ADBI Working Paper Series

CITY SIZE, MIGRATION, AND URBAN INEQUALITY IN THE PEOPLE'S REPUBLIC OF CHINA

Binkai Chen, Dan Liu, and Ming Lu

No. 723
April 2017

Asian Development Bank Institute

Binkai Chen is with the School of Economics, Central University of Finance and Economics. Dan Liu is associate professor, School of Economics, Shanghai University of Finance and Economics. Ming Lu is professor at the Department of Economics, Shanghai Jiao Tong University, and adjunct professor at Fudan University.

Suggested citation:

Chen, B., D. Liu, and M. Lu. 2017. City Size, Migration, and Urban Inequality in the People's Republic of China. ADBI Working Paper 723. Tokyo: Asian Development Bank Institute. Available: https://www.adb.org/publications/city-size-migration-and-urban-inequality-prc

Please contact the authors for information about this paper.

Email: chenbinkai@gmail.com, liu.dan@mail.shufe.edu.cn, luming1973@sjtu.edu.cn

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

Abstract

This paper examines the relationship between city size, migration, and urban income inequality using the 1% sample of the 2005 Census in the People's Republic of China (PRC). We calculate various measures of within-city income inequality for 252 PRC cities. It is found that city income inequality is significantly and positively correlated with city population size. As rural-to-urban migration is crucial in determining city size distribution in the PRC, we focus on exploring the role of massive migration into big cities on shaping this size-inequality relationship. We find that the share of migrants alone accounts for more than 40% of the city-size inequality premium. This is mainly because migration leads to a higher skill premium in larger cities by changing the skill composition of the labor force in those places. The main findings still hold after we deal with the endogeneity of migration using an instrument variable approach.

Keywords: city size, income inequality, migration, skill premium, People's Republic of China

JEL Classification: R12, O18, J31

Contents

1. INTRODUCTION

In 2011, 52.1% of the world's population was living in cities (United Nations, Department of Economic and Social Affairs, Population Division, 2012), and this number is still rising rapidly. As economies develop, economic activities generally become more concentrated in leading areas around large cities (World Bank, 2009). At the same time, inequality has increased over the past decades, both in developing and developed countries (Goldberg and Pavcnik 2007). Recently, several studies about the US have linked inequality to city size, finding that larger cities tend to have higher inequality (Baum-Snow and Pavan, 2013; Glaeser, Resseger, and Tobio 2009; Long et al. 1977). This immediately raises the question of whether this observation holds true in other countries. Additionally, could increased urbanization explain the trends in inequality noted above?

The People's Republic of China (PRC) is experiencing both rapid urbanization and rising inequality. The urbanization rate reached 54.77% by the end of 2014 (National Bureau of Statistics 2015), growing at an annual rate of 1.3 percentage points over the recent decade.[1] However, most PRC cities are still growing to gain greater productivity. Both the number of big cities with populations over 1 million and the proportion of the population living in big cities are lower than the world's average (Henderson 2007; Wang 2010; Chauvin et al. 2016). Most economists expect that PRC cities will become larger in the coming decades (Au and Henderson 2006a, 2006b; Henderson 2007; Lu 2013). If there is a connection between city size and inequality, then income inequality can be expected to grow over time. At the beginning of economic reform in the early 1980s, the Gini coefficient was less than 0.3 (Wan, Ye, and Zhuang 2012). The Gini coefficient stood at 0.474 in 2012, receding from a local peak of 0.49 in 2009.[2] Urban inequality, a major portion of overall inequality, has also kept rising over the past three decades (Ravallion and Chen 2007; Lin et al. 2010). The proportion of urban inequality that accounted for within-city inequality increased from 74% in 1992 to more than 81% in 2009.[3] However, there is no literature we are aware of that examines the drivers of within-city inequality in the PRC. We believe that city characteristics such as city size and the share of migrants in the PRC are important in this regard. In particular, if economic policies promote urbanization, and if the positive city size-inequality relationship holds in the PRC as in the US, analyzing within-city inequality becomes very relevant. Moreover, whether people feel happy is also a function of within-city income comparisons (Jiang, Lu, and Sato 2012).

To study this issue, we first constructed various within-city inequality measures for 252 PRC cities in 2005 and related them to city size. Consistent with the literature using US data (Baum-Snow and Pavan 2013; Glaeser, Resseger, and Tobio 2009; Long et al. 1977), city inequality is significantly and positively correlated with city size in the PRC within provinces. However, as opposed to Baum-Snow and Pavan (2013), who found that this correlation was due to higher within-skill group inequality in the US, the size-inequality correlation in the PRC is mainly due to higher skill premiums in larger cities. In fact, there is no significant correlation between residual inequality and

1 Computed by the authors based on the data from National Bureau of Statistics (NBS).

2 On 18 January 2013, the NBS publicly released data on the PRC's income inequality. According to the report, from 2003 to 2012, the Gini coefficient was 0.479, 0.473, 0.485, 0.487, 0.484, 0.491, 0.490, 0.481, 0.477, and 0.474. See the report at http://news.qq.com/a/20130118/001100.htm.

3 The numbers were calculated using data from the Urban Household Surveys from 1992 and 2009, respectively, by a simple decomposition of the Theil Index into a within-city component and a between-city component.

city size in the PRC. This implies that, in contrast to the story of sorting by within-skill group abilities in the US (Davis and Dingel 2012), the size-inequality relationship in the PRC is mainly due to observable skill composition. Moreover, we find that city size is positively correlated with the 90–50 percentile gap, but negatively correlated with the 50–10 percentile gap, which is the opposite of the case in the US.

Massive and unprecedented rural-urban migration into big cities can help explain the above features of the size-inequality relationship in the PRC. The share of migrants alone accounts for more than 40% of the city-size inequality premium. Even after controlling for other potential factors that could affect urban inequality and applying an instrumented variables approach to correct the potential bias due to endogenous migration flows, the share of migration still has a very significant impact on overall inequality. Further investigation into the college premium and high school premium shows that migration mainly affects overall inequality through its impacts on skill premiums. Migration accounts for more than 50% of the city-size premium for the college premium and around 30% of that for the high school premium.

Our explanation for these results is straightforward. Rural migrants are relatively less educated than local urban residents, and the average education gap between migrants and local residents increases with city size. In other words, bigger cities attracted a larger share of more uneducated workers, which increased the relative supply of unskilled workers. This would naturally lead to a higher return on skills in bigger cities, either due to skill complementarity or externalities, or both (Eeckhout et al. 2014; Combes, Démurger, and Li 2015). Finally, migration can also help explain why the upper percentile gap increases, but the lower percentile gap decreases with city size in the PRC through the above composition effects.

Our paper is organized as follows: Section 2 is a literature review; section 3 discusses data and measures of inequality; section 4 shows our empirical results; section 5 presents a mechanism analysis; the final section concludes.

2. LITERATURE REVIEW

The relationship between income inequality and city size has become a new topic in both inequality and urban studies. Our paper is most closely related to the recent research by Baum-Snow and Pavan (2013). They investigated the relationship between city size and income inequality using US data from 1979 to 2004, documenting a positive relationship between city size and wage inequality over recent decades and finding that within-group inequality in larger cities is the most important force driving this relationship. While their main focus was the contribution of city-size premiums to the nationwide increase in inequality over time, we focused our efforts on examining the cross-sectional relationship between city size and inequality using PRC data. Moreover, we emphasized the role played by rural-to-urban migration in the PRC in shaping the city-size premium, which differs from the US where urbanization has not changed significantly in recent decades.

Glaeser, Resseger, and Tobio (2009) explored the determinants of the variation in the degree of inequality across cities in the US. They found that inequality in skills can explain about one-third of the variation in urban inequality, but left the mechanism as an open question. Our paper investigates why bigger cities are more unequal in the PRC by exploring city size, migration and city inequality. In particular, our focus on rural-to-urban migration can lead to policy implications for where and how to conduct income redistribution and how to equalize welfare through public service provision in a developing country like the PRC.

Our paper is also inspired by research on migration, especially the impact of immigration on wages or wage disparities. As summarized by Card (2007), immigrants have a very small impact on native wages and inequality in the US due to imperfect substitution between immigrants and natives (also see Ottaviano and Peri 2008, 2012; Dustmann, Frattini, and Preston 2013), but clearly positive effect on overall inequality due to composition effect.

Related research on the PRC is less abundant, but largely found no negative impact of migration on native wages (Meng and Zhang 2010; Combes, Démurger, and Li 2015). Different from these papers, we emphasize how migration actually helps explain within-city inequality and city-size inequality premium.

The study of the inequality-city size relationship has also opened a new angle to further understand recent global trends in within-country inequality. Lemieux (2006) and Autor, Katz, and Kearney (2008) showed that there has been a sharp rise in US wage inequality since 1979. Han, Liu, and Zhang (2012) found that wage inequality has been also increasing in the PRC from 1989 to 2008. Yang (2005) explained why the return to education has increased during this economic transition in the PRC. Cai, Chen, and Zhou (2010) studied urban income inequality at the aggregate level from 1992 to 2003. They decomposed overall inequality into between-group inequality and within-group inequality, and found that both types of inequality increased rapidly during this period. However, most of the studies documenting rising inequality have not linked aggregate inequality to the increased concentration of the PRC's population in big cities.

Finally, spatial sorting and skill complementarity may help in understanding the relationship between city size and inequality. Combes, Duranton, and Gobillon (2008) found strong evidence of spatial sorting by skills in France, which is relevant to wage disparity across regions. Eeckhout et al. (2014) analyzed skill distribution in large and small cities based on US data. They found a higher proportion of high-skilled and low-skilled labor in big cities, and a lower proportion of medium-skilled labor, indicating that big cities promote skill complementarity. If skill complementarity really exists, this may explain why low-skilled and high-skilled labor co-locates in large cities. Inspired by the idea of skill complementarity, our study examines whether the migration of low-skilled workers into large cities can increase the skill premium, and within-city inequality widens as a consequence.

