0.023	(0.042)	0.023	(0.042)
Midgen2	-0.226	(0.188)	-0.226	(0.188)	-0.226	(0.188)	-0.226	(0.188)	-0.226	(0.188)	-0.226	(0.188)
Midgen1	-0.196	(0.211)	-0.196	(0.211)	-0.196	(0.211)	-0.196	(0.211)	-0.196	(0.211)	-0.196	(0.211)
Youngen	0.628^{***}	(0.048)	0.628^{***}	(0.048)	0.628^{***}	(0.048)	0.628^{***}	(0.048)	0.628^{***}	(0.048)	0.628^{***}	(0.048)
Cons	-1.565^{***}	(0.089)	-1.565^{***}	(0.089)	-1.565^{***}	(0.089)	-1.565^{***}	(0.089)	-1.565^{***}	(0.089)	-1.565^{***}	(0.089)
obs.	8325		8325		8325		8325		8325		8325	
censored_obs.	5169		5169		5169		5169		5169		5169	
uncensored_obs.	3156		3156		3156		3156		3156		3156	
Wald_chi2	532.09		811.37		558.54		703.93		735.06		1055.64	

Note: Significance level: 1% = ***, 5% = ** and 10% = *. For explanations of abbreviations, see Table 1.

The six Heckit tests share similar results. According to the first-stage selection results in the lower panel of the two tables, an urban worker who is young or self-confident is more likely to be employed, while both variables are insignificant to determine the employment status of a migrant. This may be because migrants move mainly because of looking for better job opportunity and higher wage in the city, therefore once they decide to move and become migrants they have a strong determination to work regardless of their age of confidence feeling. However, both urban and migrant workers who are male, more educated, and healthier, tend to have higher probability to be employed. An urban native who has the responsibility of child rearing is more likely to be employed. By contrast, a migrant rearing a child is less likely to find a job in a city because child rearing in city is more difficult and costly and may even be considered as a burden for the migrant. As for geography and age cohort, both are insignificant factors for determining employment status for the migrant; however, an urban native in the east is more likely to find a job than in the central region as the eastern urban area provides better labour conditions and job opportunities. Moreover, urban people of mid-generation are more likely to be hired, while younger generation of migrants exhibits an advantage in finding a job in the city. Thus, the employment selection behaviour for urban natives and migrants shares some commonalities but also has certain divergences.

According to the second-stage results of wage regression in Tables 3 and 4, sample correction terms derived from the first stage are significant for wage determination of both urban and migrant workers, implying the necessity to correct for sample selection bias. The negative coefficient of correction term means that the observed wage tends to underestimate the real one. Both years of education and work experience are positive and strongly significant for both urban and migrant workers, but the returns to work experience of migrant workers are nearly twice than that of urban workers implying that work experience is more important for those migrants who tend to be young, less educated and unskilled. Another reason is that the average years of work experience of the migrants is lower than that of urban workers. This result is thus consistent with the law of diminishing returns. In fact, migrants in some occupations with the same years of work experience as that of urban workers still receive a higher rate of return to work experience despite the fact that they earn less than their urban counterparts. This result supports the finding that migrant workers rely more on skill accumulation through on-the-job training. It also implies that, compared to an urban worker, migrant workers' human capital level in terms of work experience is relatively low. In fact, the average work experience of the urban is about three times that of the migrants. This may also be because migrants cannot work long in the city and have to return to their hometown voluntary or involuntary. The significant but negative estimates of squared value of work experience are consistent with the nonlinear effect suggested by the literature.

In contrast, the results show that urban workers have higher rate of returns to education than the migrants implying that urban workers enjoy more education resources than do migrant workers both in quantity and quality.[9] Despite the law of diminishing marginal returns, the rate of return to education of urban workers who have received longer years of education is still higher than that of the migrants.

Furthermore, two other human capital variables, height and health condition also show positive and significant effects on wage level for both urban and migrant workers, and their effects are relatively stronger for the migrants than for the urban workers. This may have to do with the job characteristics that migrants are mostly engaged in such as work that required more physical strength or in more risky working conditions in the manufacturing sector. All these results confirm that even under the Chinese segmented labour market environment, human capital remains an important dimension for understanding the determinants of labour income. Moreover, except for formal education in which urban workers have a greater advantage over the migrants, in other aspects of human capital such as work experience and health condition the migrants have larger rates of return. Our results suggest that human capital investment and accumulation can be an effective way to narrow the wage gap between urban and migrant workers.

