Segregated and not equal? Occupation, earnings gap between urban residents and rural migrants in Vietnam

Received 1 June 2017
Revised 25 January 2018
27 April 2018
Accepted 8 May 2018

Amy Y.C. Liu
International and Development Economics, Crawford School of Public Policy, Australian National University, Canberra, Australia

Abstract

Purpose – The purpose of this paper is to examine the earnings differentials between the locals and the rural–urban migrants in urban labor market in Vietnam.

Design/methodology/approach – Using the new Vietnam Rural-Urban Migration Survey 2013 (VRUMS2013) that is specifically designed for rural–urban migration, the author applies Appleton *et al.*'s (1999) procedure correcting for potential selectivity to decompose the offered earnings gap between the locals and the rural–urban migrants into within- and between-occupation pay differential. Bootstrapping is used to derive the standard errors for the decomposition results. The author further applies the propensity score matching (PSM) method to check whether the results are robust by restricting the sample to the "common support."

Findings – Within-job difference, particularly, the favorable treatment toward urban workers contributes significantly to the overall and total unexplained earnings gap. Further, between-job pay differential attributed to the over-representation of urban workers in high-paying job also helps to widen the gap. These results are robust restricting to the "common support" sample using PSM.

Research limitations/implications – Due to the sample size, occupations are only classified into three broad categories. Finer classification will allow a better comparison between the contributions of between and within-occupation to earning inequality. The data are only limited to a few cities and do not include other urban centers that also receive rural–urban migrants.

Practical implications – Policies to promote equal pay and alleviate within-job "discrimination," especially the preferential treatment favoring the locals (rather than to provide equal access to different jobs) are crucial for migrants' labor outcome. Moreover, this study can, to some extent, be seen as a timely contribution for the debate on household registration reform in general and in Vietnam specifically. Given China's announcement to grant permanent household registration (*hukou*) to unregistrated migrants in late 2015, investigating whether there is a two-tier labor market in the cities in Vietnam is particularly important for the ongoing debate regarding future of household registration system (*ho khau*).

Originality/value – This is the first study in Vietnam on rural–urban migration and occupation segregation – an area that has been relatively less well studied in developing/transitional countries. Vietnam is also one of the few developing countries who have household registration system in place. This has made it an interesting case. The author uses a new survey data to apply the Appleton *et al.* (1999) decomposition on the offered wage gap rather than observed wage gap. Standard errors of the decomposition results are bootstrapped and a robust check using propensity score method is conducted.

Keywords Discrimination, Vietnam, Decomposition, Propensity score matching, Occupation segregation, Rural–urban migration

Paper type Research paper

1. Introduction

Despite the apparent importance of the topic, labor market segregation between urban residents and rural migrants has received little attention outside of China. In the past decades, Vietnam has experienced a remarkable increase in its internal migrant population in the cities. The census data show that about 4.5m people (6.5 percent of the population) aged over five migrated between 1994 and 1999. Of which, 1.6m were rural–urban migrants (Dang *et al.*, 2003). It rose to 7.7 percent in 2009 (Le *et al.*, 2011). Poverty, land shortages and a lack of opportunities are the push factors that drive rural residents to leave their home village.

International Journal of Manpower
Vol. 40 No. 1, 2019
pp. 36-55

0143-7720
DOI 10.1108/IJM-06-2017-0114

The new economic opportunities (jobs and income) unleashed by the market reform in 1986, *Doi Moi* (Renovation), are one of the pulling factors that attract rural residents to migrate to urban areas (Cu, 2005).

Most research on urban–rural migration in Vietnam has been focused on understanding the patterns and the decision making of migration (e.g. Dang *et al.*, 2003; Coxhead *et al.*, 2016), as well as remittances and its impact on the welfare of rural households (e.g. Niimi *et al.*, 2008; Binci and Giannelli, 2012). Dang *et al.* (2003) and Le *et al.* (2011) examine migrants' characteristics and their relative position in the urban labor market but are mostly descriptive. Nguyen and Minh (2016) is an exception. The authors use the urban poverty survey (UPS) 2009 and Oaxaca–Blinder decomposition to estimate the average monthly wage difference between non-migrants and migrants. Noticeably, migrant workers are consistently found to work more hours and receive less pay than urban residents.

Why do rural migrant workers tend to receive a lower wage rate? Evidence in China (Meng and Zhang, 2001; Zhang, 2009; Frijters *et al.*, 2010) suggests that it can be attributed to two broad explanations. First, rural migrants may have productivity-related characteristics that adversely affect their pay. For instance, rural migrant workers tend to be less educated than their urban counterparts and the education that they receive in the rural area may be of poor quality. They tend to be younger and have less (city) experience. Second, they may be discriminated against in the job market. Frijters *et al.* (2010) use the 2008 wave of the rural–urban migration in China and Indonesia (RUMiCI) and find that 46 percent of the wage gap is unexplained. These authors argue that migrants are paid less simply because they are without urban registration status (*hukou*).

Like China, Vietnam has a household registration (*ho khau*) system. It ties individuals to their place of residence. Rural migrants have to meet strict requirements such as continuous employment and residence (as high as five years until 2005) in order to apply for *ho khau* at the destination city. Despite a gradual relaxation of the restrictions since 2007 (until recently)[1], rural migrants in the cities without urban registration still face significant challenges in finding jobs in the formal sector. Le *et al.* (2011) reports that employers were instructed by provincial officials to give priority to local residents in order to support the local economy. Large cities such as Hanoi, Hochiminh city and Da Nang require permanent *ho khau* for jobs in the public sector unless applicants are skilled workers or have graduated from top universities (World Bank, 2016). These preferential and restrictive policies of local governments may create discrimination favoring urban residents and against migrants without *ho khau*. In addition, the *ho khau* system can become intertwined with the traditional stereotypes of migrants. For instance, urban residents tend to see themselves as "more superior"[2] and attribute the rise in crime and other social problems to the influx of rural migrants (Taylor, 2004). Employers can also favor urban workers or penalize rural migrant workers without urban *ho khau* according to their taste if they perceive urban people are more superior. Further, migrants are entitled to less social welfare and public services[3], so their ability to find (better) jobs (with stable income and good working condition) could be more restricted. Differing opportunities in gaining employment in certain jobs due to *ho khau* status unrelated to productivity could result in occupational discrimination. Such discrimination can influence occupational attainment outcomes by both favouring urban residents and against migrants. Consequently, more urban residents (migrants) occupy higher (lower) paid jobs.

Despite the important link between-occupational attainment and earnings, few studies on developing countries (even in China and more so in Vietnam) explicitly connect discrimination and occupation segregation with wage determination. Meng and Zhang (2001) are among the few exceptions. They use Brown *et al.* (1980) decomposition method to examine earnings differentials between the locals and rural migrants. They find that in 1995 migrants only earned half of urban workers' hourly earnings in Shanghai with about

82 percent of the pay gap attributable to within-job discrimination. Of which, differential treatment of the two groups of workers is more important than characteristic difference.

This paper follows this line of research to examine the wage gap between urban residents and rural migrants in the cities with a special focus on occupation segregation. Its contributions are as follows. First, it goes beyond a general investigation such as Nguyen and Minh (2016). Instead, it fills the void in the literature on migration and occupation segregation that has been relatively less well studied in developing/transitional countries. As far as I am aware, this is the first of its kind in Vietnam. An empirical enquiry into the relative importance of within- and between-occupational differentials has significant policy implications: Should policies promoting equal pay within occupations, or policies aiming at providing equal access to various occupations be immediate priorities?

Second, in view of China's announcement to grant *hukou* to unregistrated migrants in late 2015, this study represents a timely contribution for the ongoing debate on household registration reform in general and in Vietnam specifically. Should *ho khau* be abolished or maintained?