3. DATA AND MEASURES OF INEQUALITY

3.1 Data

Our data come from the PRC's 2005 population survey conducted by the National Bureau of Statistics (NBS), which is the largest PRC dataset available with individual income information. The survey is nationally representative, with the respondents randomly selected from each of the PRC's 2,861 counties using a three-stage cluster sampling method (Zhang et al. 2005). Our sample is a subset of the original survey, which contains over 2,500,000 observations randomly drawn from the NBS dataset.

In 2005, the PRC still adopted the very restrictive *hukou* or household registration system, under which each person obtains an identity according to his/her parents' *hukou* identity.[4] The *hukou* system separates people into two groups: those with an urban *hukou* identity and those with a rural *hukou* identity. Only people with rural *hukou*

[4] Please refer to Liu (2005), Chan and Buckingh (2008), Chan (2009), and Lu et al. (2013) for an introduction to both the history and the reform of the *hukou* system.

are assigned farmland. For local people with rural *hukou*, it is more likely that they work on farmland since they live nearby. Since our main interest is urban inequality, we have excluded local people with rural *hukou* when calculating inequality measures. However, rural-to-urban migrant workers usually work in urban areas since they have been separated from their farmland. In this paper, all of our measures of inequality are calculated using information on individuals with urban *hukou* or rural migrants to cities.[5] A worker is identified as a migrant if he/she had a rural *hukou*, his/her registration prefecture city was not where he/she lived, and he/she had stayed there for no less than six months.[6]

In this study, we use prefecture cities in 2005 as our geographic units and the Gini coefficient as the main measure of inequality. Throughout, our focus is on male and female residents older than 18 and younger than 53.[7] Also, we only consider persons who were employed when interviewed. Only Han residents, who account for 90% of the sample, are included in the analysis. Moreover, in order to make our inequality measure more precise, we have dropped cities with sampled sizes lower than 800. The number of observations dropped account for less than 2% of the total sample size. This leaves 252 cities in our sample. Very small cities have been excluded from the sample such that only cities with a total local population equal to or above 190,000 are included.

3.2 Measures of Inequality

Our income measure is calculated using monthly income divided by weekly working hours. This is the best feasible measure since neither weekly income nor monthly working hours are available. This measure is a good proxy for the hourly wage, the focus in most of the literature on inequality (Lemieux 2006; Card 2007; Baum-Snow and Pavan 2013) if all workers worked for the same number of hours every week, or if it was the average of weekly work hours reported.

We use the Gini coefficient as our main measure of inequality in order to facilitate the comparison to the rest of the literature. The Gini coefficient is defined as $1-\frac{1}{\bar{y}}\int_y (1-F(y))^2 dy$, where *F(y)* is the share of the population with income levels less than y, and $\bar{y}$ is the average income. This measure can be interpreted as the area between the 45 degree line and the Lorenz curve divided by the triangle below the 45 degree line. It is invariant with respect to scale, which means that it would not change with a proportional increase in income for everyone or a change in population. An alternative measure that is also commonly used is the Theil index. The Theil index is defined as $T=\frac{1}{N\bar{y}}\sum_{i=1}^{N} y_i \ln\left(\frac{y_i}{\bar{y}}\right)$, where $\bar{y}$ is the mean income in the sample, y_i represents the individual's income, and *N* is the number of people in the population. Another measure of inequality is the coefficient of variation, $\frac{1}{\bar{y}}\sqrt{\int_y (y-\bar{y})^2 dF(y)}$. These

5 Ideally, we would like to focus on people who live in the central urban areas, but not affiliated counties. However, it is impossible to identify those people. By considering only people with urban identities and rural migrants, this problem can be minimized. Using the information from the Urban Household Surveys, we can show that the number of urban people living in the central urban area is highly correlated with the whole urban population, with correlation coefficient equal to 0.96 in 2005.

6 The new migrants who moved into the city for less than six months accounts for about 10% of all migrants. Our results would still hold if we include them in our analysis.

7 These age cutoffs are chosen to be as close as to those in Baum-Snow and Pavan (2013). It can be shown that our results are not affected by these age cutoffs.

two measures are also both invariant to a proportional change in income for everyone or a change in population size.

We calculate all of these measures of inequality and have listed the coefficient of correlation between each pair in Table 1. In addition, we also calculate inequality measures using information on male workers only. It can be seen from Table 1 that all of these inequality measures are fairly highly correlated. For example, the correlation coefficient between the Gini coefficient and the Theil index is 0.92. The correlation coefficient between the coefficient of variation and the Theil index is 0.88. In our analysis, we will use the Theil Index and the Gini coefficient as robustness checks to make sure that our main results do not depend on a specific measure. In general, these different measures give us very similar pictures of urban inequalities in the PRC.

Table 1: Correlations between City Income Inequality Measures

	Gini Coeff.	Theil Index	Coeff. of Variation	Residual Inequality
Gini Coeff.	1.00			
Theil Index	0.92	1.00		
Coeff. of Variation	0.69	0.88	1.00	
Residual Inequality	0.73	0.70	0.57	1.00
gini Coeff._male	0.94	0.92	0.74	0.68
Theil Index _male	0.83	0.95	0.92	0.63
Ln(pop)	0.44	0.44	0.38	0.21
Ratio of Migrants	0.55	0.58	0.44	0.27
	Gini Coeff._male	**Theil Index _male**	**Ln(pop)**	**Ratio of Migrants**
Gini Coeff.				
Theil Index				
Coeff. of Variation				
Residual Inequality				
gini Coeff._male	1.00			
Theil Index _male	0.91	1.00		
Ln(pop)	0.52	0.47	1.00	
Ratio of Migrants	0.55	0.56	0.43	1.00

Notes: This table reports the coefficients of correlation between various measures of city-level income inequality. All inequality measures were calculated using the information of income for working people with urban *hukou* or rural immigrants from 2005 mini census. The Gini Coeff._male and Theil Index_male were calculated with male workers only. Refer to the text for detailed information on how residual inequality is calculated. Ln(pop) is a proxy for the log of total city population with urban *hukou* or rural immigrants, recovered from the sample.

In 2005, the city-level Gini coefficient ranged from 0.23 to 0.48 (see Table A.1 in the appendix). Guangdong province had the most unequal cities with Shenzhen's Gini coefficient equal to 0.48 and Guangzhou's Gini coefficient equal to 0.42. Beijing and Shanghai were the sixth and seventh most unequal cities in terms of the Gini coefficient. Shaanxi province had the most equal cities. Four out of the five most equal cities were in Shaanxi. In general, cities in the eastern coastal areas were most unequal, and those in the north or west were relatively more equal. This pattern is consistent with a positive correlation between inequality and average income, with a coefficient of correlation equal to 0.26.

Overall inequality can be decomposed into two components: between-group inequality and within-group. Between-group inequality describes the distribution of income between people with different observable characteristics such as education, age, or gender, while within-group inequality describes the income distribution for people within categories of observable characteristics. Residual income inequality is commonly used as a measure of within-group inequality in the literature. A Mincer regression is typically employed to calculate residual inequality. To estimate our measures of within-group inequality, we separated people into 12 age groups and five education groups. Using these groups, we formed 60 age-education skill groups using age and education bins. Residual wage inequality was calculated from the residuals of a regression of log wages on a set of dummies identifying each of these 60 groups.[8] The Mincer regression was run for each city, while city-level residual wage inequality was measured by the Gini coefficient of the exponential of the residuals for all workers within each city.[9]

We considered two measures of between-group inequality: the college premium and the high school premium. The college premium was measured using the difference between the average log of wages for people with at least some college education and the average for all other education levels. Similarly, the high school premium is measured using the difference between the average log of wages for people with at least some high school education and the average for all other education levels. It should be noticed that a high school premium is not the return to high school education. It is the weighted average of returns to high school and college education. Both of these measures are recovered from the census based on the income measure described above.

Two measures of city size are adopted in this paper. The first is the number of residents with local urban *hukou* in the central city which comes from the PRC's City Statistical Yearbooks.[10] The main issue with using this measure is that it does not consider rural migrants without local urban *hukou* but living and working in the city areas. Therefore, for the main analysis, we used the sampled size of the sum of individuals with local urban *hukou* or rural migrants in the census as a measure for city size. This is a better measure of city size in the sense that it includes migrants, and the main portion of migrants live in the central urban areas. Since the census is a random sample of the whole population, we can use it as a proxy for the total population in the urban areas.

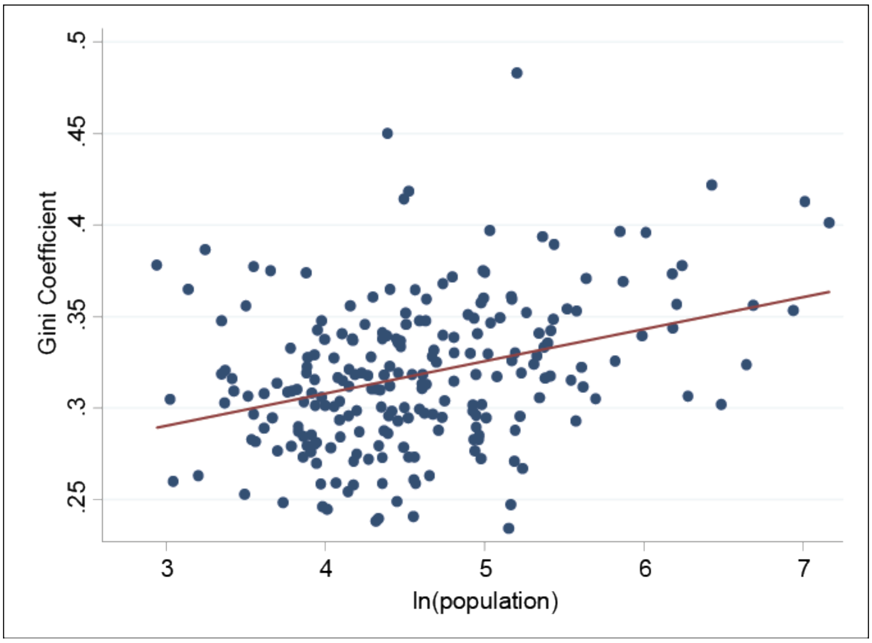
---

[8] We controlled for skills here so our results are comparable to Baum-Snow and Pavan (2013).