As for geography, its coefficients are positive and significant; and, according to the magnitude, we find that for both urban and migrant workers their wage level is higher in Eastern and Central China than in Southwest China. This is consistent with our understanding that job opportunity is better in these regions. In regard to the cohort effect, there is an evident difference between urban and migrant workers. For urban workers, there are no significant differences among different age cohorts, while for migrants, young workers tend to earn more and the highest income goes to the age group between 30 and 45 years old. This shows that after controlling for personal traits, cohort effect only exists in migrant workers and not in urban workers. We attribute this phenomenon to the relatively stable work environment in urban areas faced by urban workers with different age cohorts. However, the younger generation of migrants with more work experience tends to earn more implying either better job opportunities or labour quality of the new generation of migrants.[10]

As for firm size, consistent with the literature, larger firms tend to pay higher wages. An urban worker receives a bigger wage premium in a medium-sized company (with employees between 50 and 500 persons), while a migrant earns a higher wage premium from a big company (with employees above 500 persons). This is because migrants usually work in big assembly factories as operation workers or labourers.[11] For the type of firm ownership, both urban and migrant workers receive a higher wage premium from a foreign-owned company. However, in a private-owned company or a state-owned company only urban workers have a significant wage premium, while in a private-owned company, where migrant workers mostly worked, it pays a negative wage premium, i.e., migrant workers suffer a significant wage loss in private-owned companies. This result implies that private firms are most likely to take advantage in exploiting migrant workers.

As for industry type, urban workers tend to earn more in manufacturing; electric, water, and gas; information, computer, and software service; finance, real estate and leasing; scientific and technical service; education; health and social welfare; and domestic organization industries, while migrant workers earn more only in construction; information, computer, and software service; leasing; scientific and technical service; and education industries. The limited numbers and aspects of sectors that pay migrants better wages imply that migrants are likely segmented in the labour market (see also Meng and Zhang, 2001; Meng, 2012).

Table 5 summarizes decomposition of wage differentials. Tables 6 and 7 list unexplained and explained ratios. Group membership of workers' identity is not considered in Table 6 but is considered in Table 7.

When the urban worker is used as a reference group, the explained part is, in overall terms, larger than that of the case where the migrant is used as a reference group. Moreover, when geography, age cohort, firm size and ownership, and industry type are simultaneously added to the basic model, unexplained ratio can be reduced to 11% in the case where the urban worker is the reference group and to 15% in the case where the migrant is the reference group. The value of the unexplained ratio is close to that of Lee (2012) and Zhang et al. (2016) but smaller than that of Magnani and Zhu (2012).[12]

Columns (1) and (6) are the two baseline cases. Column (1) only considers personal traits, while column (6) considers personal traits, geography, age cohort, firm characteristics and industry type simultaneously. According to column (1), we know that personal traits as a group of explanatory variables explain 76% of wage differential. The marginal effect of other variable groups, including geography, *age cohort*, firm size and ownership, and industry type added to the case of column (1), as column (6) shows, can only increase explained ratio by 13% in Panel I (increasing from 76% in column (1) to 89% in column (6)) and by 14% in Panel II (increasing from 71% in column (1) to 85% in column (6)). Therefore, difference in personal traits is crucial to explaining the wage differential between urban and migrant workers. Thus, the results in Table 6 show that our model specification in general explains up to 85-89% of the wage gap.