Third, empirically it extends the existing literature in several ways: first, this paper uses the full decomposition method of Appleton *et al.* (1999). Similar to Brown *et al.* (1980), it can further decompose the earnings differential into within- and between-occupation wage effects. Yet, it overcomes the index number problem. Second, this paper decomposes also the offered wage difference. As migrants may self-select into low-pay jobs; or locals into professional jobs, the observed wage gap may be subject to biases in the presence of self-selection. Third, given the characteristic differences between urban residents and rural migrants may not be fully captured by the parametric model such as OLS, I use propensity score matching (PSM) to investigate the sensitivity of the decomposition results by restricting the sample to the region of "common support." Fourth, it uses new survey specifically designed for urban–rural migration, "Vietnam rural-urban Migration survey 2013" (VRUM2013), with the Vietnam Living Household Standards Survey 2012 (VHLSS, 2012) to investigate occupation segregation in the cities of destination, an important aspect of rural–urban migration.

Section 2 outlines the Appleton *et al.* (1999) decomposition method and compares it with Oaxaca (1973) and Neumark (1988). Section 3 presents the empirical results and the robustness check on the decomposition results using PSM. Concluding remarks and policy implications are presented in Section 4.

2. Decomposing the wage gap

2.1 Traditional decomposition: Oaxaca and Neumark

As Oaxaca and Neumark decompositions are commonly used in the literature, a detailed exposition is not included here. Briefly, separate standard Mincerian log wage equations are estimated for urban and rural migrants. The Oaxaca approach decomposes the wage gap into two components:

$$\ln \overline{w}_u - \ln \overline{w}_r = \beta_u(\overline{x}_u - \overline{x}_r) + (\beta_u - \beta_r)\overline{x}_r \text{(urban workers as the reference group)},$$

$$\ln \overline{w}_u - \ln \overline{w}_r = \beta_r(\overline{x}_u - \overline{x}_r) + (\beta_u - \beta_r)\overline{x}_u \text{(rural migrants as the reference group)},$$

where $\overline{w}_u$ and $\overline{w}_r$ are the means of wages of urban and migrant workers, respectively; $\overline{x}_u$ and $\overline{x}_r$ are vectors containing the respective means of the independent variables for urban residents and migrants; and β_u and β_r are the estimated coefficients. The first component captures the wage differential due to different characteristics of the two groups (explained component). The second component is the estimated effects of differences in returns to those characteristics (unexplained) which is potentially due to discrimination. Note that the

unexplained part might also reflect problems of omission of variables or unobserved factor (Jann, 2008). Hence, it may give an upper bound estimate for discrimination.

Oaxaca decomposition is often subject to the index number problem. Neumark (1988) overcome it by using the weighted average of the wage structures of urban residents and rural migrants:

$$\ln \overline{w}_u - \ln \overline{w}_r = \beta(\overline{x}_u - \overline{x}_r) + \left[(\beta_u - \beta)\overline{x}_u + (\beta - \beta_r)\overline{x}_r\right],$$

where β is the non-discriminatory wage structure. The first term is the wage gap attributable to differences in characteristics. The last two terms capture the difference between the actual and pooled returns for urban residents and rural migrants, respectively. The sum of the second and the third part that is due to favorable treatment toward the locals in which they are overpaid ("urban advantages" thereafter). The third term is the earnings gap that attributable to unfavorable treatment against rural migrants in which they are underpaid ("rural disadvantages" thereafter). Yet, the Neumark decomposition, like Oaxaca method, did not account for differences in occupational structures between the two groups.

2.2 Full decomposition: Appleton et al. (1999)

In the literature, some studies incorporate occupational dummies in wage regressions as additional explanatory variables using Oaxaca-type decomposition. This approach enables one to: control for different occupational distribution, to make some interpretation of the impact of the different occupational distributions on the urban and migrant workers by comparing the decomposition results without and with occupational dummies[4]. However, this method, as noted by Miller (1987a) and Brown *et al.* (1980), has the problem of assuming that occupations are exogenous. If occupational determination is affected by labor market discrimination, this approach would not be appropriate as it implicitly treating all differences between the occupational distributions of the two groups as explained. Even if the endogeneity problem can be dealt with simply adding controls for occupations still does not distinguish sufficiently between wage discrimination and job discrimination (Liu *et al.*, 2004; Brown *et al.*, 1980).

Instead of inferring changes between the earnings equations without and with the occupational controls, this paper follows the work of Appleton *et al.* (1999) to analyze occupational segregation directly using a behavioral model of occupational status and therefore, allows both within- and between-occupational earnings differentials to be further broken into explained and unexplained components. Also, it overcomes the index number problem.

For simplicity, let $\overline{w}_u$ and $\overline{w}_r$ refer to the mean natural log of the urban or migrant workers' hourly earnings and $\overline{p}_{uj}$ and $\overline{p}_{rj}$ be the sample proportions of the two groups in occupation j, respectively.

The mean hourly earnings of the group i can be written as the sum of earnings weighted by the proportion of workers in occupation j:

$$\overline{w}_i = \sum_{j=1}^{J} \overline{w}_{ij}\overline{p}_{ij} \quad i = u, r; J = 3.$$

Analogous to that of Neumark, Appleton *et al.* (1999) propose a non-discriminatory occupational structure. If $\overline{p}_j^*$ is the proportion of the employees in occupation j under this structure, then the difference in mean log hourly earnings can be decomposed:

$$\overline{w}_u - \overline{w}_r = \sum_{j=1}^{J} \overline{p}_j^* \left(\overline{w}_{uj}^* - \overline{w}_{rj}\right) + \sum_{j=1}^{J} \overline{w}_{uj}\left(\overline{p}_{uj} - \overline{p}_j^*\right) + \sum_{j=1}^{J} \overline{w}_{rj}\left(\overline{p}_j^* - \overline{p}_{rj}\right).$$

The first term can be decomposed using Neumark decomposition. The other two terms can be decomposed further to differences in observable characteristics and differences in returns

to these characteristics. In order to do so, a pooled and separate multinominal logit model for urban workers and rural migrants are estimated to derive the average probability in the different occupations, j. These average probabilities are denoted by $\overline{p}_{ij}^{*}$, where $i = u, r$.

The full decomposition can be written in the following form:

$$\overline{w}_u - \overline{w}_r = \sum_{j=1}^{J} \overline{p}_j^{*}(\overline{x}_{uj} - \overline{x}_{rj})\beta_j + \sum_{j=1}^{J} \overline{p}_j^{*}\overline{x}_{uj}(\beta_{uj} - \beta_j)$$

$$+ \sum_{j=1}^{J} \overline{p}_j^{*}\overline{x}_{rj}(\beta_j - \beta_{rj}) + \sum_{j=1}^{J} \overline{w}_{uj}\left(\overline{p}_{uj}^{*} - \overline{p}_j^{*}\right)$$

$$+ \sum_{j=1}^{J} \overline{w}_{rj}\left(\overline{p}_j^{*} - \overline{p}_{rj}^{*}\right) + \sum_{j=1}^{J} \overline{w}_{uj}\left(\overline{p}_{uj} - \overline{p}_{uj}^{*}\right) + \sum_{j=1}^{J} \overline{w}_{fj}\left(\overline{p}_{rj}^{*} - \overline{p}_{rj}\right).$$

The first three terms are equivalent to Neumark's within-occupation wage differentials. The sum of the last three terms captures the between-occupation differentials. The fourth and fifth terms measure the different earnings due to differences in characteristics determining occupational distribution of urban and migrant employees. The last two terms measure the contribution of differences in the proportion of urban and migrant workers in different occupations, providing an overall measure of occupational segregation, not accounted for by the characteristic differences.

