ORIGINAL ARTICLE

Are migrants discriminated against in Chinese urban labour markets?

Jason Gagnon[1], Theodora Xenogiani[2] and Chunbing Xing[3*]

*Correspondence:
xingchb@bnu.edu.cn
[1] OECD Development Centre, Paris
Cedex 16, France
Full list of author information is
available at the end of the article

Abstract

We use a nationally representative survey to investigate the incidence of discrimination against internal migrant workers in urban China, considering both migrants from rural areas (rural migrants) and those from other urban areas (urban migrants). We find that both rural and urban migrants are discriminated out of jobs with formal labour contracts. Results also suggest that urban migrants are compensated for working in the informal sector by earning higher wages. There is evidence however of wage discrimination against rural *hukou* status. A semi-parametric method suggests a larger discrimination against migrants in the upper half of the wage distribution. Discrimination against migrants seems not to decrease as their duration of stay in the urban labour market increases.

JEL codes: J7; J51; O15; N35

Keywords: China; Discrimination; Labour market; Wages; Internal migration

1 Introduction

Institutional reform, a wide and increasing rural-urban income gap[1] and the easing of internal migration restrictions, have contributed to attracting millions of workers to the fast-growing urban centres of China since the early 1980s (Cai 2000). Between 1990 and 2005 more than 100 million individuals migrated from rural to urban areas (MGI 2009) and the stock of internal migrants was estimated at 150 million in 2009 (Meng and Zhang 2010).[2] According to the 1% Chinese population survey, rural migrant workers accounted for more than 20% of the labour force in the urban labour market in 2005. Previous analyses shed little doubt as to whether this influx of labour helped fill labour shortages and spur economic growth (Liang 2001; Song and Zhang 2003). The literature is more divided as to the welfare of these migrant workers.

Research on China shows that internal migrants disproportionately take up informal sector jobs, earn less and are less likely to be covered by a social safety net (see for instance the China Labor Bulletin (2008)). The situation is exacerbated by the residence registration system (*hukou*), which largely determines one's access to local public goods and services. The *hukou* system has generated a dualistic labour market in which migrants can be discriminated on income but also on the type of jobs they have access to. To

draw inference on the existence and extent of discrimination, however, researchers need to consider the fact that migrant workers have different characteristics than urban residents, such as lower education levels. To what extent their disproportional presence in the informal sector and lower income are due to lower human capital level is an empirical question that has generated a lot of attention. At present, there seems to be no consensus on the extent of discrimination against migrant workers in the Chinese urban labour market.

The literature investigating discrimination against rural migrants mainly focuses on wages. Such an approach is incomplete, since other benefits are associated with one's *hukou* status. These benefits not only include those provided by employers and pensions, but also local public services (education, medical services, and housing), the access to which depends heavily on having a good job. While employer-provided non-wage benefits can be empirically accounted for to some extent, it is hard to monetarize the value associated with local public services. One way to deal with this issue is to look at the distribution of formal and informal jobs, since they often determine access to non-wage benefits in one's job.

This paper investigates the incidence of discrimination against internal migrants in urban China using the 2005 Chinese Population Survey, a large nationally representative dataset. It questions whether their disproportionate presence in the informal labour market is due to discrimination or to their lower human capital levels. Counterfactual empirical results show that if migrants (both rural and urban) were treated as urban residents (based on observable characteristics), there would be more with formal sector jobs. An Oaxaca-Blinder decomposition (OB, Blinder 1973; Oaxaca 1973) based on a linear probability model of having an informal sector job suggests that a sizable share of the gap between migrant workers and urban residents cannot be explained by the differences in their observable characteristics, suggesting discrimination out of the formal sector against migrant workers.

The paper then turns to the income gap. An OB decomposition, based on a standard OLS wage regression for each group, suggests that most of the income gap between rural migrants and urban residents can be explained by differences in observed characteristics, which is consistent with recent studies. However, it would be misleading to draw the conclusion that rural migrants are not discriminated against in the urban labour market. First, the typical type of job each group holds is different. Even at similar income levels, urban residents are more likely to have jobs in the formal sector than migrant workers, which are more stable and associated with better social security coverage and many benefits that are hard to monetarize. Second, discrimination might vary across the income distribution. This cannot be captured by a simple OB decomposition, which only decomposes the income gaps in means.

We therefore do two additional exercises. First, we exploit the fact that the *hukou* system has two components. One is the agricultural and non-agricultural division (or rural vs. urban), and the other is the specific location in which a *hukou* is registered. We introduce urban migrants into the analysis as a new reference group for rural migrants, as they have an urban *hukou* but are not registered in the location in which they are working. We find that urban migrants, on average, earn more than both urban residents and rural migrants. Since urban migrants differ from urban residents only in migrant status, we interpret any unexplained components of the decomposition as a premium associated

with migration. This premium can also be interpreted as compensation for the loss of *hukou*-related benefits. This is consistent with the fact that urban migrants give up various benefits associated with a local *hukou* registration, which must be compensated for when they migrate. Incomes must therefore be higher for urban migrants as long as they have the freedom to choose the location in which they work. Urban migrants are a more appropriate group to study income discrimination against rural migrants, because urban local benefits have been monetarized for urban migrants at least to some extent.[3]

As urban migrants and rural migrants differ only in *hukou* status, we interpret the unexplained component of the decomposition as discrimination against rural *hukou* status. By comparing urban migrants with both rural migrants and urban residents, results suggest that the presumed absence of income discrimination against rural migrants is in fact the net effect of a discrimination against a rural *hukou* status and a premium accrued by both rural and urban migrants in the labour market.

Second, we deepen the decomposition analysis using the semi-parametric approach proposed by DiNardo *et al.* (1996) to explain the income gap along its entire distribution. Results suggest that discrimination against migrant workers occurs in the upper half of the income distribution.

We also take into account duration in the urban labour market. When the groups of migrants are divided according to whether they have spent more or less than 5 years in a location other than the one in which their *hukou* is registered, duration does not matter in alleviating discrimination against migrant workers.

This paper contributes to the literature in the following ways. First, it uses a dataset which has advantages in identifying a nationally representative migrant sample. Second, it applies an OB method to decompose the gap between migrant and non-migrant groups in the likelihood of working in the formal or informal sector. It is worth mentioning that the Brown *et al.* (1980) method takes sectoral distribution into consideration in its decomposition analysis. However, if the income gap between different sectors is small, taking the sectoral distribution into consideration will not change the results of the simple OB decomposition of the income gaps.[4] Given the fact that the formal/informal dimension is important in its own right, we perform a simple OB decomposition to investigate the gap in formal/informal sector distributions between migrants and non-migrants. Third, it applies a non-parametric analysis of the wage distribution, rather than investigating differences in means. Fourth, it considers an additional group of individuals, namely migrants with urban (or non-agricultural) *hukou* status. This group makes up a large share of the work force in the Chinese urban labour market, but is neglected in many studies. Adding this group into the analysis helps better identify discrimination along the two dimensions of the *hukou* system: local vs. non-local and rural vs. urban (or agricultural vs. non-agricultural).

The paper is organised as follows. Section 2 provides a short discussion of the institutional context and a literature review of migrant labour outcomes in the Chinese urban labour market. Section 3 describes the 2005 1% population survey microdataset used in this paper. Section 4 presents and discusses the model specification and Section 5 reports empirical results, where rural migrants are compared not only with urban residents but also with urban migrants, both in terms of wages and the sectors in which they work. Section 6 discusses policy implications and concludes.

2   The *hukou* system and the Chinese urban labour market

The institutional framework in which the Chinese urban labour market operates has been covered extensively, with particular attention being placed on the residential registration (*hukou*) system see (Cai 2000; Deng and Gustafsson 2006; Zhao 2005). This system was originally designed in the 1950s to control migration within the country by registering household members in designated rural or urban locations. In practice, one's *hukou* status is categorized by both socio-economic eligibility (agricultural vs. non-agricultural) and registered residential location (local vs. non-local) Chan and Buckingham (2008). The first classification determines entitlement to state subsidized food grain and other prerogatives. The second classification defines one's rights to many benefits (access to health care, public education, housing, and better access to jobs) in a specific locality. *Hukou* status is determined by birth, following the status held by one's parents. To migrate permanently, one needs to change registration location.