Table 5 Summary of Income Differential Decomposition

	(1) Personal Traits	*(2) (1) + Geography*	*(3) (1) + Cohort*	*(4) (1) + Firm Type (Size and Ownership)*	*(5) (1) + Industry Type*	*(6) All Included*
Panel I. Urban as reference						
A: coefficient; explained	0.33	0.32	0.31	0.36	0.35	0.34
B: coefficient; unexplained	-0.84	-0.97	-1.42	-0.79	-1.04	-1.76
C: constant; unexplained	1.22	1.46	1.09	1.19	1.43	1.72
Panel II. Migrant as reference						
A: coefficient; explained	0.19	0.19	0.20	0.20	0.17	0.16
B: coefficient; unexplained	-0.71	-0.84	-1.32	-0.64	-0.87	-1.58
C: constant; unexplained	1.22	1.46	1.09	1.19	1.43	1.72

Table 6 Unexplained and Explained Ratios (without Group Membership)

	(1) Personal Traits	*(2) (1) + Geography*	*(3) (1) + Cohort*	*(4) (1) + Firm Type (Size and Ownership)*	*(5) (1) + Industry Type*	*(6) All Included*
Panel I. Urban as reference						
exp(A)/(exp(A)+exp(B))	76%	78%	85%	76%	80%	89%
exp(B)/(exp(A)+exp(B))	24%	22%	15%	24%	20%	11%
Panel II. Migrant as reference						
exp(A)/(exp(A)+exp(B))	71%	74%	82%	70%	74%	85%
exp(B)/(exp(A)+exp(B))	29%	26%	18%	30%	26%	15%

Table 7 Unexplained and Explained Ratios (with Group Membership)

	(1) Personal Traits	*(2) (1) + Geography*	*(3) (1) + Cohort*	*(4) (1) + Firm Type (Size and Ownership)*	*(5) (1) + Industry Type*	*(6) All Included*
Panel I. Urban as reference						
exp(A)/(exp(A)+exp(B+C))	49%	46%	66%	49%	49%	60%
exp(B+C)/(exp(A)+exp(B+C))	51%	54%	34%	51%	51%	40%
Panel II. Migrant as reference						
exp(A)/(exp(A)+exp(B+C))	42%	40%	60%	41%	40%	51%
exp(B+C)/(exp(A)+exp(B+C))	58%	60%	40%	59%	60%	49%

Among the inclusion of other variables, the marginal contribution of adding cohort variable has the greatest effect, since it increases explained ratios of column (1) by 9% in Panel I (improving it from 76% to 85%) and 11% in Panel II (improving it from 71% to 82%). By contrast, firm characteristics accounts for the least additional contribution, and geography and industry type share similar marginal effects.

Finally, let us look at the case where group membership, which stands for the classification of the urban worker and the migrant, is considered. As Table 7 shows, when group membership, the variation in the interception of regression models, is considered as the unexplained part, the explained ratio will drop sharply from 70%-89% to 42%-60%. Meanwhile, we also find that the cohort variable is least affected by the inclusion of group membership. Compared to the drop on explained part with group membership by adding geography, firm characteristics, or industry type, the cohort variable brings about the smallest drop (from 85% in Panel I of Table 6 to 66% in Panel I of Table 7 and from 82% in Panel II of Table 6 to 60% in Panel II of Table 7). These results imply that group membership has less impact on the cohort variable. Since group membership is a crucial component of unexplained ratio, we argue that the inclusion of cohort variable will increase explained ratio. By contrast, when geography, firm characteristics and industry type, are added to column (1) of Table 7, there shows no significant increase in explained ratio under group membership. This result implies that if discrimination towards migrants is reflected in the group membership, then the discrimination may largely be related to geography, firm size and ownership, and industry type. These findings are consistent with that of Meng and Zhang (2001) and Appleton *et al.* (2004) who show significant labour market segregation and discrimination in occupation and industry to migrant workers in China.

By comparing the results from Table 6 and Table 7, we can conclude that including group membership will increase the unexplained part of wage differentials between the urban and migrant workers. This group membership leads to labour market segregation in geography, firm characteristics and industry type, and this kind of market segregation and discrimination can be approximated by the institutional arrangement of the household registration system (Hukou) that creates a rural-urban divide.

5. Conclusion

Using Blinder-Oaxaca decomposition method, this study analyzes the influence of personal traits, geography, age cohort, firm characteristics and industry type on wage differential between urban and migrant workers. The results show that, without considering unexplained part resulting from group membership, up to 85-89% of wage differentials in China's labour market can be explained, which is consistent with the findings in Lee (2012) and Zhang *et al.* (2016). And, if we solely look at the influence of personal traits on wage differential, we find that they explain 71-76% of the wage differential. Among them, human capital variables such as education, work experience and health are important factors determining one's wage.