[9] We also tried running one Mincer regression for the whole sample, and calculated the residual inequality for each city using residuals from this common regression. The results were similar. In addition, we also tried alternative measures like the variance or the Theil index of residuals. The results were also similar.

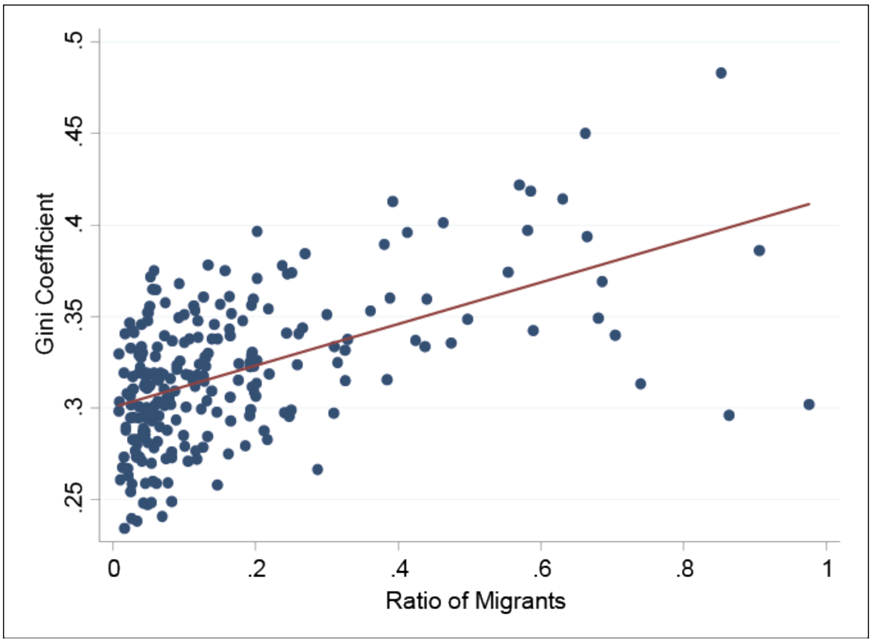
[10] There were 227 cities left when merging census data with city data, as population size is missing for some small cities. The cities with missing population size data are not "real cities" in an economic sense, as their population sizes are too small. Instead, they are administrative units, which are named as *Zizhizhou* or *Meng,* rather than cities in local terminology. Based on the 2005 census data, the average sample size of the cities with a missing population size is 389, while the average sample size of the cities with a non-missing population size is 1,299.

Figure 1: Gini Coefficient and City Size



Notes: The horizontal axis represents the log of city population and the vertical axis is the Gini coefficient income for people with urban *hukou* and rural migrants within each city.

Figure 2: Gini Coefficient and Ratio of Migrants



Notes: The horizontal axis represents the ratio of migrants in city population and the vertical axis is the Gini coefficient income for people with urban *hukou* and rural migrants within each city. Please refer to the text for detailed information.

Figure 1 shows a scatter plot for the relationship between city size and overall inequality. It shows a clear positive correlation between the Gini coefficient and city size with the coefficient of correlation of 0.44 (see Table 1), which is much stronger than the result for US cities in 2000 (Glaeser, Resseger, and Tobio 2009). This positive correlation is also seen in Table 2, which shows the average Gini coefficient for these size groups of cities. We separate cities equally into three groups according to population size and calculate the average inequality for each group. The standard deviations are reported in the parenthesis below. It can be seen that the average Gini coefficient for large cities is about 10% higher than that for small cities.[11] Figure 2 shows that there is also a positive correlation between the city-level Gini coefficients and the share of migrants. The coefficient of correlation is 0.55 (see Table 1) higher than that with city size.

Table 2: Average Income Inequality by City Size

	Small	Middle	Large
Gini coefficient	.3027	.3154	.3374
	(.0331)	(.0368)	(.0426)
Theil index	.1595	.1748	.2080
	(.0402)	(.0536)	(.0643)
90–10 percentile gap	1.4695	1.4930	1.5313
	(.1641)	(.1655)	(.2042)
90–50 percentile gap	.6036	.6762	.7854
	(.1464)	(.1478)	(.1866)
50–10 percentile gap	.8659	.8168	.7460
	(.1681)	(.1432)	(.1423)
Residual inequality	.2839	.2908	.2934
	(.0246)	(.0272)	(.0243)
College premium	.5060	.5407	.6476
	(.1454)	(.1560)	(.1983)
High school premium	.4609	.4374	.5052
	(.1439)	(.1117)	(.1431)
Ratio of Migrants	.1197	.1100	.2654
	(.1365)	(.1175)	(.2349)

Note: This table reports the average of each measure of inequality for small cities, middle cities and big size cities. Standard deviations are reported in parentheses. Cities are equally distributed into each group by population size. Percentile gaps were measured using the difference in percentiles of log (income).

4. EMPIRICAL RESULTS

We now turn to the exploration of the connection between city inequality and city size, and how migration plays an important role. We first establish a robust positive correlation between inequality and city size, and then investigate how the share of migrants affects the inequality-size relationship and discuss the potential mechanisms.

[11] An F-test was conducted to test the null hypothesis that means are the same across three groups for each measure. The null hypothesis was rejected at 5% for all measures except for lower percentile gap and residual inequality.

4.1 City Size and Inequality

The equation we estimate is straightforward:

$$\ln(\mathrm{inequality_i}) = \alpha \cdot \ln(\mathrm{city\ size_i}) + \beta \cdot \mathrm{share\ of\ migrants_i} + \mathrm{X_i}'\gamma + \varepsilon_\mathrm{i} \quad (1)$$

where $\mathrm{inequality_i}$ is the degree of income inequality in city *i*. We used the Gini coefficient as the main measure of overall inequality and the Theil index, 90–10 percentile gap and the coefficient of variation as alternative measures for robustness checks. In the main analysis, the population in the urban area is used to measure city size.[12] Our focus is β, that is, the effects of migration on inequality after controlling for city size and other observables. Furthermore, we are also interested to see how α, the city size-inequality premium, changes after controlling for the migration ratio. The vector $\mathrm{X_i}$ includes other city characteristics such as the average wage, education composition and inequality, industry composition, share of employment by state-owned enterprises (SOE), and province dummies.[13]

Table 3: City Size and Income Inequality

Dep. Var.	(1) Gini	(2) Gini	(3) Theil	(4) P90–10	(5) College Premium
Ln(pop)	0.024***	0.028***	0.037***	0.098***	0.093***
	[0.004]	[0.006]	[0.009]	[0.030]	[0.027]
_cons	0.137***	0.119*	–0.087	0.764**	–0.057
	[0.028]	[0.059]	[0.093]	[0.312]	[0.285]
R^2	0.20	0.54	0.49	0.39	0.65
Province FE	no	yes	yes	yes	yes
N	252	252	252	252	252

Dep. Var.	(6) High School Premium	(7) Residual Inequality	(8) P90–50	(9) P50–10
Ln(pop)	0.062***	0.009	0.125***	–0.027*
	[0.017]	[0.006]	[0.028]	[0.016]
_cons	0.158	0.222***	–0.406	1.170***
	[0.172]	[0.035]	[0.293]	[0.165]
R^2	0.51	0.22	0.59	0.53
Province FE	yes	yes	yes	yes
N	252	252	252	252

Notes: “gini” is within-city gini coefficient; “theil” is Theil Index; “p90–10”, “p90–50” and “p50–10” are percentile gaps. Please refer to the text for information on how to calculate skill premiums and residual inequality.

[12] Gross domestic product can be another measure of city size, but suffers more from endogeneity problem. Also, we are more interested in the effects of city expanding on inequality.

[13] We also tried to control for other variables, such as Han ratio, foreign investment, and government expenditure on education. None of these variables have a significant impact on inequality after controlling for average income. This is probably because these variables are highly correlated with average income.

Table 3 reports the unconditional correlation between each measure of inequality and city size, which can be used as a reference for later analysis. The positive correlation between inequality and city size is shown by the results in columns 1–4 with/without controlling for province-fixed effects. City size alone can explain 20% of the cross-city variation in inequality. Columns 5 and 6 show that the college premium and high school premium are both higher in bigger cities and the correlation with city size is even stronger than that for overall inequality. However, there is no significant correlation between residual inequality and city size in the PRC after controlling for province-fixed effects, as shown in Column 7. This implies that between-group inequality is a key determinant of cross-city variation in inequality. The last two columns show that the upper percentile gap is positively correlated with city size, but not the lower percentile gap.

