Session Number: Session 6 (Plenary)
Time: Friday, September 25, 9:00-10:30

Paper Prepared for the Special IARIW-SAIM Conference on
"Measuring the Informal Economy in Developing Countries"

Kathmandu, Nepal, September 23-26, 2009

Are All Migrants Really Worse Off in Urban Labour Markets? New Empirical Evidence from China

Jason Gagnon, Theodora Xenogiani, Chunbing Xing, OECD

This paper is posted on the following website: http://www.iariw.org

Forthcoming, 2009

Are *all* migrants really worse off in urban labour markets?
New empirical evidence from China

Jason Gagnon[♦], Theodora Xenogiani[♣] and Chunbing Xing[♠]

March 2009

Abstract

The rapid and massive increase in rural-to-urban worker flows to the coast of China has drawn recent attention to the welfare of migrants in urban regions. This paper focuses on rural migrant labour outcomes (earnings, sector selection) in comparison to outcomes for both urban residents and urban migrants. Data are taken from a 1/5th random draw of the 2005 1% Chinese national census survey. Contrary to popular belief, we find no earnings discrimination against rural migrants compared to urban residents. Discrimination is evident, however, in formal/informal sector choice and when comparisons are made between the two groups of migrants.

JEL Classification: O15, R23, J24, J71
Keywords: Migration, China, Discrimination, Informal Employment

♦ OECD Development Centre\Paris School of Economics
♣ OECD Directorate for Employment, Labour and Social Affairs
♠ Beijing Normal University

1. Introduction

A combination of factors in recent years, notably the increasing internal rural-urban income gap[1] and the easing of internal migration restrictions (Cai, 2000), has led to a tremendous increase in rural-to-urban migration in China, the result largely due to individuals in search of work in its coastal regions. Indeed, the number of migrants has increased significantly over the years from an estimate of 68 million in 1996 (Liang, 2001) to one of nearly 140 million in 2003 (Huang and Zhan, 2005). According to nationally representative census data from 2005, migrant workers account for more than 20% of the labour force in the urban labour market[2]. There is little doubt as to whether the influx of migrants into urban China has contributed to economic growth (Liang, 2001; Song and Zhang, 2003), but concern prevails as to the implications of the accompanying geo-demographic transformation on individual welfare.

Particular interest is raised on how migrants fare in the urban labour market, especially in comparison to their urban resident counterparts. Research shows that migrants take up jobs in informal sectors, are paid less, are less likely to be covered by urban social security systems and occasionally find it difficult to get their settled salaries from their employers on time and enforced (China Labour Bulletin, 2008). The situation is exacerbated by China's resident registration (*hukou*) system, as most rural migrants retain their *rural hukou* status, despite the fact that they may spend a significant amount of time in urban areas. Hence, the potential for discrimination on wages is evident for migrants with *rural hukou* status but it may also be present in the general ability to access formal sector jobs.

This paper sets out to answer whether migrants are really worse off in urban labour markets in China. Our starting point is a common one found in the research literature on discrimination: the notion that wage gaps may owe to either differing levels of individual human capital brought to the labour market or from differing sets of skill-prices offered on this market. Indeed, it is a relevant starting point as it carries notable policy implications. If, for instance, rural migrants are paid less as a result of lower skills brought to the urban labour market, reforming the urban labour market will have little effect on their well-being, whereas providing them with education and appropriate skills will enable them to exploit better labour market opportunities. However, if migrants have significantly lower wages compared to urban residents[3] with identical characteristics, the root cause of wage differences *could* be discrimination. Reforming the labour market, if this is the case, would lead to an increase in welfare for migrants. Moreover, looking for differences within the subgroups of migrants will add additional insight into the determinants of discrimination.

We first estimate a multinomial logit model to determine the factors determining the choice of employment sector and then apply the results to a decomposition analysis to test for the existence of wage discrimination. The Oaxaca-Blinder (OB) decomposition framework (Blinder, 1973; Oaxaca, 1973) provides a useful analytical instrument with which to face the question of discrimination and is often

[1] The most recently published ratio is estimated at 3.2:1 (Chinese National Bureau of Statistics, 2007)
[2] In some relatively developed coastal regions, such as those of Guangdong and Fujian, the share of migrants is greater than 50%.
[3] For the remainder of the text, "urban resident" is meant to mean urban resident non-migrants.

applied to compute the explained (endowment) effect and the unexplained (discrimination) effect in potential discriminatory cases of gender or black/white income differentials in the U.S. We follow the literature in this sense but the comparison between the two groups, rural migrants and urban residents, is multifaceted as it involves two dimensions of potential discrimination instead of one, and our results hinge on this dichotomy. The first dimension is *hukou* status (rural *vs*. urban), while the second is migrant status (migrant *vs*. resident). We therefore introduce another reference group -namely urban migrants[4]- to distinguish these two effects. As urban migrants are different from urban residents only in terms of migrant status, we interpret the positive unexplained figure of our OB decomposition as a premium associated with migration (or a selection effect); and as urban migrants and rural migrants differ only by *hukou* status, we interpret the unexplained figure of the OB decomposition as discrimination against rural *hukou* status.

As we are concerned by the possibility of segmentation in the urban labour market, another feature we take into account is the distribution of individuals in the formal and informal sectors as this will have an effect on income differentials. This issue is also at the heart of policy preoccupations with respect to the working conditions of rural migrants and in particular their access to formal jobs. In light of this we apply a Brown decomposition exercise (an extended version of the Oaxaca-Blinder decomposition) between and within the different groups and report the findings in Appendix 2[5]. Secondly, because sectoral segmentation can have a substantial impact on social security coverage despite having a possible limited impact on income, we also apply an OB decomposition directly to differences in sectoral distribution.

The paper uses a recent nationally representative dataset, a one-fifth random draw from the 1% census data of China, to explore the issues raised above. Results show that most of the income differential between rural migrants and urban residents can be explained by differences in individual characteristics. However by using urban migrants as a "control" group, we find that the inexistence of discrimination effect we obtain while comparing rural migrants and urban residents is the result of a counterbalance between a discrimination effect against rural *hukou* status and a premium effect accrued by migrants. As for sectoral distribution, the simple Oaxaca-Blinder decomposition indicates that migrants (both rural and urban) are discriminated along sector choice[6]. The extent of discrimination is larger for rural migrants indicating a further discrimination against rural *hukou* status.

The policy implications of our results are clear. In terms of income, rural migrants enjoy a premium associated with migration and suffer discrimination in urban labour markets simultaneously. The reason that they earn less when we compare them with urban residents is due to their lower levels of human capital. Increasing the education level of rural migrants and providing them with useful skills will help increase their income and earning opportunities. As both rural and urban migrants face unfair treatment in sector choice, reforming the labour market, especially by removing sector barriers can help increase employment in the formal sector and facilitate earnings mobility for migrants.

[4] Migrants with an urban *hukou* status moving from one urban region to another.
[5] Cai and Wang (2006) also find discrimination along gender lines in China. We therefore report gender dissagregated results in Appendix 3.
[6] With more than 80% of the difference unexplained.

The paper is organised as follows: Section 2 provides a short discussion on the institutional background and a literature review on labour market segmentation in China. Section 3 describes the data. Section 4 presents and discusses the model specification and section 5 reports our basic empirical results, where we compare rural migrants and urban residents. In section 6, we present a more in depth analysis comparing different groups. The last section concludes.

2. Institutional background and literature review

There have been several written articles providing an overview of the institutional background of China's rural-urban migration, with particular emphasis on the *hukou* registration system (Cai, 2000; Zhao, 2005; Deng and Gustafsson, 2006; de la Rupelle, 2007). Despite the several reforms to the system since the 1970s, deliberate discrimination of migrants in cities remained legal until very recently, in order to reduce competition of rural migrants in urban centres (Cai, 2000). Today, urban China faces the challenge of integrating two distinct labour force groups, and it is still unclear as to whether they are complementary or competing for the same jobs. Rural migrants, it has been shown, have higher job mobility in the urban labour market and generally have very low tenure rates than urban residents. Despite segmentation, urban residents and migrants have increasingly competed over time for labour opportunities (Knight and Yueh, 2003). Although the two groups work in seemingly segmented labour markets, competition may be increasing as more internal migration and labour market reforms gain pace.

Rural migrants in urban centres can be divided into two groups: those who have obtained an urban *hukou* registration and those who have not. Migrants with an urban *hukou* are registered officially as urban residents, a prerequisite to be covered by the urban social security system and to gain access to various forms of public assistance. Moreover, once registered as urban residents, permanent migrants forfit their rural resident status, their right to agricultural land in their community of origin as well as their voting rights on village affairs. Both anecdotal evidence and academic research (Deng and Gustafsson, 2006 for example) indicate that rural migrants who successfully obtain an urban *hukou* registration are well integrated in urban society, at least after accumulating experience in the urban labour market over time[7].

Alternatively, many rural migrants retain their rural *hukou* status, whether deliberately or not, and thus retain rights on their rural land and a voice on political affairs in their village of origin. These migrants are less integrated in urban labour markets compared to permanent migrants. They may also find it difficult to gain access to jobs in formal urban sectors due to their registration status (Zhao, 2005). As a result, non-urban-registered migrants are often paid less on average. Although they may spend a significant amount of time in urban areas, they may additionally not be covered by the urban social security system nor be entitled to various other benefits.

[7] Some rural migrants can successfully obtain a permanent urban hukou status after leaving a rural area, and therefore are often deemed "permanent migrants". It should be noted that permanent migrants are different from urban migrants with the latter having moved from other urban areas and hence holding an urban hukou. We do not consider permanent migrants in our paper.

The *hukou* system thus creates important distortions and increases inequality in the urban Chinese labour market (Whalley and Zhang, 2004) despite the fact that several papers point to migration within China as a natural mechanism for rural-urban income convergence (Lin *et al*., 2004; Du *et al.*, 2005). In fact, although migrants have been moving to urban labour markets for many years, the *hukou* system has ensured that the urban labour market remains segmented opening the possibility of discrimination against those who are not registered in urban centres.

Research shows that migrants in China are positively self-selected on the basis of (both observed and unobserved) characteristics which increases the likelihood of yielding a positive outcome in the labour market[8], yet evidence of discrimination against rural migrants in the urban labour market has been documented. As shown using data from the 2002 China Household Income Project (CHIP), migrants themselves perceive to be discriminated against in urban labour markets (Demurger *et al*., 2008). Both casual observation and existing research (Meng and Zhang, 2001; Cai *et al*., 2003; UNDP, 2005) indicate that a significant share of migrants take up jobs in the informal sector, are paid less and are also less likely to be covered by urban social security systems (Wei, 2007). The lack of social security coverage is likely to contribute to an important decrease in welfare; a report by the China Labour Bulletin (2008), for instance, reported that the current wage gap between urban and rural regions would increase from 3-fold to 6-fold in real terms, if we considered the benefits accrued from social security. Even worse is that migrants occasionally find it difficult to get their settled salaries from their employers on time and enforced[9]. Despite extensive reforms in minimum wage legislation (see *The 1994 Labour Law*), the large number of migrants working informally ensures that the minimum wage is not binding.

Many empirical papers have looked at the rise in urban labour market segmentation and urban-rural disparity in China (for instance Whalley and Zhang, 2004; Knight and Li, 2005). Although the relative disadvantages of rural migrants raises concern, little has been done to study the determinants of their labour market outcomes and better understand labour market outcome differentials between them and urban residents[10].

Another related question is whether the urban labour market is segmented with respect to the *hukou* system. Given that an individual's *hukou* status may be correlated with the individual characteristics of a migrant such as education level, experience and ability, it is difficult to tell whether poor performance in the labour market is due to these characteristics (*e.g.* low human capital) or the *hukou* status.

Démurger *et al.* (2008), decompose annual earnings differences between urban residents and rural

[8] For instance, selection has been documented on the basis of their level of education, their age, their health status or their gender (Kikuchi et al., 2000; Wu, 2008).

[9] The China Labour Bulletin report (2008) claims that in 2004 there were 114 997 labour dispute lawsuits filed by migrants.

[10] Among the very few, Meng and Zhang (2001) find that educated urban residents are more likely to have a white-collar job or to work in wholesale or retail trade occupations. Moreover, despite wage discrimination against migrants (which can be as high as 50%), they find that 82% of the discrimination is due to inequality between sectors. Meng (2001) finds that migrants with higher levels of education and urban labour experience are more likely to be self-employed in the informal sector. Shi and Zhang (2006) find that the return to education in the urban labour market is around 5.4%, and show that education is important in determining higher wages for migrants in urban centres.

migrants into four categories (a sectoral effect, a wage effect, an hours worked effect and a population effect) and find that migrant workers have a comparative advantage in working in the private sector while the opposite holds for urban residents. Moreover, the population effect, the underlying individual characteristics of urban residents and migrants, is significantly important, signaling that pre-market rather than on-market factors prevail.