On the other hand, if group membership is considered, the explained ratios drop prevalently in all the cases to 42-60%. However, among them, the case where the cohort variable is added is least affected. This also suggests that the inclusion of firm characteristics and industry type variables without considering group membership is likely to underestimate the effect of discrimination as the migrants are subject to labour market discrimination and segregation by firms and industries. Likewise, the inclusion of group membership can be a good approximation for the estimation of total effect of discrimination in the Chinese labour market. In China, group membership between urban and migrant workers is mainly due to the institutional arrangement of the Hukou system.

In summary, our findings on the one hand, imply that wage differential of China's labour market is largely accounted for by the difference in human capital level, since personal traits such as education level, work experience, height and health condition are all crucial to determining wages. However, despite the fact that migrants are subject to less educational resources and opportunity to access education, they still have significant higher rates of return to health and work experience. Thus, policies towards improvement in human capital investment and accumulation of the migrants can be an effective means to narrow the wage gap between rural and urban workers. For example, equal access to education for the children of migrants, better health care coverage for the migrant workers, and providing more on-the-job training for the migrants.

Furthermore, we also find that the addition of the cohort variable helps to increase explained ratio, whereas other factors such as geography, firm characteristics and industry type are more accountable for the discrimination towards migrants who are poorly paid due to labour market segmentation under different group membership between the urban and migrant workers. Thus, model specification without considering group membership is likely to underestimate the effect of discrimination. The cohort effect may represent better labour quality of new generations of migrants or better working conditions due to government policy such as the implementation of the Labor Contract Law in 2008. However, the effect of group membership actually reflects the institutional arrangement of household registration (Hukou) system that not only discriminates against the migrants in their identity, wage compensation and social welfare entitlement, but also on their children's education opportunity and admittance, which significantly imposes a negative effect on the future generation of migrants. Thus, an institutional reform to abolish the Hukou system is perhaps a critical policy to close the income gap between the rural and urban divide and a fundamental of the reform should focus on how to give equal access of public goods and social services for citizens within the same city.[13]

Appendix

Proof of equation (3)

Consider the case where migrant is used as a reference group, wage differential can be expressed as:

$$\overline{\mathrm{INC}}_u - \overline{\mathrm{INC}}_m$$
$$= (e^{\overline{X}_u'\hat{\beta}_u} - e^{\overline{X}_m'\hat{\beta}_m})$$
$$= e^{\overline{X}_m'\hat{\beta}_m}(e^{(\overline{X}_u'\hat{\beta}_u - \overline{X}_u'\hat{\beta}_m)+(\overline{X}_u'\hat{\beta}_m - \overline{X}_m'\hat{\beta}_m)} - 1)$$

$= \overline{\mathrm{INC}}_m(e^{(\overline{X}_u'\hat{\beta}_u - \overline{X}_u'\hat{\beta}_m)+(\overline{X}_u'\hat{\beta}_m - \overline{X}_m'\hat{\beta}_m)} - 1)$ which can be alternatively expressed as $(\overline{\mathrm{INC}}_u - \overline{\mathrm{INC}}_m)/\overline{\mathrm{INC}}_m = e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m} - 1.$

This equation describes wage differential between urban and migrant workers as a deviation from the average wage level of migrant workers, $\overline{\mathrm{INC}}_m$. We treat $\overline{X}_u' - \overline{X}_m'$, $\hat{\beta}_u - \hat{\beta}_m$ and $(\overline{X}_u'\hat{\beta}_u - \overline{X}_u'\hat{\beta}_m) + (\overline{X}_u'\hat{\beta}_m - \overline{X}_m'\hat{\beta}_m)$ as three sources of the deviation so we decompose the deviation as three parts influenced by the three factors. We give each part an equal weight so $e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}$ can be expressed as a sum of:

(i) $e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}$ where $\overline{X}_u' = \overline{X}_m'$, $\hat{\beta}_u \neq \hat{\beta}_m$ and $(\overline{X}_u'\hat{\beta}_u - \overline{X}_u'\hat{\beta}_m) + (\overline{X}_u'\hat{\beta}_m - \overline{X}_m'\hat{\beta}_m) \neq 0$,