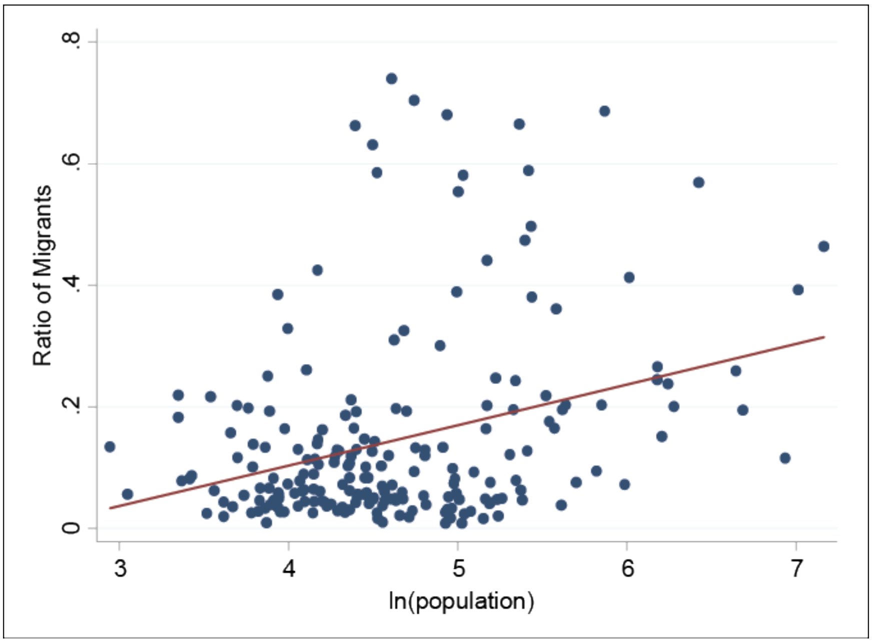
4.2 The Role of Migration

In the following analysis, we focus on the role of migration on shaping the above size-inequality relationship in the PRC. The PRC is experiencing rapid urbanization and its urban system is being reshaped by rural-to-urban and interregional migration. As in other countries, the PRC also has an urban system consisting of cities of different sizes. During the pre-reform years, the PRC government controlled interregional migration through the *hukou* system, and intentionally moved people to inland areas to support the development of lagging regions. This made the PRC's spatial distribution of population deviate from the rank-size correlation, namely Zipf's Law. However, since the economic reforms began, especially after the 1990s, the PRC has experienced a drastic rise in urbanization and interregional migration. Large cities have attracted more migrants (Gao, Lu, and Sato 2015). Figure 3 shows that city size and the share of migrants are positively correlated to each other. Thus, the spatial distribution of the PRC population has been shifting closer to the rank-size correlation, following Zipf's Law.[14]

From the simple ordinary least squares (OLS) results reported in Table 4, it can be seen that the city size-inequality relationship still exists after controlling for the migrant ratio (column 2), but the coefficient on city size drops by about 40%. Also, the R-squared increases from 0.20 to 0.33. This implies that migration is indeed closely associated with inequality and the city-size inequality premium. However, we cannot make further inference from this simple regression due to the following two concerns: first, since migration itself is endogenously determined, any omitted variable that affects both migration and inequality could lead to potential upward bias in the estimated coefficient on the migrant share; second, inequality itself could affect migration directions, which would raise the issue of reverse causality. We deal with these two issues by both trying to control for other factors that were documented to be related to inequality and migration, and adopting an instrumental variable approach.

[14] See Lu et al. (2013), Lu and Wan (2014), and Chen and Lu (2015), for references on the evolution of the PRC's urban system.

Figure 3: Migration and City Size



Notes: The vertical axis represents the ratio of migrants in city population and the horizontal axis is the log of urban population. Please refer to the text for detailed information.

Migration is potentially correlated with many economic factors that also affect the income distribution. Column 3 includes mean wages in a city to control for the development level. Following the literature on US inequality (Baum-Snow and Pavan 2013, Glaeser, Resseger, and Tobio 2009), we also considered education and industry composition. In column 3, we control for both education composition and inequality. The share of workers with some college education and the share of workers with some high school education are used to capture the differences in education composition across cities. We also construct the Gini coefficient of the education level as a measure of education inequality.[15] There are seven education categories based on the highest level of education completed: below primary school, primary school, middle school, high school, professional college, university, graduate school, and above. One to seven is used to indicate each category.[16]

Education compositions can affect income inequality in two ways. First, it has been widely documented that within-group income inequality is higher for those with higher education (Lemieux 2006). At the same time, education composition may affect between-group inequality. A change in education composition means that the supply of workers with various education attainments changes, and thus it would affect the relative earnings of each group of people. For example, if there were a greater number of college graduates, we would expect the college premium to be lower, holding other conditions equal. Education inequality is expected to increase city income inequality as shown in Glaeser, Resseger, and Tobio (2009). Column 3 shows that the high school ratio and education inequality both increase city income inequality. The share of people with at least some college education is not significantly correlated with overall city

[15] We also checked the results using the Theil index and the coefficient of variation. Both gave us similar results.

[16] An alternative measure of education inequality is to recover years of schooling from this indicator. The problem is that it is unclear whether people who claim high school-level education graduated from high school, or if they were actually drop-outs.

income inequality. This is unlike what we have learned from the US (Lemieux 2006; Autor, Katz, and Kearney 2008; Glaeser, Resseger, and Tobio 2009). The difference is consistent with our previous findings in Table 3 that the cross-city variation in inequality was mainly driven by between-group inequality in the PRC, but by within-group inequality in the US. It could be partially explained by PRC college graduates having a lower probability of opening their own business than other education groups (Lu and Ni 2014).

Column 4 of Table 4 includes industry composition, which can be relevant to local inequality for two reasons. First, the education composition of workers employed by each industry can be different. Some industries require higher-skilled labor than others. Second, the return to education could be different across industries. Also, the return to unobservable skills could vary across industry. Following Glaeser, Resseger, and Tobio (2009), we use the share of employment in the agricultural sector,[17] and the share of employment in the manufacturing sector as our proxies for industry compositions for our main analysis.[18] Column 4 suggests that a higher manufacturing ratio is accompanied by a lower within-city income inequality.

Moreover, since ownership structure was relevant to both migration and income inequality in the process of urbanization in the PRC (Xing and Li 2012; Liu, Park, and Zhao 2010), we controlled for the share of state-owned firms as the measure of firm ownership composition. Ownership restructuring can also affect both between-group inequality and within-group inequality. On the one hand, it could increase within-group inequality. In a state-owned firm, people are paid according to observable characteristics, such as experience. This means that ownership restructuring through developing more private firms and linking salaries with performance can increase within-group inequality. On the other hand, it could also increase between-group inequality. In state-owned firms, income could depend less on education, but more on experience and social connections. Based on the results from the Mincer regression using our data, the return on education is on average 10% lower in state-owned firms than other firms, and the return on experience is around 26% higher. So ownership privatization may raise education returns and increase between-education inequality. As expected, Column 5 shows that the share of state-owned firms is negatively correlated with income inequality, but insignificantly after other variables are controlled for. This doesn't necessarily mean that institutional reform didn't have any impact on inequality. It is more likely that the impacts were already reflected in educational composition and industry composition.[19] The last column reports the results when only male samples are considered, as in Baum-Snow and Pavan (2013). Neither the sign pattern nor the magnitudes change much. This further confirms our above findings.[20]

[17] Here, the agriculture sector refers to the agriculture-related activities around the suburb areas around cities, such as planting vegetables.

[18] We also examined industry compositions at the one-digit level (see Table A.2 in the appendix), as in Baum-Snow and Pavan (2013).

[19] In fact, the last column in Table A.2 shows that the negative correlation is significant when other factors are not controlled for.

[20] It should be noted that if the estimations of equations (1) to (5) were repeated for male workers, we would still find that city size can explain the inequality variation to a great extent. Again, about 36% of the city size-inequality relationship can be explained by migrant ratio. But we do not report these stepwise regressions to save space. Stepwise results are available upon request.

Table 4: City Inequality and Migration: The Dependent Variable is Gini Coefficient

Sample	(1) All	(2) All	(3) All	(4) All	(5) All	(6) Male Only
Ln(pop)	0.024***	0.014***	0.019***	0.022***	0.021***	0.023***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Migrant ratio		0.090***	0.090***	0.118***	0.117***	0.161***
		[0.028]	[0.028]	[0.033]	[0.036]	[0.029]
Mean wage			–0.021	–0.026	–0.025	–0.028
			[0.022]	[0.020]	[0.020]	[0.017]
High school			0.238***	0.219**	0.221***	0.263***
			[0.084]	[0.080]	[0.080]	[0.064]
College			0.052	0.019	0.023	0.042
			[0.052]	[0.049]	[0.049]	[0.042]
Edu gini			0.985***	0.882***	0.877***	0.722***
			[0.188]	[0.201]	[0.196]	[0.178]
Ag ratio				–0.010	–0.006	–0.009
				[0.036]	[0.045]	[0.048]
Manuf ratio				–0.080**	–0.079**	–0.083**
				[0.032]	[0.032]	[0.032]
Soe ratio					–0.005	–0.018
					[0.027]	[0.028]
_cons	0.137***	0.195***	–0.034	–0.008	–0.005	–0.013
	[0.028]	[0.027]	[0.103]	[0.090]	[0.090]	[0.080]
Provinces	no	no	yes	yes	yes	yes
R^2	0.20	0.33	0.61	0.63	0.63	0.64
N	252	252	252	252	252	252

Note: * significant at 10% level; ** significant at 5% level; *** significant at 1% level; standard errors are reported in parentheses and clustered at province level. “migrant ratio” is the ratio of migrants in city population; “mean wage” is the log of average individual wage; “high school” is the ratio of high school graduates; “college” is the ratio of workers with college education or above; “edu gini” is education Gini coefficient; “Ag ratio” is the share of workers in agriculture sector and “Manuf ratio” is the share of workers in the manufacturing sector; “Soe ratio” is share of workers employed be state-owned firms. The detailed definitions of variables are explained in the text.