The lack of access to social benefits in one's job is likely to contribute to an important decrease in welfare; a report by the China Labor Bulletin (2008), for instance, reported that the current wage gap between urban and rural regions would increase from 3- to 6-fold in real terms, if the benefits accrued from social security were considered. Due to significant differences in employment opportunities and welfare and benefit entitlements, there is a strong incentive for rural residents to change their *hukou* registration from agricultural to non-agricultural, which requires approval from the state. Both the process and the number of such moves are tightly controlled by the government. Temporary migrants who cannot change registration location also need official approval to move. To migrate without authorization, people are vulnerable to round-ups and deportation, and cannot access many other local rights.

Despite the system, growth in the manufacturing (export) sector in China over the last twenty years has attracted a significant number of migrant workers to fast-growing urban centres. Policy makers both of the central and local governments have since faced a significant challenge in integrating new arrivals. *hukou* policies have become increasingly flexible since the 1980s. A major change was the decentralisation of the system's management, with many local governments obtaining full managerial power to determine their own policies towards migrants. It has also become easier for workers and households to transfer their registrations to other locations (in particular to small and medium-sized cities), and temporary residence permits are being granted more often. Notably, it has become possible for people to migrate and obtain a job without a valid permit. This paper focuses on migrants who have not changed their *hukou* status, since those that have changed it cannot be identified in the Chinese Population Survey. This group is nevertheless, very small.

Deliberate discrimination of migrants in cities remained legal until very recently, with the aim of reducing competition in urban centres (Cai 2000) and avoiding the creation of slums common in big countries urbanising in a relatively short period of time. Migrants are less likely to access good jobs and earn less on average than non-migrants (Zhao 2005). Local public services remain either inaccessible or expensive for them. Although rural migrants (those with an agricultural *hukou* status) may stay for long periods in urban areas, most are not covered by the urban social security system nor entitled to various other social and economic benefits. It is also worth mentioning that most of these benefits

are determined by the specific location in which the *hukou* is registered, meaning that migrants from other urban areas also face such problems.

Earlier studies on China using OB decompositions or related methods find some evidence of discrimination against rural migrants in the urban labour market. For instance, Meng and Zhang (2001) find that 51% of the wage gap between urban residents and migrants is due to unexplained factors, while Dinh and Maurer-Fazio (2004) find 25%, Wang (2005) 43% and Deng (2007) 60%. These findings are challenged by more recent studies. Using the 2002 China Household Income Project data, both Démurger *et al.* (2009) and Messinis (2013) find that the main source of disparity between urban residents and rural migrants is related to pre-market variables (education opportunities) rather than those directly related to the performance in the labour market. Using the 2005 China Urban Labor Survey (CULS),[5] Lee (2012) finds that there is only a small amount of discrimination between migrant and urban workers in terms of wages. However, when total labour remuneration is taken into account, the extent of the discrimination against migrant workers is larger. It is still difficult to put a monetary value on many of the *hukou*-related benefits, which include access to local public schools, medical services, and secure and decent jobs. Existing studies do not take into consideration these benefits. They also use data from different regions at different times and ultimately derive different conclusions, making any comparison a difficult task (Zhao 2005).

In comparison to existing literature, this paper uses a national dataset with better coverage of migrants at a similar moment in time. It also investigates differences beyond wages, looking at differences in the probability of having an informal job. Finally it investigates differences not only at the mean but across the whole wage distribution.

3   Data and summary statistics

Data used in this analysis come from a one-fifth random draw of the 2005 1% population survey, which is administered by the Chinese National Bureau of Statistics (NBS). The sample size of around 2.3 million individuals covers 31 provinces, municipalities, and autonomous regions, and is representative of mainland China. The sample studied in this analysis includes individuals aged from 16 to 60 and who are not currently in school. Employers, household workers, and observations with no or zero declared income are also not considered.

The 2005 Chinese Population Survey has advantages for studying migrant labour market outcomes relative to household surveys. 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 its inherent sampling process. For example, surveys that base their sampling frame on neighbourhoods and communities may under-represent migrants who arrived recently and do not live in registered housing and those who live at the construction site collectively. Census data does not suffer from such problems, since it aims at covering the entire population. It therefore provides a good coverage of the migrant population within China - including the floating population.

This paper focuses solely on individuals working in urban areas. The area in which an individual is currently residing or working is classified into three categories in the survey: city, town, and village. In this analysis, city and town are defined as urban areas and villages are defined as rural areas. These definitions are consistent with those used by the NBS to produce rural and urban statistics.[6]

Two questions in the questionnaire are used to identify migrants in the urban labour market: (1) In which location is your *hukou* registered? and (2) How long ago did you leave this location? Migrants are defined as those who have left their registered *hukou* location for more than six months. Another question asks the type of *hukou* (agricultural or non-agricultural) and helps divide migrants further into two categories. As individuals with an agricultural *hukou* usually come from rural areas, we refer to this group as rural migrants. Similarly, we refer to migrants with non-agricultural *hukou* as urban migrants as most of them come from other urban areas, and refer to residents who have local urban *hukou* as urban residents.

In addition, the questionnaire asks the reason for migrating. Most migrants said they had moved for work, and this is especially true for rural migrants (the share is approximately 61%). We restrict our sample to those who migrated for work or business-related reasons. The definitions of informal employment are based on an individual's declared employment status and the availability of a formal labour contract, consistent with the standard ILO guidelines (ILO 1993, 2003). The possible answers as to one's employment status include the following: employee, employer (who employs other persons), self-employed (who are neither employed by others nor do they employ others) and household worker. Only employees and the self-employed are retained for the analysis.

Those declaring themselves to be self-employed are categorised as being part of the informal sector. Employees were then further divided depending on their declared contract status, which had the following possible answers: fixed term contract, no fixed term contract (long term contract) and no contract. Individuals answering that they do not have a contract are considered to have informal sector jobs, while those answering that they have a contract, regardless of whether it is fixed or not, are considered to have formal sector jobs. Although it is required by law that the employees and employers sign a contract, many employees do not have contracts due to lack of inspection and poor enforcement.[7] These two definitions of informal employment (self-employment and the absence of a formal labour contract) are mutually exclusive. They provide a certain degree of heterogeneity within the informal sector, yet both are characterised by the lack of social security coverage.

Income data is captured using income earned in the previous month. In cases where income was not earned monthly, respondents were asked to calculate a monthly income. To calculate the hourly income, we divide the monthly income by monthly working hours, calculated as hours worked last week × 4. Educational level was divided into four levels, following ISCED standards: (1) primary and below, (2) junior middle, (3) senior middle, and (4) college and above.

Summary statistics are reported in Table 1. There are 219712 urban residents, 94621 rural migrants, and 22214 urban migrants in our analysis, accounting for 65%, 28%, and 7% of the whole sample, respectively. Compared to migrant workers, urban residents are older, primarily married and relatively highly educated. Self-employed workers are typically older than the other groups, less educated, and male, while showing higher incomes than workers without a formal contract.

