

# **Hukou-based Labor Market Discrimination and Ownership Structure in Urban China: A Double-selectivity Approach**

**Yang Song**

**School of Economics, Renmin University of China**

## **1. Introduction**

It has been reported that since 2003, Chinese factories in a few coastal cities have been facing a labor shortage in that firms have difficulty hiring enough workers as they want. As time goes by, the phenomenon of labor shortage has not disappeared but has spread to more cities. In July 2010, the media reported that factories in Guangdong have even hired many illegal workers from Vietnam and elsewhere (Chan, 2010). In contrast, we still observe hundreds of millions of people staying in the rural area and earning very low wage. Some scholars called the low-wage workers in the rural area as rural labor surplus (Knight et al., 2011). Why do these rural workers choose to work in the agriculture sector and earn very little rather than migrating to cities for better-paying jobs?

Previous studies have provided several reasons for the co-existence of urban labor shortage and rural labor surplus, such as credit market imperfection with costly migration, and the large amount of cost of living for rural-to-urban migrants in cities due to the household registration system (*hukou*) (Chau et al., 2012; Knight et al., 2011). Nonetheless, there is another possible reason for this puzzle that most of previous studies have ignored: rural-to-urban migrants face labor market discrimination in cities due to their *hukou* status, which reduces their expected payoffs associated with migrating to urban areas. In January 1958, the central government issued the Regulations of household Registration which required that each person born in China should be classified either rural *hukou* or urban *hukou* in a given location. This paper attempts to provide empirical evidence on the labor market discrimination against rural *hukou* holders in Chinese cities, so as to offer another explanation for the simultaneous urban labor shortage and rural labor surplus.

Specifically, this paper examines the extent of labor market discrimination against rural *hukou* holders in state-owned enterprises and private firms,

respectively. Based on previous literature, the urban labor market in China is segmented involving state-owned enterprises (SOE) and private sector (Song and Li, 2010; Chen et al., 2005). Specifically, comparable workers earn more in the SOE sector than in the private sector (Dong and Xu, 2009). Furthermore, the high-wage jobs in the SOE sector are rationed in the sense that many workers who are willing to and also able to work in the SOE sector cannot find jobs there since there are not enough jobs available in the SOE sector (Knight and Song, 2005).

To my knowledge, this paper is the first one to test the hukou-based discrimination in different ownership sectors. Since the Chinese government still has some influence over the pay in state-owned enterprises (SOEs), SOEs provide higher wages and benefits as well as better job security compared to private firms, while the wages in private firms are largely determined by the market forces. Hence, the results of this study can help us better understand whether the hukou-based discrimination is driven by market behaviors or by institutional forces.

An important methodological contribution of this study is to adopt a

double-selectivity approach to account for selectivity both in sector and hukou status determination. The first selectivity comes from the possibility that stated-owned enterprises selectively hire people with certain characteristics. Another selection problem arises because a person's hukou status is not randomly assigned. Currently, the most important determinant of a person's *hukou* type when the person was born is parental *hukou* types. Local governments of Chinese cities can grant urban hukou to some rural hukou holders who meet the criteria specified by each local government in their own jurisdictions. In general, selectivity bias may arise when a model is applied to a non-random sub-sample of the underlying population. In our context, people who are working in the SOE sector and who have urban hukou are not drawn randomly from an underlying population. The selectivity issue of hukou status has never been dealt with or even considered in previous literature. This paper will be the first one to apply the double-selection framework to study labor market discrimination in China.

The structure of this paper is as follows. Section 2 provides some institutional background of the current Chinese labor market with particular attention to the labor market segmentation and the hukou system. In

Sections 3 and 4, I will review previous literature and elaborate the contributions of this paper. Section 5 describes the methodology of studying wage discrimination in two ownership sectors and provides various model specifications. In Section 6, I present the data description and report the econometric results. Section 7 concludes.

## **2. Institutional background of the Chinese labor market**

### **2.1 Labor market segmentation between SOEs and private firms**

China has the usual dimension of labor market segmentation – some jobs being much better than others, with not enough employment in the better job segments for all who would like these jobs and are capable of performing them. As many scholars have pointed out, the main feature of the current labor market in urban China is a segmented labor market involving state-owned enterprises and private enterprises (Song and Li, 2010; Knight and Song, 2005).

Specifically, state-owned enterprises (SOEs) still provide higher wages and

benefits as well as better job security than private firms. Zhu (2010) argued that SOEs pay the wages higher than market-clearing level, while the wages in non-SOEs are set at the market-clearing level.

Furthermore, the jobs in SOEs are strictly rationed, in that many workers in private enterprises who are willing and able to work in SOEs cannot find a job in SOEs (Demurger et al., 2012). In recent years, a considerable number of people applied for very limited jobs vacancies in SOEs. Accordingly, even though many applicants possess the job qualifications, they could not obtain jobs in SOEs due to relatively fewer vacancies and intense competition.

## **2.2 Unique hukou system**

In addition to the labor market segmentation, the current Chinese labor market has a unique household registration system (*hukou*).

Every Chinese citizen holds either an urban or rural hukou in a particular location. Currently, the most important determinant of a person's *hukou* type when the person was born is parental *hukou* types. One's *hukou* status would

remain unchanged no matter where the individual physically moved, unless he or she goes through a formal procedure of *hukou* conversion. The criteria to convert the *hukou* status from rural to urban vary from one place to another, since local governments have now received full power and discretion to make their own *hukou* policies within their administrative jurisdictions. Some local governments admit local urban *hukou* to people who have salaried jobs or own a house in their administrative areas, or possess some types of occupational skills fitting the requirements stipulated by local governments (Chan and Buckingham, 2008). In Guangdong, the government converts various criteria, such as schooling, working skills, contributions to social insurance, volunteering and blood donation, into credit points, which accumulate to help obtaining local urban *hukou* (Cai, 2011).