3. Empirical results

3.1 Data

I use the VHLSS2012 to compile the sample for urban residents. Ideally, it should be restricted to individuals who are living in the cities where they were born and registered for permanent residency. Unfortunately, VHLSS2012 does not collect information on respondents' birthplace; or explicitly identify the rural–urban migrants living in the cities. Nonetheless, the VHLSS2012 contains information about respondents' household registration status. Urban residents here refer to those who registered their *ho khau* in the same commune in the cities where they were living at the time of the interview. This definition may include urban-to-urban migrants as well as long-term rural–urban migrants who had already registered their household status in the city where they have been residing. If migrants are paid less relative to urban residents and the wage gap persists over time[5], by including these long-term migrants may bias the locals-migrants earnings gap downward.

The migrant sample is from the VRUM2013. Based on the rural household sample of the VHLSS2012, VRUMS2013 surveys those individuals who have migrated from rural to urban areas of Hanoi, Hochiminh city, Binh Duong & Dong Nai. As the focus of the paper is on the earnings differential and occupation, I opt to keep the long-term migrants in the cities in the migrant sample to ensure sufficient samples for different occupation cells. By including them, I may over-estimate the earnings of the migrants[6].

As long-term migrants are included in the both samples, it is the lower bound of the wage gap that is estimated here. Additionally, rural–urban migrants and urban residents are: wage earners in their main job in the previous 12 months; aged between 16 and 65 years, inclusive; and supplied earnings data. Further, I restrict the urban sample to those who were in Hanoi, Hochiminh City, and Binh Duong & Dong Nai. Hence, I cannot draw any inferences about wage differentials between urban residents and migrants in Vietnam in general. Further, if the VRUM2013 is less successful in capturing migrants at the bottom of the earnings distribution, the earnings gap would possibly be under-estimated.

In total, there are 1,712 wage earners (724 urban residents and 988 migrants). I also deflate the earnings data of the VRUMS2013 by the urban CPI to make it comparable to the VHLSS2012[7]. I categorize respondents' job into: professionals and officer workers (professionals thereafter); production workers: semi-skilled and skilled workers, manual and low-skilled laborers, and machine operator; others: workers in services, sales, or military. "Others" serves as the base group. The first category refers to as white collar jobs, whereas the last two categories together as blue collar jobs.

As shown in Table I, on average, migrants have lower log hourly earnings[8], shorter potential experience, fewer years of schooling than urban workers. Most of them work in private sector and fewer of them are married. More migrants are females.

Table II reveals that the earnings disparity between the locals and migrant professionals is noticeably the largest. Briefly, urban professionals are paid an hourly wage rate of 68 percent more than their migrant counterparts[9] (3.72 vs 3.2 in log differentials). Migrants holding a job other than professional jobs have longer years of schooling than the locals; and more migrant professionals work for (domestic) private firms and located in Hochiminh City.

In terms of the occupational distribution (Table III), about 46 percent of the urban wage earners vs 21 percent of migrants are professionals and office workers. Most migrants (71 percent) compare to 40 percent of the urban workers work as production workers.

	Urban		Migrants		Difference	
Variables	Mean	SD	Mean	SD	Mean	SD
Log hourly earnings	3.35	0.66	2.88	0.54	0.47***	0.03
Potential experience	19.60	11.85	12.86	9.52	6.75***	0.52
Potential exp^2	524.56	538.74	255.81	372.28	269.07***	22.01
Gender	0.53	0.50	0.58	0.49	−0.05*	0.02
Married	0.69	0.46	0.60	0.49	0.09***	0.02
Years of schooling	12.22	4.14	10.80	3.98	1.42***	0.19
State	0.38	0.49	0.16	0.36	0.22***	0.02
Private	0.50	0.50	0.62	0.49	−0.12***	0.02
Foreign-invested firms	0.12	0.01	0.22	0.01	−0.10***	0.01
Hochiminh City	0.47	0.50	0.55	0.50	−0.08***	0.02
Hanoi	0.33	0.47	0.28	0.45	0.05***	0.02
Other	0.20	0.01	0.17	0.01	0.03	0.02
No. of observations	724		988			

Notes: I used a *t*-test to examine the difference in mean of each variables for the two groups and present the results in the last column. Significance in the *t*-statistic indicates that I can reject the null hypothesis that the means of the two groups are not different. *,**,***Indicate 10, 5 and 1 percent significant levels, respectively

Table I.
Summary statistics for the main variables used in earning equations, by migration status

	Professional				Productive workers				Others			
	Urban		Migrants		Urban		Migrants		Urban		Migrants	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Log hourly wage	3.72	0.63	3.20	0.58	3.05	0.46	2.80	0.49	2.99	0.60	2.73	0.53
Potential experience	17.01	10.94	8.33	6.39	22.25	12.01	14.03	9.72	20.69	12.56	14.64	11.00
Potential exp^2	408.62	445.62	109.94	176.76	638.73	591.19	291.03	395.82	584.08	588.29	333.69	440.75
Gender	0.52	0.50	0.50	0.50	0.54	0.50	0.59	0.49	0.56	0.50	0.70	0.46
Married	0.73	0.45	0.53	0.50	0.69	0.46	0.62	0.49	0.62	0.49	0.65	0.48
Years of schooling	15.38	2.15	14.44	2.44	9.24	3.47	9.62	3.72	10.20	3.29	11.45	3.53
State	0.56	0.50	0.35	0.48	0.19	0.39	0.09	0.29	0.31	0.47	0.20	0.40
Private	0.34	0.48	0.45	0.50	0.64	0.48	0.66	0.47	0.65	0.48	0.71	0.46
Hochiminh city	0.39	0.49	0.60	0.49	0.51	0.50	0.54	0.50	0.64	0.48	0.49	0.50
Hanoi	0.50	0.50	0.31	0.46	0.17	0.38	0.25	0.43	0.24	0.43	0.44	0.50

Table II.
Summary statistics for the main variables used in earning equations in different sectors

Table III. Occupation distribution, hours worked and earning of rural migrants and urban residents

	Occupation distribution				Hours worked a month by occupation			
	Urban residents		Rural migrants		Urban residents		Rural migrants	
	Freq.	%	Freq.	%	Mean	SD	Mean	SD
Professionals	335	46.27	211	21.36	186.86	32.09	207.64	41.60
Productive workers	287	39.64	697	70.55	200.05	43.42	237.59	64.00
Others	102	14.09	80	8.10	216.37	64.68	262.16	76.51
All	*724*	*100*	*988*	*100.00*	*196.24*	*43.70*	*233.19*	*62.82*
White collar	335	46.27	211	25.84	186.86	32.09	207.64	41.60
Blue collar	389	53.73	777	74.16	204.33	50.31	240.12	65.77
	Monthly earnings by occupation				Hourly earnings by occupation			
	Urban residents		Rural migrants		Urban residents		Rural migrants	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Professionals	9,209.55	7,216.50	6,714.10	12,049.63	50.32	38.89	31.74	55.19
Productive workers	4,517.90	2,192.38	4,288.53	1,946.94	23.52	12.72	18.80	15.37
Others	4,494.18	2,667.24	4,402.75	1,707.23	25.20	26.22	17.93	11.43
All	*6,685.42*	*5,696.29*	*4,815.79*	*5,897.38*	*36.16*	*32.12*	*21.49*	*29.22*
White collar	9,209.55	7,216.50	6,714.10	1,2049.63	50.32	38.89	31.74	55.19
Blue collar	4,511.68	2,322.70	4,300.29	1,922.95	23.97	17.28	18.71	15.01

Many migrants not only have low-end jobs, but also, in each occupation, they also work longer hours and receive less pay. On average, they work over 37 hours more in a month than their urban counterparts. White collar workers have the largest monthly wage gap of 2,495 thousand dongs (about $117)[10]. Understandably, migrants also receive lower hourly wages.