(ii) $e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}$ where $\overline{X}_u' \neq \overline{X}_m'$, $\hat{\beta}_u = \hat{\beta}_m$ and $(\overline{X}_u'\hat{\beta}_u - \overline{X}_u'\hat{\beta}_m) + (\overline{X}_u'\hat{\beta}_m - \overline{X}_m'\hat{\beta}_m) \neq 0$, and

(iii) $e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}$ where $\overline{X}_u' \neq \overline{X}_m'$, $\hat{\beta}_u \neq \hat{\beta}_m$ and $(\overline{X}_u'\hat{\beta}_u - \overline{X}_u'\hat{\beta}_m) + (\overline{X}_u'\hat{\beta}_m - \overline{X}_m'\hat{\beta}_m) = 0$.

Thus, $(\overline{\mathrm{INC}}_u - \overline{\mathrm{INC}}_m)/\overline{\mathrm{INC}}_m = [e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}|_{\overline{X}_u' = \overline{X}_m'} + e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}|_{\hat{\beta}_u = \hat{\beta}_m} + e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)}e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}|_{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m) = -(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}]/3 - 1 \propto e^{\overline{X}_u'(\hat{\beta}_u - \hat{\beta}_m)} + e^{(\overline{X}_u' - \overline{X}_m')\hat{\beta}_m}$.

Notes

* Yih-chyi Chuang 莊奕琦 is a distinguished professor of the Department of Economics at the National Chengchi University, Taipei, Taiwan. He received his Ph.D. in Economics from University of Chicago in 1993. He was a visiting professor at Georgetown University in 1999 and Fulbright scholar at Fairbank Center for East Asian Research of Harvard University in 2003. From 2005 to 2016, he was invited as a guest professor at Free University of Berlin in Germany during the summer/winter terms. From 2012 to 2015, he served as the Dean of College of Social Sciences and the Director of Taiwan Studies Center at NCCU. His major teaching and research areas are theories of growth and development, education and labour market, and international trade and investment. He can be reached at <ycchuang03@gmail.com>.

** Eric Yan (顏扶昌) received his Ph.D. degree in Economics from the Dept. of Economics, Monash Business School, Monash University, Melbourne, Australia in 2012. Since then, he had worked as a postdoctoral researcher at the Dept. of Economics, National Chengchi University and Institute of Economics, Academia Sinica in Taiwan. He can be reached at <fulgent945@hotmail.com>.

1. The number of floating population increased from 25 million people in 1990 to 37 million persons in 1997. According to the National Bureau of Statistics of China, there were around 140 million rural-to-urban migrants in 2008.
2. In the early 1980s, the household responsibility system, which replaced the production team system as the agriculture production and accounting unit, allowed households to contract land, machinery and other facilities from the collective farms. By the end of 1983, 94.2% of production teams had adopted the system. See, e.g., Lin (1987).
3. The plan, also referred to as Family Planning Policy, was implemented in 1979 and called for each family to have one child only in order to curb a then-surging population and limit the demands for resources that may slowdown the development of the whole economy.
4. As illustrated in the Twelve Five-Year Plan (2010-2015) and Third Plenary Session of the 18th Central Committee of CCP in November, 2013, the major issues of future economic reform in China are to change the development strategy leaning towards more inward-oriented and deepen the market orientation by using urbanization as a vehicle to narrow rural-urban income gap.
5. For example, Bound and Johnson (1992), Mincer (1991), Allen (1991), and Krueger (1993) relate wage inequality to technology; Beaudry and Green (2005) relate wage inequality to human capital; Forbes (2001) relates wage inequality to trade that spreads technology; and Altonji and Blank (1999) and Heckman (1998) consider discrimination to be the cause of wage inequality.
6. Oaxaca & Ransom (1994) show that the decomposition approaches can be further generalized as:

$$\ln \overline{W}_U - \ln \overline{W}_M = (\overline{\boldsymbol{X}}_U - \overline{\boldsymbol{X}}_M)\hat{\beta}^* + \overline{\boldsymbol{X}}_U(\hat{\beta}_U - \hat{\beta}^*) + \overline{\boldsymbol{X}}_M(\hat{\beta}^* - \hat{\beta}_M)$$