### **3. Past research on labor market discrimination against rural hukou**

The discrimination that migrating rural *hukou* holders face is well documented in the literature (Meng, 2012). Most of previous studies have used the Blinder-Oaxaca decomposition technique to conclude there is some

degree of labor market discrimination against rural hukou in urban China. The basic idea of the Blinder-Oaxaca decomposition is that differences in average wages can be decomposed to differences in characteristics (sometimes called endowment) and differences in returns to these characteristics (also called coefficients). The latter is conventionally used as an estimate of discrimination (Oaxaca, 2007). Gravemeyer, Gries and Xue (2010) employed a Blinder-Oaxaca decomposition analysis to estimate the wage differential between workers with urban and rural hukou in Shenzhen city, and found that on average and all else being equal, 52.9 per cent of the higher wage of urban *hukou* holders is due to higher returns to the different characteristics and attributes, while 47 per cent is due to the higher endowments in characteristics such as education. This would imply that 52.9 percent of the average wage difference between the two groups cannot be explained by observed characteristics. Two recent studies tested the hukou-based wage discrimination using more broader measures of remuneration including insurance contributions and as well as bonuses earned (Lee, 2012; Frijters et al., 2010). The results suggest that migrants still face a significant amount of labor market discrimination.

Nonetheless, very few studies have analyzed the discrimination in state-owned enterprises (SOE) and private firms separately. Chen and Hoy (2011) conducted a qualitative research on discriminatory behaviors in state-owned and private manufacturing companies in Shanghai, and argued that the wage discrimination is larger in state-owned enterprises than in private sectors. However, they failed to provide quantitative evidence on the exact extent of labor market discrimination in two ownership sectors.

To my knowledge, only one previous study explored quantitatively the differential effects of hukou status on wages between different economic sectors. Gagnon *et al* (2011) utilized part of the Census data in 2005 and analyzed the wage differential between urban and rural hukou holder in three sectors, including formal sector, self-employment, and “no contract” sector. “No contract” sector is defined as jobs without a written labor contract, and formal sector is defined as wage and salary jobs with a written labor contract. Using this division, they found that the wage differential between urban and rural hukou holders is much larger in the formal sector than in the other two sectors. More specifically, the wage differential is negligible for self-employment and “no contract” workers. Furthermore,

even for the very small wage differential, more than 90% of it could be explained by the human capital variables. That is, the hukou-based wage discrimination in self-employment and “no contract” sector is almost non-existent. In contrast, urban hukou holders earn 60% more than rural hukou holders in the formal sector, and nearly 35% of this wage differential cannot be explained by observed characteristics.

As can be seen, although the study above compared the wage discrimination in different sectors, it did not consider the division between SOEs and private firms. A considerable share of private firms has labor contracts with their employees and thus belongs to the formal sector defined as in Gagnon *et al* (2011). Therefore, we could conclude that no previous study has provided quantitative evidence on labor market discrimination against rural hukou holders in different urban sectors.

#### **4. My contributions**

To my knowledge, this paper is the first attempt to explore the extent of wage discrimination against rural hukou in SOEs and private firms separately. Wage discrimination is defined as the situation in which two

groups of workers who have the same characteristics receive different wages (Becker, 1971; Aigner and Cain, 1977).

Additionally, this paper makes a methodological contribution to the empirical literature on labor market discrimination in China by adopting a double-selectivity approach to account for selectivity both in sector and hukou status determination. The first selectivity comes from the possibility that stated-owned enterprises selectively hire people with certain characteristics. Another selection problem arises because the hukou status is inherited from the parents and thus not randomly assigned. By the special institutional feature of the SOE sector, we employ an excluded instrument in the sector determination equation, which is a dummy variable denoting whether the person got this job through government recommendation. Since the government in China still has much control over SOEs, the government sometimes introduces some people to work in state-owned enterprises, but this kind of government reference is very rare in private firms. In addition, we use the number of siblings as the excluded instrument for the hukou determination equation. Due to different population policies implemented in urban and rural areas, rural hukou holders in the same age cohort have more

siblings than urban hukou holders, so the number of siblings can serve as an exogenous variation in the hukou status for identification.

Finally, this paper will provide more convincing empirical evidence on labor market discrimination by several robustness checks. In addition to using the original Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973), I will employ two alternative decomposition techniques developed by Cotton (1988) and by Oaxaca and Ransom (1988; 1994), respectively. Moreover, since the decomposition technique relies heavily on the linearity of the first stage regression, we also utilize the propensity score matching as another robustness check, by constructing more comparable groups of urban and rural hukou holders in each ownership sector. Specifically, we regard having an urban hukou as a “treatment” and estimate the average effect of treatment on the treated (ATT) based on one-to-one matching (also called the nearest-neighbor matching).

## **5. Methodology and model specification**

The most direct way to obtain estimates of the differential effects of hukou status on wages between state-owned enterprises (SOE) and private firms is

to include two dummy variables denoting the firm ownership and hukou status as well as an interaction of these two dummy variables in a Mincer (1974) type wage regression. However, this specification ignores the fact that the wage structures for urban and rural hukou holders in SOEs and private firms may differ from one another. Allowance for different wage structures for two types of workers in two economic sectors requires estimation of four separate wage regressions.

In addition, since workers with each hukou status and in each ownership sector are not randomly drawn from the population, two selection issues need to be considered in order to consistently estimate four wage equations.

In summary, we have a system of six equations in total, with two selection equations and four wage equations as follows. Let  $S_i^*$  and  $U_i^*$  be two latent unobserved continuous variables to determine sector allocation and hukou status, respectively, for person  $i$ .  $W_j^i$  ( $i=U$  or  $R$ ,  $j= S$  or  $P$ ) denote the wages for urban and rural hukou holders in the SOE sector and private firms, respectively.

$$S_i^* = X_{li}' d_{li} + u_{li}, \text{ SOE sector determination.} \quad (1)$$

$$U_i^* = X_{2i}' d_{2i} + u_{2i}, \text{ Urban hukou determination.} \quad (2)$$

$$\ln W_S^U = X_S^U g_S^U + v_S^U. \quad (3)$$

$$\ln W_S^R = X_S^R g_S^R + v_S^R. \quad (4)$$

$$\ln W_P^U = X_P^U g_P^U + v_P^U. \quad (5)$$

$$\ln W_P^R = X_P^R g_P^R + v_P^R. \quad (6)$$

The error terms in the six equations above are assumed to have the following variance-covariance matrix.