3.2 Mincerian earnings equations

I use the OLS to estimate the Mincerian log hourly earnings equation for the pooled sample[11] and for urban and migrant workers separately. I include in these earnings equations some typical variables, such as year of schooling, potential experience[12] and its square term, as well as dummies on gender, marital status[13], destination city[14] (Hanoi, Hochiminh city and Binh Duong & Dong Nai)[15] and migrant status (for the pooled sample)[16]. Following Zhang and Wu (2013), I also include three ownership dummies (state sector, (domestic) private sector or foreign-invested firms[17]) as the explanatory variables.

Table AI confirms the difference between the locals and migrants in terms of education, whether working for the state sector and the cities of destination. They are strongly significant at the one percent level as shown in the column "Difference." Lower returns to migrants' education may reflect the lower quality of education in rural areas. Urban residents receive lower wages if they work for the state sector relative to those with a job in foreign-invested firms.

3.3 Conventional decomposition of wage gap

The Oaxaca decomposition result confirms that migrants earn less than the locals (Table IV). The earnings difference is significant at the one percent level. Using urban wage structure as the reference, the difference in characteristics only accounts for 35 precent of the overall gap. There is not much difference in the result if the rural wage structure is used instead[18]. These results are different from Nguyen and Minh (2016). They find that the observed wage gap between the two groups is no longer significant once the observed characteristics of workers are controlled for. The differences of the results might not come as a surprise as the two studies differ in many aspects. For example, Nguyen and Minh (2016) does not control for factors such as occupational attainment and ownership distribution. Also, they use the UPS data collected in 2009, about five years earlier than the VRUMS2013. Vietnam was characterized by the immediate aftermath of the global financial crisis and expansionary

	Observed earnings gap	SE	%
ln $\overline{W}_u$ 3.352			
ln $\overline{W}_r$ 2.876	0.476***	0.0299	
Oaxaca			
Urban wage structure			
Characteristics	0.165***	0.0244	34.73
Returns	0.310***	0.0310	65.27
Rural wage structure			
Characteristics	0.158***	0.0268	33.29
Returns	0.317***	0.0301	66.71
Neumark			
Weighted wage structure			
Skill difference	0.232***	0.0232	48.83
Urban advantage	0.103***	0.0097	21.64
Rural disadvantage	0.140***	0.0130	29.53

Notes: The Neumark decomposition results is generated by the nldecomp command and bootstrapped 100 times in Stata. *,**,***Significant at 10, 5, and 1 percent levels, respectively

Table IV. Conventional decomposition of urban-migrant wage gap

fiscal and monetary policies. Provided that urban and migrant workers have responded to the macroeconomic conditions differently between the two periods, the relative wage structure between the urban and migrant workers might have changed[19].

The results of Neumark (1988) are also in line with that of Oaxaca, with the unexplained component accounts for 51 percent of the earnings gap. The urban advantage is less important than the rural disadvantage (22 and 30 percent respectively).

Decomposition by occupation. Table V reports conventional decomposition results separately for three occupation types. Positive and significant wage gaps favoring urban residents are apparent across different occupations with professionals having the larger earnings differences (0.52 in log). Irrespective of which decomposition method is used, differences in returns are more important than differences in characteristics in explaining the observed earnings gap across the board. Moreover, Neumark method reveals that the key driver behind the differences in return between the two groups of workers vary across occupations.

Analysis based on separate occupational-specific wage equations does not take into account the difference in occupational distribution between the locals and migrant workers. Hence, I use the Appleton *et al.* (1999) decomposition method to investigate: How much would the relative economic position of migrants have been worsened (improved) if more migrants were to take up blue collar jobs (professional jobs)?

3.4 Decomposition of Appleton et al. and selectivity

Strictly speaking, there are two sources of sample selection when the OLS focuses on wage earners only: wage earners are only observed when they work; the selective decision to work in the wage sector. Possible endogeneity of occupational choice further complicates the selection problem. For simplicity, as the paper focuses on occupation, Lee's (1983) two-stage procedure is used to correct for endogenous selection into a particular occupation, conditioned on one's decision to work in the wage sector.

In the first stage, the predicted probability P_{ij} generated by the multinominal logit occupation equation is used to compute the correction terms, λ_{ij}, for occupation j for migrants and urban dwellers separately. The correction term is included in the respective equation as an additional regressor in the second stage.

	Professional			Productive workers			Other		
	Actual earnings gap	SE	%	Actual earnings gap	SE	%	Actual earnings gap	SE	%
Professional									
ln $\overline{W}_u$ 3.722									
ln $\overline{W}_r$ 3.199	0.523***	0.0528							
Laborer									
ln $\overline{W}_u$ 3.050									
ln $\overline{W}_r$ 2.795				0.254***	0.0329				
Others									
ln $\overline{W}_u$ 2.989									
ln $\overline{W}_r$ 2.735							0.254***	0.0843	
Oaxaca									
Urban wage structure									
Characteristics	0.168***	0.0802	32.12	0.045**	0.0250	17.86	−0.129**	0.700	−50.77
Returns	0.355***	0.0888	67.88	0.208***	0.0348	82.14	0.384***	0.0886	150.77
Rural wage structure									
Characteristics	0.136***	0.0441	25.95	0.022	0.0278	8.83	−0.039	0.0818	−15.29
Returns	0.387***	0.0573	74.05	0.231***	0.0373	91.17	0.293***	0.0953	115.29
Neumark									
Weighted wage structure									
Skill difference	0.247***	0.044	47.18	0.075***	0.0192	29.67	−0.013	0.0611	−5.10
Urban advantage	0.169***	0.026	32.41	0.052***	0.0087	20.51	0.150***	0.0295	58.90
Rural disadvantage	0.107***	0.017	20.41	0.127***	0.0209	49.82	0.118***	0.0270	46.20

Note: *,**,***Significant at 10, 5, and 1 percent levels, respectively

Table V. Conventional decomposition of occupation-specific urban-migrant wage

The estimated occupation-specific wage equations corrected for selection bias (Table AIII) shows that males receive even higher wages relative to their female counterparts across all occupations. It is noteworthy that the estimated rate of return to schooling for the professionals is no longer significant (vs Table AII where selectivity is not accounted for)[20]. Note also that λ is positive and significant for urban professionals. Significant self-selection is also found for migrant productive workers and urban residents with "other jobs" but it is negative.

In the presence of selectivity bias, the observed wage differential will deviate from the unobserved wage differential in terms of wage offers (Miller, 1987b). Following the discussion on selectivity correction, the average observed wage can be written as $\ln \overline{\text{wage}_j} = \overline{x}_j\hat{\beta}_j + \overline{\theta}_j\hat{\lambda}_j$ for urban and migrant workers where $\overline{\theta}_j$ is the mean Lee selective correction term and $\widehat{\lambda}_j$ is the parameter estimates for the earnings equation. For simplicity, I suppress the subscript j and use the subscript u and r to denote whether an individual is a local or a migrant. The earnings difference can be written as:

$$\overline{w_u} - \overline{w_r} = \overline{x_u}\widehat{\beta_u} + \overline{\theta_u}\widehat{\lambda_u} - \left(\overline{x_r}\widehat{\beta_r} + \overline{\theta_r}\widehat{\lambda_r}\right)$$

$$= \left(\overline{x_u}\widehat{\beta_u} - \overline{x_r}\widehat{\beta_r}\right) + \left(\overline{\theta_u}\widehat{\lambda_u} - \overline{\theta_r}\widehat{\lambda_r}\right).$$

The average observed wage gap can be written as the sum of the average offered wage gap and the difference in average selectivity bias between groups. Therefore, observed wage difference may overstate or understate the difference in average wage offers. Suppose discrimination reduces migrants' wage offer, if the locals and migrants have the same reservation wage, the lower (average) wage offers for migrants

would be offset by their larger selectivity bias, giving rise to a narrow observed wage gap. Re-arranging:

$$\overline{w_u} - \overline{w_r} - \left(\overline{\theta_u}\widehat{\lambda_u} - \overline{\theta_r}\widehat{\lambda_r}\right) = \overline{x_u}\widehat{\beta_u} - \overline{x_r}\widehat{\beta_r},$$

the offered average wage gap on the right hand side is expressed as the observed average wage gap minus the selectivity terms for different groups. It is the offered wage gap that I apply Appleton *et al.*'s procedure. The result is reported in Table VI.