Many of the above cited studies use the Oaxaca-Blinder framework to test for discrimination in the labour market (the unexplained share of the income differential). However, many of them use data from different regions at different times and ultimately derive different conclusions, making any comparison a difficult task. For instance, Meng and Zhang (2001) find that 51% of the wage differential between urban residents and migrants is due to unexplained factors (discrimination, loosely speaking) while Dinh and Maurer-Fazio (2004) find 25% and Wang (2005) 43%, all using different datasets which focus on different regions of the country. Deng (2007), using the China Household Income Project (CHIP) data collected by the CASS[11] and which reasonably covers the country, finds that 60% of the income differential originates from unexplained factors. As commented by Zhao (2005), "The datasets that have been used in existing research papers are quite varied", and in some cases they are outdated. In all cases, it can be said that the results cannot be generalized as the data is not representative of the entire population of the country.

In this paper we wish to understand the reason for wage differences in urban labour markets between migrants and urban residents. We use a Oaxaca-Blinder model to decompose the difference between skill levels and skill-prices. Our research contributes to the literature by using a more representative dataset, the $1/5^{th}$ random draw from the 2005 1% national census, which allows us to distinguish between work in the formal and the informal sectors. This is a salient feature when studying migration, as other sample datasets may not fully capture all migrants, especially those working informally. Discrimination can appear in many forms. By introducing a sectoral breakdown along informal-formal labour market segmentation, we gain a more complete and realistic picture of rural migrant labour outcomes in urban labour markets.

3. Data and summary statistics

3.1 The data

The data we use is a one-fifth random draw from the 1% census data of China administered by the National Bureau of Statistics (NBS) in 2005. The sample size is around 2.3 million individuals covering 31 provinces, municipalities and autonomous regions. We use the sampling rule to extrapolate the total population in China, which is 1.29 billion, slightly lower than the 1.31 billion referenced in the China Statistical Yearbook (NBS, 2006)[12]. Generally speaking our data is representative of mainland China. This is confirmed by Figure A1 in Appendix 1, where we show the weighted sample shares by province compared to the corresponding shares we get from the Chinese Statistical Yearbook (CSY). The difference we find is negligible.

[11] Chinese Academy of Social Sciences.
[12] There are at least two possible reasons for the difference. First, our data does not include Hongkong, Macau, and Taiwan, while the CSY does. Second, military staff may be underrepresented in our sample.

The 1% census data has great advantages for studying migrant labour market outcomes. An ordinary household survey may be less likely to obtain a representative sample of migrants due to the floating nature of migrants and due to the sampling process. For example, surveys sampling migrants from neighbourhoods or communities may under-represent those who arrived recently and those who live at the construction site collectively. Census data does not suffer from such problems.

Another advantage of our census data is that, not only can we identify rural-to-urban migrants but also urban-to-urban migrants, which allows us to have a more detailed and comprehensive picture of labour market integration in the urban labour market as it offers an alternative comparison group to rural migrants.

3.2 Definitions

Two questions in the questionnaire are used to identify migrants: (1) *Where is your hukou registration place*? and (2) *How long have you left your hukou registration place?* We define migrants as those who have left their *hukou* registration place for more than half a year and disaggregate them into four categories according to the place where they were living at the time of the survey (city, town or village) and their *hukou* type (rural or urban). Namely, we define these categories as follows:

- rural-to-rural: individuals with rural *hukou* status who have moved to another rural area (village or town).
- rural-to-urban: individuals with rural *hukou* status who have moved to an urban area (city).
- urban-to-rural: individuals with urban *hukou* status who have moved to a rural area
- urban-to-urban: people with urban *hukou* status who have moved to another urban area.

The sample sizes of these four types of migrants are 38.7, 159.5, 12.5, and 116.8 thousand respectively. Given the fact that our sample is a $1/5^{th}$ draw from the 1% national census, we can estimate the corresponding totals for these four groups, which are 19.4, 79.7, 6.3, and 58.4 million respectively[13]. The total number of migrants is approximately 164 million, nearly 12.5% of the entire population. The shares of the above four types of migrants are 1.5%, 6.1%, 0.5%, and 4.5% of the total population[14].

We also disaggregate migrants by reason for leaving their *hukou* registration place (see *Table A* 2 in Appendix 1). Across all groups, most migrants move for work, and this is especially true for rural-to-urban migrants (the share approximately 61 percent). People also migrate for other reasons; for rural-to-rural migrants, the second largest reason is marriage (which accounts for 19%), while for rural-to-urban migrants, the second most important reason is to be with their families (relatives). Marriage is also a major

[13] The unweighted results are 14.2 millions, 66.4 million, 3.8 million and 43.7 million respectively, summing up to 128 million.
[14] It is difficult to compare the magnitude of migrants to other papers since different researchers use different criteria to identify migrants depending on data availablility. For example, Cai and Wang (2003) find there are 131 million rural migrants using the 2006 Chinese agricultural census, which is in fact larger than the one we use (99.1 million). The difference between this figure and our figure of 128 million can be explained by comparing how migrants are defined. The definition of a migrant in their study is based on whether individuals "*spent at least one month outside their home counties*".

reason for this type of migration. The number of urban-to-urban migrants is also large, but the share of those who move for job opportunities is relatively small (only 20%).

As we are mainly concerned with *urban* labour market outcomes, we focus on the following three groups: rural-to-urban migrants, urban-to-urban migrants, and urban residents (non-movers). In addition, we restrict our sample to those out of school and aged 16 to 60 and those who migrate to look for work and for business reasons.

This paper also accounts for the evidence linked to the segmented nature of the Chinese urban labour market[15] by extending the analysis to formal and informal employment. Our data allows for two definitions of informal employment: (a) self-employment and (b) the absence of a formal labour contract[16]. As it will be shown, these two mutually exclusive definitions allow for some degree of heterogeneity within informal employment. Both are characterised by the lack of social security coverage. Employees with a formal contract, whether long-term or short-term, are defined as being formally employed.

3.3 A few summary statistics

Summary statistics are reported in *Tables 1* and *2*. Our data show that urban residents (non-movers) are gender balanced, quite educated, mostly all married and that many are working in the formal sector, and particularly in the public sector. Self-employed workers are typically older, less educated and male, while showing higher income earnings than 'no contract' employees.

Rural migrants are younger, less educated and more likely to be male. Again, we observe differences depending on the definition of informal employment we use. The self-employed are older and less educated while formal workers have similar characteristics as the 'no contract' employees. Urban migrants are similar to rural migrants in terms of age, gender balance and marriage status but have much higher education levels and also earn the most. In terms of informal employment, self-employed workers earn more than workers without a contract for both types of migrants, but the opposite is observed for urban residents, where waged employees earn more than the self-employed. One possible explanation is the high number of individuals working in the public sector and forming the group of so-called iron rice-bowl jobs.

Apart from comparing average earnings, we also plot income distributions for each group (see *Figure 1* to *Figure 4*). When we compare the income distributions of urban residents, rural migrants and urban migrants (*Figure 1*) that of rural migrants has the smallest dispersion. Taking the income distribution of urban residents as a benchmark, that of rural migrants falls disproportionally on the lower-medium part of the benchmark distribution. This is just the opposite for urban migrants. It has greater dispersion, and it falls disproportionally on the upper-half of the benchmark distribution. Although the distributions have different shapes and positions, they are generally "normal". This paper looks at the mean income

[15] For example, Meng and Miller (1995) emphasize segmentation based on occupation while Demurger *et al.* (2008) emphasize segmentation based on ownership.
[16] For the remainder of the text we will reference this group as the 'no contract' group.

differential as opposed to the income differential along the whole range of the income distribution and attempts to explain the difference between these three income distributions.

In *Figure 2* to *Figure 4*, we disaggregate the income distribution of each group into three categories: formal employment, self-employment, and 'no contract' employees. There are significant differences between formal and informal employment, and between the self-employed and the group of 'no contract' employees. This is an indication that segmentation exists between formal and informal employment, and there is some degree of heterogeneity within the informal sector.

4. Model specification

We use two two main empirical strategies in this paper. First we employ a multinomial logit model to infer simple correlations related to sector choice and also to generate summary statistics for counterfactual predictions. Three multinomial logit models are estimated, one each for urban residents, urban migrants and rural migrants, featuring employment sector as the dependent variable. The employment sectors are defined as: *j* = formal, self-employment and no-contract.

The formal model is as follows:

$$\Pr(y_i = j) = \frac{\exp(X_i \beta_j)}{1 + \sum_{j}^{J} \exp(X_i \beta_j)} \qquad (j = 1,2,3) \tag{1}$$

where X_i is a vector of explanatory variables related to sector *j*.

In our next step, we use a Oaxaca-Blinder decomposition framework accounting for wage differences between rural migrants and urban residents[17]. The income differential between the two groups can be decomposed into two parts: one due to differences in individual skill levels (the so-called *endowment effect*) and the other due to the differences in the skill-prices individuals face in the labour market (the *price effect*). The Oaxaca-Blinder model is estimated in two steps. First, two more regressions –one for each group– are estimated, by assuming wages for each group can be defined as follows:

$$W_i^g = \alpha^g + X_i^g \beta^g + \varepsilon_i^g \qquad (g = u, m) \tag{2}$$

where W_i^g refers to the income (in log form) of individual i where $g = u, m$ refers to urban residents and rural migrants. X_i^g is a vector of independent variables, including education, age, marital status, gender, province dummies, industry dummies and occupation dummies. α^g is the intercept for group *g*.

[17] For the moment we only consider differences between rural migrants and urban residents, and let the analysis on urban migrants for later.

If the model is estimated using an OLS model, we can state $\overline{W}^g = \hat{\alpha}^g + \overline{X}^g \hat{\beta}^g$, with the over bar " ¯ " on *W* and *X* refering to sample means, and $\hat{\alpha}^g, \hat{\beta}^g$ the OLS estimates for α^g, β^g. Differencing out the mean wages for both groups, the typical Oaxaca-Blinder model is then as follows:

$$\overline{W}^u - \overline{W}^m = (\alpha^u - \alpha^m) + (\overline{X}^u - \overline{X}^m)\beta^u + (\beta^u - \beta^m)\overline{X}^m \quad (3)$$

The second term on the right-hand side $(X^u - X^m)\beta^u$ is the wage differential due to differing individual characteristics (such as human capital) in the absence of discrimination. The third term, $(\beta^u - \beta^m)X^m$, measures the proportion of the relative wage differential due to discrimination. Discrimination is measured as the residual, or the unexplained difference in the regression coefficients. We also conduct a Brown *et al.* (1980) decomposition, which also considers formal and informal sectors. The full Brown model is elaborated in Appendix 2.

In addition to analysing income differentials using a Oaxaca-Blinder decomposition technique, we also look at the differences with respect to sector choice. In particular we decompose differentials in sectoral distributions into *endowment* and *price* effects[18]. For this purpose we estimate a linear probability model (instead of a multinomial logit model), and then we apply the results to the Oaxaca-Blinder decomposition. We do this by defining a dummy variable for work in the formal sector (equal to zero if it is work in the informal sector).

5. Basic empirical results

As indicated, our empirical strategy contains several steps. First, we estimate the three multinomial logit models to determine characteristics consistent with sector employment and then use these results to predict counterfactual (and factual) sectoral distributions. Second, we estimate income equations for different sectors for both urban residents and rural migrants using simple OLS regressions, and calculate Oaxaca-Blinder decompositions for each sector using results from the regressions. A Brown decomposition analysis is also performed to ensure robustness and the results relating to it are reported in the Appendix 2.

5.1. Results of the sector choice model

Table 3 presents the marginal effects of the multinomial logit regression for sector choice. For urban residents, age, education, gender, and marital status all have significant effects on sector choice. Taking formal employment as the base category, the probability of being self-employed takes an inverted U-shape with respect to age. Individuals with higher levels of education, women and unmarried individuals are less likely to be self-employed. The effects of these variables on the probability of being in the 'no contract' category are quite different. First, as people age, the probability of not having a formal contract decreases

[18] or sector-choice structure effect.

and after the age of 35, the probability changes little with age. The more educated are also less likely to be without a contract. The effect of college education on self-employment and 'no contract' work is quite different however. College education compared to senior middle school seems to have only a marginal (and even negative) effect for people working without a labour contract. As the number of college graduates increased tremendously in the last several years, the pressure to find a good job has also consequently increased and our results may reflect the current situation of many college graduates. Finally, women and unmarried individuals are more likely to work outside the protective confines of a formal contract.