$$\Sigma = \begin{bmatrix} \mathbf{S}_{u1} & \mathbf{r} & a_{11} & a_{12} & a_{13} & a_{14} \\ \mathbf{r} & \mathbf{S}_{u2} & a_{21} & a_{22} & a_{23} & a_{24} \\ a_{11} & a_{21} & \mathbf{S}_S^U & 0 & 0 & 0 \\ a_{12} & a_{22} & 0 & \mathbf{S}_S^R & 0 & 0 \\ a_{13} & a_{23} & 0 & 0 & \mathbf{S}_P^U & 0 \\ a_{14} & a_{24} & 0 & 0 & 0 & \mathbf{S}_P^R \end{bmatrix}. \quad (7)$$

## 5.1 No selectivity correction

In the absence of selection issues, we can employ OLS estimation to consistently estimate the four wage regressions from equation (3)-(6). This simple specification indeed assumes that all of the off-diagonal terms in (7) are zero. Given the OLS results, I then decompose the average log wage

differential into the endowment effect which can be explained by productivity-related characteristics, and the unexplained coefficient effect which is commonly used as the estimation for wage discrimination. Specifically, the decomposition of the wage differential between urban and rural hukou holders in a given sector  $j$  (SOE or private) can be written as follows.

$$\begin{aligned}\overline{\ln W_j^U} - \overline{\ln W_j^R} &= \overline{X_j^U}' \hat{g}_j^U - \overline{X_j^R}' \hat{g}_j^R \\ &= (\overline{X_j^U} - \overline{X_j^R})' g_j^* + [\overline{X_j^U}' (\hat{g}_j^U - g_j^*) + \overline{X_j^R}' (g_j^* - \hat{g}_j^R)].\end{aligned}\quad (8)$$

In the equations above,  $\overline{\ln W_j^U}$  and  $\overline{\ln W_j^R}$  are the predicted mean (log) wages for urban and rural hukou holders in sector  $j$  (S or P),  $\overline{X}'$  is the mean vector of wage determining variables,  $\hat{g}$  is a vector of estimated coefficients (wage structure), and  $g^*$  is the non-discriminatory wage structure. The first term above  $(\overline{X_j^U} - \overline{X_j^R})' g_j^*$  is attributed to different endowments (productivity-related characteristics) in two groups of workers and is usually called the endowment effect. The second term in the square bracket  $[\overline{X_j^U}' (\hat{g}_j^U - g_j^*) + \overline{X_j^R}' (g_j^* - \hat{g}_j^R)]$  is typically interpreted as the part of the wage gap that is associated with differences in returns (coefficients)

to the characteristics between the two groups, which often serves as the estimate for discrimination (Liu et al., 2000; Oaxaca, 2007).

A practical concern associated with the adoption of the Blinder-Oaxaca approach is called the index number problem. This refers to the fact that the decomposition of the wage gap is not unique, depending on the choice of the non-discriminatory wage structure, that is, the choice of  $g^*$ . The original version of the Blinder-Oaxaca approach assumes that  $g^*$  takes extreme values equal to the wage structure of one reference group in the study. For example, if  $g^* = \hat{g}^U$ , then we assume that the wage structure for urban hukou holders would be adopted in the absence of discrimination (Liu et al., 2000). This is a very common assumption in that the wage structure of the dominant group is usually considered as the hypothetical wage structure in the absence of discrimination. However, if one believes that the current wage structure for rural hukou holders represents what the competitive wage structure would be in the absence of discrimination, then  $g^* = \hat{g}^R$ . As mentioned earlier, most of studies on the Chinese labor market use this original version of the Blinder-Oaxaca decomposition method (Gagnon et al., 2011; Qu and Zhao, 2011; Gravemeyer et al., 2010; Liu et al., 2000), and

employ the two extreme wage structures to approximate the wage structure in the absence of discrimination.

Cotton (1988) argued that derived weights using averages of coefficients for two groups will be more accurate than coefficients for one reference group. According to this approach, the non-discriminatory wage structure in sector  $j$  is  $\hat{g}_j^* = f_j^U \hat{g}_j^U + f_j^R \hat{g}_j^R$ , where  $f_j^U$  and  $f_j^R$  are the proportions of urban and rural hukou holders in sector  $j$ , respectively. Thus, Cotton's weights are sometimes referred as population-weighted non-discriminatory wage structure (Liu et al., 2000).

An alternative way to decompose the wage differential is to obtain the competitive wage structure ( $g^*$ ) from a regression with the pooled sample (Oaxaca and Ransom, 1988; 1994). As they argued, if by the absence of discrimination one means the complete cessation of existing discrimination activity, then it might be reasonable to suppose that the resulting competitive wage structure would be the parameter vector estimated with the pooled sample of two groups of workers.

However, there is no consensus in international literature about which weighting scheme is unambiguously better than another (Powers et al., 2011). Thus, this paper will report the decomposition results using four alternative non-discriminatory wage structures, respectively, including the wage structure of urban hukou holders, rural hukou holders, population-weighted wage structure, and the wage structure from the pooled sample.

## **5.2 Double selectivity approach**

This section takes into account the two selection issues and assumes that the two selection processes are correlated ( $r \neq 0$ ). In this case we adopt the double-selectivity approach to estimate the system of equations (1)-(6) with the stochastic specification in (7). Tunali (1986) shows that if the error terms have a joint normal distribution, then a two-step estimation procedure can be applied to correct for the double-selectivity issue, which is very similar to Heckman's two step method (Heckman, 1979). Specifically, the two selection equations are estimated using a bivariate probit model, and augmented selectivity correction terms are derived as additional regressors in four wage equations.