In terms of wages, urban migrants earn the most of the three groups, followed by urban residents and rural migrants, who earn about half of the hourly wage of urban migrants. Rural migrants are the most likely not to have a formal labour contract, with a share of about 48%, followed by urban migrants (34%) and finally urban residents (30%). Both

Table 1 Summary statistics

	Total (1)	Formal contract (2)	Self-employed (3)	Without contract (4)
A. urban residents				
Age	38.0	38.3	39.1	37.2
Education levels (%)				
Primary and below	5.7	3.8	14.6	6.0
Junior middle school	30.1	24.1	52.6	33.0
Senior middle school	33.9	35.7	28.5	32.4
College and above	30.3	36.4	4.3	28.6
Female (%)	41.8	41.4	37.1	44.4
Not married (%)	12.0	11.5	6.7	15.2
Left the *hukou* within 0.5–5 years	-	-	-	-
Monthly income (Yuan)	1133.1	1286.5	857.9	940.2
Hourly income (Yuan)	6.6	7.6	4.3	5.4
No unemployment insurance (%)	55.4	36.0	94.1	78.3
No pension (%)	37.3	19.1	78.5	56.9
No medical insurance (%)	36.3	18.9	81.7	52.6
Obs.	219712	128509	25832	65371
B. rural migrants				
Age	30.2	28.8	35.7	29.2
Education levels (%)				
Primary and below	19.6	12.6	30.4	20.8
Junior middle school	62.5	62.9	57.2	64.2
Senior middle school	15.9	21.3	11.6	13.6
College and above	1.9	3.2	0.7	1.4
Female (%)	41.4	45.2	29.3	43.1
Not married (%)	36.5	43.4	8.5	41.7
Left the *hukou* within 0.5–5 years	73.1	73.9	60.4	77.1
Monthly income (Yuan)	997.9	1144.7	994.0	892.2
Hourly income (Yuan)	4.8	5.7	4.6	4.2
No unemployment insurance (%)	92.9	82.8	99.5	97.9
No pension (%)	85.8	66.8	97.7	95.3
No medical insurance (%)	82.5	62.2	94.9	92.7
Obs.	94621	32947	16536	45138
C. urban migrants				
Age	31.8	31.3	36.6	30.5
Education levels (%)				
Primary and below	4.6	2.4	9.8	5.5
Junior middle school	29.4	20.3	45.7	36.4
Senior middle school	36.5	36.7	33.6	37.3
College and above	29.5	40.6	10.8	20.7
Female (%)	41.1	40.2	35.2	45.0
Not married (%)	36.7	39.4	10.7	43.9
Left the *hukou* within 0.5–5 years	70.3	69.5	62.7	74.7
Monthly income (Yuan)	1688.5	2090.2	1351.6	1218.0
Hourly income (Yuan)	9.3	11.9	6.8	6.5
No unemployment insurance (%)	71.8	53.7	94.4	89.7
No pension (%)	57.0	34.5	80.6	81.2
No medical insurance (%)	59.6	37.5	85.4	82.5
Obs.	22214	11482	3272	7460

Source: 1% population survey of China (2005).

rural and urban migrants are more likely to be self-employed than urban residents, and self-employed workers have higher incomes than workers without a formal contract.

The data also show that more urban residents have formal labour contracts, relative to migrants. It is also evident from the data that migrants are less covered by social security benefits, since they generally have informal sector jobs that do not include unemployment insurance, pension benefits, or medical coverage.

4 OB and DFL decompositions

The methodology for this paper is based on two decomposition models, one based on the probability of having an informal sector job and the other on incomes.

We first investigate differences between the three groups in the probability of having an informal sector job. Three multinomial logit models are estimated, one for each group (rural migrants, urban residents, and urban migrants), with the dependent variable being a categorical variable representing each type of job, and a set of control variables. Types of job are defined as:*j = formal jobs*, *self-employment*, and *jobs without a formal labour contract*, with the latter two representing the informal sector. The results of these models are used to calculate counterfactual shares of each employment sector for each migrant group as if they were treated as another group.

Because the OB decomposition method requires two groups instead of three, a linear probability model (LPM) is then estimated to identify the main factors associated with having an informal sector job:

$$Prob^g(inf = 1|X) = X_i^g \beta^g \tag{1}$$

where *inf* is a binary variable that equals one if an individual has a job in the informal sector. Superscript $g = (ur; rm; um)$ refers to urban residents, rural migrants and urban migrants. X_i^g is a vector of control variables, including education, age, age squared, marital status, gender, and provincial dummies.

We then compute OB decompositions for the gaps in formal-informal sector distributions between different groups g based on the results derived from $\bar{Prob}^g(inf = 1) = \bar{X}^g \hat{\beta}^g$, with $\bar{Prob}^g(inf = 1)$ and $\bar{X}$ referring to sample means, and $\hat{\beta}^g$ representing the OLS estimates of β^g. The OB model comparing urban residents and rural migrants, for instance, would be as follows:

$$\bar{Prob}^u(inf = 1) - \bar{Prob}^m(inf = 1) = \left(\bar{X}^u - \bar{X}^m \right) \hat{\beta}^u + \left(\hat{\beta}^u - \hat{\beta}^m \right) \bar{X}^m \tag{2}$$

The second term on the right-hand side, $\left(\bar{X}^u - \bar{X}^m \right) \hat{\beta}^u$, represents the gap in probability due to individual characteristics (such as human capital) in the absence of discrimination. The term $\left(\hat{\beta}^u - \hat{\beta}^m \right) \bar{X}^m$, measures the relative gap in probability due to unexplained factors. Discrimination, according to the OB literature, is assumed to be the unexplained difference in the regression coefficients.[8]

We then turn to income gaps. The OB decomposition method is also applied to decompose the income gaps between different groups, replacing $\bar{Prob}^g(inf = 1)$ by income levels $\bar{W}$. In addition, we decompose the income gaps in the whole distribution. The income distribution for different groups g is modelled as follows:

$$\theta^g(W) = \int f^g(W|X)\phi^g(X)dX \tag{3}$$

where $\theta^g(W)$ is the distribution of income for group g, $f^g(W|X)$ is the conditional distribution, and $\phi^g(X)$ is the distribution of X for group g. To decompose the difference in income distributions between rural migrants and urban residents, for example, we need to construct a counterfactual distribution:[9]

$$\theta^{cf}(W) = \int f^u(W|X)\phi^m(X)dX \tag{4}$$

which is a combination of the characteristics of the migrants and the wage structure for urban residents. We define a real-valued functional, $v(\theta(.))$, that can be thought of as a set of rules mapping different distributions $\theta(.)$ to different real numbers. The $v(.)$ can be a rule calculating various statistical measures of $\theta(.)$, such as variance, Theil indices, and percentiles. The change in the distribution $v(.)$ can be divided into two parts:

$$v\left(\theta^u\right) - v\left(\theta^m\right) = \left[v\left(\theta^u\right) - v\left(\theta^{cf}\right)\right] + \left[v\left(\theta^{cf}\right) - v\left(\theta^m\right)\right] \tag{5}$$

We can then decompose any functional of the distribution into explained and unexplained parts. This exercise would be identical to the OB decomposition if we considered the difference in means.

DiNardo *et al.* (1996) point out that the counterfactual distribution in equation (5) can be constructed through a re-weighting method by doing the following. First, run a probit model with both rural migrants and urban residents with the dependent variable being an indicator for urban residents (yes=1/no=0) and the same set of control variables as before. Second, calculate a weighted variable $\hat{\rho} = (1 - \hat{p})/\hat{p}$ based on the probability of being an urban resident ($\hat{p}$) from the results of the probit model. Third, estimate the distribution function using the observations of urban residents and the weighted variable $\hat{\rho}$.

The weighted variable plays a fundamental role. Since rural migrants are less educated, observations of urban residents with lower education levels will be given more weight so that we can get closer to the characteristic distribution of rural migrants. The probability of a less educated individual being an urban resident ($\hat{p}$) is relatively lower and the weight assigned is therefore relatively higher.