The signs of the coefficients for rural migrants are the same as for urban residents (except when looking at gender) but the magnitude of the coefficients differs quite a bit. The results in *Table 3* indicate that the urban labour market treats urban residents and rural migrants with identical individual characteristics differently, at least with respect to their choice of sector. To see this difference more clearly, we use the multinomial logit model results for urban residents to predict the counterfactual sectoral distribution for rural migrants. From *Table* 4, we see that if treated as urban residents based on observable characteristics, rural migrants would have a different sectoral distribution. Most importantly, approximately 10 percent of rural migrants would be reallocated from the 'no contract' category to formal employment, whereas the share of the self-employed would not change much. We can expect this to have an effect on the income differential between rural migrants and urban residents.

5.2. Income determination for different subgroups

We next consider income determination for formal work, self-employment and ´no contract´employees, for both urban residents and rural migrants. The aim of this exercise is to bring out the price difference between the sectors and between rural migrants and urban residents. The results pertaining to the OLS regressions are presented in *Table 5* and we highlight some of the more important results in by plotting income-age profiles and returns to education in *Figure 5* and *Figure 6*. For urban residents (left panel of *Table 5*), the income-age profiles vary quite a bit depending on sector of work; for the formal employment and the 'no contract' employees groups income levels increase with age, while for the self-employed income first increases and then decreases with age. For rural migrants, however, all three groups show no significant differences in the income-age profile. In addition they are quite similar to that of self-employed urban residents. Their income level begins to decrease around 25 to 30 years of age. As shown in *Figure 6*, the returns to education are also different; for urban residents there are no significant differences in returns to education between formal employment and 'no contract' work, whereas the self-employed have the lowest returns to education. The same is true for rural migrants, with the formally employed having the highest return to education among them. Beyond this, the urban labour market also rewards gender and marital status differently. Women's earnings are significantly lower than those of men. The differential is much higher for the self-employed than for the formally employed and 'no contract' employees. There are only slight differences in coefficients between urban residents and rural migrants along gender lines.

5.3. Oaxaca-Blinder decomposition results

Using the results from the OLS exercise, we turn to the Oaxaca-Blinder decomposition for rural migrants and urban residents. We start by analyzing the entire sample and then to the sub-samples based on formal and informal sector definitions. The results are shown in the first column of *Table 6.* If sector choice is not taken into account, nearly 100% of the income differential between rural migrants and urban residents can be attributed to differences in their characteristics (the endowment effect). We then compare rural migrants and urban residents with respect to formal and informal employment. The results are nearly the same. Within formal employment, 83% of the income differential is due to the endowment effect. For the self-employed, it's 92%, and for 'no-contract' employees, 160%. The decomposition results for 'no-contract' employees show that if rural migrants had the same level of human capital as their urban resident counterparts and were paid as if they are, their income would not only be higher than their actual income, but also higher than the income of urban residents.

Loosely speaking, the differential between rural migrants and urban residents is mainly caused by the differences in human capital levels they bring to the urban labour market. However, we should be very careful in interpreting these results. In particular, our results indicate that human capital levels are very important, but it does not mean that the differences in returns to human capital (both in terms of income and sector choice) are not important.

6. Discussion

6.1. Are urban residents the right reference group?

It is useful at this point to go back to the two criteria used to identify rural migrants. The first is *hukou* type, with migrants having rural *hukou* status and urban residents having an urban *hukou* status. The second criteria is whether surveyed individuals left their registered *hukou* place[19]. It is thus possible that rural migrants obtain a premium for moving (due to positive self-selection into migration), but at the same time be discriminated against dueto their rural *hukou* status. These two opposing effects may produce a close-to-zero "unexplained effect" as was found in the previous section. The advantage of our dataset is that it not only allows us to identify rural migrants but also urban migrants. By comparing the income differential between these two groups, we can attempt to separate out the pure *hukou* effect as both groups are migrants.

The last two columns in *Table 3* report the results of the multinomial logit regression results for sector employment for urban migrants. Taking formal employment as the base category, the probability of being self-employed increases with age and the rate of increase is larger in earlier years. On the contrary, the probability of being without a contract decreases with age. Education is negatively related to the probability of being in informal employment. The effect of higher education on the probability of informal employment is higher for urban migrants than for rural migrants. As shown in *Table 4*, a counterfactual prediction shows that if urban migrants were treated as rural migrants at least with respect to choice of

[19] This twofold criteria makes our research different from most papers that focus on the income differential between two naturally divided groups (based on gender or skin colour for example).

sector, there would be less urban migrants in the formal sector (from 51.7% to 44%), and more in self-employed and 'no contract' groups (from 14.7% to 16.2%, and from 33.6% to 39.8% respectively).

OLS regression results are presented in the last three columns of Table 5. Highlighting some of the results from these regressions in *Figures 5* and *6* (age and education coefficients), it becomes obvious that urban migrants have a higher return to age than rural migrants, at least for formal employment and 'no-contract' informal employment. Most importantly, urban migrants have the highest returns to education for each level of education in all three sectors.

The second column of *Table 6* reports the overall Oaxaca-Blinder decomposition results for rural and urban migrants by sector. Ignoring the differences in sector distribution, the regular decomposition results indicate that nearly 60% of the income differentials between these two groups can be explained by difference in their characteristics and the remaining 40% is due to differences in skill-prices. However we find significant heterogeneity across sectors. In the formal sector, differences in endowments can explain 67% of the income differential whereas in the 'no-contract' group (self-employment group) the share goes down to 56% (13.4%). It is interesting that the explained and unexplained shares of the income differential between rural and urban migrants do not change much if we consider only recent migrants or only the non-recent migrants.

By comparing rural migrants with urban migrants instead of urban residents, we obtain quite different results. Rural migrants fare worse than the urban migrants not only because they have low levels of human capital but also because they are treated differently due to their rural *hukou* status.[20] These results indicate that migrants do receive some premium for migrating, and this holds true both for rural and urban migrants.

6.2. Premiums for migrants: urban residents versus urban migrants

To evaluate the existence and magnitude of a "migrant premium", we compare urban migrants with urban residents. Both groups have urban *hukou* status with the only observable difference being that the first group is made up of migrants. Hence if we find that urban migrants have higher income levels not only because of their higher human capital levels, but also because of the different skill-prices, this might provide evidence of a premium for migration.

The simple Oaxaca-Blinder decomposition in column 3 of *Table 6* indicates that the premium not only exists but also that it is important. The simple overall decomposition shows that 83.5% of the income difference is unexplained. For the decompositions by sectors, the unexplained shares are 61.2%, 66.4%, and 140% for the formal sector, self-employed, and no-contract employees respectively.

Although urban migrants have large premiums in income determination within each sector, they seem to detain less of an advantage to some extent in the choice of sector. The results shown in *Table 4* indicate

[20] Clearly, agricultural *hukou* status has various implications. For example, we use age as a proxy for potential experience. But what really matters is urban labour market experience, which is not available in the data. For rural migrants, age is definitely a bad proxy.

that if treated as urban residents in sector allocation, urban migrants should figure more prominently in the formal sector (57% instead of 52%), and fewer in the self-employed group (10.5% instead of 14%)[21]. This implies that if urban migrants were treated as urban residents in terms of sector employment, they would moreover have higher income levels. This negative effect is overcome by the large migration-related premiums migrants receive in terms of earnings.

6.3. Does duration of migration matter?

We can think of the migrant premium referenced above as a net average effect for migrants as a whole. However, premiums for migrants may differ substantially and the heterogeneity may come not only from the type of hukou they have, but also their migration duration or otherwise their urban labour market experience. The Harris and Todaro (1970) framework suggests that migrants may first enter the informal labour market while they wait and perhaps gather experience for an opportunity at a formal sector job. Unfortunately, it is very difficult to find a comparable proxy for urban labour market experience for all three groups. In order to evaluate the assimilation effect, we apply our method to decompose the income differentials between recent and non-recent migrants.

For both rural and urban migrants, the duration of their migration episode is important. More than 50% of migrants (rural and urban) have less than 3 years of local urban labour market experience (we call this group *recent migrants*). It seems true for both rural and urban migrants, that the self-employed tend to have longer migration duration, and the 'no contract' employees tend to be more recent migrants. To see the effect of duration more clearly, we split rural migrants and urban migrants into recent and non-recent sub-groups (*Table* 2).

To get the decomposition results, we estimate multinomial logit models for recent and non-recent migrants separately and OLS regressions for the different sectors. The multinomial logit results for sector of employment are reported in *Table 7* and OLS income determination results in *Table 8*. Recent and non-recent migrants show different patterns both in terms of income determination and sector choice, and this is the case for both rural and urban migrants. The Oaxaca-Blinder decomposition results, reported in *Table 6*, show that for rural migrants (column 5), 47% of the differential between recent and non-recent migrants is unexplained. In the formal sector and the no-contract employee group however, the unexplained shares are lower than the overall percentages, which are around 40%. For the self-employed, 88% of the difference between recent and non-recent migrants is unexplained. This means that the assimilation effect is more evident for the self-employed. This is the expected result as setting up a business in an urban area requires financial and social capital which takes time to accumulate. The case for urban migrants is similar.

For sectoral distribution analysis we turn to our counterfactual predictions (*Table 4*). If recent migrants were treated as non-recent migrants, there would be more recent migrants in formal sectors or self-employment, and this is true for both rural and urban migrants. Nevertheless the difference between the

[21] This also explains the negative percentages we obtained for the inter-sectoral differences in the Brown decomposition analysis (See Table A3 in the second appendix).

actual and counterfactual sector distributions is not very large, and this is closely related to the fact that actual sector distributions of recent and non-recent migrants are similar. As a result, a Brown decomposition analysis shows that the share of the inter-sectoral difference in the total difference is relatively low: -6.3% and 6.8% for rural and urban migrants, respectively (Table A3 in the second appendix). Most of the difference is due to intra-sectoral differentials, 58.6% and 46.1% of which can be explained by differences in characteristics.

6.4. Decomposing sectoral distribution differentials

Our analysis has not, until now, revealed any major role played by sectoral segmentation. The largest inter-sectoral share we found in the Brown decomposition analysis was 39.5% (urban residents *vs.* rural migrants in *Table A3*) with all other comparisons less than 10%. We argue that these results do not imply that discrimination against migrants in terms of sectoral choice is unimportant. Even a small degree of discrimination can be quite significant, given the substantial differences that exist in terms of social security coverage, working conditions and pay between the formal and the informal sectors.

To show this we estimate linear probability models (not reported) and use the results to calculate Oaxaca-Blinder decompositions based on sector choice. These decomposition results are reported in *Table 9*. In the first two columns, we consider the broad definition of informal employment, including both self-employed and 'no-contract' employees and compare rural migrants with urban residents first. It is clear that the sectoral distribution differential is large. The fraction of informal employment for rural migrants is 33.7 percentage points higher than that for urban residents. The decomposition result shows that only 17% of the difference can be explained by differences in characteristics, while 83% remains unexplained. This is in contrast to the results we had derived both with the Oaxaca-Blinder and Brown decomposition results for *income* differentials, which indicated the dominant role of the endowment effect and a minor role played by sector segmentation. The results here show that there is a significant share of differential caused by discrimination in terms of sector choice, which means migrants may be even more worse off in terms of social security coverage and working conditions, even conditional on their characteristics.

Interestingly (and likely by coincidence) the unexplained share of the sector distribution differential between urban migrants and urban residents is also 83%. However, taking into account that the overall sector distribution differential for these two groups is smaller than that between rural migrants and urban residents (16.6 as opposed to 33.7 percentage points), the extent of discrimination against urban migrants is smaller. It should be noted that urban migrants enjoy the highest average income level, and they are better (not worse) off conditional on their characteristics in terms of income. However, the results here show that they are still discriminated in terms of sector choice.

What follows naturally is to compare rural migrants and urban migrants. The results are also as expected. As these two groups are both migrants, differences in characteristics can explain a larger share (around 50%) of the sector-choice differential. Still, half of the differential is due to unexplained factors. We interpret this as discrimination against rural *hukou* status. We also consider recent and non-recent migrants

separately but results do not change much. The final decomposition exercise for these two groups is to compare recent and non-recent migrants for each group separately. The results indicate a large share of discrimination against recent migrants. As differentials between recent and non-recent migrants are not very large, the decomposition results is of minor importance for us. However, because we use a broad definition of informal employment, the small differentials may be caused by composition change within the broad informal employment definition.

In the next four columns of *Table 9*, we consider self-employment and 'no-contract' employees separately. The general pattern is similar to the one found in the first two columns, but with a slight variation. There are at least two points worth mentioning. First, due to the small fraction of self-employment in all three groups, the difference between groups is relatively small, especially when we compare non-recent rural and urban migrants. Second, when we compare recent and non-recent migrants, there are larger sector distribution changes than under the broad informal definition. For both rural and urban migrants, the fraction of self-employed is larger for the non-recent migrants than for recent migrants, and the fraction of 'no-contract' employees is smaller.