where $\hat{\beta}^*$ is the real non-discriminated coefficients of wage structure, which by definition is a weighted average of $\hat{\beta}_U$ and $\hat{\beta}_M$, i.e., $\hat{\beta}^* = \boldsymbol{\Omega}\hat{\beta}_U + (\boldsymbol{I} - \boldsymbol{\Omega})\hat{\beta}_M$ and $\boldsymbol{\Omega}$ is a matrix of weights and $\boldsymbol{I}$ is an identity matrix.

7. See Appendix for detailed derivation.
8. We choose monthly wage above RMB500 as the threshold for full-time worker because RMB500 is the lower bound of minimum monthly wage among those cities and provinces under survey.
9. Urban citizens are guaranteed to received formal education and used to have educational subsidy from their work unit (Danwei), while children of migrants without Hukou cannot gain entry to a school in the city.
10. In 2008, the Labor Contract Law was implemented in China. Since then, the government set the minimum wage and adjust at an average annual growth rate of 15% to 20%, which has significantly increased the wage of young migrant workers.
11. For example, Faxconn, the giant electronics manufacturing subcontractor and the world's second largest private employer after Wal-Mart, employed some 1.4 million workers in China in 2013. Foxconn's Longhua facility in the Shenzhen Special Economic Zone alone hired some 300,000 Chinese migrant workers to do the assembling of IT products, especially for the Apple Company.
12. Magnani and Zhu (2012) also point out that controlling for occupation and industry variables in decomposition may underestimate discrimination effects.
13. In recent years, China's Ministry of Public Security has been considering reforms to the controversial household registration (Hukou) system, including replacing temporary residence permits held by migrant workers in cities with permanent ones. A new measure of the reform requires temporary residents to obtain certain points before becoming permanent residents. China's Ministry of Public Security announced that it had issued 28.9 million new urban residency permits in 2016. However, the new point system will still involve salary, tax payment, education level and years of residence of applicants, which would result in new unequal relations among citizens. Thus, in our view the fundamental of the reform is how to give equal access of public goods and social services for citizens within the same city.

References

Allen, S.G. (1991), "Technology and the Wage Structure", mimeo, North Carolina State University.

Altonji, J.G. and Blank R. (1999), "Race and Gender in the Labor Market", in Orley C. Ashenfelter and David Card (eds), *Handbook of Labor Economics, Vol. 3C*, Amsterdam: Elsevier Science B.V., pp. 3143-3259.

Appleton, S., Knight J., Song L. and Xi Q. (2004), "Contrasting Paradigms: Segmentation and Competitiveness in the Formation of the Chinese Labor Market", *Journal of Chinese Economic and Business Studies*, Vol. 2, No. 3, pp. 185-205.

Beaudry, P. and Green, D. (2005), "Changes in U.S. Wages, 1976-2000: Ongoing Skill Bias or Major Technological Change?", *Journal of Labor Economics*, Vol. 23, No. 3, pp. 609-648.

Blinder, A.S. (1973), "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources*, Vol. 8, No. 4, pp. 436-455.

Bound, J., and Johnson, G. (1992), "Changes in the Structure of Wages during the

1980s: An Evaluation of Alternative Explanations", *American Economic Review*, Vol. 82, No. 3, pp. 371-392.

Chan, K.W. and Buckingham, W. (2008), "Is China Abolishing the Hukou System?", *China Quarterly*, Vol. 195 (September), pp. 582-606.

Demurger, S., Gurgand, M., Li S. and Yue, X. (2009), "Migrants as Second-Class Workers in Urban China? A Decomposition Analysis", *Journal of Comparative Economics*, Vol. 37, No. 4, pp. 610-628.

Deng, Q. and Li, S. (2010), "Wage Structures and Inequality among Local and Migrant and Urban Workers", in Xin Meng, Chris Manning, Li Shi and Tadjuddin Noer Effendi (eds), *The Great Migration: Rural-Urban Migration in China and Indonesia*, Cheltenham, UK: Edward Elgar Publishing Ltd.