Let “S” be a dummy variable equal to 1 if  $S_i^* \geq 0$  and 0 otherwise. Similarly, we define “U” as a dummy variable equal to 1 if  $U_i^* \geq 0$  and 0 if  $U_i^* < 0$ . S and U indicate whether an individual is working in a state-owned enterprise (SOE) and has an urban hukou, respectively. After estimating the bivariate probit model for the two selection equations, we can obtain the following selectivity terms. Set  $C_1 = X_{1i}'d_{1i}$ ,  $C_2 = X_{2i}'d_{2i}$ ,  $M_1 = \frac{C_1 - rC_2}{\sqrt{1-r^2}}$ , and

$$M_2 = \frac{C_2 - rC_1}{\sqrt{1-r^2}}. \Phi_2 \text{ is the bivariate standard normal distribution function,}$$

$f$  is the univariate standard normal density function, and  $\Phi$  is the cumulative standard normal distribution. The selectivity terms generated by

$$\text{the bivariate estimation are } \hat{I}_S^U = \frac{f(C_1)\Phi(M_2)}{\Phi_2(C_1, C_2; r)}; \hat{I}_U^S = \frac{f(C_2)\Phi(M_1)}{\Phi_2(C_1, C_2; r)};$$

$$\hat{I}_S^R = \frac{f(C_1)\Phi(-M_2)}{\Phi_2(C_1, -C_2; -r)}; \hat{I}_R^S = -\frac{f(C_2)\Phi(M_1)}{\Phi_2(C_1, -C_2; -r)}; \hat{I}_P^U = -\frac{f(C_1)\Phi(-M_2)}{\Phi_2(-C_1, C_2; -r)};$$

$$\hat{I}_U^P = -\frac{f(C_2)\Phi(-M_1)}{\Phi_2(-C_1, C_2; -r)}; \hat{I}_P^R = -\frac{f(C_1)\Phi(-M_2)}{\Phi_2(-C_1, -C_2; r)}; \hat{I}_R^P = -\frac{f(C_2)\Phi(-M_1)}{\Phi_2(-C_1, -C_2; r)}.$$

In the second step, I estimate four augmented wage equations for urban and rural hukou holders in two ownership sectors as specified below.

$$\ln W_S^U = X_S^U g_S^U + a_{11} \hat{I}_S^U + a_{12} \hat{I}_U^S + w_S^U, \quad (9)$$

$$\ln W_S^R = X_S^R g_S^R + a_{21} \hat{I}_S^R + a_{22} \hat{I}_R^S + w_S^R, \quad (10)$$

$$\ln W_P^U = X_P^U g_P^U + a_{31} \hat{I}_P^U + a_{32} \hat{I}_U^P + w_P^U, \quad (11)$$

$$\ln W_P^R = X_P^R g_P^R + a_{41} \hat{I}_P^R + a_{42} \hat{I}_R^P + w_P^R. \quad (12)$$

Although the non-linearity of the selection equations can in principle help the identification for our system of equations, it is better to include some excluded variables that are assumed to be relevant only for one selection equation and excluded from the other selection and the wage equations (Rabe, 2011). The excluded instrument that identifies SOE sector determination equation is a dummy variable indicating whether the individual is recommended by the government for this job. Since the government in China still has much control over SOEs, the government sometimes introduces some people to work in state-owned enterprises, but this kind of government reference is very rare in private firms. These government introductions are usually made through various methods, such as a phone call, a letter, or a face-to-face conversation, and can help some people to find a job in SOEs. Thus, whether or not to have government reference for a job is a good indicator for the sector determination.

Moreover, since state-owned enterprises usually have established wage schemes for workers with different characteristics such as education and experience, it is unlikely that the government reference would directly affect workers' wages in the SOE sector.

In addition, we use the number of siblings as the excluded instrument for the hukou determination equation. Due to different population policies implemented in urban and rural areas, rural hukou holders in the same age cohort have more siblings than urban hukou holders. Specifically, the One-child policy in China implemented from 1978 requires that all urban hukou couples have only one child, but rural hukou couples can have two children (Greenhalgh, 1986). Accordingly, for people born after 1978, many rural hukou holders have one sibling, but most of urban hukou holders have no siblings at all. Furthermore, from 1950s to 1970s, the government policy encouraged rural people to have more children by requiring each family in rural areas turn over a fix quantity of agriculture products to the government. This policy led rural people to have children and increase agriculture production. As a consequence, older people with rural hukou tend to have more siblings than those with urban hukou in the same age cohort (Zhai,

1991). These policies above made the number of siblings serve as an exogenous variation in the hukou status for identification. Finally, it is hard to think of the channels through which the number of siblings can have direct impact on either the wages or the sector allocation.

Given the estimates from augmented wage regressions (9)-(12), I can decompose the average wage differentials between urban and rural hukou holders in state-owned enterprises and private firms, respectively. However, the decomposition becomes more complicated when there are selectivity correction terms. The complexity arises from the way to treat the selection term when we decompose the average (log) wage differential. Previous literature has provided two major ways to deal with the selectivity term in decomposition. Dolton and Makepeace (1986) treat the selectivity-correction terms in the same way as the other regressors in the wage functions, while Reimers (1983) proposes that the selectivity-correction terms can be used to modify the wage differentials, and then the modified wage differential is decomposed to endowment effect and coefficient effect. As Beller and Blau (1988) argued, the adjustment for selectivity bias is important, particularly when one is interested in potential

differences in employer treatment of two groups. In this case, the focus should be on wage offers rather than observed wages (Beller and Blau, 1988; Reimers, 1983). Therefore, I adopt Reimers's strategy to modify the wage differential, and then decompose the adjusted wage differential into the endowment effect and coefficient effect. Equation (13) and (14) provide the formulae for the wage decomposition in the SOE sector and private sector, respectively ( $g^* = \hat{g}^U$ ).

$$\overline{(\ln W_S^U - a_{11}\bar{I}_S - a_{12}\bar{I}_U)} - \overline{(\ln W_S^R - a_{21}\bar{I}_S - a_{22}\bar{I}_R)} = (\bar{X}_S^U - \bar{X}_S^R)' \hat{g}_S^U + \bar{X}_S^R' (\hat{g}_S^U - \hat{g}_S^R) \quad (13)$$

$$\overline{(\ln W_P^U - a_{31}\bar{I}_P - a_{32}\bar{I}_U)} - \overline{(\ln W_P^R - a_{41}\bar{I}_P - a_{42}\bar{I}_R)} = (\bar{X}_P^U - \bar{X}_P^R)' \hat{g}_P^U + \bar{X}_P^R' (\hat{g}_P^U - \hat{g}_P^R). \quad (14)$$

## 6. Data description and econometric results

I utilize 2008 wave of the Rural-Urban Migration in China (RUMiC) survey to analyze the proposed question. The Longitudinal Survey on Rural Urban Migration in China (RUMiC) consists of three parts: the Urban Household Survey, the Rural Household Survey and the Migrant Household Survey. It was initiated by a group of 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), which provides the

Scientific Use Files. The financial support for RUMiC was obtained from the Australian Research Council, the Australian Agency for International Development (AusAID), the Ford Foundation, IZA and the Chinese Foundation of Social Sciences.