7. Conclusions

In this paper we use a nationally representative sample to investigate how rural migrants fare in the urban labour market in China. This paper is different from the existing literature in several important ways. First, our data is nationally representative and is better at capturing a representative sample of rural migrants. Second, we distinguish among different groups of migrants instead of only comparing rural migrants and urban residents. In particular, we add urban migrants in our analysis, and this allows us to separate the rural *hukou* effect from the migrant premium effect. Third, we consider sectoral segmentation in terms of formal and informal employment, which is an important dimension of labour market outcomes of rural migrants in Chinese cities.

The main finding in this paper is in stark difference to those in the existing literature. When we compare rural migrants with urban residents, nearly 100% of the difference can be explained by differences in characteristics, which means the skill-price effect is almost negligible. The comparison between rural migrants and urban migrants gives a totally different picture however. 40% of the income differential is unjustified. Given the fact that *hukou* status is the only difference between these two groups conditional on other personal characteristics (and migrant status), this exercise is more appropriate for us to detect discrimination against rural *hukou* status. By comparing urban migrants with urban residents, we find significant "migrants premiums". We conjecture that the "no discrimination" effect we derive from the comparison between rural migrants and urban residents may be the result of a counterbalance between the discrimination effect against rural *hukou* and the premium effect for migrants.

Another result based on the Brown decomposition analysis, is that sector segmentation plays a minor role in explaining the income differential. Sectoral segmentation, however, may be important in terms of social security and working conditions and therefore the sector distribution differential in its own right is worth studying (and this is overlooked by many researchers). The simple Oaxaca-Blinder decomposition

indicates that migrants (both rural and urban) are discriminated in sector choice (more than 80% is unexplained). The extent of discrimination is larger for rural migrants indicating a further discrimination against rural *hukou* status taken the magnitude of the differential. The decomposition for differences between sector distributions complements the income decomposition in a very important way.

There are of course limitations in the approach taken in this paper. A first difficulty has to do with choosing the appropriate reference group. We take a step forward by using urban migrants as an additional reference group. However, this method is also not without problemes as rural migrants and urban migrants may be different in unobservable characteristics other than their *hukou* status, even conditional on the characteristics we control for. Another difficulty is measurement error which is especially salient when we are comparing urban residents (migrants) with rural migrants. Finally, age (even potential experience) is a poor proxy for urban labour market experience for rural migrants.

The policy implications of our results are clear. In terms of income, rural migrants enjoy migrant premiums and suffer discrimination at the same time. Generally speaking, however, the reason they may be worse off when compared to urban residents is due to their lower levels of human capital. Increasing education levels of rural migrants, and providing them with training and relevant urban labour market skills will help increase their earning opportunities. As both rural and urban migrants face unfair treatment in sector choice, reforming the labour market, notably removing sector barriers, may help increase formal employment.

Discrimination against rural *hukou* status is evident in our study and ideally a complete cancelation of the system would eventually lead to a more equal treatment on the labour market. This may not be practically feasible however and in fact remains a central focus of debate in China. What is perhaps more pressing is to ensure that migrants have access to basic social services, even in cases where they are employed informally. Presently those without such coverage face exorbitant costs for health services and in sending their children to urban schools. In fact, many migrants leave their children at home in rural China, in effect putting more pressure on household members left behind and adding to the social strain caused by migration. For migrants who systematically move for jobs, obtaining urban social security coverage is futile as social security systems are for the most part non-portable.

References

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

Brown, Philip H (2006). "Parental Education and Investment in Children's Human Capital in Rural China", *Economic Development and Cultural Change*, University of Chicago Press, vol. 54(4), pages 759-89.

Brown, R. S., Moon, M., and Zoloth, B. S. (1980). "Incorporating Occupational Attainment in Studies of Male/Female Earnings Differentials", *The Journal of Human Resources*, 15, 3-28.

Cai, Fang (2000), "The Invisible Hand and Visible Feet: Internal Migration in China", *World Economy and China*, No. 5.

Cai, Fang and D.Wang (2003) "Migration as Marketization: What Can We Learn from China's 2000 Census Data?", *The China Review*, 3(2):73-93.

Cai et al. (2003)

Cai, Fang and M.Wang (2006), "Gender Wage Differentials in China's Urban Labour Market", UNU-WIDER Research Paper No.2006-146.

China Labour Bulletin (2008), «Wages in China», www.china-labour.org.hk, February.

De la Rupelle, Maëlys (2007), "Le contrôle des migrations en Chine : un atout économique?", mimeo.

Démurger, Sylvie, M.Gurgand, L.Shi and Y.Ximing (2008), "Migrants as second-class workers in urban China? A decomposition analysis", PSE Working Papers 2008-20, PSE (Ecole normale supérieure).

Deng, Quheng and B.Gustafsson (2006), "China's Lesser Known Migrants", IZA Discussion Paper No. 2152.

Deng, Quheng, (2007) , "Earnings Differential between Urban Residents and Rural Migrants:Evidence from Oaxaca-Blinder and Quantile Regression Decompositions", *Chinese Journal of Population Science*, No. 2, (in Chinese)

Dinh, Ngan and M.Maurer-Fazio (2004), "Differential Rewards to, and Contributions of, Education in Urban China's Segmented Labor Markets", *The Pacific Economic Review*, Volume 9, No.3, pp. 173-189.

Ding, C., Y.Song, and Y.Zenou (2008), "Let's not throw the baby out with the bath water: The role of urban villages in housing rural migrants in China", *Urban Studies*, 45, 313-330.

Du et al. (2005) Is Migration Helping China's Poor? Paper prepared for the Conference on "Poverty, Inequality, Labour Market and Welfare Reform in China," Australia National University, August 25-27, 2004.

Huang, Ping and S.Zhan (2005), "Internal Migration in China: Linking it to Development", Prepared for the regional conference on migration and development in Asia, in Lanzhou, China, 14-16 March.

Kikuchi, Masao, A.Maruyama and T.Wang (2000), "Rural-Urban Migration and Labor Markets in China: A Case Study in a Northeastern Province", *The Developing Economies*, Vol.38(1), pp.80-104.

Knight, John and L.Yueh (2003), "Job Mobility of Residents and Migrants in Urban China," Economics Series Working Papers 163, University of Oxford, Department of Economics.

Knight, John and L.Yueh (2004), "Urban Insiders versus Rural Outsiders: Complementarity or Competition in China`s Urban Labour Market?", Economics Series Working Papers 217, University of Oxford, Department of Economics.

Knight, John and S.Li (2005), "Wages, firm profitability and labor market segmentation in urban China", *China Economic Review*, Vol. 16(3), pp. 205-228.

Liang, Zai (2001), "The Age of Migration in China", *Population and Development Review*, Vol. 27, No. 3 pp. 499-524.

Lin, Justin, G.Wang and Y.Zhao (2004), "Regional Inequality and Labor Transfers in China", *Economic Development and Cultural Change*, 52(3), 87-603.

Maurer-Fazio, Margaret, J.Hughes and D.Zhang (2004), "The Economic Status of China's Ethnic Minorities", Mimeo

Meng, Xin and P.Miller, (1995), "Occupational Segregation and Its Impact on Gender Wage Discrimination in China's Rural Industrial Sector", Oxford Economic Papers, 47(1), pp. 136-155.

Meng, Xin (2001), "The Informal Sector and Rural-Urban Migration: A Chinese Case Study", *Asian Economic Journal*, 15:1, pp.71-89.

Meng, Xin and J.Zhang (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*, Elsevier, vol. 29(3), pages 485-504.

Oaxaca, Ronald (1973), "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, Vol. 14, No. 3, pp. 693-709.

Zhan, Shaokang, S.Zhenwei and W.Blas (2002), "Economic Transition and Maternal Health Care for Internal Migrants in Shanghai, China", *Health Policy and Planning*, Vol.17(suppl.1), 47-55.

Shi, Xiaoxia and H.Zhang (2006), "Education, Experience and Income of Rural Migrants", *World Economic Papers*, No. 2. (in Chinese).

Song, S. and H.Zhang (2003), "Rural-urban migration and urbanization in China: Evidence from time-series and cross-section analyses", *China Economic Review*, Elsevier, vol. 14(4), pages 386-400.

UNDP (2005), "China Human Development Report 2005: Development with Equity ", United Nations Development Programme.

Wang, Meiyan, (2005), "Employment Opportunities and Wage Gaps in the Urban Labor Market:A Study of the Employment and Wages of Migrant Laborers," *Social Sciences in China*, No. 5. (in Chinese)

Whalley, John and S.Zhang (2004), "Inequality Change in China and (Hukou) Labour Mobility Restrictions", NBER Working Papers 10683, National Bureau of Economic Research, Inc.

Wei, Yan (2007), "Rural-Urban Migrant Workers in China: The Vulnerable Group in Cities", Paper presented at the 6th Berlin Roundtables on Transnationality on "Population Politics and Migration" held on 14-19 February 2007, Berlin, Germany.

Wu, Zheren (2008), "Self-selection and Earnings of Migrants: Evidence from Rural China", Osaka School of International Public Policy Discussion Paper 08-25.

Zhao Zhong (2005), "Migration, Labour Market Flexibility, and Wage Determination in China: A Review", *Labour and Demography*, 0507009, EconWPA.