An alternative approach to construct the counterfactual distribution is based on quantile regressions (*the QR approach*). This paper uses the DFL approach as it is computationally faster.[10]

5 Empirical results
5.1 Discrimination out of formal sector jobs
Table 2 presents the estimated coefficients associated with the probability of having an informal sector job for each group using a multinomial logit model. Education has a consistently significant impact on the type of job one has, with highly educated workers having a lower probability of working in informal sectors. Table 3 presents a counterfactual exercise providing the distribution of each group by type of job as if they had similar coefficients as their counterparts. Using the coefficients estimated for urban residents (as reported in the first two columns, Table 2), more rural migrants would have formal labour contracts (from 35% to 47%), the share of those without contracts would decrease by about the same amount while the share of the self-employed would largely remain unchanged. Urban migrants also seem to face discrimination out of the formal sector, but to a lesser extent than rural migrants. If urban migrants were treated as urban residents,

Table 2 Marginal effects of multinomial logit regressions

	Urban residents		Rural migrants		Urban migrants	
	Self-employed	Without contract	Self-employed	Without contract	Self-employed	Without contract
Age	-0.007	-0.100***	0.143***	-0.022***	0.140***	-0.078***
	(0.007)	(0.005)	(0.009)	(0.006)	(0.020)	(0.013)
Age2	-0.000	0.001***	-0.002***	0.000***	-0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.225***	0.114***	-0.543***	-0.143***	-0.111**	0.109***
	(0.015)	(0.010)	(0.023)	(0.015)	(0.046)	(0.032)
Junior middle school	-0.537***	-0.098***	-0.563***	-0.497***	-0.463***	-0.342***
	(0.026)	(0.024)	(0.027)	(0.022)	(0.094)	(0.085)
Senior middle school	-1.509***	-0.521***	-1.037***	-0.971***	-1.180***	-0.923***
	(0.027)	(0.024)	(0.037)	(0.027)	(0.095)	(0.084)
College and above	-3.449***	-0.694***	-1.847***	-1.412***	-2.234***	-1.600***
	(0.039)	(0.024)	(0.104)	(0.055)	(0.105)	(0.087)
Unmarried	-0.475***	0.121***	-1.103***	0.059**	-0.780***	0.203***
	(0.035)	(0.020)	(0.042)	(0.024)	(0.081)	(0.047)
N	219712		94621		22214	

Notes: *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Samples include individuals aged 16–60 years and who are out of school.

Observations of migrants who migrated for reasons not related to employment were dropped.

The base category in the multinomial logit is formal employment.

Regional dummies are included in the regressions but not reported.

Standard errors are reported below coefficients.

Source: 1% population survey of China (2005).

Table 3 Sectoral Distributions based on Multinomial Logit Regression Results

	Actual distributions (%)			Predicted distributions (%)					
	Urban residents	Rural migrants	Urban migrants	Rural migrants	Urban migrants	Urban residents	Urban migrants	Urban residents	Rural migrants
				As urban residents		As rural migrants		As urban migrants	
Formal contract	58	35	52	47	57	38	44	47	39
Self-employed	12	17	15	17	11	26	16	22	18
Without contract	30	48	34	36	32	36	40	30	43
Total	100	100	100	100	100	100	100	100	100

Notes: The sample includes those aged 16–60 years old and who are out of school.

Observations with no or zero income declared were dropped.

Source: 1% population survey of China (2005).

there would be more of them in the formal sector (57% rather than 52%), and fewer in self-employment (11% rather than 15%). If instead urban migrants had the same coefficients as rural migrants, there would be fewer urban migrants in the formal sector (from 52% to 44%) and more workers without formal labour contracts (from 34% to 40%), indicating potential discrimination against rural *hukou* status.

To further investigate the incidence of discrimination in the type of job one has, the probability of having an informal sector job is estimated using a linear probability model, and the decomposition results are reported in Table 4. The results in panel A are calculated based on the counterfactual probability $\bar{X}^m \hat{\beta}^u$, while those in panel B based on $\bar{X}^u \hat{\beta}^m$. In the first three columns of panel A, the total sample is divided by formal and informal jobs, the latter including self-employment and jobs without formal labour contracts. First, the gap in the share of workers having an informal sector job between rural migrants and urban residents is large. The share of rural migrants having an informal sector job is 34 percentage points higher than that of urban residents. The OB decomposition result shows that only 17% of the difference can be explained by differences in observed characteristics, while 83% is unexplained. Using different counterfactual probabilities (panel B) gives similar results (14% explained and 86% unexplained).

The difference in the probability of having an informal sector job between urban migrants and urban residents is smaller than it is for rural migrants (17% vs. 34%). However, 83% of the difference is also unexplained despite the fact that urban migrants enjoy the highest average income level. Thus, the results suggest significant discrimination against both types of migrants out of formal sector jobs. Observed characteristics explain a relatively larger share (around 50%) of the probability of having an informal sector job when rural migrants are compared to urban migrants. However, half of the difference is nevertheless due to unexplained factors and we interpret this as evidence of some discrimination against rural *hukou* status.

Using the same method, we also decompose the gaps in the probability of being self-employed (columns 4–6) *or* without a formal contract (columns 7–9), the results also indicate discrimination against both rural and urban migrant workers out of formal sector jobs.

5.2 Income discrimination against rural migrants, revisited

5.2.1 OB decomposition results

We then run straightforward OLS income regressions on the determinants of hourly incomes by type of job (formal work, self-employment and without a formal labour contract) for each type of worker (urban residents, rural migrants, urban migrants). Controls include age, age squared, four levels of education, marital status and gender and the results are presented in Table 5.[11]

There are significant differences in the coefficients of each group. The returns to education, for example, are generally higher for urban residents than for rural migrants (see panels A and B). As for urban migrants (panel C), the returns to education are lower than those for urban residents, but higher than those for rural migrants. Within each group in panels A, B, and C, the returns to education are the highest for formal contract jobs, followed by the jobs without contracts, and then the self-employed. Incomes also differ according to gender, marital status, and age. Earnings for women are significantly lower than those of men. The gap is larger for the self-employed than for the formally

Table 4 Oaxaca-Blinder decomposition of the gap in the average probability of having an informal sector job

	Self-employed and without a contract			Self-employed			Without a contract		
A: W=1	coef.	s.e.	%	coef.	s.e.	%	coef.	s.e.	%
Rural migrants vs. urban residents									
Difference	0.337	0.002		0.087	0.001		0.250	0.002	
Explained	0.059	0.001	17	0.035	0.001	41	0.023	0.001	9
Unexplained	0.279	0.002	83	0.051	0.001	59	0.227	0.002	91
Urban migrants vs. urban residents									
Difference	0.166	0.003		0.057	0.002		0.109	0.003	
Explained	0.028	0.001	17	-0.005	0.001	-9	0.034	0.001	31
Unexplained	0.138	0.003	83	0.062	0.002	109	0.075	0.003	69
Urban migrants vs. rural migrants									
Difference	-0.171	0.004		-0.030	0.002		-0.141	0.003	
Explained	-0.087	0.003	51	-0.006	0.003	20	-0.081	0.003	57
Unexplained	-0.085	0.005	49	-0.024	0.003	80	-0.061	0.005	43
Non-recent migrants									
Difference	-0.191	0.006		-0.072	0.005		-0.119	0.006	
Explained	-0.100	0.006	52	-0.048	0.005	66	-0.052	0.006	44
Unexplained	-0.092	0.008	48	-0.024	0.007	34	-0.067	0.008	56
Recent migrants									
Difference	-0.162	0.004		-0.015	0.003		-0.147	0.004	
Explained	-0.079	0.004	49	0.007	0.003	-46	-0.086	0.004	59
Unexplained	-0.082	0.006	51	-0.022	0.004	146	-0.060	0.006	41
B: W=0	coef.	s.e.	%	coef.	s.e.	%	coef.	s.e.	%
Rural migrants vs. urban residents									
Difference	0.337	0.002		0.087	0.001		0.250	0.002	
Explained	0.047	0.004	14	-0.093	0.003	-107	0.140	0.004	56
Unexplained	0.291	0.004	86	0.180	0.003	207	0.111	0.004	44
Urban migrants vs. urban residents									
Difference	0.166	0.003		0.057	0.002		0.109	0.003	
Explained	-0.034	0.004	-21	-0.067	0.002	-117	0.033	0.003	30
Unexplained	0.200	0.005	121	0.124	0.003	217	0.076	0.004	70
Urban migrants vs. rural migrants									
Difference	-0.171	0.004		-0.030	0.002		-0.141	0.003	
Explained	-0.098	0.004	57	-0.028	0.003	94	-0.070	0.004	50
Unexplained	-0.073	0.005	43	-0.002	0.003	6	-0.071	0.005	50
Non-recent migrants									
Difference	-0.191	0.006		-0.072	0.005		-0.119	0.006	
Explained	-0.086	0.007	45	-0.048	0.005	67	-0.038	0.006	32
Unexplained	-0.105	0.009	55	-0.024	0.007	33	-0.081	0.008	68
Recent migrants									
Difference	-0.162	0.004		-0.015	0.003		-0.147	0.004	
Explained	-0.105	0.005	65	-0.021	0.003	138	-0.085	0.005	58
Unexplained	-0.056	0.006	35	0.006	0.004	-38	-0.062	0.006	42

The decomposition results are based on a linear probability model.