In the empirical work, a multinomial logit model with three occupation categories is first estimated. The explanatory variables included are gender dummy, marital status, year of schooling, age and its square term, ownership dummies (state sector, private sector and foreign-invested firms) and three destination city dummies (Hanoi, Hochiminh city, and Binh Duong & Dong Nai). They are expected to influence employers' willingness to hire and an individual's occupational choice. Identification here is achieved by including additional variables: number of children (less than 16 years old) and its square term[21]. Theoretically the number of children may affect individual's participation in a particular occupation as it may capture the pressure for having a more stable job or more flexible work arrangement that often associates with family responsibilities (Brown *et al.*, 1980). Yet, it does not directly affect wages. The detailed results of the multinominal logits model used to derive the predict probabilities are not reported here[22].

The offered earnings gap is 0.58 in log and is significant at 5 percent level (Table VI). The sum of (2), (3), (6) and (7) provides the contribution of unexplained part. It accounts for 81 percent of the overall gap and is also significant at one percent level[23]. This result is in line with the traditional decomposition results. Yet, Appleton *et al.* decomposition offers a richer understanding by unpacking the unexplained portion of the pay gap further into within and between-occupational earnings differences.

Over 59 percent of the overall earnings gap is due to within-occupational earnings differences between the two groups, (1) to (3). The positive sum implies that migrants indeed earned less than urban workers within the same occupation. Further, deviation in returns, (2) + (3), explains 58 percent (97 percent) of the overall earnings differences (of the whole

	Offered earnings gap	%
Offered wage gap		
$\left(\ln \overline{W_u} - \overline{\lambda_u}\tau_u\right) - \left(\ln \overline{W_r} - \overline{\lambda_r}\tau_r\right) = 3.359 - 2.784$	0.575**	100.00
Earnings differences due to within-occupational differences		
Characteristics (1)	0.010	1.79
Deviation in returns to urban workers (2)	0.239***	41.59
Deviation in returns to rural migrants (3)	0.092	16.04
Earnings differences due to between sector (sectoral location) differences		
Characteristics (4) + (5)	0.101	17.56
Deviation urban workers' predicted and actual occupation composition not accounted for by characteristic differences (6)	0.077***	13.36
Deviation rural migrants' predicted and actual occupation		
Composition not accounted for by characteristic differences (7)	0.056	9.66
Explained (1) + (4) + (5)	0.111	19.35
Unexplained (2) + (3) + (6) + (7)	0.464***	80.65

Note: *,**,***Significant at 10, 5, and 1 percent levels, respectively

Table VI. Full decomposition of the overall urban-migrant earnings gap, accounting for selectivity

within-occupational earnings gap). In particular, deviation in urban workers' returns or within-job favorable treatment toward urban workers, which is statistically significant at one percent level, contribute 42 percent to the overall pay gap. This is much higher than the 16 percent contribution of deviation in migrants' returns or within-job unfavorable treatment against migrant workers.

The last three terms indicate that the share of the earnings gap which may be attributed to between-occupational differences. The positive sum implies that earnings differences between occupations are more favorable to urban workers than migrants. Specifically, it suggests that the overall earnings gaps would have been 41 percent narrower if both groups were equally distributed across different occupations; and that urban workers earned more because on average more urban workers have high-paying occupations. Urban workers' concentration in high-paying jobs accounts for 13 percent of the overall job and it is statistically significant. While the inter-occupational earnings differential is attributable to workers' characteristic difference accounts for 18 percent of the between-job differential, it is not found to be statistically significant. Similarly, the over-representation of migrant workers in low paying occupations is also not significant. Nonetheless it helps to keep the overall pay gap 10 percent wider than it otherwise would be.

The results so far indicate that the occupational segregation effect reinforces the overall discrimination within-job pay; and the within-job wage gap drives most of the overall gap. These results are generally in line with Brown *et al.* (1980)'s decomposition results of Meng and Zhang (2001) on China[24]. The result of Appleton *et al.* (1999) reveals information that is otherwise masked by traditional decompositions: favorable treatment toward urban workers drives the within-job earnings gap; and favorable allocation of occupations toward urban workers significantly contribute to the between-job earnings differences.

The wage gap not correcting for selectivity is 0.48 in log (not shown here), representing a gap of 21 percent narrower than the offered pay gap. The widening offered gap indicates that if wage equation is not adjusted for selectivity the observed wage gap would be under-estimated. The gap is still mostly attributed to within-occupational difference without correcting for selectivity. With regard to between-job differences, urban workers being disproportionately located in high-paying occupations remain the most important factor.

3.5 Robust check – PSM and common support

Recall urban and migrant workers are quite different in some of their characteristics. Using observations where the characteristics of the two groups vastly differ from their "support" to estimate the earnings equations could yield misleading results. OLS estimation may render the estimated gap not reliable if the model is not specified correctly, or the assumption of linearity is violated, not to mention if the explanatory variables have measurement error (Briggs, 2004). Further, if the common support assumption is violated, the estimate of the unexplained part of the earnings gap could be biased upward (Ñopo, 2008).

Following Zhang and Meng (2007), I apply PSM approach and restrict to observations with a common support to alleviate the concern that the overall estimated earnings difference between the two groups is due to a lack of comparability or selection bias on observables (i.e. the "out-of-support" comparison).

Assuming urban residents as the control group and migrants as the treatment group, the basic idea of PSP is to match urban residents and migrants based on a propensity score which is simply a conditional probability. I first estimate a logit regression on a migrant dummy[25]. I include a vector of explanatory variables[26], e.g. experience and its square term, gender, married, years of schooling, three ownership dummies (state, private and foreign-invested firms), and three city dummies (Hanoi, Hochiminh city, and Binh Duong & Dong Nai). These variables are included based on the significance of the mean difference of various characteristics between the two groups (Table I). Table AIV presents the logit results[27].

After fitting the propensity score generated by the logit regression, an urban resident is matched with a migrant. As a result, the two samples generated from the matching procedure are statistically comparable across the set of controlled characteristics. I then experiment with several commonly used matching methods[28]. Below I report the results of the five-nearest neighborhood matching.

The two-sample *t*-test shows no significant differences in all the covariates between the locals and migrants after matching, indicating that the balancing is effective[29]. The standardised bias calculates the percentage reduction in the bias before and after matching. Bertrand (2009) regards a standardised bias in excess to 20 percent as being too large. Rosenbaum and Rubin (1985) suggest a bias under 10 percent. Caliendo and Kopeinig (2008) recommend a maximum about from 3 to 5 percent. Table AV shows that the covariates are well balanced with a bias after matching of no more than 5 percent. That is, the covariate means of the migrants have close resemblance to that of the locals.

Regarding the overall measures of covariate balance the null hypothesis of joint significance of the covariates of the likelihood ratio cannot be rejected before matching (LR χ^2: 423.70; *p*-value: 0.000) but it is rejected after matching (LR χ^2 is 7.57; *p*-value: 0.578). Moreover, the Pseudo R^2 reduces from 0.182 before matching to 0.003 after matching. The low Pseudo R^2, the insignificant LR after matching, the low mean and median bias after matching all suggest that the specification of the propensity score estimator is effective.

Additionally, the propensity scores across the treatment and control groups have a sizeable overlap (common support condition). Only eight observations (the treated group) are off the common support. Note that the PSM method only addresses selection on the observables. Hence, I apply the full decomposition with selection correction procedure on the restricted sample. The results (not shown here) are very similar to those with the full sample without restriction (Table VI).