Figure 1: Distribution for urban residents, rural migrants and urban migrants

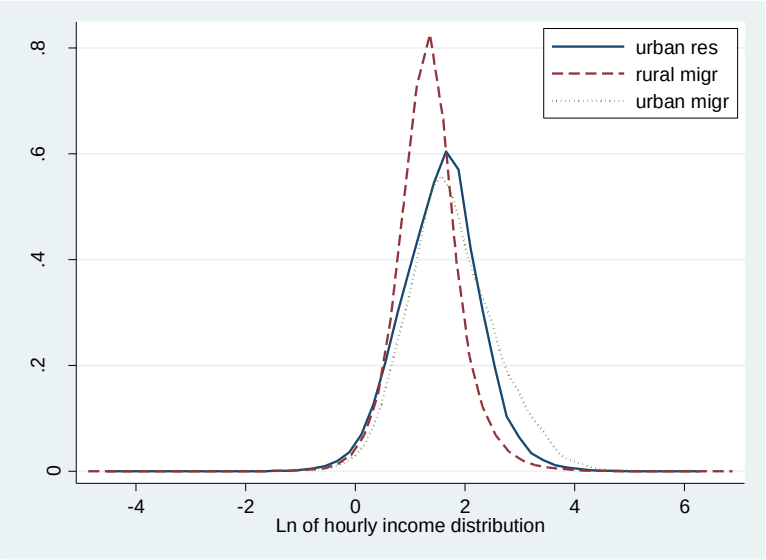


Figure 2: Distributions for formal and informal employment, urban residents

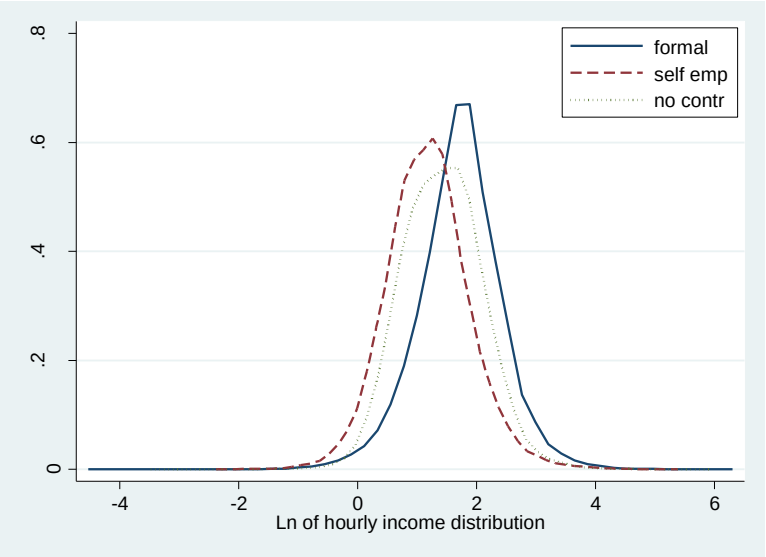


Figure 3 Distributions for formal and informal employment, rural migrants

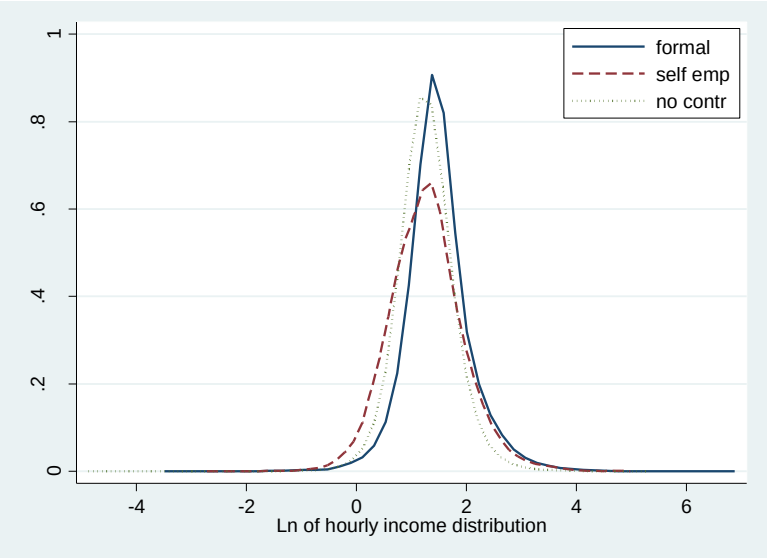


Figure 4 Distributions for formal and informal employment, urban migrants

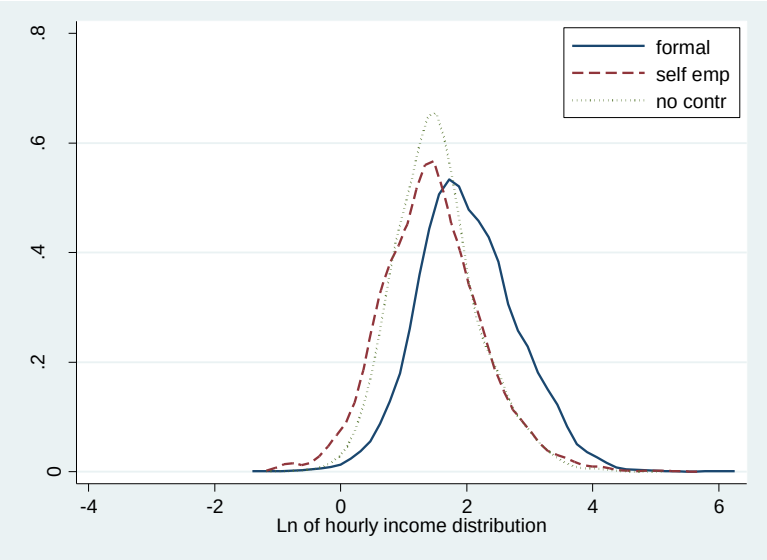
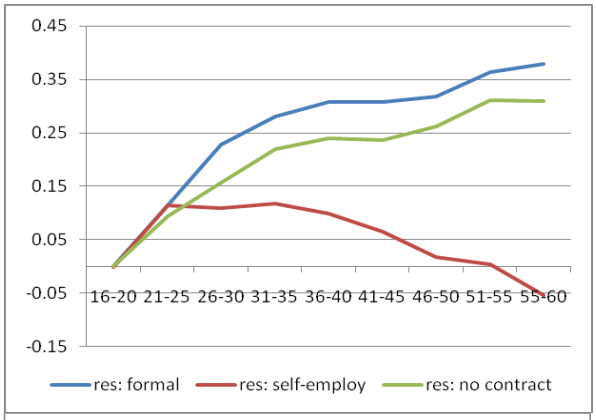


Figure 5 Ln(income)-age profile



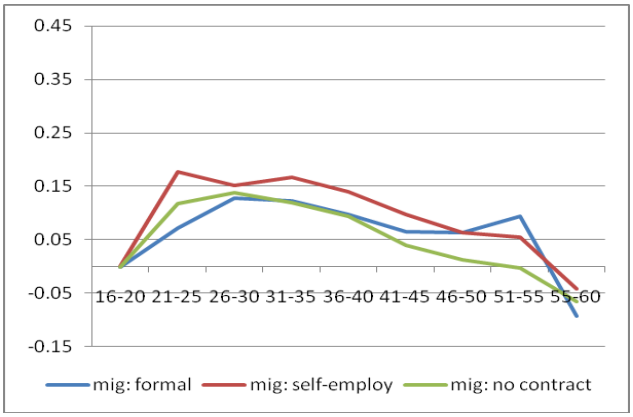
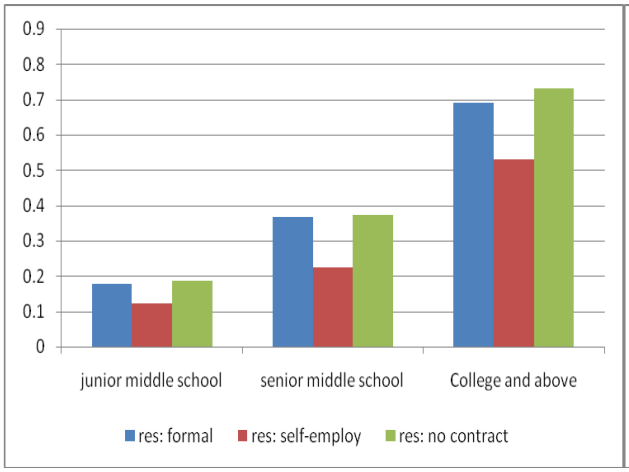
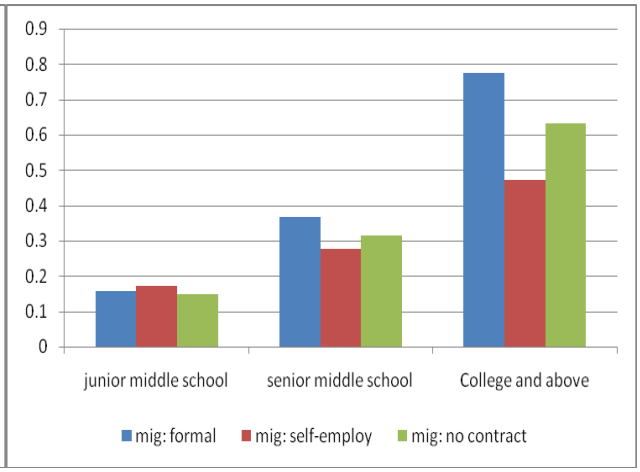


Figure 6 Returns to education





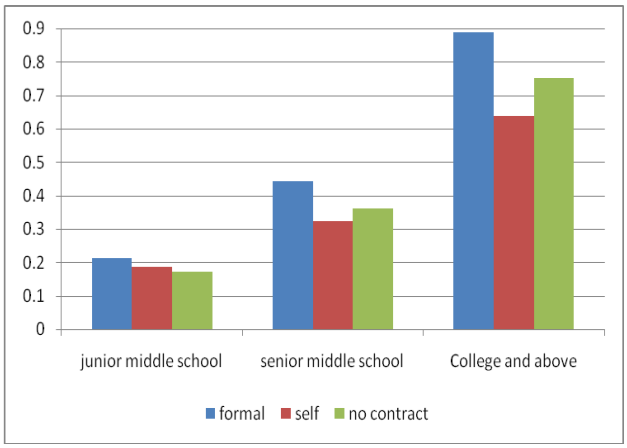


Table 1 summary statistics

	Urban residents				Rural migrants				Urban migrants			
	total	formal	self-emp	no contr	total	formal	self-emp	no contr	total	formal	self-emp	no contr
age	39.9	40.7	39.1	37.1	30.5	29.6	35.8	29.1	32.5	32.4	36.8	30.5
education levels												
Primary and below	0.10	0.11	0.15	0.06	0.21	0.17	0.31	0.22	0.06	0.04	0.10	0.06
junior middle school	0.37	0.35	0.53	0.34	0.62	0.62	0.57	0.64	0.32	0.26	0.46	0.38
senior middle school	0.31	0.32	0.28	0.32	0.15	0.18	0.11	0.13	0.36	0.36	0.34	0.36
College and above	0.22	0.22	0.04	0.29	0.02	0.03	0.01	0.01	0.26	0.34	0.10	0.20
# of people in the household	3.65	3.61	3.88	3.71	4.75	5.31	3.17	4.89	3.47	3.48	2.95	3.71
gender	1.50	1.53	1.37	1.44	1.44	1.52	1.30	1.43	1.44	1.45	1.36	1.46
marital status.	0.11	0.11	0.06	0.14	0.34	0.38	0.08	0.42	0.33	0.34	0.10	0.43
left the hukou within 0.5-3 years					0.57	0.56	0.45	0.63	0.54	0.52	0.49	0.59
occupation												
Manager, officials, …	0.04	0.05	0.02	0.04	0.02	0.03	0.02	0.01	0.06	0.10	0.04	0.02
technician (professional)	0.23	0.28	0.04	0.23	0.02	0.03	0.02	0.02	0.15	0.19	0.05	0.13
administrative staff	0.15	0.16	0.01	0.17	0.03	0.05	0.01	0.03	0.09	0.12	0.01	0.08
service staff	0.26	0.18	0.64	0.27	0.34	0.25	0.64	0.29	0.45	0.33	0.71	0.47
related to ag, forestry, fishery ect.	0.04	0.06	0.04	0.02	0.02	0.04	0.02	0.01	0.01	0.01	0.02	0.00
related to manufacture/transport/ect.	0.27	0.27	0.26	0.28	0.56	0.60	0.30	0.63	0.25	0.25	0.17	0.29
employment status												
employee	0.82	0.91	0.00	1.00	0.74	0.78	0.00	1.00	0.71	0.76	0.00	1.00
employer	0.03	0.06	0.00	0.00	0.05	0.14	0.00	0.00	0.09	0.19	0.00	0.00
self-employed	0.13	0.00	1.00	0.00	0.19	0.00	1.00	0.00	0.17	0.00	1.00	0.00
household worker	0.02	0.03	0.00	0.00	0.03	0.08	0.00	0.00	0.03	0.05	0.00	0.00
ownership												
public sector	0.26	0.28	0.00	0.34	0.01	0.01	0.00	0.01	0.05	0.07	0.00	0.04
SOE	0.26	0.39	0.00	0.13	0.04	0.05	0.00	0.04	0.10	0.16	0.00	0.07
collective owned enterprises	0.05	0.06	0.00	0.06	0.03	0.05	0.00	0.03	0.04	0.05	0.00	0.05
Family business (registered)	0.19	0.07	0.82	0.15	0.35	0.22	0.79	0.28	0.35	0.21	0.89	0.27
private enterprises	0.12	0.10	0.00	0.23	0.34	0.36	0.00	0.47	0.32	0.33	0.00	0.45
other work unit	0.04	0.04	0.00	0.04	0.13	0.26	0.00	0.10	0.10	0.16	0.00	0.08
others	0.07	0.06	0.18	0.06	0.09	0.05	0.21	0.07	0.05	0.03	0.11	0.05
monthly income	1058	1188	848	902	973	1100	982	878	1527	1905	1231	1133
hourly income	6.12	7.04	4.23	5.19	4.61	5.38	4.57	4.07	8.25	10.62	6.11	5.92
type of contract												
Fixed short term	0.21	0.34		0.00	0.34	0.94		0.00	0.43	0.81		0.00
long term contract	0.41	0.66		0.00	0.02	0.06		0.00	0.10	0.19		0.00
no contract	0.38	0.00		1.00	0.64	0.00		1.00	0.47	0.00		1.00
no unemployment insurance	0.69	0.63	0.94	0.78	0.95	0.88	1.00	0.98	0.76	0.63	0.95	0.90
no pension	0.45	0.38	0.77	0.55	0.89	0.76	0.98	0.96	0.61	0.44	0.79	0.81
no medical insurance	0.46	0.41	0.81	0.52	0.85	0.71	0.94	0.93	0.64	0.48	0.84	0.83

Table 2 summary statistics for recent and non recent migrants

	Rural migrants		Urban migrants	
	Recent	**Non recent**	**Recent**	**Non Recent**
age	28.3	33.5	30.7	34.5
education levels				
Primary and below	0.19	0.25	0.05	0.06
junior middle school	0.65	0.58	0.32	0.33
senior middle school	0.15	0.15	0.36	0.35
College and above	0.02	0.02	0.27	0.25
# of people in the household	5.27	4.06	3.60	3.31
gender	1.47	1.40	1.45	1.42
marital status.	0.45	0.19	0.41	0.23
employment status				
employee	0.80	0.65	0.75	0.67
employer	0.03	0.07	0.08	0.11
self-employed	0.15	0.25	0.15	0.19
household worker	0.02	0.04	0.02	0.03
occupation				
Manager, officials, …	0.01	0.03	0.05	0.08
technician (professional)	0.02	0.03	0.15	0.15
administrative staff	0.03	0.04	0.08	0.09
service staff	0.32	0.38	0.46	0.43
related to ag, forestry, fishery ect.	0.02	0.03	0.01	0.01
related to manufacture/transport/ect.	0.60	0.51	0.25	0.24
ownership				
public sector	0.01	0.01	0.05	0.05
SOE	0.04	0.04	0.10	0.10
collective owned enterprises	0.03	0.03	0.04	0.03
self-employed	0.32	0.41	0.34	0.36
private enterprises	0.37	0.30	0.32	0.31
other work unit	0.15	0.11	0.11	0.10
others	0.08	0.11	0.05	0.05
monthly income	901	1072	1371	1715
hourly income	4.28	5.07	7.43	9.24
type of contract				
Fixed short term contract	0.34	0.34	0.42	0.44
long term contract	0.02	0.02	0.09	0.12
no contract	0.64	0.64	0.49	0.45
no unemployment insurance	0.95	0.94	0.77	0.75
no pension	0.89	0.88	0.63	0.59
no medical insurance	0.85	0.85	0.66	0.62