The survey covers three groups of Chinese households: 5000 migrant households who worked in 15 designated cities (migrant survey) in nine provinces or metropolitan areas; 5000 urban local incumbent households in the same cities (urban household survey), and 8000 rural households (rural household survey). Most of the migrants have rural hukou while the urban sample largely takes urban hukou. This feature gives us enough observations for both urban and rural hukou holders. For the research purpose of this paper, I will only use the urban sample and migrant sample in 2008 wave of the survey. The nine provinces or metropolitan areas covered in the survey are Shanghai, Guangdong, Jiangsu, Zhejiang, Anhui, Hubei, Sichuan, Chongqing and Henan. The first four of these are the largest migrant destinations, and the remaining five are the largest migration sending areas. The RUMiC survey of migrant workers in urban cities, to the best of our knowledge, is the only random sample of migrant workers for China so far.

The survey records detailed individual information, such as monthly earnings, hours of work, demographic characteristics, and work and employment information.

For the purpose of this study, I restrict the sample to employees aged from 18 to 60 in urban areas. Furthermore, people working at foreign-owned companies and self-employed people are excluded from our sample. The remaining sample consists of 5,012 urban hukou holders and 3,795 rural hukou holders. The total sample size is 8807. The summary statistics of the key variables in the data set are presented in Table 1.

*Table 1 Inserted Here*

As can be seen from the table, among 5,012 urban hukou holders, more than half of them (3,198) are working in state-owned enterprises, whereas only a very small proportion of rural hukou holders are working in this high wage sector. In addition, the average monthly earnings are 2294.6 (yuan) and 1529.8 in the SOE sector for urban and rural hukou holders, respectively, so the average earning differential is about 50%. On the other hand, the

average differential in terms of monthly earnings is about 30% in the private firms.<sup>1</sup> Given this evidence so far, we cannot conclude that the wage discrimination is more severe in SOEs than private firms because the wage differentials could be driven by differences in productivity-related characteristics. That is, the raw wage differential presented here is only the average one but does not control for workers' characteristics.

Table 1 shows that on average, urban hukou holders are older than rural hukou holders indicating that most of rural migrants in cities are young people aged around 30. Furthermore, more than 60% of rural hukou holders are males, while the gender composition is more balanced for urban hukou holders. In terms of human capital variables, urban hukou holders have more education than rural hukou holders, and the average education is slightly higher in state-owned enterprises than in the private sector. The data set also contains a variable to reflect people's social capital measured by the

---

<sup>1</sup> In the data set, some firms' ownership are identified as collective-owned. This is a special firm ownership in China due to the planned economy in which some firms were owned by a community, a city, etc. Since the transformation from planned economy to market economy in 1978, most of the collective-owned firms have become more market-oriented and privatized especially after 1990s (Sun, 2008). Therefore, although some of the original collective-owned firms are still officially called collective, they are very similar to private enterprises in practice. So I put the collective-owned firms in the dataset to the category of private enterprises.

number of greetings sent in the past Chinese new year festival. As seen in Table 1, urban hukou holders in the SOE sector have the highest social capital, and urban hukou holders have more social capital than rural hukou holders in each sector. For a job-related characteristic, workers with urban hukou tend to work in larger firms than those with rural hukou. In addition, the SOE firms are typically larger than private firms.

The excluded variable to identify the sector selection equation is a dummy variable indicating whether the individual is recommended by the government for this job. Table 1 indicates that nearly half of the urban hukou holders in the SOE sector are recommended by the government for this job, while this ratio is much smaller in private sector. This is because the SOE sector has very close relationship with the government in China, and the government sometimes introduces some people to the state-owned enterprises, but this kind of government reference is very rare in the private firms. Moreover, the excluded variable to identify the hukou determination equation is the number of siblings. The fact that urban hukou holders on average have more children than rural hukou holders is mostly because that urban hukou holders are older than migrant rural hukou holders, and were

not influenced by the One-child policy. Remember that this table only reports unconditional means, rather than means conditional on age. Thus, we cannot know whether this instrument is valid or not given this table.

As Giulietti et al. (2012) claimed, in the RUMiC dataset, hourly wages are more prone to measurement error (being calculated by combining monthly wages and weekly hours worked), so I choose to use monthly wages as our dependent variable.

## **6.1 Evidence on wage discrimination and ownership structure**

### **6.1.1 No selectivity correction**

In the first step, the wage equation is estimated for urban and rural hukou holders separately within each sector without dealing with the selectivity issue and the results are reported in Table 2. The dependent variables in all of wage equations are the natural logarithm of monthly earnings measured in nominal Chinese currency (yuan) in 2007. The independent variables include potential labor market experience and its square, completed years of schooling, social capital, seniority, health status, weight, height, school

performance, gender, marriage status, a dummy variable indicating job training, firm size, and a set of dummies including occupation, industry, and province dummies to control for the difference in cost of living in different regions.

*Table 2 Inserted Here*

Table 2 presents the coefficients of the major human capital variables, and the coefficients of other variables are not reported. As the table shows, the return to schooling is around 3 percent in China and is very similar between urban and rural hukou holders in each sector, which is consistent with recent studies (Song, 2012; Demurger et al., 2012). In addition, it turns out that the coefficients on human capital variables are all statistically significant for rural hukou holders and mostly larger than the coefficients for urban hukou holders in both sectors, whereas the coefficients on social capital display the opposite pattern. This may imply that the wage structure is more market-oriented for rural hukou holders, while social networking is very important for the wages of urban hukou workers. Comparing the state-owned enterprises (SOE) with private firms, we can find that the wage

structure is very similar for rural hukou holders in these two sectors. For urban hukou holders, the return to major human capital variables is higher in private firms than in SOEs, including years of schooling and work experience. However, it is noteworthy that the return to social capital and firm tenure is higher in SOEs than private firms for urban hukou holders.