4. Conclusions

Correcting for the occupational self-selection, migrants earn 17 percent less than the locals in terms of their (offered) hourly earnings despite several reforms of the *ho khau* system. It is the unexplained rather than explained part, and within-job earnings differential (59 percent) rather than between-job difference that account for most of the pay gap.

The positive and sizable within-job earnings differential on the deviations in urban workers' returns imply that urban workers' advantage is more important than the unfavorable treatment against migrants. Between-job pay differential accounts for about 41 percent of the overall gap. The over-representation of urban workers in high-paying jobs not accounted for characteristic differences alone is the key contributor. These qualitative results are largely unchanged even after restricting the samples to those within common support.

Overall, the empirical findings indicate that occupation matters. It is an important indicator for quality of life, particularly, for migrants who earn less than their urban counterparts across different occupations. The results highlight the urgency to promote equity in pay. To this end, the enforcement of anti-discrimination legislation, safeguarding the principle of equal pay for equal work, combined with an institutional framework that creates incentives to comply, is essential to supporting migrants and urban workers to compete equally for the same job. Moreover, difference in job distribution while less important cannot be ignored. Therefore, anti-discrimination policy interventions, for example, relaxing and removing barriers such as *ho khau* in cities and ensuring equal access to good jobs is important to dismantle occupational discrimination. In the longer term, increasing investment to improve quality of (higher) education in rural areas is essential to paving the way for new policy that directly deals with positive discrimination, to be applied where certain groups are under-represented in certain occupations with employers encouraged to take a positive action in recruitment and promotion.

The result of the paper is not without caveat. Due to the sample size of the VRUMS2013, occupations are only classified into three broad categories. Finer classification will allow a better comparison between the contributions of between- and within-occupation to earning inequality. The data do not include other urban centers that also receive rural–urban migrants (e.g. Hai Phong and Da Dang). For future work, it is crucial to have a larger scale rural–urban migrant survey. In addition, a larger sample would enable researchers to divide the sample into different gender groups. This would allow better understanding especially among migrants in terms of the mechanisms leading to differences in occupational distribution as females may face different opportunities and challenges in finding jobs in the cities.

Notes

1. Vietnam's reform of *ho khau* is more like a zig-zag than a straight line (World Bank, 2016). In 2007, the Law on Residence (Decree 108/2005/ND-CP 19 August) relaxed the requirements for migrants to apply for permanent residence in the cities. e.g., the duration required for residence is reduced to one year. Uninterrupted employment is also no longer required before applying for urban *ho khau*. However, the Law on Residence effective in 2014 (Decree 31/2014/ND-CP 18 April) again requires migrant applicants to have uninterrupted employment for at least two years prior to the application of local *ho khau* in Hanoi, Hai Phong, Da Nang, Hochiminh City and Can Tho. In Hanoi, they must also own a house or a long-term rental contract and must have lived there continuously for at least three years (La, 2015).
2. Migrants are often referred to as *nha que* (country bumpkins).
3. They often do not have equal access to public services such as health, housing and education for their children in the city (Le *et al.*, 2011).
4. For instance, Arulampalam *et al.* (2007) and Miller (1987a) argue that estimates of the pay gap with and without such controls provide lower and upper bounds on the extent of discrimination.
5. Zhang (2009) finds that the wage gap between urban and long-term migrants remains over time.
6. Data examination show that the earnings of migrants are indeed higher if these long-term migrants were included in the migrant sample.
7. A rise of 6.28 percent between 2012 and 2013 was reported (*Business Times*, 2014).
8. The definition of earnings includes not only wages, but bonuses, allowances, as well as in-cash and in-kind payments. It is measured in a thousand dongs.
9. $(e^{3.72}-e^{3.2}/e^{3.2}) = 0.68$.
10. $1 approximately equals to 21,340 dongs at the time of writing.
11. The null hypothesis of no structural difference between the two groups is rejected at the one percent significance level. The Chow statistics is $F(k,N1+N2-2\times k) = F(10,1692) = 18.07$.
12. Potential experience is defined as age – years of schooling – school entry age (i.e. six years old). I also experimented with "city work experience" in migrants' earnings equation. The estimated coefficients remain more or less the same. However, missing values on "city work experience" led to the use of potential experience instead.
13. Married is coded as one and zero otherwise.
14. "Binh Duong and Dong Nai" is the reference group.
15. The decomposition analysis here does not focus on individual categories of a dummy variable. Instead, it only focuses on the contribution of the sum which is invariant to the choice of the omitted reference group (Yun, 2005). Therefore, normalization to avoid the invariant problem is not relevant in this context.
16. Ethnic dummies are excluded from the model as over 98 percent of migrants are ethnic Kinh.

17. Foreign-invested firm is the base group. It includes joint venture or pure foreign-owned enterprise.

18. Using occupation dummies as additional explanatory variables (Section 2.2), the positive and significant coefficients for both groups clearly show that occupation plays a role in determining their earnings outcomes. Relative to the decomposition results without occupation dummies controlling for occupation decreases the portion of the unexplained part irrespective of which wage structure is used. Using the urban wage structure the unexplained part decreases from 65 to 61 percent once occupation dummies are controlled for, i.e. about 4 percent can be traced back to the part of the pay gap that reflects the concentration of migrants in low-paying jobs (Hirsch and Jahn 2015).

19. To rule out model specification/methodology is the culprit I estimated a model specification as close to that of Nguyen and Minh (2016) as the VRUM2013 data is allowed (i.e. only controlling for education dummies, age and its square, gender dummy, marital status, location dummies) using monthly wage as the dependent variable and the same decomposition method. A positive and significant pay gap (0.31) remains with only 40 percent attributed to the explained part (97 percent in Nguyen and Minh, 2016).

20. The data reveals that how "professionals" is defined may have a role to play. Recall (Section 3.1) "professionals" includes "professional only" and "office workers". The estimated rates of return to schooling with selection correction is significant for urban workers with "professionals only" type of occupation. For migrants, the insignificant result, however, does not appear to relate to how "professionals" is defined. The returns to schooling remain positive but insignificant even when the earnings equation is estimated separately for "professionals only" and "office workers." Potentially this may reflect migrants with "professional only" jobs receive poorer quality of (higher) education in rural areas. Among migrant "officer workers" about 67 percent have tertiary education (vs 17 percent of their urban counterparts). Yet, on average, they still are not rewarded significantly more for their qualification. Perhaps also these migrants may have settled as office workers even though they have tertiary education as their ability to find better jobs could be more restricted than their urban counterparts.

21. Missing values on parents' occupation have precluded me from using it for identification purpose. Nonetheless, I experimented with alternative instruments such as how many other members have the same occupation as the respondent; and a dummy indicating if a worker has at least one other member with the same job category. Both variables are significant at the one percent level for migrants but are insignificant for the locals.

22. Briefly, education is important in increasing (decreasing) a rural migrant's chance of being a professional (production worker) than the reference group 'other jobs'. It also increases the locals' chance of having a professional job. In particular, the coefficients associated with no. of children and its square (the instruments) are significant at one percent and 5 percent irrespective of the occupation type.

23. The decomposition result does not change qualitatively using alternative instruments (footnote 21). Within-job difference still explains most of the overall gap.

24. Brown *et al.* (1980) decomposition result corrected for selectivity is not shown here. Briefly, it is substantially affected by the index number problem. For instance, the total (within-) explained effect switches from 71 (59) percent to 31 (11) percent of the wage gap if migrants' earnings is used as the non-discriminatory norm rather than that of the locals'. The between-unexplained effect also changes from 8 to 14 percent respectively.

25. Logit and probit models usually yield similar results (Caliendo and Kopeinig, 2008, p. 5).

26. Only variables that affect the participation decision and the outcome variable should be used to estimate the propensity score (Caliendo and Kopeinig, 2008).