Table 3 marginal effects of multinomial logit regressions

	Urban residents		Rural migrants		Urban migrants	
	Self-employed	No contract	Self-employed	No contract	Self-employed	No contract
	dp/dx	dp/dx	dp/dx	dp/dx	dp/dx	dp/dx
	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.
Age 16-20 omitted						
age: 21-25	0.022***	-0.144***	0.059***	-0.046***	0.053***	-0.061***
	(0.005)	(0.009)	(0.006)	(0.006)	(0.015)	(0.014)
age: 26-30	0.025***	-0.193***	0.094***	-0.071***	0.091***	-0.123***
	(0.005)	(0.009)	(0.007)	(0.008)	(0.015)	(0.016)
age: 31-35	0.019***	-0.220***	0.115***	-0.088***	0.115***	-0.128***
	(0.005)	(0.009)	(0.007)	(0.008)	(0.015)	(0.017)
age: 36-40	0.019***	-0.227***	0.132***	-0.103***	0.124***	-0.136***
	(0.005)	(0.009)	(0.007)	(0.009)	(0.016)	(0.019)
age: 41-45	0.017***	-0.224***	0.142***	-0.100***	0.125***	-0.173***
	(0.005)	(0.009)	(0.007)	(0.010)	(0.016)	(0.020)
age: 46-50	0.009*	-0.219***	0.154***	-0.123***	0.116***	-0.128***
	(0.005)	(0.009)	(0.007)	(0.012)	(0.017)	(0.022)
age: 51-55	0.001	-0.232***	0.138***	-0.100***	0.116***	-0.175***
	(0.005)	(0.010)	(0.008)	(0.015)	(0.017)	(0.026)
age: 55-60	0.008	-0.204***	0.120***	-0.103***	0.125***	-0.075**
	(0.006)	(0.011)	(0.010)	(0.020)	(0.020)	(0.035)
Primary and below omitted						
Junior middle school	-0.033***	-0.013**	-0.029***	-0.089***	-0.029***	-0.063***
	(0.002)	(0.005)	(0.002)	(0.005)	(0.007)	(0.018)
Senior middle school	-0.089***	-0.082***	-0.050***	-0.177***	-0.073***	-0.169***
	(0.002)	(0.005)	(0.004)	(0.006)	(0.008)	(0.018)
College and above	-0.214***	-0.071***	-0.109***	-0.237***	-0.147***	-0.285***
	(0.002)	(0.005)	(0.011)	(0.014)	(0.009)	(0.019)
unmarried	-0.032***	0.032***	-0.121***	0.089***	-0.077***	0.079***
	(0.002)	(0.004)	(0.004)	(0.006)	(0.007)	(0.010)
Female	-0.017***	0.029***	-0.049***	-0.001	-0.014***	0.029***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.007)
Observations	219712		94621		22214	

Note: the base category is formal employment. region dummies not reported. *** p<0.01, ** p<0.05, * p<0.1

Table 4 actual and counterfactual sectoral distributions based on multinomial logit regression results

	actual		Predicted based on mlogit			rural migrants			urban migrants		
						actual		predicted	actual		predicted
	urban	rural migr	urban migr	Rural migrant as urban resident	Urban migrant as urban resident	Urban migrant as rural migrant	non recent	recent	Recent as non recent	non recent	recent	Recent as non recent
Formal	0.585	0.348	0.517	0.457	0.571	0.440	0.345	0.351	0.359	0.535	0.501	0.524
Self-employ	0.118	0.175	0.147	0.170	0.105	0.162	0.232	0.133	0.168	0.167	0.130	0.139
No contract	0.298	0.477	0.336	0.373	0.324	0.398	0.423	0.517	0.474	0.298	0.369	0.337

Table 5 OLS regression results: dependent variable=log(hourly income)

	Urban residents			Rural migrants			Urban migrants		
	formal	Self-emp	No contr	formal	Self-emp	No contr	formal	Self-emp	No contr
Age 16-20 omitted									
age: 21-25	0.114***	0.113**	0.093***	0.073***	0.177***	0.117***	0.117***	0.006	0.130***
	(0.016)	(0.045)	(0.013)	(0.009)	(0.036)	(0.007)	(0.027)	(0.119)	(0.023)
age: 26-30	0.227***	0.108**	0.157***	0.128***	0.151***	0.138***	0.276***	0.096	0.191***
	(0.017)	(0.045)	(0.014)	(0.011)	(0.036)	(0.009)	(0.029)	(0.118)	(0.027)
age: 31-35	0.280***	0.117**	0.219***	0.123***	0.167***	0.119***	0.296***	0.067	0.175***
	(0.017)	(0.045)	(0.015)	(0.013)	(0.037)	(0.009)	(0.031)	(0.119)	(0.030)
age: 36-40	0.307***	0.098**	0.239***	0.098***	0.140***	0.094***	0.304***	-0.005	0.142***
	(0.017)	(0.046)	(0.015)	(0.014)	(0.037)	(0.010)	(0.034)	(0.121)	(0.033)
age: 41-45	0.307***	0.064	0.236***	0.065***	0.098**	0.040***	0.279***	-0.007	0.097***
	(0.017)	(0.046)	(0.015)	(0.016)	(0.038)	(0.012)	(0.036)	(0.122)	(0.036)
age: 46-50	0.318***	0.017	0.262***	0.064***	0.064	0.012	0.182***	-0.012	0.087**
	(0.017)	(0.046)	(0.015)	(0.020)	(0.040)	(0.015)	(0.039)	(0.125)	(0.039)
age: 51-55	0.363***	0.004	0.311***	0.095***	0.054	-0.003	0.197***	0.041	0.003
	(0.018)	(0.047)	(0.016)	(0.026)	(0.043)	(0.018)	(0.046)	(0.130)	(0.048)
age: 55-60	0.379***	-0.053	0.309***	-0.092**	-0.042	-0.065***	0.267***	-0.096	-0.038
	(0.019)	(0.050)	(0.019)	(0.036)	(0.052)	(0.025)	(0.065)	(0.145)	(0.061)
female	-0.132***	-0.225***	-0.164***	-0.136***	-0.249***	-0.159***	-0.164***	-0.198***	-0.145***
	(0.003)	(0.009)	(0.004)	(0.006)	(0.011)	(0.005)	(0.012)	(0.028)	(0.013)
Primary and below omitted									
junior middle school	0.178***	0.121***	0.186***	0.158***	0.171***	0.150***	0.214***	0.187***	0.171***
	(0.009)	(0.013)	(0.009)	(0.009)	(0.012)	(0.006)	(0.039)	(0.046)	(0.029)
senior middle school	0.366***	0.225***	0.374***	0.368***	0.277***	0.314***	0.444***	0.324***	0.360***
	(0.009)	(0.014)	(0.009)	(0.011)	(0.017)	(0.008)	(0.038)	(0.048)	(0.030)
College and above	0.691***	0.529***	0.731***	0.775***	0.472***	0.632***	0.888***	0.638***	0.751***
	(0.009)	(0.023)	(0.010)	(0.018)	(0.059)	(0.019)	(0.039)	(0.060)	(0.033)
unmarried	0.001	-0.061***	-0.034***	-0.015	-0.001	-0.055***	0.003	0.153***	-0.032*
	(0.006)	(0.020)	(0.008)	(0.009)	(0.022)	(0.007)	(0.017)	(0.052)	(0.019)
R-squared	0.424	0.147	0.423	0.291	0.148	0.235	0.435	0.205	0.389
N	128509	25832	65371	32947	16536	45138	11482	3272	7460

Note, province dummies and constants not reported. Industry and occupation controlled.

Table 6 Oaxaca-Blinder decomposition for income differential (based on OLS regression results)

		urban residents vs. rural migrants	rural vs urban migrants	Urban residents vs. Urban migrants	recent vs. non recent migrants	
					rural	urban
Industry and occupation controlled						
difference		0.238	-0.488	-0.250	0.107	0.183
Explained (%)		101.7	59.0	16.5	53.2	50.6
Unexplained (%)		-1.7	41.0	83.5	46.8	49.4
formal employment	Difference	0.252	-0.604	-0.351	0.121	0.179
	Explained (%)	82.6	67.2	38.8	59.3	55.7
	Unexplained (%)	17.4	32.8	61.2	40.7	44.3
self employed	Difference	-0.096	-0.236	-0.333	0.077	0.086
	Explained (%)	92.3	13.4	33.6	11.8	14.0
	Unexplained (%)	7.7	86.6	66.4	88.2	86.0
no contract	Difference	0.153	-0.302	-0.149	0.118	0.188
	Explained (%)	160.0	56.0	-40.0	59.6	46.6
	Unexplained (%)	-60.0	44.0	140.0	40.4	53.4
recent migrants	Difference		-0.448			
	Explained (%)		58.5			
	Unexplained (%)		41.5			
non recent migrants	Difference		-0.524			
	Explained (%)		57.5			
	Unexplained (%)		42.5			

Table 7 marginal effects for multinomial logit results (migration duration)

	Rural migrants				Urban migrants			
	Non Recent		Recent		Non Recent		Recent	
	Self-employed	No contract	Self-employed	No contract	Self-employed	No contract	Self-employed	No contract
	dp/dx	dp/dx	dp/dx	dp/dx	dp/dx	dp/dx	dp/dx	dp/dx
	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.	s.e.
Age 16-20 omitted								
age: 21-25	0.024	-0.072***	0.044***	-0.029***	0.065	-0.063	0.050***	-0.056***
	(0.018)	(0.015)	(0.005)	(0.007)	(134.035)	(35.751)	(0.014)	(0.017)
age: 26-30	0.082***	-0.106***	0.063***	-0.041***	0.120	-0.119	0.078***	-0.118***
	(0.019)	(0.016)	(0.006)	(0.009)	(247.253)	(64.895)	(0.015)	(0.020)
age: 31-35	0.119***	-0.126***	0.072***	-0.051***	0.157	-0.129	0.091***	-0.117***
	(0.019)	(0.016)	(0.006)	(0.010)	(324.047)	(93.749)	(0.016)	(0.023)
age: 36-40	0.147***	-0.151***	0.082***	-0.050***	0.167	-0.123	0.098***	-0.139***
	(0.019)	(0.017)	(0.006)	(0.011)	(345.079)	(104.560)	(0.016)	(0.025)
age: 41-45	0.163***	-0.137***	0.087***	-0.062***	0.165	-0.171	0.100***	-0.159***
	(0.019)	(0.018)	(0.006)	(0.013)	(341.120)	(87.419)	(0.017)	(0.028)
age: 46-50	0.179***	-0.174***	0.098***	-0.066***	0.154	-0.121	0.092***	-0.118***
	(0.020)	(0.020)	(0.007)	(0.017)	(318.297)	(93.774)	(0.017)	(0.031)
age: 51-55	0.158***	-0.167***	0.087***	-0.018	0.162	-0.158	0.088***	-0.184***
	(0.021)	(0.023)	(0.008)	(0.021)	(335.117)	(89.152)	(0.019)	(0.037)
age: 55-60	0.163***	-0.198***	0.051***	-0.000	0.206	-0.094	0.046*	-0.036
	(0.024)	(0.030)	(0.011)	(0.029)	(426.133)	(147.299)	(0.026)	(0.051)
Primary and below omitted								
Junior middle school	-0.041***	-0.071***	-0.021***	-0.097***	-0.031	-0.047	-0.027***	-0.084***
	(0.005)	(0.007)	(0.003)	(0.007)	(64.458)	(41.861)	(0.010)	(0.027)
Senior middle school	-0.081***	-0.143***	-0.035***	-0.189***	-0.087	-0.137	-0.062***	-0.206***
	(0.007)	(0.009)	(0.004)	(0.008)	(179.533)	(118.145)	(0.010)	(0.027)
College and above	-0.210***	-0.153***	-0.054***	-0.276***	-0.177	-0.238	-0.122***	-0.334***
	(0.022)	(0.021)	(0.011)	(0.018)	(366.213)	(228.109)	(0.011)	(0.028)
marit	-0.158***	0.122***	-0.090***	0.053***	-0.085	0.098	-0.069***	0.055***
	(0.009)	(0.009)	(0.004)	(0.008)	(174.859)	(41.425)	(0.008)	(0.015)
r3	-0.065***	0.013**	-0.037***	-0.013***	-0.018	0.014	-0.011**	0.041***
	(0.004)	(0.005)	(0.002)	(0.004)	(36.313)	(10.786)	(0.005)	(0.010)
Observations	40080		54541		10407		11807	