Another finding from Table 2 is that for urban hukou holders, males with the same observed characteristics earn about 30% more than females, which suggests an evidence of gender discrimination. For rural hukou holders, the gender wage premium is about 10%. We can conclude that the gender discrimination is more severe among urban hukou holders than among those with rural hukou.

Table 3 reports the decomposition of the overall wage differential into the explained and unexplained components in terms of log differentials using equation (8). The four alternative non-discriminatory wage structures are used as weights for the wage decomposition, including wage structures for urban and rural hukou holders, respectively, population-weighted and pooled-sample wage structures. The raw log wage differential is higher in

SOE sector than in private sector between urban and rural hukou holders. Furthermore, the entire wage differential in the private sector between urban and rural hukou holders is due to people's productivity-related characteristics, meaning that there is no any discrimination in the private sector. In contrast, there is some proportion of wage differential in the SOE sector that can not be explained by observed characteristics, although it is very small.

*Table 3 Inserted Here*

### **6.1.2 Double-selectivity correction**

Table 4 presents the results of two selection equations using a bivariate probit model. For the sector determination equation, the estimated coefficients are standard ones in that people who have more education and work experience are more likely to work in the SOE sector. In addition, the coefficient on the excluded variable (Gov. reference) is positively significant, partly verifying my hypothesis that people with government reference are more likely to be hired in SOEs because the Chinese government still has some control over SOEs in terms of hiring decisions.

For the hukou determination equation, we can see that the excluded variable (number of siblings) also has expected effect on the probability of having an urban hukou. More siblings are associated with a smaller probability of being an urban hukou holder. That is, urban hukou holders with the same observed characteristics have fewer number of siblings than those with rural hukou. As the table shows,  $\rho$  is significantly positive implying that the inclination to obtain urban hukou and the inclination to work in the SOE sector are positively related to each other.

*Table 4 Inserted Here*

After estimating the selection equations, I run four augmented wage equations with with Tunali double-selectivity terms.<sup>2</sup> Accordingly, the wage decomposition results with double-selectivity correction are reported in Table 5.

*Table 5 Inserted Here*

---

<sup>2</sup> The results of augmented wage regressions are available upon request.

The decomposition results show that the wage discrimination is unambiguously more severe in the SOE sector than in private firms, regardless of the weight used for non-discriminatory wage structure. For example, if we take urban hukou holders' wage structure as the non-discriminatory wage structure (which is the most frequently used weight in previous literature), the unexplained log wage differential is 0.384 in SOEs and 0.049 in private firms. If we translate the log differential into percentage points, the results from Tunali double-selectivity approach suggest that for observationally-equivalent workers, urban hukou holders earn about 50% more than rural hukou holders do in the SOE sector, but only 5% more in the private sector.

Compared with the results in table 3, the wage discrimination is aggravated by the selectivity adjustment, especially in the SOE sector. This is because rural hukou holders are positively selected into the SOE sector, while there is a negative correlation between the inclination to obtain an urban hukou and the wage rate in the SOE sector. In other words, mean wage offers are above mean observed wages among urban hukou holders and below mean observed wages among rural hukou holders.

To recapitulate, the econometric results presented in this section suggest that rural hukou holders face more wage discrimination in the SOE sector in urban China. Estimates without considering the two selection issues tend to under-estimate the extent of labor market discrimination against rural hukou holders.

## **6.2 Robustness check: propensity score matching**

As a robustness check, we examine the wage discrimination using the propensity score matching technique. We first run a logit regression to determine each person's probability of being an urban hukou holder. Then, for comparable urban and rural hukou holders with the nearest propensity scores, we calculate the wage differential in the SOE and private sector separately. This differential for comparable workers is also called the average treatment effect on the treated (ATT), which is equal to the average monthly income for urban hukou holders minus the counterfactual average income if they held rural hukou. The results from propensity score matching

are highly comparable to those from the decomposition analysis.<sup>3</sup> Specifically, the estimated log wage differentials (i.e., ATT) for comparable urban and rural hukou holders are 0.381 (significant at 10% level) in the SOE sector and 0.075 (not significant) in private firms.

## 7. Conclusions

The results of this paper demonstrate that rural-to-urban migrants with rural hukou face much wage discrimination in the high-wage SOE sector in urban China, which reduces their expected payoffs associated with migrating to the city. Specifically, the econometric results show that for observationally-equivalent workers, urban hukou holders earn about 50% more than rural hukou holders do in the SOE sector, but only 5% more in the private sector (urban-weighted decomposition).

We have to acknowledge that the regression-based decomposition technique has limitations on estimating the extent of discrimination since it is not possible to include all relevant individual characteristics and job-related features in regressions. However, this paper has tried to minimize the

---

<sup>3</sup> We discard part of the sample in the control group with extreme propensity scores to satisfy the overlap condition. Also, the balancing tests of the two groups were verified, and the results are available upon request.

possibility that unobserved characteristics play a large role in explaining the wage differential by including as many control variables as the data set allows, such as school performance, health status, job training, firm size, and a set of dummies including occupation, industry, and province dummies.

To make the results more convincing, I have conducted various robustness checks including trying alternative non-discriminatory wage structure and using the propensity score matching technique. All of our robustness checks lend the support to our major finding that rural hukou holders face more wage discrimination in the SOE sector than in the private sector.

## **References:**

- Aigner, Dennis and Glen Cain, 1977. "Statistical Theories of Discrimination." *Industrial & Labor Relations Review* 30 (2): 175-187.
- Arrow, Kenneth, 1973. "The Theory of Discrimination." in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, Princeton University Press, pp. 3-33.
- Becker, Gary S, 1971. *The Economics of Discrimination*, 2nd ed. Chicago: University of Chicago.