Table 5 shows that our baseline results, that is, that city size and inequality are positively correlated and migration plays an important role on shaping this correlation, hold regardless of how we measure inequality. We re-estimate the models using the Theil index, coefficient of variation and the 90–10 percentile gap. The results are consistent with those reported in Table 4. Again, larger cities are found to have higher inequality and city size and province-fixed effects can explain almost half of the variation in the Theil index. The migrant ratio partly explains why large cities are more unequal in income. The sign and significance of other control variables don’t change much either.

Table 5: City Inequality and Migration Using Alternative Measures of Inequality

Dep. Var.	(1) Thiel	(2) Thiel	(3) Thiel	(4) Thiel	(5) Cov	(6) p90–10
Ln(pop)	0.037***	0.027***	0.023***	0.027***	0.091**	0.085***
	[0.009]	[0.009]	[0.007]	[0.007]	[0.033]	[0.023]
Migrant ratio		0.093***	0.135***	0.170***	0.317**	0.462**
		[0.033]	[0.036]	[0.048]	[0.145]	[0.222]
Mean wage			–0.016	–0.017	0.014	–0.140
			[0.028]	[0.024]	[0.083]	[0.083]
High school			0.209**	0.177**	0.036	1.220***
			[0.094]	[0.086]	[0.274]	[0.401]
College			0.050	–0.009	–0.290	0.351
			[0.048]	[0.047]	[0.206]	[0.259]
Edu gini			0.807***	0.650**	0.127	5.264***
			[0.220]	[0.242]	[1.800]	[1.225]
Ag ratio				–0.038	–0.212	0.130
				[0.051]	[0.129]	[0.244]
Manuf ratio				–0.113**	–0.309*	–0.396**
				[0.047]	[0.152]	[0.158]
Soe ratio				–0.001	–0.037	0.167
				[0.032]	[0.120]	[0.196]
_cons	–0.087	–0.022	–0.179	–0.145	–0.034	–0.373
	[0.093]	[0.088]	[0.140]	[0.123]	[0.432]	[0.563]
Provinces	yes	yes	yes	yes	yes	yes
R^2	0.49	0.53	0.54	0.56	0.35	0.52
N	252	252	252	252	252	252

Note: * significant at 10% level; ** significant at 5% level; *** significant at 1% level; standard errors are reported in parentheses and clustered at province level. “migrant ratio” is the ratio of migrants in city population; “mean wage” is the log of average individual wage; “high school” is the ratio of high school graduates; “college” is the ratio of workers with college education or above; “edu gini” is education Gini coefficient; “Ag ratio” is the share of workers in agriculture sector and “Manuf ratio” is the share of workers in the manufacturing sector; “Soe ratio” is share of workers employed be state-owned firms. The detailed definitions of variables are explained in the text.

Finally, to deal with the potential endogeneity of migration flows, we adopt an instrumental variables (IV) approach by using historical population size as an instrument for contemporaneous migration. The idea behind this approach is that migrants tend to prefer larger cities for better job opportunities and higher income. The historical population of each city in 1953 was to a great extent exogenous to current inequality. Also, the three decades from 1953 to 1982 were the planned era, when population flows were strictly controlled. Interregional migration was not based on market forces because the PRC government moved people to inland and rural areas (Gao, Lu, and Sato 2015). To reduce the possibility that the historical city size may affect inequality through other current variables, we controlled for all the current variables including current population size in every regression. Table 6 shows that our major findings are robust to using different measurements of inequality or alternative IVs. The migrant ratio has a larger coefficient in the IV estimations than the corresponding OLS results in Table 4 and 5. Moreover, we also used the migration pattern in the history as alternative IV, as in Card (2000). We explore migration pattern in 2000 and 1990, and the main results are robust to this.[21] Finally, we considered the endogeneity issues of all other explanatory variables, as discussed in Glaeser, Resseger, and Tobio (2009), and adopted a similar approach by using the same variables in the history as control variables in the regression. Specifically, we use the 1990 census to construct educational composition, educational inequality, migrant ratio, and industry composition. The results still remain.[22]

Table 6: City Inequality and Migration Using Instrument Variables

Sample	All	All	All	All	Male Only	Male Only
	IV = Pop53	IV = Pop53	IV = Pop53	IV = Pop82	IV = Pop53	IV = Pop82
Dep. Var.	(1) Gini	(2) Theil	(3) p90–10	(4) Gini	(5) Gini	(6) Gini
Migrant ratio	0.456***	0.507***	2.129***	0.402***	0.481***	0.397***
	[0.142]	[0.146]	[0.677]	[0.113]	[0.151]	[0.111]
Ln(pop)	0.012*	0.017**	0.033	0.013**	0.015**	0.017***
	[0.006]	[0.008]	[0.032]	[0.006]	[0.007]	[0.005]
Mean wage	–0.122**	–0.117**	–0.602**	–0.106**	–0.123**	–0.097**
	[0.055]	[0.057]	[0.244]	[0.044]	[0.058]	[0.044]
High school	0.472***	0.433**	2.446***	0.436***	0.470***	0.421***
	[0.151]	[0.173]	[0.771]	[0.129]	[0.131]	[0.103]
College	0.040	0.004	0.505*	0.037	0.041	0.040
	[0.054]	[0.063]	[0.285]	[0.049]	[0.055]	[0.047]
Edu gini	1.710***	1.504***	9.269***	1.583***	1.349***	1.183***
	[0.452]	[0.449]	[2.786]	[0.389]	[0.344]	[0.273]
Ag ratio	–0.096	–0.131*	–0.250	–0.080	–0.102	–0.077
	[0.075]	[0.077]	[0.359]	[0.069]	[0.078]	[0.071]
Manuf ratio	–0.206***	–0.239***	–0.985***	–0.185***	–0.219***	–0.184***
	[0.076]	[0.092]	[0.331]	[0.063]	[0.074]	[0.058]
Soe ratio	0.230**	0.231***	1.320**	0.192**	0.197*	0.139*
	[0.096]	[0.084]	[0.545]	[0.082]	[0.100]	[0.082]
_cons	–0.048	–0.186	–0.628	–0.045	0.002	–0.003
	[0.113]	[0.143]	[0.640]	[0.109]	[0.099]	[0.092]
Provinces	yes	yes	yes	yes	yes	yes
First-stage F	16.13	16.13	16.13	9.62	14.81	8.92
R^2	0.32	0.42	0.13	0.41	0.42	0.51
N	232	232	232	232	232	232

Note: * significant at 10% level; ** significant at 5% level; *** significant at 1% level; standard errors are reported in parentheses and clustered at province level. The population in 1982 is used as instrument variable for migration rate in the first six columns and the population in 1953 is used as instrument variable in the last column. "migrant ratio" is the ratio of migrants in city population; "mean wage" is the log of average individual wage; "high school" is the ratio of high school graduates; "college" is the ratio of workers with college education or above; "edu gini" is education Gini coefficient; "Ag ratio" is the share of workers in agriculture sector and "Manuf ratio" is the share of workers in the manufacturing sector; "Soe ratio" is share of workers employed be state-owned firms. The detailed definitions of variables are explained in the text.

[21] We did not use this in the main analysis because we do not have information on migration patterns in earlier years.

[22] Detailed results are available upon request.

5. MORE DISCUSSIONS ON THE ROLE OF MIGRANTS ON WITHIN-CITY INEQUALITY

Migrants could increase within-city inequality through their impact on both between-group inequality and within-group inequality. First, migrants can increase within-group inequality, since migrants usually get paid less than local workers due to *hukou* restrictions and other administrative issues (see Afridi et al. 2009; Huang 2010; Chan and Zhang 1999). Second, migrants are on average less educated or have received low-quality education. More migrants, all else equal, would increase the supply of unskilled workers and thus increase the skill premium, which could further increase inequality. Moreover, the ratio of migrants tends to be higher in bigger cities. This could contribute to the city-size premium in inequality. In the following analysis, we further discuss the mechanisms through which the migrant ratio affects city size-inequality relationship.

5.1 Large Cities Have More Low-income Migrants

To show this, a key feature of the density distributions in Figure 4.A is that the proportions of both poor people and rich people increase with city size. First, we investigate the entire distribution of income by city size. Cities are separated into three groups according to their population size: top quintile, bottom quintile, and those in the middle quintiles.[23] Next, we subtract city mean income from individual income and pool the observations in each group together. Figure 4 shows the kernel density distribution of the demeaned income in each group. This is especially true for the top quintile group. There are significantly more workers with income much lower than the mean income and more workers with income much higher than the mean income as well. Interestingly, when migrant samples are excluded from the distribution analysis in Figure 4.B, the large cities do not have as many poor people as in Figure 4.A. In addition, Figure 5 shows how migrants change income distributions in small and big cities. Obviously, the bigger proportion and magnitude of low-income migrants together drag down median income much more in big cities than in small cities. This confirms that there are more low-income migrants in the large cities. This feature tends to widen within-city inequality more in larger cities. The above patterns also help us understand why the upper percentile gap increases with city size and the lower percentile gap decreases with city size.