Note: the base category is formal employment. region dummies not reported. *** p<0.01, ** p<0.05, * p<0.1

Table 8 OLS Regression results for migration duration: dependent variable=log(hourly income)

	rural migrants						urban residents				
	Non recent			recent			Non recent			recent	
	formal	Self emp	No contr	formal	Self emp	No contr	formal	Self emp	No contr	formal	Self emp
Age 16-20 omitted											
age: 21-25	0.064***	0.245***	0.118***	0.065***	0.137***	0.110***	0.023	-0.25	0.047	0.116***	0.022
	(0.024)	(0.072)	(0.016)	(0.009)	(0.041)	(0.007)	(0.069)	(0.320)	(0.058)	(0.029)	(0.126)
age: 26-30	0.109***	0.206***	0.138***	0.114***	0.100**	0.119***	0.113	-0.245	0.091	0.313***	0.117
	(0.027)	(0.072)	(0.018)	(0.013)	(0.043)	(0.011)	(0.070)	(0.313)	(0.060)	(0.033)	(0.129)
age: 31-35	0.097***	0.192***	0.098***	0.111***	0.137***	0.110***	0.150**	-0.274	0.069	0.311***	0.077
	(0.028)	(0.072)	(0.019)	(0.015)	(0.044)	(0.012)	(0.072)	(0.314)	(0.064)	(0.039)	(0.131)
age: 36-40	0.060**	0.170**	0.070***	0.098***	0.099**	0.092***	0.145**	-0.339	0.009	0.329***	-0.022
	(0.029)	(0.072)	(0.020)	(0.017)	(0.045)	(0.013)	(0.073)	(0.314)	(0.066)	(0.043)	(0.134)
age: 41-45	0.021	0.1	0.017	0.073***	0.100**	0.032**	0.101	-0.393	-0.003	0.340***	0.062
	(0.031)	(0.073)	(0.021)	(0.021)	(0.047)	(0.015)	(0.075)	(0.315)	(0.070)	(0.046)	(0.137)
age: 46-50	-0.012	0.07	-0.039	0.108***	0.06	0.028	0.046	-0.457	-0.021	0.186***	0.114
	(0.036)	(0.075)	(0.025)	(0.026)	(0.050)	(0.019)	(0.078)	(0.318)	(0.073)	(0.052)	(0.143)
age: 51-55	0.032	0.09	-0.070**	0.127***	0.003	0.014	0.038	-0.359	-0.07	0.225***	0.113
	(0.043)	(0.078)	(0.030)	(0.034)	(0.057)	(0.023)	(0.086)	(0.320)	(0.085)	(0.062)	(0.151)
age: 55-60	-0.167***	-0.016	-0.094**	-0.045	-0.101	-0.083***	0.214**	-0.591*	-0.077	0.149	0.307
	(0.057)	(0.084)	(0.041)	(0.048)	(0.078)	(0.032)	(0.106)	(0.327)	(0.101)	(0.091)	(0.209)
Female	-0.192***	-0.283***	-0.207***	-0.094***	-0.200***	-0.129***	-0.201***	-0.240***	-0.163***	-0.130***	-0.158***
	(0.010)	(0.015)	(0.008)	(0.007)	(0.016)	(0.005)	(0.018)	(0.040)	(0.022)	(0.016)	(0.039)
Primary and below omitted											
Junior middle school	0.153***	0.171***	0.156***	0.155***	0.165***	0.135***	0.182***	0.072	0.169***	0.254***	0.373***
	(0.014)	(0.016)	(0.009)	(0.011)	(0.017)	(0.007)	(0.053)	(0.062)	(0.045)	(0.057)	(0.069)
Senior middle school	0.376***	0.277***	0.324***	0.352***	0.271***	0.298***	0.413***	0.196***	0.356***	0.483***	0.503***
	(0.017)	(0.023)	(0.013)	(0.013)	(0.026)	(0.010)	(0.053)	(0.066)	(0.046)	(0.056)	(0.071)
College and above	0.791***	0.412***	0.724***	0.744***	0.499***	0.544***	0.874***	0.558***	0.766***	0.918***	0.755***
	(0.029)	(0.087)	(0.031)	(0.023)	(0.080)	(0.025)	(0.054)	(0.085)	(0.051)	(0.057)	(0.086)
unmarried	-0.013	-0.001	-0.075***	-0.002	0.011	-0.032***	-0.011	0.128	-0.053*	0.039*	0.181***
	(0.016)	(0.035)	(0.012)	(0.011)	(0.029)	(0.009)	(0.024)	(0.082)	(0.030)	(0.023)	(0.065)
R-squared	0.314	0.151	0.253	0.263	0.148	0.218	0.438	0.241	0.405	0.428	0.193
N	13819	9312	16949	19128	7224	28189	5570	1736	3101	5912	1536

Note, province dummies and constants not reported. Industry and occupation controlled.

Table 9 Oaxaca-Blinder decomposition results for sector choice (based on Linear Probability Model)

		Different definitions of informal employment					
		Self-employed + no contract		Self-employed		No contract	
rural migrants vs urban residents							
	difference	0.337	100	0.087	100	0.250	100
	explained	0.059	17	0.034	39	0.025	10
	unexplained	0.278	83	0.053	61	0.226	90
urban migrants vs urban residents							
	difference	0.166	100	0.057	100	0.109	100
	explained	0.028	17	-0.005	-9	0.033	30
	unexplained	0.138	83	0.062	109	0.076	70
urban migrants vs rural migrants							
Total	difference	-0.171	100	-0.030	100	-0.141	100
	explained	-0.087	51	-0.007	23	-0.080	57
	unexplained	-0.084	49	-0.023	77	-0.061	43
recent migrants	difference	-0.189	100	-0.065	100	-0.124	100
	explained	-0.099	53	-0.038	59	-0.061	49
	unexplained	-0.090	47	-0.027	41	-0.063	51
non recent migrants	difference	-0.153	100	-0.005	100	-0.149	100
	explained	-0.072	47	0.013	-267	-0.084	57
	unexplained	-0.082	53	-0.017	367	-0.064	43
recent vs non recent migrants							
rural migrants	difference	0.014	100	-0.087	100	0.101	100
	explained	-0.007	-52	-0.059	67	0.052	51
	unexplained	0.021	152	-0.028	33	0.050	49
urban migrants	difference	0.049	100	-0.027	100	0.076	100
	explained	0.014	29	-0.024	91	0.039	51
	unexplained	0.035	71	-0.002	9	0.037	49

APPENDIX

Appendix 1

Table A 1

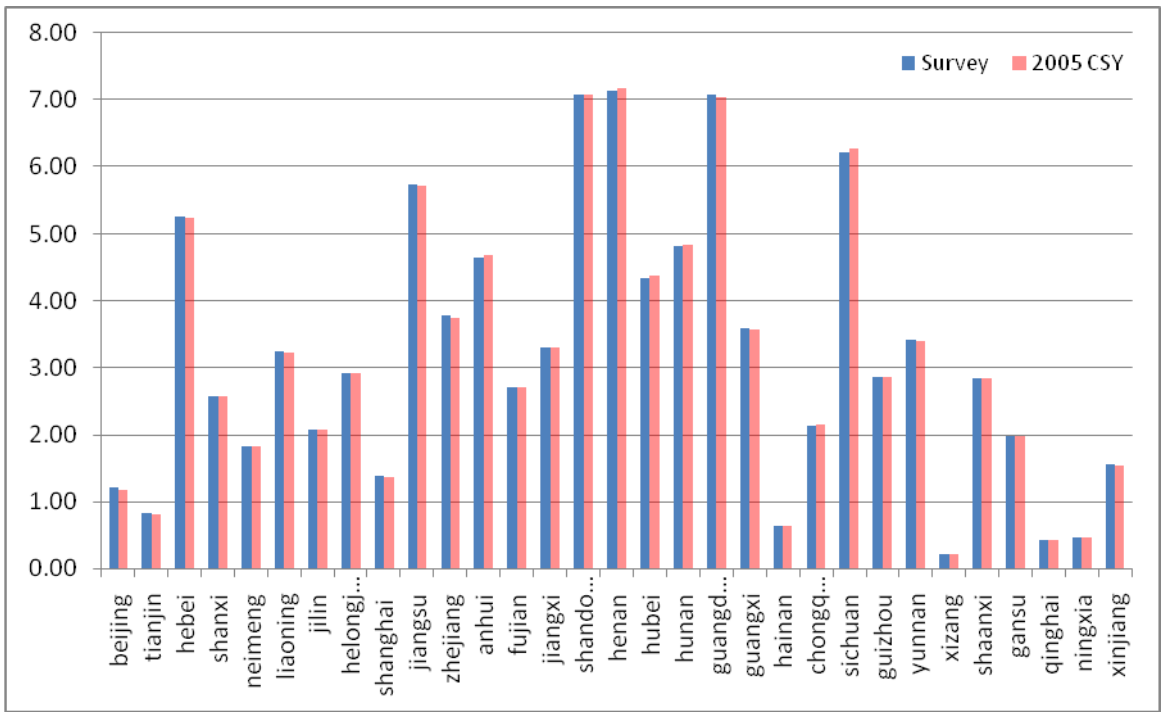
	sample size	calculated national population	Population from CSY
	2,585,481	1,292,740,500	1,307,560,000
male	50.15	50.15	51.53
female	49.85	49.85	48.47

Table A 2

Why did you leave your *hukou* registration place?	four types of migrants			
	rural-rural	rural-urban	urban-rural	urban-urban
For job or business	49.45	60.81	18.53	19.65
Job change	0.4	0.61	7.92	6.22
Employed	0.11	0.16	2.66	1.52
Training	1.26	4.05	1.55	4.01
Move house (change living place)	2.01	2.75	4.12	22.5
Marriage	18.77	5.18	11.83	9.14
Move with relatives	11.8	15.32	16.38	15.41
Move to live with relatives or friends	10.5	6.74	14.55	8.97
Temporary *hukou* change	1.06	0.37	9.62	4.08
On a business trip	0.37	0.37	0.26	0.24
Others	4.28	3.64	12.58	8.25
Weighted sample size	38,724	159,497	12,545	116,790
Unweighted sample size	28,495	132,840	7,646	87,315

Note: categories 2, and 3 are mainly for those worked in the public sector.

Figure A 1 National representativeness of the sample



Appendix 2: Brown Decomposition

Brown *et al*. (1980) analysis

As seen in the descriptive part of this paper, migrants and residents differ in important ways when looking at sector distribution, and considering the fact that the income distribution between sectors varies substantially, the differences in sectoral distribution between the two groups may have a strong impact on observed income. In addition to the Oaxaca-Blinder decomposition, we use the decomposition method of Brown *et al.* (1980) to take into account this double selection.

Formally, we assume wages for the different groups of rural migrants and urban residents to be defined as:

$$W_{ij}^{\ g} = \alpha_j^{\ g} + X_{ij}^{\ g}\beta_j^{\ g} + \varepsilon_{ij}^{\ g} \tag{4}$$

where $W_{ij}^{\ g}$ refers to the income (in log form) of individual i in sector *j*, where $j = 0,1,2$ refers to formal employment, self-employed, and no contract employees. $g = u,m$ refers to urban residents and rural migrants. $X_{ij}^{\ g}$ is a vector of independent variables, including education, age, marital status, gender, province dummies, industry dummies and occupation dummies. $\alpha_j^{\ g}$ is the intercept for category *j* and group *g*. If the model is estimated using an OLS model, we have $\overline{W}_j^{\ g} = \hat{\alpha}_j^{\ g} + \overline{X}_j^{\ g}\hat{\beta}_j^{\ g}$, the over bar "$\bar{}$" for W and X refers to sample means, and $\hat{\alpha}_j^{\ g}, \hat{\beta}_j^{\ g}$ are the OLS estimates for $\alpha_j^{\ g}, \beta_j^{\ g}$. Let $P_j^{\ g}$ be the fraction of individuals in sector *j* for *g* group, then we have $\overline{W}^g = \sum_j (P_j^{\ g}\hat{\alpha}_j^{\ g} + P_j^{\ g}\overline{X}_j^{\ g}\hat{\beta}_j^{\ g})$. Finally we can decompose the income differential between urban residents and rural migrants as follows:

$$\begin{aligned}\overline{W}^u - \overline{W}^m = &\underbrace{\sum_j P_j^{\ m}(\hat{\alpha}_j^{\ u} - \hat{\alpha}_j^{\ m})}_{\text{I}} + \underbrace{\sum_j P_j^{\ m}\overline{X}_j^{\ m}(\hat{\beta}_j^{\ u} - \hat{\beta}_j^{\ m})}_{\text{WD}} + \underbrace{\sum_j P_j^{\ m}\hat{\beta}_j^{\ u}(\overline{X}_j^{\ u} - \overline{X}_j^{\ m})}_{\text{PD}} \\ &+ \underbrace{\sum_j \overline{W}_j^{\ u}(P_j^{\ u} - \hat{P}_j^{\ m})}_{\text{QD}} + \underbrace{\sum_j \overline{W}_j^{\ u}(\hat{P}_j^{\ m} - P_j^{\ m})}_{\text{OD}}\end{aligned} \tag{5}$$

$\hat{P}_j^{\ m}$ is the *predicted* fraction of rural migrants in sector *j* assuming they are treated as urban residents based on their observable characteristics. Following Brown *et al.* (1980), we define I and WD as unjustified differences in intra-sectoral incomes, PD as the justifiable intra-sectoral income differentials, and QD and OD as the justifiable and unjustifiable fractions of sectoral segregation. So, terms PD and QD capture the share of the income differential that is due to the differences in characteristics between urban residents and rural migrants, and the other terms capture the share of income differential which is unexplained.