- Beller, Andrew, and Francine Blau. 1988. "Trends in Earnings Differentials by Gender, 1971-1981." *Industrial and Labor Relations Review* 41 (4): 513-529.
- Blinder, A, 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal of Human Resources* 8 (4): 436-455.
- Cai, Fang, 2011. "Hukou System Reform and Unification of Rural-urban Social Welfare." *China & World Economy* 19 (3): 33-48.
- Chan, Kam Wing. 2010. "The Household Registration System and Migrant Labor in China: Notes on a Debate." *Population and Development Review* 36 (2): 357-364.
- Chan, Kam Wing and Will Buckingham. 2008. "Is China Abolishing the *Hukou* System?" *China Quarterly* 195: 582-606.
- Chen, Yu and Caroline Hoy. 2011. "Explaining Migrants' Economic Vulnerability in Urban China." *Asian Population Studies* 7 (2): 123-136.
- Chen, Y., Démurger, S., & Fournier, M. 2005. Earnings differentials and ownership structure in Chinese enterprises. *Economic Development and Cultural Change* 53(4): 933-958.
- Cotton, Jeremiah. 1988. "On the Decomposition of Wage Differentials." *Review of Economics and Statistics* 70 (2): 236-243.
- Demurger, S., S. Li, J. Yang. 2012. "Earnings Differentials between the Public and Private Sectors in China: Exploring Changes for Urban Local Residents in the 2000s." *China Economic Review* 23: 138-153.

- Dolton, P.J. and Makepeace, G.H. 1986. "Sample Selection and Male-female Wage Differentials in the Graduate Labour Market", *Oxford Economic Papers*: 143-57.
- Dong, Xiao-yuan, and Lixin Colin Xu. 2009. "Labor Restructuring in China: Toward a Functioning Labor Market." *Journal of Comparative Economics*, 37(2): 287-305.
- Frijters, P., Lee, L., & Meng, X. 2010. "Jobs, Working Hours, and Remuneration Packages of Migrants and Urban Residents." In *The Great Migration: Rural-Urban Migration in China and Indonesia*, ed. Effendi Tadjuddin Nur, Li Shi, Manning Chris, Meng Xin. Cheltenham, U.K: Edward Elgar Publishing Limited.
- Gagnon, Jason, Theodora Xenogiani, and Chunbing Xing, 2011. "Are All Migrants Really Worse off in Urban Labour Markets: New Empirical Evidence from China, IZA Discussion Paper No. 6268.
- Giulietti, Corrado, Guangjie Ning, and Klaus F. Zimmermann. 2012. "Self-employment of Rural-to-urban Migrants in China." *International Journal of Manpower* 33 (1): 96-117.
- Gravemeyer, Stefan, Thomas Gries and Jinjun Xue. 2010. "Income Determination and Income Discrimination in Shenzhen." *Urban Studies*: 1-19.
- Greenhalgh, S. 1986. "Shifts in China's population policy, 1984-86: Views from the central, provincial, and local levels. " *Population and Development Review*: 491-515.
- Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*

47(1):153–161.

Knight, J, L. Song. 2005. *Towards a Labour Market in China*. Oxford: Oxford University Press.

Lee, L. (2012). “Decomposing wage differentials between migrant workers and urban workers in urban China's labor markets.” *China Economic Review* 23 (2): 461-470.

Liu, P., X. Meng, J. Zhang. 2000. “Sectoral Gender Wage Differentials and Discrimination in the Transitional Chinese Economy.” *Journal of Population Economics* 13: 331-352.

Meng, X. 2012. “Labor Market Outcomes and Reforms in China. ” *The Journal of Economic Perspectives* 26(4): 75-101.

Mincer, J. 1974. “Schooling, Experience, and Earnings.” New York: National Bureau of Economic Research.

Nancy, Chau, Ravi Kanbur, and Yu Qin. 2012. “Do Public Work Schemes Deter or Encourage Outmigration? Empirical Evidence from China.” CEPR Discussion Paper 8778.

Neuman, S. and Oaxaca, R.L. 2004. “Wage Decompositions with Selectivity Corrected Wage Equations: A Methodological Note.” *Journal of Economic Inequality* 2: 3-10.

Oaxaca, R. 1973. “Male–female Wage Differentials in Urban Labour Markets, ” *International Economic Review* 9(3): 693–709.

- Oaxaca, R. and R. Ransom. 1988. "Searching for the Effect of Unionism on the Wages of Union and Nonunion Workers." *Journal of Labor Research* 9: 139-148.
- Oaxaca, R. and Ransom, R. 1994 "On Discrimination and the Decomposition of Wage Differentials, *Journal of Econometrics* 61: 5-21.
- Oaxaca, R. 2007. "The Challenge of Measuring Labor Market Discrimination against Women. *Swedish Economic Policy Review* 14: 199-231.
- Powers, Daniel A, Hirotoishi Yoshioka, and Myeong-Su Yun. 2011. "Mvdcmp: Multivariate Decomposition for Nonlinear Response Models." *The Stata Journal*: 11 (4): 556-576.
- Qu, Z., Z. Zhao. 2011. "Evolution of the Chinese Rural-Urban Migrant Labor Market from 2002 to 2007." IZA Discussion Paper No. 5421.
- Rabe, B. 2011. "Dual-earner Migration: Earnings Gains, Employment and Self-selection." *Journal of Population Economics* 24 (2): 477-497.
- Reimers, C. W. 1983. "Labor Market Discrimination against Hispanic and Black Men." *Review of Economics and Statistics* 65: 570-579.
- Song, Yang. 2012. "Poverty Reduction in China: The Contribution of Popularizing Primary Education." *China & World Economy* 20 (1): 105-122.
- Song, Jin, Shi Li. 2010. "Ownership and Earnings Inequality in Urban China." *Frontiers of Economics in China* 5(4): 582-603.
- Sun, Jian. 2008. "The Study of China's Full Employment of Rural Labor." Doctoral

- Dissertation, Renmin University of China.
- Tunali, I. 1986. “A general Structure for Models of Double-selection and an Application to a Joint Migration/earnings Process with Remigration.” *Research in Labor Economics* 8 (B), 235–282.
- Zhai, Z. 1991. “The Economic Reasons for Rural Population Growth. ” *Population Research*, No. 4 (in Chinese).
- Zhang, Y., Chen, J. and Wong, P. 2011 “Effect of Trade Unions on Industrial Labor Income in China.” *Asian Politics & Policy* 3: 95–114.
- Zhu, Ling. 2010. “Improving Equity and Sustainability of Social Protection System in China.” *Chinese Journal of Population Science*, No. 5 (in Chinese).