5.2 Large Cities Have More Less-educated Migrants

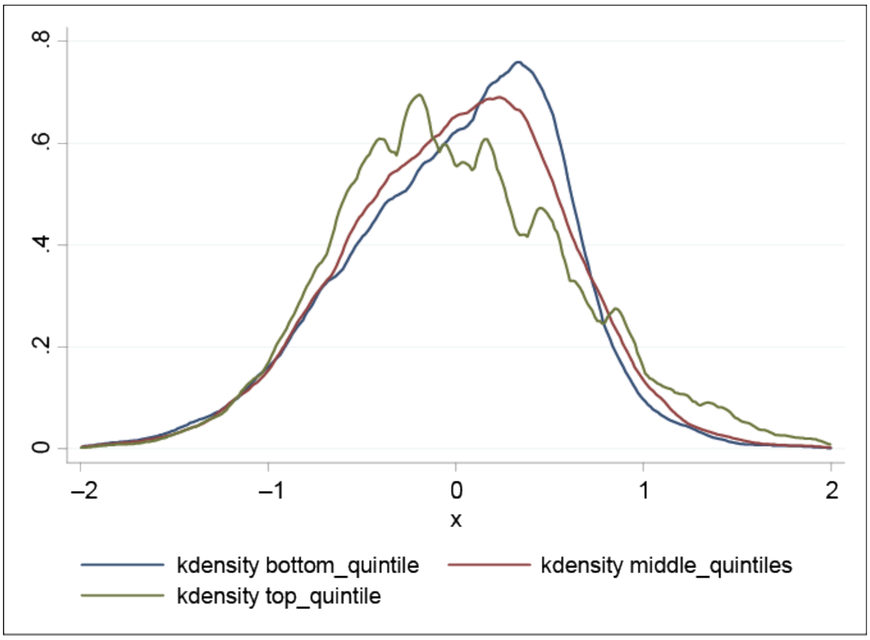
Figure 6 shows the education compositions for each city size group. Cities were first put into 10 size bins equally according to population. The share of workers with secondary education or lower was calculated for each city. Next, we took the average ratio across cities within each bin. Similarly, we calculated the average ratio of workers with some college education for each bin. It can be seen from the red and blue lines in Figure 6, which takes into consideration both local urban residents and migrants, that the fraction of college or more decreases with city size, while the fraction of below high school education increases with city size. This is very different from the picture for the US in Baum-Snow and Pavan (2013), which shows a positive correlation between the fraction of college graduates and city size.

---

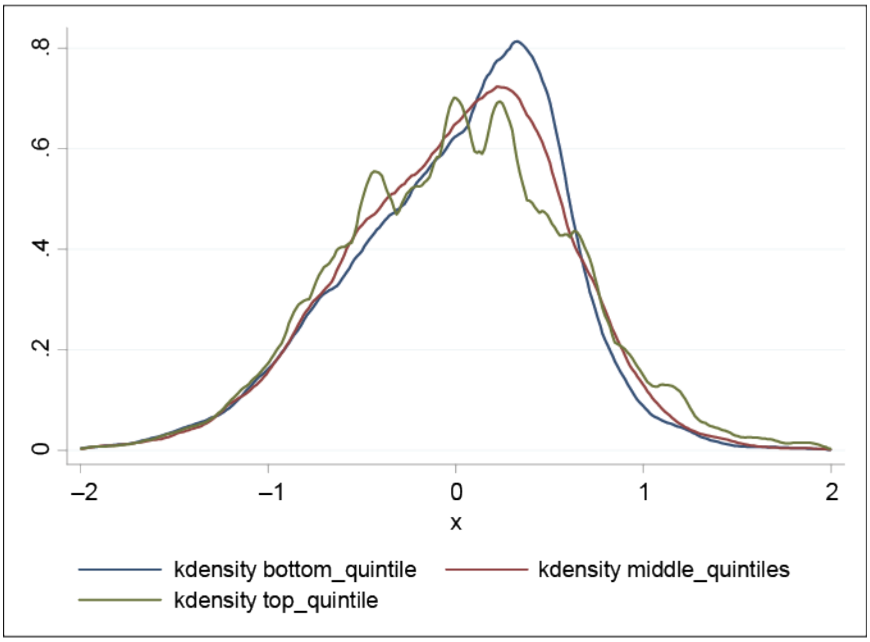
[23] We also tried separating cities into three equally sized bins. The pattern presented below still exists, but it is slightly weaker.

Figure 4: Density of Income Distribution

4.A With Migrants



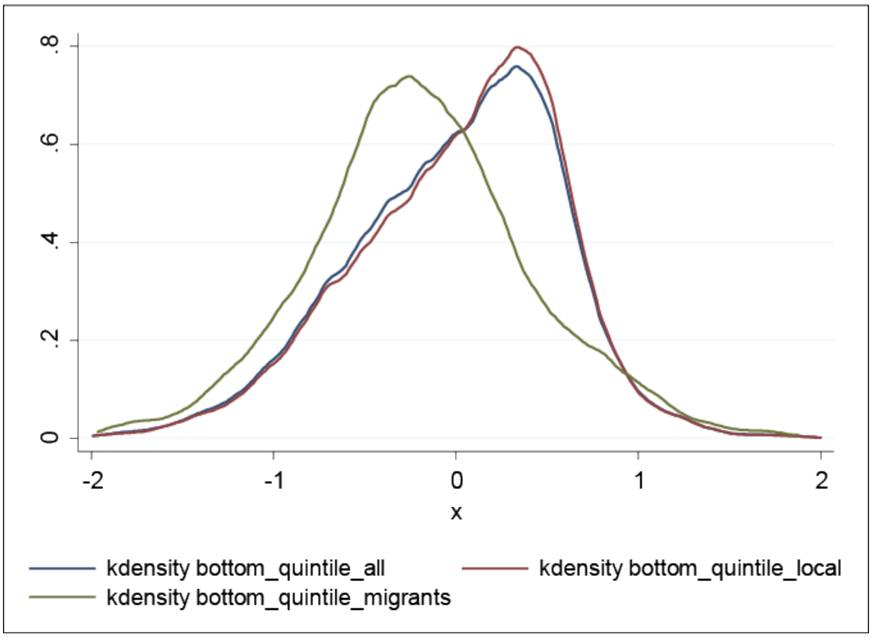
4.B Without Migrants



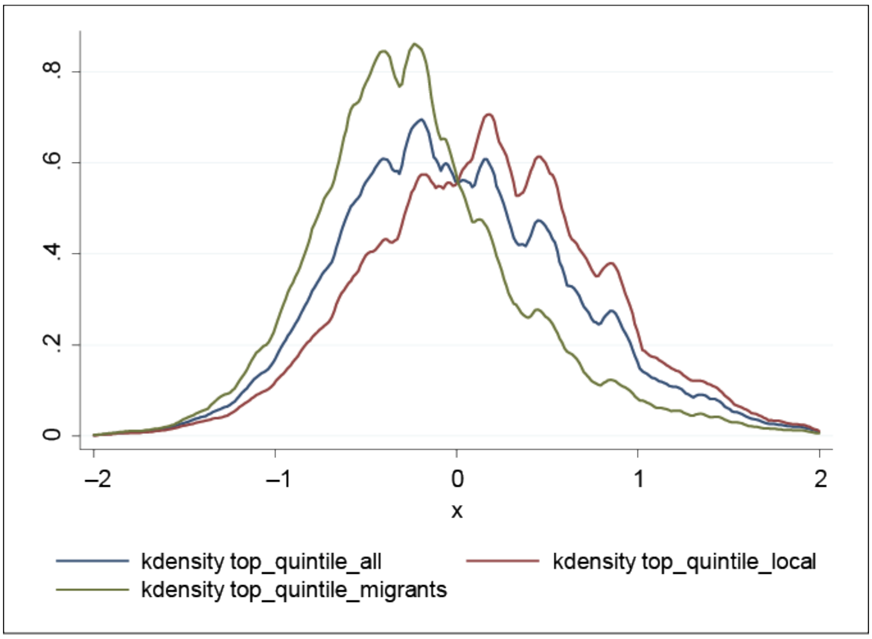
Note: The figures above show the kernel density of income distribution for three groups of cities by size: top quintile, bottom quintile and those in the middle. City mean income is subtracted from individual income used here. Figure 4.A is based on information on both local residents with urban *hukou* and migrants; Figure 4.B is based only on local urban residents.

Figure 5: Densities of Income Distributions for Locals, Migrants and All

5.A Income Distributions for Bottom Quintile Cities



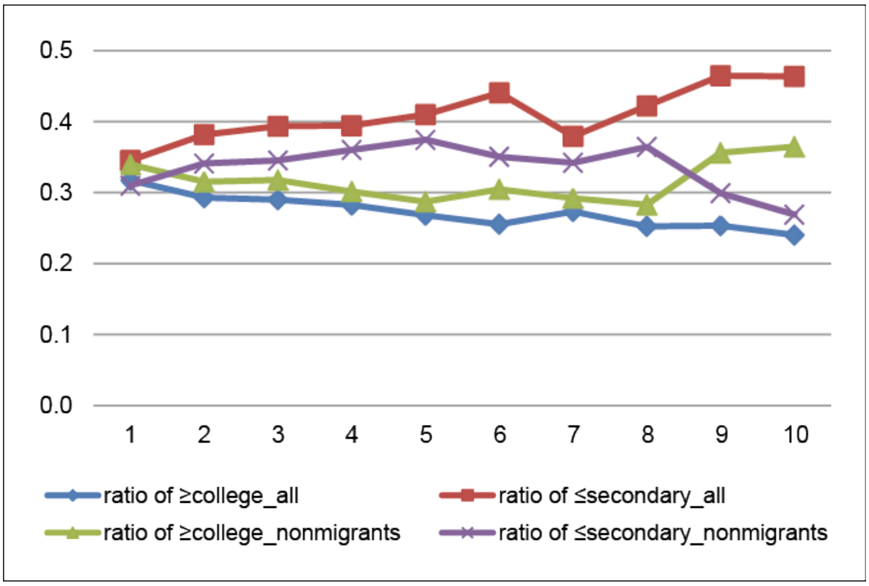
5.B Income Distributions for Top Quintile Cities



Note: The figures above show the kernel density of income distribution for three groups of cities by size: top quintile, bottom quintile and those in the middle. City mean income is subtracted from individual income used here. Figure 5.A is based on observations in bottom quintile cities; Figure 5.B is based on observations in top quintile cities. “all” includes both local residents with urban *hukou* and migrants, “local” includes local residents with urban *hukou*, and “migrants” includes only rural migrants.