Forbes, K. (2001), "Skill Classification Does Matter: Estimating the Relationship between Trade Flows and Wage Inequality", *Journal of International Trade & Economic Development*, Vol. 10, No. 2, pp. 175-209.

Galbraith, J.K. (1985), *The Anatomy of Power*, Boston, MA: Houghton Mifflin Company.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, No. 1, pp. 153-161.

Heckman, J. (1998), "Detecting Discrimination", *Journal of Economic Perspectives*, Vol. 12, No. 2, pp. 101-116.

Hübler, O. (2016), "Height and Wages", in John Komlos and Inas R. Kelly (eds), *The Oxford Handbook of Economics and Human Biology*, New York: Oxford University Press.

Jones, F.L. and Kelley, J. (1984), "Decomposing Differences between Groups: A Cautionary Note on Measuring Discrimination", *Sociological Methods and Research*, Vol. 12, No. 3, pp. 323-343.

Komlos, J. and Baur, M. (2003), "From the Tallest to (One of) the Fattest: The Enigmatic Fate of the American Population in the 20th Century", CESifo Working Paper No. 1028.

Knight, J. and Song, L. (2003), "Increasing Urban Wage Inequality in China", *The Economics of Transition*, Vol. 11, No. 4, pp. 597-619.

Krueger, A. (1993), "How Computers Changed the Wage Structure: Evidence from Microdata, 1984-1989", *Quarterly Journal of Economics*, Vol. 108, No. 1, pp. 33-60.

Lee, L. (2012), "Decomposing Wage Differentials between Migrant Workers and Urban Workers in Urban China's Labor Markets", *China Economic Review*, Vol. 23, No. 2, pp. 461-470.

Lin, J.Y. (1987), "The Household Responsibility System Reform in China: A Peasant's Institutional Choice", *American Journal of Agriculture Economics*, Vol. 69, No. 2, pp. 410-415.

Liu, P.W., Meng, X., and Zhang, J. (2000), "Sectoral Gender Wage Differentials and Discrimination in the Transitional Chinese Economy", *Journal of Population Economics*, Vol. 13, No. 2, pp. 331-352.

Magnani, E. and Zhu, R. (2012), "Gender Wage Differentials among Rural–urban Migrants in China", *Regional Science and Urban Economics*, Vol. 42, No. 5, pp. 779-793.

Meng, X. (2012), ”Labor Market Outcomes and Reforms in China”, *Journal of Economic Perspectives*, Vol. 26, No. 4, pp. 75-101.

Meng, X. and Zhang, J. (2001), “The Two-Tier Labor Market in Urban China: Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai”, *Journal of Comparative Economics*, Vol. 29, No. 3, pp. 485-504.

Mincer, J. (1991), “Human Capital, Technology, and the Wage Structure: What Do Time Series Show?”, NBER Working Paper No. 3581, Cambridge, MA: National Bureau of Economic Research.

Naughton, B. (2007), *The Chinese Economy: Transitions and Growth*, Cambridge, MA: MIT Press.

Oaxaca, R. (1973), “Male-female Differentials in Urban Labor Markets”, *International Economic Review*, Vol. 14, No. 3, pp. 693-709.

Oaxaca, R. and M. Ransom (1994), “On Discrimination and the Decomposition of Wage Differentials”, *Journal of Econometrics*, Vol. 61, No. 1, pp. 5-21.

Persico, N., Postlewaite, A. and Silverman, D. (2004), “The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height”, *Journal of Political Economy*, Vol. 112, No. 5, pp. 1019-1053.

Rozelle, S., Dong, X.-Y., Zhang, L. and Mason, A. (2002), “Gender Wage Gaps in Post-Reform Rural China”, *Pacific Economic Review*, Vol. 7, No. 1, pp. 157-179.

Wang, F.L. (2005), *Organizing through Division and Exclusion: China’s Hukou System*, Stanford, CA: Stanford University Press.

Zhang, L., Sharpe, R.V., Li, S. and Darity, W.A. (2016), “Wage Differentials between Urban and Rural-Urban Migrant Workers in China”, *China Economic Review*, Vol. 41 (December), pp. 222-233.