27. The participation equation in the PSM is not a determinants model. Hence, the correlation of the covariates with treatment status (migrant or not) is more informative than *t*-values and the adjusted R^2 (Khandker *et al.*, 2010, p. 58). The correlation of experience, state sector dummy and years of schooling with migrant status is over 30 percent, 25 percent and 17 percent, respectively.

28. The three, four and five-nearest neighborhood matching with replacement and stratification matching generate the lowest bias across all the covariates. For other matching methods attempted, years of schooling or marital status is the only covariate that generates a maximum standardised bias ranging from about 6 to 9 percent.

29. Before matching there are significant large differences in many covariates between the two groups.

References

Appleton, S., Hoddinott, J. and Krishnan, P. (1999), "The gender wage gap in three African countries", *Economic Development and Cultural Change*, Vol. 47 No. 2, pp. 289-312.

Arulampalam, W., Booth, A. and Bryan, M. (2007), "Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution", *Industrial and Labour Relations Review*, Vol. 60 No. 2, pp. 163-186.

Bertrand, O. (2009), "Effects of foreign acquisitions on R&D activity: evidence from firm-level data for France", *Research Policy*, Vol. 38 No. 6, pp. 1021-1031.

Binci, M. and Giannelli, G.C. (2012), "Internal vs. international migration: impacts of remittances on child well-being in Vietnam", IZA DP No. 6523, Institute for the Study of Labour, Bonn, April.

Briggs, D.C. (2004), "Causal inference and the Heckman Model", *Journal of Educational and Behavioral Statistics*, Vol. 29 No. 4, pp. 397-420.

Brown, R.S., Moon, M. and Zoloth, B.S. (1980), "Incorporating occupational attainment in studies of male-female earnings differentials", *Journal of Human Resources*, Vol. 15 No. 1, pp. 3-28.

Business Times (2014), "Vietnam CPI rises 0.69% in January: GSO", *Business Times*, January 14, available at: http://businesstimes.com.vn/vietnam-cpi-rises-0-69-january-gso/

Caliendo, M. and Kopeinig, S. (2008), "Some practical guidance for the implementation of propensity score matching", *Journal of Economic Surveys*, Vol. 22 No. 1, pp. 31-72.

Coxhead, I., Nguyen, V.C. and Vu, L.H. (2016), "Migration in Vietnam: new evidence from recent surveys", MPRA Paper No. 70217, Munich, March 20.

Cu, C.L. (2005), "Rural to Urban migration in VietNam", in Thanh, H. H. and Sakata, S. (Eds), *Impact of Socio-Economic Changes on the Livelihoods of People Living in Poverty in Vietnam*, (Joint Studies on Economic Development Policies in ASEAN and Neighbouring Countries (ASEDP) Series, no. 71), Institute of Developing Economies, Japan External Trade Organisation (IDE-JETRO) and ASEDP, Hanoi, March, pp. 1-40.

Dang, N.A., Tacoli, C. and Hoang, X.T. (2003), "Migration in Vietnam: a review of information on current trends and patterns, and their policy implications", Paper presented at the Regional Conference on Migration, Development and Pro-Poor Policy Choices in Asia, Dhaka, June.

Frijters, P., Lee, L. and Meng, X. (2010), "Jobs, working jours, and remuneration packages for migrants and urban residents", in Meng, X., Manning, C., Li, S. and Effendi, T.N. (Eds), *The Great Migration: Rural-Urban Migration in China and Indonesia*, Chapter 3, Edward Elgar Publishing, Cheltenham, pp. 47-73.

Jann, B. (2008), "The Blinder–Oaxaca decomposition for linear regression models", *The Stata Journal*, Vol. 8 No. 4, pp. 457-458.

Khandker, S., Gayatri, B. and Samad, H. (2010), "Handbook on impact evaluation: quantitative methods and practices", *The International Bank for Reconstruction and Development*, The World Bank, Washington, DC, pp. 58.

La, H.A. (2015), "Housing gaps between urban-in migrants and local residents: the case of Vietnam", paper prepared for The study of rural-urban migration in Vietnam with insight from China and Indonesia Conference, January 13–14, Hanoi.

Le, B.D., Tran, G.L. and Nguyen, T.P.T. (2011), "Social protection for rural-urban migrants in Vietnam: current situation, challenges and opportunities", CSP Research Report No. 08, Institute for Social Development Studies, Hanoi, January.

Lee, L.F. (1983), "Generalized econometric Models with Selectivity", *Econometrica, Econometric Society*, Vol. 51 No. 2, pp. 507-512.

Liu, P.W., Zhang, J. and Chong, S.C. (2004), "Occupational segregation and wage differentials between natives and immigrants: evidence from Hong Kong", *Journal of Development Economics*, Vol. 73 No. 1, pp. 395-413.

Meng, X. and Zhang, J. (2001), "Two-tier labour markets in urban China: occupational segregation and wage differentials between urban residents and rural migrants in Shanghai", *Journal of Comparative Economics*, Vol. 29, pp. 485-504.

Miller, P.W. (1987a), "The wage effect of the occupational segregation of women in Britain", *Economic Journal*, Vol. 97 No. 388, pp. 885-896.

Miller, P.W. (1987b), "Gender differences in observed and offered wages in Canada, 1980", *Canadian Journal of Economics*, Vol. 20 No. 2, pp. 225-244.

Neumark, D. (1988), "Employer's discriminatory behaviour and the estimation of wage discrimination", *Journal of Human Resources*, Vol. 23, pp. 279-295.

Nguyen, C.V. and Minh, T.P. (2016), "Are migrants in large cities underpaid? Evidence from Vietnam", *IZA Journal of Migration*, Vol. 5 No. 20, pp. 1-23.

Niimi, Y., Pham, T.H. and Reilly, B. (2008), "Determinants of remittances: recent evidence using data on internal migrants in Vietnam", Policy Research Working Paper No. 4586, World Bank Development Research Group, Trade Team, Washington, DC, April.

Ñopo, H. (2008), "Matching as a tool to decompose wage gaps", *Review of Economics and Statistics*, Vol. 90 No. 2, pp. 290-299.

Oaxaca, R. (1973), "Male–female wage differentials in urban labour markets", *International Economic Review*, Vol. 14 No. 3, pp. 693-709.

Rosenbaum, P. and Rubin, D. (1985), "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score", *The American Statistician*, Vol. 39 No. 1, pp. 33-38.

Taylor, P. (2004), "Introduction: social inequality in a social state", *Social Inequality in Vietnam and the Challenges to Reform Edited by Philip Tayler*, Vietnam Update Series, Institutes of Southeast Asian Studies, Singapore, pp. 1-40.

VHLSS (2012), *Vietnam Household Living Standard Survey 2012*, General Statistics Office of Vietnam, Hanoi.

World Bank (2016), *Vietnam's Household Registration System*, Hong Duc Publishing House, Hanoi, May.

Yun, M.-S. (2005), "A simple solution to the identification problem in detailed wage decompositions", *Economic Inquiry*, Vol. 43 No. 4, pp. 766-772.

Zhang, D. and Meng, X. (2007), "Assimilation or dissimilation? The labour market performance of rural migrants in Chinese cities", Paper for the 6th Conference on the Chinese economy, Clermont-Ferrand.

Zhang, D.D. (2009), "Assimilation or dissimilation? The labour market performance of rural migrants in Chinese cities", PhD Dissertation, Australian National University, Canberra.

Zhang, Z. and Wu, X. (2013), "Registration status, occupational segregation, and rural migrants in urban China", Population Studies Center Research Reports No. 13-809, Institute for Social Research, University of Michigan, MI, November.

Further reading

Caliendo, M., Hujer, R. and Thomsen, S. (2005), "The employment effects of job creation schemes in Germany: a microeconometric evaluation", Discussion Paper No. 1512, IZA, Bonn.