What can contribute to this difference in education compositions between the PRC and the US? The answer is rural-urban migration. It has been seen that overall there is a higher fraction of migrants in bigger cities. The purple and green lines in Figure 6 are based on the calculation that excludes migrants from the sample, only considering local urban people. It can be seen that, on average, local urban workers are more educated than migrants for all city bins. In addition, for the largest two size bins, local urban workers are, on average, much better educated. The fraction of college or more increases with city size, while the fraction of secondary education or below decreases with city size for local urban residents. One conclusion that can be drawn from these figures is that migration has increased the relative supply of less-educated workers, especially in big cities.

Figure 6: Education Composition by Size Group



Note: The above figure shows the average education compositions for cities in each size bin. The blue and red lines include both local urban residents and migrants; the green and purple lines exclude migrants from the calculation.

5.3 Less-educated Migrants Have Changed the Education Premium

To further confirm that migration can affect city inequality and the city-size premium through changing education compositions and, consequently, education premiums, we investigated the relationship between migration and skill premiums directly. The results are reported in Table 7.

First, we directly examined how the migrant ratio affects the education premium in column (1) to (6). We estimated the college and high school premium by using observations within each city particularly, and then match them with other city-level variables. The simple OLS results reported in Columns 1 and 4 show that migration alone accounts for more than 50% of the city size premium in college premium, and 32% of that in high school premium. It can be seen that a higher migrant ratio is associated with a higher skill premium when other factors are controlled for, no matter if OLS or IV estimation is used. The coefficient on the migrant ratio is also slightly larger in college premium regressions than that in high school premium regressions. These results are further confirmed by running Mincer regressions and allowing the high school return or the college return to vary with city size and share of migrants, as shown in Table A.3, where we controlled for age, marital status, sex, occupation, industry, and city-fixed effects in all regressions. The above findings are consistent with the evidence on complementarity between skilled native workers and unskilled migrants found in the literature (see Eeckhout et al. 2014; Ottaviano and Peri 2012).

Table 7: Skill Premiums, Percentile Gaps and Migration

	College Premium			**High School Premium**		
Dep. Var.	**(1) OLS**	**(2) OLS**	**(3) IV**	**(4) OLS**	**(5) OLS**	**(6) IV**
Migrant ratio	0.447***	0.636***	1.453***	0.175**	0.523***	0.949***
	[0.048]	[0.160]	[0.397]	[0.074]	[0.097]	[0.336]
Ln(pop)	0.046***	0.044***	0.023	0.043**	0.015	0.004
	[0.014]	[0.013]	[0.023]	[0.018]	[0.013]	[0.015]
Mean wage		–0.208**	–0.453***		–0.029	–0.157
		[0.076]	[0.125]		[0.079]	[0.123]
High school		0.367	0.906*		0.463*	0.744***
		[0.344]	[0.491]		[0.230]	[0.279]
College		0.323	0.364*		0.594**	0.615***
		[0.240]	[0.204]		[0.233]	[0.207]
Edu gini		2.381**	4.315**		2.216***	3.224***
		[0.998]	[1.718]		[0.716]	[1.192]
_cons		0.050	–0.193		–0.236**	–0.362***
		[0.158]	[0.183]		[0.101]	[0.138]
Other controls	no	yes	yes	no	yes	yes
Province FE	yes	yes	yes	yes	yes	yes
First-stage F	–	–	16.13	–	–	16.13
R^2	0.72	0.75	0.70	0.53	0.67	0.65
N	232	232	232	232	232	232

	P90–50		**P50–10**	
Dep. Var.	**(7) OLS**	**(8) IV**	**(9) OLS**	**(10) IV**
Migrant ratio	0.425***	2.148***	0.043	–0.019
	[0.128]	[0.660]	[0.132]	[0.418]
Ln(pop)	0.084***	0.038	–0.006	–0.005
	[0.018]	[0.026]	[0.017]	[0.019]
Mean wage	–0.152**	–0.669***	0.049	0.067
	[0.070]	[0.240]	[0.078]	[0.133]
High school	0.939**	2.076***	0.410	0.369
	[0.410]	[0.630]	[0.248]	[0.283]
College	0.161	0.247	0.262	0.258
	[0.178]	[0.276]	[0.227]	[0.208]
Edu gini	4.251***	8.329***	1.085	0.940
	[1.272]	[2.513]	[0.673]	[1.062]
_cons	0.208	–0.303	0.035	0.053
	[0.187]	[0.355]	[0.190]	[0.214]
Other controls	yes	yes	yes	yes
Province FE	yes	yes	yes	yes
First-stage F	–	16.13	–	16.13
R^2	0.68	0.44	0.62	0.62
N	232	232	232	232

"IV" =instrumental variable; "OLC" = ordinary least squares.

Note: * significant at 10% level; ** significant at 5% level; *** significant at 1% level; standard errors are reported in parntheses and clustered at province level. The population in 1982 is used as instrument variable for iv results "migrant ratio" is the ratio of migrants in city population; the definitions of variables are the same as in the above tables; other controls include industrial compositions and SOE ratio.

Second, we examined how migration is associated with upper and lower percentile gaps differently. The results are shown in Table 7. Columns (7) to (10) show that a higher ratio of migrants is positively associated with the upper percentile gap (p90–50) and insignificantly associated with the lower percentile gap (p50–10). This implies that migrants, most of whom are less educated, are likely to be complementary to high-income workers, while dragging down median income by either adding relatively low-income workers into city population, or bringing competition to median-income workers, or both.

6. CONCLUDING REMARKS

In this paper, we examined the relationship between city size, migration and income inequality using a sub-sample of the 2005 population survey in the PRC. We calculated various measures of city-level income inequality for 252 PRC cities, using samples of migrants and those with urban *hukou*. There are two baseline findings. First, overall income inequality is higher in larger cities. This is mainly due to a higher skill premium and not to residual inequality in larger cities, which is different from the US where the opposite holds. Second, rural-to-urban migration matters for explaining these features of the size-inequality relationship in the PRC.

In addition, we investigated how migration affects city inequality and the city size-inequality premium. We found that migration increased the skill premium by changing the skill composition of city populations. Since migrants are less educated than local urban workers and this difference is even more noteworthy in larger cities, migration decreases the relative supply of skilled workers in bigger cities more than in smaller ones, which leads to higher skill premium and thus a higher between-group inequality in bigger cities. However, it should be noticed that this is only one of the potential channels through which migration can affect the city-size premium. Our discussion here does not exclude other channels, such as within-education inequality due to wage discrimination on migrants.

This paper contributes to the literature in the following respects. First, this is a first attempt to investigate the issue of within-city inequality and city development in a developing country like the PRC. The PRC's rapid urbanization process and its ongoing change to the urban system led us to look into the effects of migration on local inequality. Second, we further examine the potential channels through which migration affects city inequality. This provides implications for policy issues: large cities should not restrict the inflow of migrants as they do, because migrants can be complements to skilled workers and raise skill premium, which benefits both the local residents and migrants. The higher inequality associated with more migrants in the urban area is mainly due to a natural change in skill compositions. The correct policy is to equalize public service access for low-income workers to reduce welfare inequality. In this way, the negative effects of income inequality in large cities could be mitigated. However, if relatively poor migrants cannot get equal access to public services, the urban public service system serves to widen welfare inequality, rather than narrow welfare inequality as it should do.

REFERENCES

Afridi, F., S. X. Li, and Y. Ren. 2010. Social Identity and Inequality: The Impact of China’s Hukou System. IZA DP No. 6417.

Au, C-C., and J. V. Henderson. 2006a. Are Chinese Cities Too Small? Review of Economic Studies 73(3): 549–576.

———. 2006b. How Migration Restrictions Limit Agglomeration and Productivity in China. Journal of Development Economics 80: 350–388.

Autor, D., F. Katz, and M. Kearney. 2008. Trends in U.S. Wage Inequality: Revising the Revisionists. Review of Economics and Statistics: 300–323.

Baum-Snow, N., and R. Pavan. 2013. Inequality and City Size. The Review of Economics and Statistics 95(5): 1535–1548.

Bernard, A. B., and B. J. Jensen. 1999. Exceptional Exporter Performance: Cause, Effect, or Both? Journal of International Economics 47(1): 1–25.

Cai, H., Y. Chen, and L. Zhou. 2010. Income and Consumption Inequality in Urban China: 1992–2003. Economic Development and Cultural Change 58: 385–413.

Card, D. 2009. Immigration and Inequality. American Economic Review 99(2): 1–21.

Card, D., and J. DiNardo. 2000. Do Immigrant Inflows Lead to Native Outflows? The American Economic Review Papers and Proceedings 90(2): 360–367.

Chan, K. W. 2009. The Chinese Hukou System at 50. Eurasian Geography and Economics 50(2): 197–221.

Chan, K. W., and L. Zhang. 1999. The Hukou System and Rural-Urban Migration: Processes and Changes, The China Quarterly 160(1): 818–855.

Chan, K. W., and W. Buckingham. 2008. Is China Abolishing the Hukou system? The China Quarterly 195(1): 582–605.

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio. 2016. What is Different about Urbanization in Rich and Poor Counties? Cities in Brazil, China, India and the United States, Journal of Urban Economics: 1–33.

Chen, Z., and M. Lu. 2015. Urban Systems and Urban Development in the People’s Republic of China, Asian Development Bank Institute (ADBI) Working Paper 552. Asian Development Bank Institute, http://www.adb.org/publications/urban-systems-and-urban-development-prc.

Combes, P-P., G. Duranton, and L. Gobillon. 2008. Spatial Wage Disparities: Sorting Matters! Journal of Urban Economics 63(2): 723–742.

Davis, D., and J. Dingel. 2012. A Spatial Economy of Knowledge. National Bureau of Economic Research (NBER) Working Paper 18188.