The sample includes individuals aged 16–60 years and who are out of school.

Observations with no or zero income declared were dropped.

Results in panel A (W=1) are calculated based on the counterfactual $\bar{x}^m \hat{\beta}^u$, and those in panel B (W=0) based on $\bar{x}^u \hat{\beta}^m$.

Source: 1% population survey of China (2005).

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

	Formal contract	Self-employed	Without a contract
A: Urban residents			
Age	0.034***	0.017***	0.020***
	(0.002)	(0.004)	(0.002)
Age2	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Female	-0.161***	-0.259***	-0.189***
	(0.003)	(0.008)	(0.004)
Junior middle school	0.372***	0.121***	0.213***
	(0.009)	(0.013)	(0.010)
Senior middle school	0.650***	0.233***	0.483***
	(0.009)	(0.014)	(0.010)
College and above	1.070***	0.573***	1.021***
	(0.009)	(0.023)	(0.010)
Unmarried	-0.009	-0.065***	-0.075***
	(0.007)	(0.019)	(0.008)
R-squared	0.349	0.127	0.368
N	128509	25832	65371
B: Rural migrants			
Age	0.039***	0.021***	0.037***
	(0.002)	(0.004)	(0.002)
Age2	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)
Female	-0.181***	-0.293***	-0.194***
	(0.006)	(0.011)	(0.004)
Junior middle school	0.195***	0.170***	0.160***
	(0.009)	(0.012)	(0.006)
Senior middle school	0.441***	0.279***	0.349***
	(0.011)	(0.017)	(0.008)
College and above	0.948***	0.527***	0.746***
	(0.018)	(0.058)	(0.019)
Unmarried	-0.022**	-0.025	-0.049***
	(0.010)	(0.022)	(0.007)
R-squared	0.213	0.127	0.195
N	32947	16536	45138
C: Urban migrants			
Age	0.077***	0.006	0.049***
	(0.005)	(0.012)	(0.005)
Age2	-0.001***	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)
Female	-0.187***	-0.229***	-0.162***
	(0.012)	(0.028)	(0.013)
Junior middle school	0.252***	0.189***	0.177***
	(0.040)	(0.047)	(0.030)
Senior middle school	0.532***	0.348***	0.419***
	(0.040)	(0.049)	(0.030)
College and above	1.089***	0.687***	0.916***
	(0.040)	(0.059)	(0.032)
Unmarried	0.025	0.157***	-0.003
	(0.017)	(0.051)	(0.020)
R-squared	0.382	0.175	0.340
N	11482	3272	7460

Notes: *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.
Samples include those aged 16–60 years old and who are out of school.
Observations with no or zero income declared were dropped.
Province dummies and a constant are included in the regressions but not reported.
Standard errors are reported in brackets.
Source: 1% population survey of China (2005).

employed and workers without a formal labour contract, but there are only small differences in coefficients among urban residents, rural migrants, and urban migrants along gender lines.

Using these results, an OB decomposition is computed between rural migrants and urban residents. The results are shown in the first three columns of Table 6. If the type of

Table 6 Oaxaca-Blinder decomposition of income gaps

		Urban residents vs. rural migrants			Rural migrants vs. urban migrants			Urban residents vs. urban migrants		
A: W=1		coef.	s.e.	%	coef.	s.e.	%	coef.	s.e.	%
Total difference	Difference	0.238	0.002		-0.488	0.006		-0.250	0.006	
	Explained	0.232	0.003	97	-0.255	0.005	52	-0.078	0.003	31
	Unexplained	0.006	0.003	3	-0.233	0.006	48	-0.172	0.005	69
Formal employment	Difference	0.252	0.004		-0.604	0.008		-0.351	0.008	
	Explained	0.220	0.004	87	-0.351	0.008	58	-0.141	0.005	40
	Unexplained	0.033	0.005	13	-0.252	0.009	42	-0.210	0.007	60
Self employed	Difference	-0.096	0.007		-0.236	0.015		-0.333	0.015	
	Explained	-0.096	0.006	100	-0.032	0.009	14	-0.110	0.007	33
	Unexplained	0.000	0.008	0	-0.204	0.016	86	-0.222	0.014	67
No contract	Difference	0.153	0.004		-0.302	0.008		-0.149	0.008	
	Explained	0.212	0.004	139	-0.150	0.006	50	-0.003	0.006	2
	Unexplained	-0.060	0.005	-39	-0.152	0.008	50	-0.146	0.007	98
Recent migrants	Difference	0.267	0.003		-0.475	0.007		-0.208	0.007	
	Explained	0.257	0.003	96	-0.253	0.005	53	-0.060	0.004	29
	Unexplained	0.010	0.003	4	-0.222	0.007	47	-0.148	0.005	71
Non recent migrants	Difference	0.160	0.004		-0.510	0.011		-0.350	0.011	
	Explained	0.163	0.003	102	-0.246	0.009	48	-0.120	0.006	34
	Unexplained	-0.003	0.004	-2	-0.264	0.012	52	-0.230	0.009	66
B: W=0		coef.	s.e.	%	coef.	s.e.	%	coef.	s.e.	%
Total difference	Difference	0.238	0.002		-0.488	0.006		-0.250	0.006	
	Explained	0.168	0.005	71	-0.317	0.007	65	-0.157	0.006	63
	Unexplained	0.070	0.005	29	-0.171	0.007	35	-0.094	0.006	37
Formal employment	Difference	0.252	0.004		-0.604	0.008		-0.351	0.008	
	Explained	0.228	0.008	90	-0.414	0.010	69	-0.211	0.009	60
	Unexplained	0.025	0.008	10	-0.190	0.011	31	-0.140	0.009	40
Self employed	Difference	-0.096	0.007		-0.236	0.015		-0.333	0.015	
	Explained	-0.059	0.006	62	-0.053	0.015	22	-0.141	0.010	42
	Unexplained	-0.037	0.009	38	-0.184	0.019	78	-0.192	0.015	58
No contract	Difference	0.153	0.004		-0.302	0.008		-0.149	0.008	
	Explained	0.111	0.007	73	-0.185	0.009	61	-0.102	0.009	69
	Unexplained	0.042	0.007	27	-0.117	0.010	39	-0.047	0.010	31
Recent migrants	Difference	0.267	0.003		-0.475	0.007		-0.208	0.007	
	Explained	0.175	0.005	66	-0.324	0.008	68	-0.135	0.007	65
	Unexplained	0.092	0.006	34	-0.151	0.008	32	-0.073	0.007	35
Non recent migrants	Difference	0.160	0.004		-0.510	0.011		-0.350	0.011	
	Explained	0.117	0.009	73	-0.289	0.012	57	-0.232	0.011	66
	Unexplained	0.044	0.010	27	-0.221	0.014	43	-0.118	0.012	34