This decomposition procedure is easy to implement. Note that the terms I+WD+PD in equation (5) can also be written as follows: $\sum_j P_j^m[(\hat{\alpha}_j^u - \hat{\alpha}_j^m) + \overline{X}_j^m(\hat{\beta}_j^u - \hat{\beta}_j^m) + \hat{\beta}_j^u(\overline{X}_j^u - \overline{X}_j^m)]$. The terms in brackets are simply the standard Oaxaca-Blinder decomposition for the *j* sector, among which $(\hat{\alpha}_j^u - \hat{\alpha}_j^m) + \overline{X}_j^m(\hat{\beta}_j^u - \hat{\beta}_j^m)$ is the unexplained part and $\hat{\beta}_j^u(\overline{X}_j^u - \overline{X}_j^m)$ is the explained part of the income differential in sector *j*. Hence we can first decompose the income differential of each sector, and sum them up using P_j^m as weights to get the intra-sectoral explained and unexplained differences. To get the two parts of inter-sectoral differences, we need to estimate the counterfactual sector distribution $\hat{P}_j^m$.

To model the choice of sectors and for the calculation of $\hat{P}_j^m$, we estimate the following multinomial logit models:

$$P_{ij}^u = prob(y_i^u = j) = \frac{1}{1+\sum_{j=1}^{2}\exp(X_i^u \beta_j^u)} \quad \text{if} \quad j=0$$
$$P_{ij}^u = prob(y_i^u = j) = \frac{\exp(X_i^u \beta_j^u)}{1+\sum_{j=1}^{2}\exp(X_i^u \beta_j^u)} \quad \text{if} \quad j=1,2 \qquad (6)$$

Where j=0, 1, 2 refers to formal employment, self-employed, and no contract employees and we use formal employment as the base category. Using the sample of urban residents, we can have the estimates of β_j^u,denoted by $\hat{\beta}_j^u$. Then for rural migrants *i* with characteristics vector X_i^m, if they face the same sector choice structure as urban residents, their predicted probability of being in sector *j* will be:

$$\hat{P}_{ij}^m = prob(y_i^m = j) = \frac{1}{1+\sum_{j=1}^{2}\exp(X_i^m \hat{\beta}_j^u)} \quad \text{if} \quad j=0$$
$$\hat{P}_{ij}^m = prob(y_i^m = j) = \frac{\exp(X_i^m \hat{\beta}_j^u)}{1+\sum_{j=1}^{2}\exp(X_i^m \hat{\beta}_j^u)} \quad \text{if} \quad j=1,2 \qquad (7)$$

It follows that the overall predicted probability of being in sector *j* for rural migrants is: $\hat{P}_j^m = \sum_{i\in M}\hat{P}_{ij}^m$.

Then we apply the decomposition procedure of (2) to the differential between urban residents and urban migrants and between rural migrants and urban residents:

$$\overline{W}^u - \overline{W}^{um} = \mathrm{I}^1 + \mathrm{WD}^1 + \mathrm{PD}^1 + \mathrm{QD}^1 + \mathrm{OD}^1$$
$$\overline{W}^m - \overline{W}^{um} = \mathrm{I}^2 + \mathrm{WD}^2 + \mathrm{PD}^2 + \mathrm{QD}^2 + \mathrm{OD}^2,$$

where *um* stands for urban migrants. Because $\overline{W}^u - \overline{W}^m = (\overline{W}^u - \overline{W}^{um}) - (\overline{W}^m - \overline{W}^{um})$, by introducing urban migrants, we can separate both the explained and unexplained part in (2) into two parts: the *hukou* effect and the migration effect.

Table A3 Brown decomposition

	urban residents vs. rural migrants		rural vs. urban migrants		urban residents vs. urban migrants		recent vs. non recent migrants			
							rural		urban	
total differential	0.238	100.0	-0.488	100.0	-0.250	100.0	0.107	100.0	0.183	100.0
intra-sectoral	0.144	60.5	-0.449	91.9	-0.281	112.1	0.114	106.3	0.171	93.4
explained	0.174	73.0	-0.271	55.6	-0.067	26.6	0.063	58.6	0.084	46.1
unexplained	-0.030	-12.5	-0.177	36.3	-0.214	85.5	0.051	47.7	0.086	47.4
inter-sectoral	0.094	39.5	-0.040	8.2	0.031	-12.2	-0.007	-6.3	0.012	6.8
explained	0.056	23.6	-0.022	4.4	0.001	-0.5	-0.007	-6.3	0.002	0.9
unexplained	0.038	15.9	-0.018	3.7	0.029	-11.7	0.000	0.0	0.011	5.9
total explained	0.230	96.6	-0.293	60.0	-0.065	26.1	0.056	52.3	0.086	47.0
total unexplained	0.008	3.4	-0.195	40.0	-0.185	73.8	0.051	47.7	0.097	53.2

Based on the results from the multinomial logit regressions (*Tables 3* and *7*), the Brown decomposition results are presented in *Table A3***Error! Not a valid bookmark self-reference.** First, intra-sectoral differences can explain more than 60% of the total income differentials. The remaining 40% can be explained by inter-sectoral difference. In total, 97% of the difference in income between rural migrants and urban residents can be explained by differences in their personal characteristics, operating either through income determination within sectors, or through the choice of sector. As was the case with the Oaxaca-Blinder decomposition analysis, we should be wary on how to interpret these results. For example, the *unexplained inter-sectoral* figure represents about 16% of the total income differential, which is not a minor contribution. Nearly 92% of the total differential can be attributed to the intra-sectoral differential and 8% is due to inter-sectoral differential (4.4% explained and 3.7% unexplained). Because of the minor role of sectoral choice, it is not suprising that the percentages of total explained and unexplained (60% and 40%) are almost identical to the results derived from the Oaxaca-Blinder decomposition (*Table 6*).

In comparing urban and rural migrants, although the total share of inter-sectoral difference is relatively small (-12.2%), it is still worth noting that most of the inter-sectoral difference is unexplained, and this is a strong sign that urban migrants are treated differently. Nonetheless this negative effect is overcome by the large premia migrants receive in terms of income determination. Consistent with the Oaxaca-Blinder exercise, the Brown decomposition shows that a large share of income differential between urban residents and urban migrants is due to unexplained factors (74%).

Appendix 3: Gender Analysis

Table A4 Oaxaca-Blinder decomposition by gender

		urban residents vs. rural migrants	rural vs. urban migrants	Urban residents vs. Urban migrants	recent vs. non recent migrants	
					rural	urban
Male						
difference		0.223	-0.488	-0.265	0.132	0.183
Explained (%)		85	59	26	50	47
Unexplained (%)		15	41	74	50	53
formal employment	Difference	0.207	-0.583	-0.376	0.156	0.200
	Explained (%)	65	66	43	61	56
	Unexplained (%)	35	34	57	39	44
self employed	Difference	-0.092	-0.233	-0.325	0.094	0.089
	Explained (%)	93	15	33	7	3
	Unexplained (%)	7	85	67	93	97
no contract	Difference	0.158	-0.297	-0.139	0.135	0.183
	Explained (%)	136	56	-29	56	39
	Unexplained (%)	-36	44	129	44	61
recent migrants	Difference		-0.509			
	Explained (%)		56			
	Unexplained (%)		44			
non recent migrants	Difference		-0.459			
	Explained (%)		60			
	Unexplained (%)		40			
Female						
difference		0.261	-0.487	-0.226	0.022	0.155
Explained (%)		121	58	4	-71	50
Unexplained (%)		-21	42	96	171	50
formal employment	Difference	0.294	-0.604	-0.310	0.017	0.128
	Explained (%)	100	67	34	-160	53
	Unexplained (%)	0	33	66	260	47
self employed	Difference	-0.036	-0.297	-0.333	0.018	0.052
	Explained (%)	109	28	33	-14	-8
	Unexplained (%)	-9	72	67	114	108
no contract	Difference	0.153	-0.317	-0.165	0.059	0.165
	Explained (%)	185	57	-42	36	39
	Unexplained (%)	-85	43	142	64	61
recent migrants	Difference		-0.560			
	Explained (%)		61			
	Unexplained (%)		39			
non recent migrants	Difference		-0.427			
	Explained (%)		54			
	Unexplained (%)		46			

Table A5 Actual and counterfactual sectoral distribution based on multinomial logit model by gender

	actual			Predicted based on mlogit			rural migrants			urban migrants		
	urban	rural migr	urban migr	Rural mig as Urban res	Urban m As urban res	Urb m As rur m	actual		predi	actual		predi
							non recent	recent	Recent as non recent	non recent	recent	Recent as non recent
MALE												
Formal	0.589	0.326	0.525	0.474	0.581	0.436	0.331	0.322	0.331	0.533	0.516	0.522
Self-employ	0.127	0.211	0.162	0.197	0.121	0.186	0.261	0.168	0.208	0.180	0.145	0.162
No contract	0.284	0.463	0.313	0.329	0.298	0.378	0.409	0.510	0.461	0.287	0.339	0.316
FEMALE												
Formal	0.579	0.380	0.506	0.431	0.553	0.435	0.370	0.386	0.395	0.539	0.480	0.528
Self-employ	0.104	0.124	0.126	0.130	0.084	0.133	0.182	0.089	0.117	0.145	0.111	0.109
No contract	0.316	0.496	0.368	0.439	0.364	0.432	0.448	0.525	0.488	0.316	0.408	0.363

Table A6 Brown decomposition by gender

	urban residents vs. rural migrants		rural vs. urban migrants		urban residents vs. urban migrants		recent vs. non recent migrants			
							rural		urban	
Male										
total differential	0.211	100.0	-0.466	100.0	-0.268	100.0	0.137	100.0	0.178	100.0
intra-sectoral	0.146	68.9	-0.433	93.1	-0.296	110.5	0.139	101.5	0.179	100.1
explained	0.146	69.3	-0.260	55.8	-0.092	34.2	0.076	55.8	0.084	46.8
unexplained	-0.001	-0.4	-0.173	37.2	-0.204	76.2	0.063	45.7	0.095	53.3
inter-sectoral	0.066	31.1	-0.032	6.9	0.028	-10.5	-0.002	-1.5	0.000	-0.1
explained	0.052	24.6	-0.014	3.0	0.003	-1.0	-0.002	-1.8	0.001	0.7
unexplained	0.014	6.4	-0.018	3.9	0.025	-9.4	0.000	0.3	-0.001	-0.8
total explained	0.199	94.0	-0.274	58.8	-0.089	33.2	0.074	54.0	0.085	47.5
total unexplained	0.013	6.0	-0.192	41.2	-0.179	66.8	0.063	46.0	0.094	52.5
Female										
total differential	0.261	100.0	-0.487	100.0	-0.233	100.0	0.026	100.0	0.129	100.0
intra-sectoral	0.183	70.2	-0.460	94.4	-0.266	114.3	0.039	148.2	0.129	100.2
explained	0.248	95.0	-0.281	57.7	-0.050	21.6	0.000	1.1	0.056	43.8
unexplained	-0.065	-24.8	-0.179	36.7	-0.216	92.7	0.039	147.1	0.073	56.4
inter-sectoral	0.078	29.8	-0.027	5.6	0.033	-14.3	-0.013	-48.2	0.000	-0.2
explained	0.062	23.6	-0.011	2.2	0.003	-1.3	-0.008	-30.8	0.001	1.0
unexplained	0.016	6.2	-0.016	3.4	0.030	-12.9	-0.005	-17.4	-0.001	-1.2
total explained	0.309	118.6	-0.291	59.9	-0.047	20.3	-0.008	-29.7	0.058	44.7
total unexplained	-0.049	-18.6	-0.195	40.1	-0.186	79.7	0.034	129.7	0.071	55.3