Dustmann, C., T. Frattini, and I. Preston. 2013. The Effect of Immigration along the Distribution of Wages. Review of Economic Studies 80: 145–173.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny. 2014. Spatial Sorting. Journal of Political Economy 122(3): 554–620.

Gao, H., M. Lu, and H. Sato. 2015. Inclusive Urban Employment: How Does City Scale Affect Job Opportunities for Different People? Asian Economic Papers 14(2): 98–128.

Glaeser, E. L., M. Resseger, and K. Tobio. 2009. Inequality in Cities. Journal of Regional Science 49(4): 617–646.

Goldberg, P. K., and N. Pavcnik. 2007. Distributional Effects of Globalization in Developing Countries. Journal of Economic Literature 45(1): 39–82.

Gustafsson, B., S. Li, and T. Sicular. 2008. Inequality and Public Policy in China. Cambridge: Cambridge University Press.

Han, J., R. Liu, and J. Zhang. 2012. Globalization and Wage Inequality: Evidence from Urban China. Journal of International Economics 87(2): 288–297.

Helpman, E., O. Itskhoki, and S. Redding. 2010. Inequality and Unemployment in a Global Economy. Econometrica 78(4): 1239–1283.

Henderson, J. V. 2007. Urbanization in China: Policy Issues and Options, Reports for the China Economic Research and Advisory Program, download from: http://www.econ.brown.edu/faculty/henderson/finalfinalreport-2007050221.pdf.

Huang, Y. 2010. Urbanization, Hukou System and Government Land Ownership: Effects on Rural Migrant Works and on Rural and Urban Hukou Residents, Background Paper for the Perspectives on Global Development 2010: Wealth.

Jiang, S., M. Lu, and H. Sato. 2012. Identity, Inequality, and Happiness: Evidence from Urban China. World Development 40(6): 1190–1200.

Lemieux, T., 2006. Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? American Economic Review 96(3):461–498.

Lin, T., J. Zhuang, D. Yarcia, and L. Fen. 2010. Income Inequality in the People's Republic of China 1990–2005, In J. Zhuang (ed.), Poverty, Inequality, and Inclusive Growth in Asia: Measurement, Policy Issues, and Country Studies. London: Anthem Press.

Liu, Z. Q. 2005. Institution and Inequality: The Hukou System in China. Journal of Comparative Economics 33(1): 133–157.

Long, J. E., D. W. Rasmussen, and C. T. Haworth. 1977. Income Inequality and City Size. The Review of Economics and Statistics 59(2): 244–246.

Lu, M. 2013. The Power of Space: Geography, Politics and Urban Development, (in Chinese), Shanghai People's Press.

Lu, M., and G. Wan. 2014. Urbanization and Urban System in China: Research Findings and Policy Recommendations, Journal of Economic Survey 28(4): 671–685.

Lu, M., and P. Ni. 2014. Education and Entrepreneurship in China, working paper.

Lu, M., Z. Chen, Y. Wang, Y. Zhang, Y. Zhang, and C. Luo. 2013. China's Economic Development: Institutions, Growth and Imbalances, Edward Elgar.

Lu, M., Z. Chen, Z. Zhao, X. Zhu, and X. Xu. 2013. China's Regional Development: Review and Prospect. Routledge, Taylor & Francis Group.

Meng, X., and D. Zhang. 2010. Labour Market Impact of Large Scale Internal Migration on Chinese Urban 'Native' Workers. IZA Discussion Paper 5288.

National Bureau of Statistics. 2015. The Statistical Report of National Economic and Social Development of the People's Republic of China (2014), (in Chinese), downloadable at http://www.stats.gov.cn/tjsj/zxfb/201502/t20150226_685799.html

Park, A., X. Song, J. Zhang, and Y. Zhao. 2006. Returns to Skill, Labor Market Transition, and The Rise of Wage Inequality in Urban China, 1988 to 2003. Working Paper.

Ottaviano, G. I. P., and G. Peri. 2008. Immigration and National Wages: Clarifying the Theory and the Empirics, NBER Working Paper No. 14188.

———. 2012. Rethinking the Effect of Immigration on Wages. Journal of the European Economic Association 10: 152–197.

Ravallion, M., and S. Chen. 2007. China's (Uneven) Progress against Poverty. Journal of Development Economics 82: 1–42.

United Nations, Department of Economic and Social Affairs, Population Division, 2012, World Urbanization Prospects: The 2011 Revision, CD-ROM Edition.

Wan, G., M. Lu, and Z. Chen. 2007. Globalization and Regional Income Inequality in China, Empirical Evidence from within China. Review of Income and Wealth 53(3): 35–59.

Wan, G., J. Ye, and J. Zhuang. 2012. On Regional and Inter-Household Inequality in China: The Threat of Inequality and the Move to Inclusive Growth, in Wing Thye Woo, Ming Lu, Jeffrey D. Sachs and Zhao Chen (eds.), A New Economic Growth Engine for China: Escaping the Middle-Income Trap by Not Doing More of the Same, Imperial College Press, and World Scientific: 155–182.

Wang, X. 2010. Urbanization Path and City Scale in China: An Economic Analysis, (in Chinese), Economic Research Journal No. 10:20–32.

World Bank. 2009. World Development Report 2009: Reshaping Economic Geography.

Xing, C., and S. Li. 2012. Residual Wage Inequality in Urban China, 1995–2007, China Economic Review 23(2): 205–222.

Yang, D. T. 2005. Determinants of Schooling Returns During Transition: Evidence from Chinese Cities. Journal of Comparative Economics 33: 244–264.

Zhang, W., X. Li, and H. Cui, 2005. China's Inter-census Survey in 2005. NBS, People's Republic of China.

Zheng, S., R. B. Peiser, and W. Zhang. 2009. The Rise of External Economies in Beijing: Evidence from Intra-urban Wage Variation. Regional Science and Urban Economics 49: 449–459.

APPENDIX

Table A1: Summary Statistics for Some Composition Variables

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
College ratio	252	0.27	0.08	0.04	0.47
High school ratio	252	0.32	0.06	0.15	0.46
Migrant ratio	252	0.16	0.18	0.01	0.97
Share of state-owned firms	252	0.54	0.16	0.05	0.87
Manufacturing ratio	252	0.22	0.13	0.02	0.83
Gini	252	0.32	0.04	0.23	0.48
Theil	252	0.18	0.06	0.09	0.52
College premium	252	0.56	0.18	0.01	1.26
High school premium	252	0.46	0.13	0.09	0.89

"Obs" = observations; "Std. Dev." = standard deviation.

Notes: The "college ratio" is the share of individuals with some college education in the considered sample. The "high school ratio" is the share of individuals with some high-school education but no college education. The "migrant ratio" is the share of migrants in city population.

Table A2: Inequality and Other Variables

	Gini	Gini	Gini	Gini	Gini	Gini
Ln(pop)	0.024***	0.022***	0.018***	0.018***	0.014**	0.013***
	[0.004]	[0.004]	[0.004]	[0.005]	[0.006]	[0.003]
Mean wage		0.021				
		[0.024]				
High school			0.022			
			[0.044]			
College			–0.106**			
			[0.040]			
Edu gini			0.722***			
			[0.132]			
Ag ratio				–0.043		
				[0.034]		
Manuf ratio				0.050		
				[0.036]		
One-digit industries					yes	
Soe ratio						–0.111***
						[0.027]
_cons	0.137***	0.093	0.025	0.114*	0.258	0.205***
	[0.028]	[0.059]	[0.068]	[0.059]	[0.214]	[0.056]
R^2	0.20	0.21	0.40	0.24	0.36	0.38
N	252	252	252	252	252	252

Notes: * significant at 10% level; ** significant at 5% level; *** significant at 1% level; "Ln(pop)" is the log of city population; "mean wage" is the log of average individual wage; "high school" is the ratio of high school graduates; "college" is the ratio of workers with college education or above; "edu gini" is education Gini coefficient; "Ag ratio" is the share of workers in agriculture sector and "Manuf ratio" is the share of workers in the manufacturing sector; "Soe ratio" is share of workers employed be state-owned firms. The detailed definitions of variables are explained in the text.

Table A3: Skill Premium, City Size, and Migration from Mincer Regressions

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) OLS	(6) OLS	(7) OLS	(8) IV
High school	0.159***	–0.099***	0.015	0.085***				
	[0.002]	[0.014]	[0.015]	[0.019]				
High *ln(pop)		0.029***	0.009***	–0.004				
		[0.002]	[0.002]	[0.003]				
High *migration			0.203***	0.334***				
			[0.010]	[0.021]				
College					0.302***	–0.250***	–0.029*	0.038
					[0.003]	[0.015]	[0.017]	[0.023]
College * ln(pop)						0.063***	0.026***	0.014***
						[0.002]	[0.002]	[0.003]
College*migration							0.385***	0.514***
							[0.013]	[0.029]
_cons	3.101***	3.105***	3.127***	3.136***	3.233***	3.164***	3.178***	3.189***
	[0.347]	[0.347]	[0.347]	[0.345]	[0.284]	[0.283]	[0.282]	[0.282]
Other controls	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.35	0.35	0.35	0.35	0.46	0.47	0.47	0.48
N	212,685	212,685	212,685	204,592	174,093	174,093	174,093	165,995

“IV” =instrumental variable; “OLC” = ordinary least squares.

Note: * significant at 10% level; ** significant at 5% level; *** significant at 1% level; Standard errors are clustered at city level; The dependent variable is the log of hourly wage; columns 1-4 only use observations with junior high or high school education; columns 5-8 only use observations with high school or college education; “high school” and “college” are dummy variables indicating whether the individual has high school or college education; ln(pop) is the log of city population; “migration” is the share of migrants in city population; all regressions control for sex, age dummies, marital status, industry, occupation, city fixed effects; In columns 4 and 8, the population in 1953 is used as IV for migration.