Notes: Results are based on the income regressions of Tables 5.
Samples include those aged 16–60 years old and who are out of school.
Observations with no or zero income declared were dropped.
Results in panel A (W=1) are calculated based on the counterfactual $\bar{X}^m \hat{\beta}^u$, and those in panel B (W=0) based on $\bar{X}^u \hat{\beta}^m$.
Source: 1% population survey of China (2005).

employment sector is not taken into account, nearly all of the income gap (97%) between rural migrants and urban residents can be attributed to differences in observed individual characteristics. If we use an alternative counterfactual income ($\bar{X}^u \hat{\beta}^m$), the explained share will decrease, but still as high as 70%. Running separate OB decompositions for each sector does not alter the results by much. According to these OB decompositions, the gap between rural migrants and urban residents is mainly due to differences in human capital.
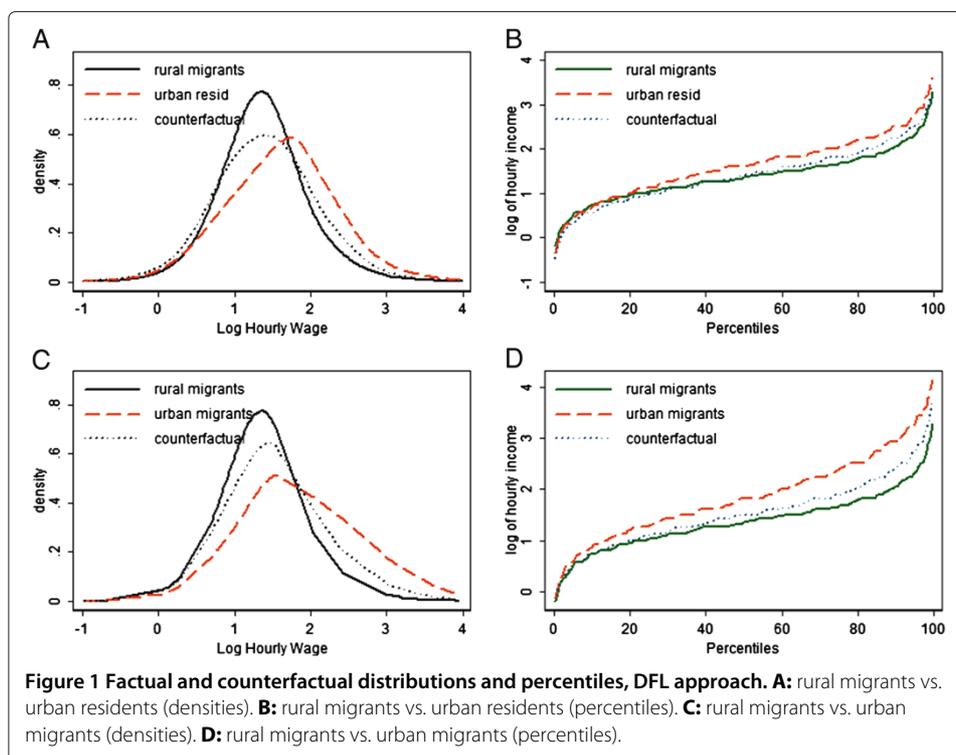
The results may again be the net effect of a premium accrued from migration and discrimination associated with a rural *hukou* status. Urban migrants with urban *hukou* status are used to disentangle the two effects.

The middle three columns of Table 6 report OB decomposition results for the income gap between rural and urban migrants, two groups that differ only in rural/urban *hukou* status. The general results indicate that 52% of the income gap between these two groups can be explained by observed differences in their individual characteristics and 48% remains unexplained. We also find some heterogeneity across sectors. In the formal sector, observed differences in characteristics can explain 58% of the income gap whereas in the group without formal contracts the share of explained gap is only 50%. Amongst the self-employed workers, only 14% of the gap in this group can be explained by observable individual characteristics, suggesting discrimination against rural *hukou* status. These patterns remain if alternative counterfactual incomes are used (see panel B, Table 6).

Urban migrants are then measured up against urban residents to evaluate discrimination associated with being a migrant (not registered locally) but with a non-agricultural *hukou* status. Since urban migrants earn more than urban residents, the OB decomposition in the last three columns of Table 6 suggests that migrants do indeed gain a premium. The decomposition exercise shows that 69% of the income difference is unexplained, on average, between urban migrants and urban residents. But this result is sensitive to the choice of the counterfactual income (it becomes 37% in panel B, Table 6). Moreover, decomposition results broken down by sector show that the unexplained shares are 60%, 67% and 98% for workers in the formal sector, the self-employed, and workers with no formal labour contract respectively. These unexplained shares become smaller in panel B, but it is still safe to draw the conclusion that if urban migrants were treated the same as urban residents, their incomes would decrease but would still be higher than those of urban residents with the same observable characteristics.

There may be two reasons for this result. First, urban migrants are positively selected in terms of unobservable characteristics. Second, urban migrants ask for compensation for giving up secure jobs. Suppose an urban worker is considering a move to another city. She must find a better job, otherwise she will not migrate. So compared to those who do not migrate with identical characteristics, she must have a higher income to compensate for the *hukou* related benefits she gives up.

To investigate the effect of *hukou* status on discrimination more deeply, we look across the income distribution to see if the effect differs at different points. Panel A of Figure 1 shows the actual income distributions of rural migrants and urban residents and the counterfactual distribution supposing rural migrants were paid as if they were urban residents. To obtain the counterfactual distribution, we first run a probit model including both rural migrants and urban residents with the dependent variable equalling 1 if the individual is an urban resident. The predicted weights $\hat{\rho} = (1 - \hat{p})/\hat{p}$ are then applied to

Figure 1 Factual and counterfactual distributions and percentiles, DFL approach. A: rural migrants vs.
urban residents (densities). **B:** rural migrants vs. urban residents (percentiles). **C:** rural migrants vs. urban
migrants (densities). **D:** rural migrants vs. urban migrants (percentiles).

the urban resident sample. The results indicate that even if rural migrants were paid like
urban residents according to their observable characteristics, their average income level
would not increase significantly. This is consistent with the OB decomposition results.
However, their income dispersion would increase. This means that rural migrants in the
upper half of the income distribution would earn higher incomes, while those at the lower
half would earn less. This pattern is also clear in panel B of Figure 1, where the actual and
counterfactual percentiles are reported. For the middle level percentiles, the actual and
counterfactual income levels for rural migrants are not significantly different.

We also compare rural migrants with urban migrants (rather than urban residents)
using the DFL approach to construct counterfactual distributions. Results, shown in Panel
C and D of Figure 1, indicate that if rural migrants were paid like urban migrants, their
income distribution would move further to the right, but remain lower than the distribu-
tion of urban migrants. This effect is larger for individuals in the upper half of the income
distribution.

In conclusion, there seems to be some, although not overwhelming support for discrim-
ination against those with rural *hukou* status, while there seems to be more support for a
premium garnered by being a migrant.

5.3 Does migration duration matter?

The Harris and Todaro (1970) framework suggests that migrants may first enter the infor-
mal labour market, accumulating experience before obtaining an opportunity for a formal
sector job. The income premium earned by migrants observed above likely varies accord-
ing to the amount of time migrants have spent in the urban labour market. In order to
evaluate this labour market assimilation effect, we separate migrant samples according to

their duration in the urban labour market and then investigate the differences in the distribution by job type and income, as in the exercises presented earlier. To define the two groups, we use a cutoff of 5 years in the urban labour market.[12]

Descriptive statistics are presented in Table 1. For both rural and urban migrants, the duration of their migration episode is important. More than 70% of migrants (rural and urban) have less than 5 years of local urban living experience. Both rural and urban self-employed migrants tend to have resided longer in the host labour market, while workers with no formal labour contract tend to be more recent arrivals.