Chan, A. and Wang, H.Z. (2004/2005), "The impact of the state on comparing Taiwanese factories in China and Vietnam", *Pacific Affairs.*, Vol. 77 No. 4, pp. 629-646.

Swinkels, R. and Turk, C. (2006), "Explaining ethnic minority poverty in Vietnam: a summary of recent trends and current challenges", Draft Background paper for CEM/MPI meeting on Ethnic Minority Poverty, World Bank, Hanoi, September 28.

Appendix

Table AI. Results of earnings equations for all, urban and migrant workers

Independent variables	All		Urban workers		Migrant workers		Difference	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Male	0.206***	0.024	0.202***	0.038	0.216***	0.031	0.013	0.049
Potential experience	0.026***	0.004	0.019***	0.006	0.027***	0.006	0.008	0.009
Potential exp^2	0.000***	0.000	0.000***	0.000	0.000***	0.000	0.000	0.000
Married	0.075***	0.029	0.110***	0.047	0.035	0.036	−0.074	0.059
Years of schooling	0.081***	0.004	0.098***	0.006	0.067***	0.005	−0.031***	0.008
State	−0.098***	0.041	−0.240***	0.068	0.028	0.052	0.268***	0.085
Private	−0.184***	0.035	−0.241***	0.066	−0.151***	0.040	0.090	0.075
Hochiminh city	0.068***	0.033	0.140***	0.054	0.005	0.042	−0.135***	0.067
Hanoi	0.028***	0.039	0.108*	0.061	−0.070	0.050	−0.178***	0.078
Migrant	−0.314***	0.027						
Constant	2.022***	0.071	1.867***	0.106	1.883***	0.084	0.016	0.133
No. of observations	1,712		724		988		1,712	
Adjusted R^2	0.4029		0.4115		0.2740		0.4151	
F-statistics	126.40		57.17		42.39		64.91	

Notes: The differences are derived from a pooled regression of rural migrants and urban workers. In addition to the variables listed here, all the variables are interacted with a dummy variable indicating if an individual is a rural migrant. The coefficients and *t*-ratio for these interaction terms are reported in the column labeled "Difference." *,**,***Significant at 10, 5, and 1 percent levels, respectively

	Professional				Productive workers				Other			
	Urban workers		Migrant workers		Urban workers		Migrant workers		Urban workers		Migrant workers	
Independent variables	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Male	0.168***	0.060	0.259***	0.072	0.258***	0.051	0.244***	0.035	−0.042	0.112	0.119	0.121
Potential experience	0.032***	0.012	0.044***	0.021	0.019***	0.009	0.027***	0.006	0.029	0.016	0.029	0.020
Potential exp^2	−0.001***	0.000	−0.001	0.001	0.000*	0.000	0.000***	0.000	0.000*	0.000	−0.001	0.000
Married	0.095	0.078	0.020	0.085	0.083	0.060	0.043	0.042	−0.153	0.133	−0.018	0.128
Years of schooling	0.116***	0.015	0.101***	0.016	0.050***	0.009	0.053***	0.006	0.095***	0.018	0.050***	0.021
State	−0.456***	0.108	−0.089	0.105	−0.076	0.088	0.065	0.067	0.040	0.262	−0.499***	0.234
Private	−0.359***	0.110	−0.200**	0.104	−0.136*	0.078	−0.111***	0.043	−0.302	0.262	−0.719***	0.212
Hochiminh city	0.302	0.100	0.014***	0.134	0.026	0.062	0.014	0.043	0.159	0.150	−0.060	0.226
Hanoi	0.187*	0.102	−0.070	0.147	−0.057	0.080	−0.052	0.053	0.285	0.181	0.126	0.231
Constant	1.669***	0.241	1.482***	0.283	2.261***	0.142	1.940***	0.090	1.783***	0.352	2.496***	0.420
No. of observations	335		211		287		697		102		80	
Adjusted R^2	14.87		8.24		8.98		24.18		7.95		4.07	
F-statistics	0.2721***		0.2368***		0.2008***		0.2306***		0.3824***		0.2589**	

Note: *,**,***Significant at 10, 5, and 1 percent levels, respectively

Table AII.
Results of occupation-specific earnings equation for urban and migrant workers, without accounting for selectivity

Table AIII.
Results of occupation-specific earnings equation for urban and migrant workers, accounting for selectivity

	Professional				Productive workers				Other			
	Urban		Migrants		Urban		Migrants		Urban		Migrants	
Independent variables	Coef.	SE.	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Male	0.191***	0.0686	0.466***	0.1529	0.257***	0.0551	0.280***	0.0419	−0.046	0.1124	0.224	0.1723
Potential exp.	0.036*	0.0213	0.051***	0.0244	0.019***	0.0081	0.029***	0.0078	0.022	0.0152	0.047*	0.0290
Potential exp^2	−0.001	0.0006	−0.001	0.0008	0.000**	0.0002	−0.001***	0.0002	0.000	0.0003	−0.001	0.0006
Married	0.104	0.0875	0.094	0.1257	0.081	0.0520	0.057	0.0454	−0.178	0.1262	0.049	0.1542
Yrs of schooling	0.034	0.0334	−0.109	0.1286	0.053***	0.0139	0.004	0.0122	0.073***	0.0183	0.086***	0.0326
State	−0.491***	0.1128	−0.345*	0.1943	−0.063	0.0971	−0.152*	0.0912	0.209	0.2446	−0.134	0.3578
Private	−0.343***	0.1094	−0.135	0.1568	−0.129**	0.0698	−0.138***	0.0386	−0.174	0.2154	−0.336	0.3517
Hochiminh city	0.275***	0.0903	−0.182	0.2759	0.032	0.0661	−0.060	0.0446	0.310***	0.1550	0.314	0.3651
Hanoi	0.170*	0.0982	−0.009	0.2290	−0.050	0.0828	−0.097*	0.0564	0.388***	0.1933	0.562	0.4076
Lambda	0.400***	0.1555	1.099*	0.6430	0.034	0.1748	−0.687***	0.1545	−0.379**	0.1985	−0.967***	0.4969
Constant	3.083***	0.5856	5.501***	2.4950	2.252***	0.1286	2.246***	0.1168	1.343***	0.4879	−0.675	1.6971
No. of obs.	335		211		287		697		102		80	
$\chi^2(10)$	106.3***		46.11***		99.37***		197.12***		62.49***		24.02***	

Note: *,**,***Significant at 10, 5, and 1 percent levels, respectively

Independent variables (migrant = 1)	All Coef.	SE
Male	0.477***	0.1153
Potential experience	−0.147***	0.0202
Potential exp^2	0.001**	0.0004
Married	0.550***	0.1392
Years of schooling	−0.211***	0.0190
State	−1.047***	0.1919
Private	−0.542***	0.1685
Hochiminh city	0.674***	0.1596
Hanoi	0.783***	0.1858
Constant	4.023***	0.3239
No. of observations	1,712	
Pseudo R^2	0.1814	
LR χ^2 (9)	423.28	

Note: *,**,***Significant at 10, 5, and 1 percent levels, respectively

Table AIV.
Logit selection model for propensity score

Variable	Mean Treated	Control	% Bias	% reduction |bias|	*t*-value
Gender	0.58	0.57	1.3	85.5	0.29
Experience	12.87	12.91	−0.4	99.4	−0.10
Experience square	256.31	255.96	0.1	99.9	0.02
Married	0.60	0.59	1.7	91.3	0.36
Years of schooling	10.85	10.64	5.0	85.7	1.10
State	0.16	0.15	1.3	97.5	0.34
Private	0.63	0.64	−2.2	90.8	−0.51
Foreign-invested firms	0.55	0.52	4.8	69.7	1.07
Hanoi	0.28	0.26	3.8	67.6	0.87

Table AV.
Checking matching quality using the two-sample *t*-test and the standardised bias

Corresponding author
Amy Y.C. Liu can be contacted at: yukchu.liu@anu.edu.au