## Tables

**Table 1 Summary Statistics of Key Variables**

	SOE		Private	
	<u>Urban hukou</u>	<u>Rural hukou</u>	<u>Urban hukou</u>	<u>Rural hukou</u>
Age	40.19 (9.75)	32.08 (10.79)	38.01 (10.16)	30.73 (10.24)
Years of schooling	12.53 (3.28)	9.65 (2.57)	11.88 (3.23)	9.36 (2.48)
Experience	21.67 (11.41)	16.25 (11.64)	20.12 (11.79)	15.12 (11.13)
Monthly earnings	2294.63 (1969.25)	1529.76 (748.40)	2018.49 (2450.10)	1500.99 (957.43)
Hourly wage	15.08 (23.36)	7.40 (4.28)	12.76 (22.54)	6.72 (4.47)

Male	0.58 (0.49)	0.65 (0.48)	0.55 (0.50)	0.61 (0.49)
Number of siblings	3.36 (1.48)	2.96 (1.46)	3.43 (1.59)	2.97 (1.36)
Gov. reference	0.45 (0.50)	0.03 (0.16)	0.16 (0.36)	0.003 (0.06)
Social capital	35.58 (46.51)	28.72 (46.86)	33.19 (44.03)	31.93 (44.89)
N	3,198	658	1,814	3,137

**Table 2 Log Wage Regressions without Selectivity Correction**

	SOE		Private	
	Urban hukou	Rural hukou	Urban hukou	Rural hukou
Years of schooling	0.028*** (0.005)	0.023*** (0.008)	0.033*** (0.007)	0.036*** (0.004)
Experience	0.002 (0.005)	0.018*** (0.006)	0.009 (0.006)	0.021*** (0.003)
Exp. Sq	-0.0002 (0.0001)	-0.0005*** (0.0001)	-0.0005*** (0.0001)	-0.0005*** (0.0001)
Male	0.228*** (0.035)	0.088* (0.047)	0.296*** (0.048)	0.092*** (0.023)
Social capital	0.0016*** (0.0002)	0.0002 (0.0003)	0.001*** (0.0004)	-0.0001 (0.0002)
Firm tenure	0.013*** (0.001)	0.015*** (0.004)	0.009*** (0.002)	0.019*** (0.002)
Constant	7.756*** (0.532)	5.652*** (0.583)	6.502** (0.653)	5.786*** (0.271)
N	2,679	639	1,455	3,040
Adjusted R-sq	0.1973	0.2275	0.1939	0.1725

Note: \*\*\* significant at 1%; \*\* significant at 5%; \* significant at 10%. Figures in parentheses are standard errors. The dependent variables in all of wage equations are the natural logarithm of monthly earnings in 2007. All regressions also control for health status, weight, height, school performance, gender, marriage status, a dummy variable indicating job training, firm size, and a set of dummies including occupation, industry, and province dummies.

**Table 3 Wage Decomposition without Selectivity Correction**

		<b>Urban-weighted</b>	<b>Rural-weighted</b>	<b>Pooled sample weighted</b>	<b>Population-weighted</b>
<b>SOE</b>	Total diff	0.294 (0.022)	0.294 (0.022)	0.294 (0.022)	0.294 (0.022)
	Explained	0.287 (0.063)	0.289 (0.101)	0.292 (0.021)	0.287 (0.055)
	Unexplained	0.008 (0.065)	0.005 (0.103)	0.003 (0.013)	0.007 (0.058)
<b>Private</b>	Total diff	0.162 (0.020)	0.162 (0.020)	0.162 (0.020)	0.162 (0.020)
	Explained	0.289 (0.031)	0.226 (0.090)	0.170 (0.016)	0.250 (0.058)
	Unexplained	-0.126 (0.035)	-0.064 (0.092)	-0.008 (0.009)	-0.087 (0.060)

Note: The main entries represent the decomposition of log wage differential, in which wage is measured by monthly earnings. Figures in parentheses are standard errors.

**Table 4 Probit Estimates for Selection Equations**

	<b><u>Bivariate Probit</u></b>
<b>SOE sector determination</b>	
Years of schooling	0.088 *** (0.009)
Experience	0.048 *** (0.008)
Exp. Sq	-0.001 *** (0.0002)
Male	-0.113 * (0.063)
Firm tenure	0.023 *** (0.003)
Gov. reference	0.727 *** (0.063)
Constant	-2.947 *** (0.743)
N	5359
P-value	0.000

---



---

***Hukou status determination***

Years of schooling	0.241 *** (0.018)
Experience	0.140 *** (0.015)
Exp. Sq	-0.002 *** (0.0003)
Male	-0.436 *** (0.114)
Firm tenure	0.044 *** (0.006)
Number of siblings	-0.028 ** (0.014)

---

N	5359
P-value	0.000

---

$\rho$	0.177 *** (0.050)
--------	----------------------

---

Note: \*\*\* significant at 1%; \*\* significant at 5%; \* significant at 10%. Figures in parentheses are standard errors. All regressions also control for variables that are included in the wage regressions.

**Table 5 Wage Decomposition with Double Selectivity**

**Correction**

		<b>Urban- weighted</b>	<b>Rural- weighted</b>	<b>Pooled sample weighted</b>	<b>Population- weighted</b>
<b>SOE</b>	Total diff (adjusted)	1.096 (0.393)	1.096 (0.393)	1.096 (0.393)	1.096 (0.393)
	Explained	0.712 (0.311)	-0.071 (0.205)	0.877 (0.316)	0.579 (0.260)
	Unexplained	0.384 (0.200)	1.167 (0.449)	0.218 (0.082)	0.517 (0.219)
<b>Private</b>	Total diff (adjusted)	0.322	0.322	0.322	0.322

---

	(0.112)	(0.112)	(0.112)	(0.112)
Explained	0.273	-0.038	0.263	0.077
	(0.118)	(0.252)	(0.057)	(0.165)
Unexplained	0.049	0.361	0.060	0.246
	(0.222)	(0.301)	(0.059)	(0.239)