The first two columns in Table 7 present the distribution by job type for both recent and non-recent rural migrants. Job type changes significantly as urban area experience increases, but most changes happen within the informal sector. The share of rural migrants in the formal sector remains almost unchanged. On the other hand, the share of self-employed migrants increases from 14% to 26%. The pattern for urban migrants is similar to that of rural migrants. Despite the increase in urban area experience, the probability of an urban migrant having a formal sector job only increases slightly (by 2 percentage points). The share of self-employed migrants, however, increases from 13% to 18%.

Table 8 reports results from a multinomial logit model based on job type for both recent and non-recent migrants, showing that the probability of being self-employed increases as people age. It is therefore expected that if recent migrants (both rural and urban) are treated as non-recent migrants, a greater number of them will be self-employed, but the share of formal sector employment will remain relatively unchanged. The results suggest that although migration duration matters for the type of jobs held by migrants, it plays a minor role in helping migrants obtain a formal sector job. In addition, the OB decomposition based on a linear probability model in Table 4 suggests that the discrimination against rural migrants does not decrease with duration in an urban area. We reach this conclusion by comparing rural and urban migrants with the same amount of urban area experience, and the results indicate that the unexplained part of the gap (both in terms of absolute difference and in relative share) does not decrease.

We also look at the determinants of income for rural and urban migrants depending on duration, and the results are reported in Table 9. For recent migrants, incomes increase relatively more quickly and the gap between women and men is lower. These results combined with those in Table 5 are used to perform OB decompositions. The results in Table 6 indicate that most of the income gap between both recent and non-recent migrants and

Table 7 Sectoral distribution of migrants based on multinomial logit model (by migration duration)

	Rural migrants			Urban migrants		
	Actual distribution		Predicted distribution	Actual distribution		Predicted distribution
	Non-recent	Recent	Recent as non-recent	Non-recent	Recent	Recent as non-recent
Formal contract	34	35	36	53	51	54
Self-employed	26	14	18	18	13	15
Without contract	41	50	46	29	36	32
Total	100	100	100	100	100	100

Samples include individuals aged 16–60 years and who are out of school.
Observations of migrants who migrated for reasons not related to employment were dropped.
Source: 1% population survey of China (2005).

Table 8 Marginal effects of multinomial logit results (by migration duration)

Rural migrants	Non-recent		Recent	
	Self-employed	No contract	Self-employed	No contract
Age	0.129***	-0.013	0.127***	-0.028***
	(0.018)	(0.014)	(0.011)	(0.007)
Age2	-0.001***	0.000	-0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.382***	-0.079**	-0.612***	-0.169***
	(0.039)	(0.032)	(0.028)	(0.018)
Junior middle school	-0.486***	-0.413***	-0.590***	-0.526***
	(0.044)	(0.039)	(0.035)	(0.027)
Senior middle school	-1.000***	-0.903***	-1.057***	-0.988***
	(0.059)	(0.050)	(0.048)	(0.033)
College and above	-2.325***	-1.351***	-1.615***	-1.428***
	(0.197)	(0.107)	(0.123)	(0.065)
Unmarried	-0.701***	0.286***	-1.211***	-0.020
	(0.079)	(0.051)	(0.050)	(0.028)
N	25482		69139	
Urban migrants	Non-recent		Recent	
	Self-employed	No contract	Self-employed	No contract
Age	0.124***	-0.036	0.156***	-0.081***
	(0.036)	(0.028)	(0.025)	(0.015)
Age2	-0.001***	0.000	-0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.084	0.086	-0.123**	0.113***
	(0.078)	(0.062)	(0.057)	(0.037)
Junior middle school	-0.314**	-0.269*	-0.569***	-0.413***
	(0.144)	(0.139)	(0.124)	(0.108)
Senior middle school	-1.024***	-0.759***	-1.287***	-1.032***
	(0.147)	(0.139)	(0.125)	(0.108)
College and above	-2.089***	-1.394***	-2.346***	-1.724***
	(0.169)	(0.145)	(0.137)	(0.110)
Unmarried	-0.455***	0.314***	-0.884***	0.154***
	(0.145)	(0.088)	(0.097)	(0.055)
N	6608		15606	

Notes: *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.
Samples include individuals aged 16–60 years and who are out of school.
Migrants that migrated for reasons not related to employment were dropped.
The base category in the multinomial logit is formal employment.
Regional dummies are included in the regressions but not reported.
Standard errors are reported in brackets.
Source: 1% population survey of China (2005).

urban residents is still due to observable characteristics. By comparing rural and urban migrants, the unexplained part of the income gap between the two groups does not decrease as duration in an urban area increases.

These results regarding duration spent in an urban area indicate that discrimination, both for job type and income, does not decrease as duration in urban areas for migrants increases.

6   Conclusions

In this paper, a nationally representative sample of individuals is used to investigate whether migrants are discriminated in terms of income and out of jobs with formal labour

Table 9 OLS income regression results by migration duration: dependent variable=log (hourly income)

	Rural migrants		Urban migrants	
	Non-recent	Recent	Non-recent	Recent
Age	0.021***	0.034***	0.033***	0.065***
	(0.003)	(0.002)	(0.008)	(0.004)
Age2	-0.000***	-0.000***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.283***	-0.167***	-0.237***	-0.168***
	(0.008)	(0.004)	(0.018)	(0.010)
Junior middle school	0.180***	0.184***	0.208***	0.229***
	(0.009)	(0.005)	(0.038)	(0.027)
Senior middle school	0.406***	0.404***	0.462***	0.511***
	(0.012)	(0.007)	(0.038)	(0.027)
College and above	0.955***	0.863***	1.094***	1.072***
	(0.028)	(0.014)	(0.039)	(0.028)
Unmarried	-0.076***	-0.012*	-0.035	0.047***
	(0.013)	(0.006)	(0.026)	(0.015)
R-squared	0.206	0.181	0.394	0.379
N	25482	69139	6608	15606

Notes: *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.
The dependent variable is the log of hourly income.
Samples include those aged 16–60 years old and who are out of school.
Observations with no or zero income declared were dropped.
Province dummies and a constant are included in the regressions but not reported.
Standard errors are reported in brackets.
Source: 1% population survey of China (2005).

contracts in the urban labour market in China. It contributes to the existing literature in several important ways. First, as opposed to other papers investigating discrimination in China, the data are nationally representative and provide a better coverage of rural and urban migrants. Second, it distinguishes between rural and urban migrants in order to independently identify discrimination against rural *hukou* status and discrimination against non-local *hukou* status. Third, it breaks down jobs according to whether they include formal labour contracts, and differentiates between those without a formal labour contract and self-employed workers.

An OB decomposition of the fitted probability of having an informal sector job shows that migrants (both rural and urban) are discriminated against for jobs with formal labour contracts. The extent of discrimination is larger for rural migrants indicating further discrimination against those with rural *hukou* status.

A comparison of rural migrants and urban residents shows that nearly all of the income gap can be explained by differences in individual characteristics, suggesting that discrimination is almost negligible. A comparison between rural migrants and urban migrants, isolating the migration effect but not the rural/urban *hukou* status effect, shows that 40% of the income gap is unexplained by observed characteristics. By comparing instead urban migrants with urban residents, this time isolating the *hukou* effect (agricultural/non-agricultural), we find evidence of discrimination in income. Since urban migrants earn the most of all groups, we therefore presume that the absence of discrimination from the comparison between rural migrants and urban residents is a net effect of discrimination against rural *hukou* status and a premium accrued generally by migrating.

There are limitations in the approach taken in this paper. The first is in the choice of the appropriate reference group. Although this paper takes a step forward by using urban migrants as an additional reference group, there may be unobservable characteristics between rural migrants and urban migrants. Moreover, while self-selection into the labour force could be dealt with in this paper, another source of endogeneity may stem from choice of sector. In fact, the results of our OB decompositions for urban migrants suggests that the self-employed may indeed be self-selected.

The policy implications of our results suggest that, since rural migrants enjoy a premium as migrants yet face discrimination at the same time, the reason they may be worse off when compared to urban residents is due to their lower levels of human capital. Increasing the 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 discrimination in access to jobs with formal labour contracts, reforming the labour market, notably removing barriers to accessing public goods, may help migrants access better jobs.

This study suggests evidence of discrimination in the urban labour market against those with rural *hukou* status. In addition, migrants without social security coverage face high costs for health services. In terms of schooling and childcare, many migrants leave their children home in the rural parts of China, in effect putting more pressure on household members left-behind and adding to the already existing social strain caused by migration. Hence it is important to ensure that migrants have access to basic social services, even in cases where they are employed informally.

Endnotes

[1] The ratio in 2007, for instance, was 3.3 to 1 (China Statistical Yearbook, National Bureau of Statistics of China 2008).

[2] This number is only slightly below the number of total international migrants in the world, which in 2005 was estimated at 191 million (United Nations 2009).

[3] This analysis depends on the assumption that people with different *hukou* registrations have the same (unobserved) ability. However, people with local *hukou* have higher innate ability (this may be true, for example, for those who obtained *hukou* through special achievement). In this case, the income gap would be a lower estimate for the true *hukou* premium. It might also be true that people born with local *hukou* status have lower ability or less entrepreneurial spirit compared to migrants without local *hukou* status. These possibilities need to be taken into consideration when interpreting results.

[4] The major difference between the formal and informal sector comes from non-wage benefits, which are hard to quantify. For example, formal jobs are more secure, offer better conditions and benefits which may not be reflected in the wage. Workers may even be willing to accept a lower wage to have a formal sector job. These aspects cannot be dealt with using the Brown *et al.* (1980) method. We also decompose the income gaps using the Brown method and the results are similar to those obtained using the regular OB decomposition.

[5] The China Urban Labor Survey covers only five cities: Shanghai, Wuhan, Shenyang, Fuzhou, and Xian, all of which are provincial capitals.

[6] Cities (*shiqu*) refer to the city proper of regions of different administrative levels. Prefecture-level cities or higher level cities directly controlled by the central government usually govern surrounding counties, including rural areas. In cases of county level cities, city proper refers to the areas where the county government is located and some nearby neighbourhoods. Towns (zhen) refer to where the town level government is

located. Zhen is a level of administrative government approved by higher level government. In Chinese statistics, urban (or *chengzhen*) refers to city and town. Villages (xiangcun) refer to rural areas.

[7] The contracts should have the following terms: the term (length) of the contract; tasks; working time and vacations allocated; compensation; social security (such as unemployment insurance; pension; medical insurance; and housing). Some employers (usually in the private sector) do not sign contracts mainly to circumvent the need to pay insurance for employees.

[8] We also calculate the OB decomposition using an alternative counterfactual income $\bar{X}^u \hat{\beta}^m$, the gap can then be decomposed into: $\left(\bar{X}^u - \bar{X}^m\right) \hat{\beta}^m + \left(\hat{\beta}^u - \hat{\beta}^m\right) \bar{X}^u$.

[9] Similarly, we can construct an alternative counterfactual distribution for the decomposition: $\theta^{cf1}(W) = \int f^u(W|X)\phi^m(X)dX$.

[10] By running quantile regressions at different quantiles using data of urban residents, we obtain a detailed description of its conditional distribution. The estimated skill price structure (conditional distribution) can then be applied to data (skill distribution) of rural migrants by multiplying the quantile coefficient matrix of urban residents by the data matrix of rural migrants. This approach estimates the conditional distribution explicitly, with some risk of imposing strong restrictions on its structure. No effort is needed to estimate the composition change parametrically. Chernozhukov *et al.* 2013 show that the quantile approach and DFL reweighting approach are equally valid under correct specifications. We choose DFL because it is computationally fast. We also tried to use the QR approach to construct the counterfactual distributions. They produce similar results.

[11] We also do exercises controlling for industry and occupation dummies, and both the regression results and the decomposition results are similar to those without controlling for industry and occupation.

[12] We also use a 3-year cutoff as a robustness check, and the results are similar.

Author details
[1]OECD Development Centre, Paris Cedex 16, France. [2]OECD Directorate for Employment, Labour and Social Affairs, Paris Cedex 16, France. [3]Beijing Normal University, Beijing, China.

References
Blinder A (1973) Income discrimination: reduced form and structural estimates. J Hum Resour 8(4):436–455
Brown RS, Moon M, Zoloth BS (1980) Incorporating occupational attainment in studies of male-female earnings differentials. J Human Resour 15(1):3–28
Cai F (2000) The invisible hand and visible feet: internal migration in China. World Econ China 5:24–31
Chan KW, Buckingham W (2008) Is China abolishing the Hukou system? China Quarterly 195:582–606
Chernozhukov V, Fernandez-Val I, Melly B (2013) Inference on counterfactual distributions. Econometrica 81(6):2205–2268
China Labor Bulletin (2008) Incomes in China. www.china-labor.org.hk
Démurger S, Gurgand M, Li S, Yue X (2009) Migrants as second-class workers in urban China? A Decomposition Analysis. J Comp Econ 37(4):610–628
Deng Q (2007) Earnings differential between urban residents and rural migrants: evidence from Oaxaca-Blinder and quantile regression decompositions. Chin J Population Sci 2:8–16. (in Chinese)
Deng Q, Gustafsson B (2006) China's Lesser Known Migrants. IZA, Bonn. IZA Discussion Paper No. 2152
DiNardo J, Fortin N, Lemieux T (1996) Labor market institutions and the distribution of wages, 1973–1992: a semi-parametric approach. Econometrica 64(5):1001–1044

Dinh N, Maurer-Fazio M (2004) Differential rewards to, and contributions of, education in urban China's segmented labor markets. Pac Econ, Rev 9(3):173–189

Harris J, Todaro M (1970) Migration, unemployment, and development: a two-sector analysis. Am Econ Rev 60:126–142

ILO (1993) Resolution concerning statistics of employment in the informal sector, adopted by the Fifteenth International Conference of Labour Statisticians In: Fifteenth International Conference of Labour Statisticians, Geneva, 19-28 January 1993. Report of the Conference (Geneva, ILO)

ILO (2003) Guidelines concerning a statistical definition of informal employment, endorsed by the Seventeenth International Conference of Labour Statisticians (November-December 2003) In: Seventeenth International Conference of Labour Statisticians, Geneva, 24 November-3 December. Report of the Conference (Geneva, doc. ICLS/17/2003/R.)

Lee L (2012) Decomposing income differentials between migrant workers and urban workers in urban China's labor markets. China Econ Rev 23(2):461–470

Liang Z (2001) The age of migration in China. Popul Dev, Rev 27(3):499–524

Meng X, Zhang D (2010) Labor market impact of large scale internal migration on Chinese urban 'Native' workers. IZA, Bonn. IZA Discussion Paper No. 5288

Meng X, Zhang J (2001) The two-tier labor market in urban China: occupational segregation and income differentials between urban residents and rural migrants in Shanghai. J Comp Econ 29(3):485–504

Messinis G (2013) Returns to education and rural migrant income differentials in China: IV quantile treatment effects. China Econ Rev 26:39–55

MGI (2009) Preparing for China's urban billion. McKinsey Global Institute, March. http://www.mckinsey.com/insights/urbanization/preparing_for_urban_billion_in_china

National Bureau of Statistics Of China (2008) China statistical yearbook. China Statistics Press, Beijing

Oaxaca R (1973) Male-female income differentials in urban labor markets. Int Econ, Rev 14(3):693–709

Song S, Zhang H (2003) Rural-urban migration and urbanization in China: evidence from time-series and cross-section analyses. China Econ Rev Elsevier 14(4):386–400

United Nations (2009) International migration report 2006: a global assessment. Department of Economic and Social Affairs, New York

Wang M (2005) Employment opportunities and income gaps in the urban labor market: a study of the employment and incomes of migrant laborers. Soc Sci China 5:36–46. (in Chinese)

Zhao Z (2005) Migration, labor market flexibility, and income determination in China: a review. Dev Econ 43:285–315