

# Productivity Gains from Agglomeration and Migration in the People's Republic of China between 2002 and 2013

PIERRE-PHILIPPE COMBES, SYLVIE DÉMURGER, AND SHI LI\*

---

We evaluate the evolution of productivity gains in cities in the People's Republic of China between 2002 and 2013. In 2002, rural migrants exerted a strong positive externality on the earnings of urban residents, which were also higher on average in cities with access to foreign markets through a seaport. In 2007 and 2013, city size (measured in terms of both employment density and land area) was the crucial determinant of productivity. Market access, whether internal or external, played no direct role. Rural migrants still enhanced urban residents' earnings in 2007 and 2013, though the effect was less than half that in 2002. Urban gains and their evolution over time are very similar on a total and a per hour earnings basis. Finally, skilled workers and females experienced slightly larger gains than unskilled workers and males.

*Keywords:* agglomeration economies, migration, cities, urban development, wage disparities

*JEL codes:* J31, O18, O53, R12, R23

---

## I. Introduction

Although the empirical literature on productivity gains associated with larger cities has matured over the past decade (Combes and Gobillon 2015), evidence about the characteristics and role of cities in developing economies is still largely incomplete (Chauvin et al. 2017). This stands in sharp contrast with urbanization rates that are much lower but increasing rapidly in many developing economies compared with developed countries where urbanization is fairly stable. Therefore, quantifying gains and costs from such a rapidly evolving spatial

---

\*Pierre-Philippe Combes: CNRS Research Professor at Université de Lyon, CNRS, GATE L-SE UMR 5824; Professor at Sciences Po; and Research Fellow at the CEPR. E-mail: ppcombes@gmail.com; Sylvie Démurger (corresponding author): CNRS Research Professor at Université de Lyon, CNRS, GATE L-SE UMR 5824; and Research Fellow at IZA. E-mail: demurger@gate.cnrs.fr; Shi Li: Professor, Business School, Beijing Normal University and Research Fellow at IZA. E-mail: lishi@bnu.edu.cn. This paper was part of a research program on Amenities and Agglomeration in China (NSFC Grant No. 71503023). We would like to thank the participants at the Asian Development Review Conference on Urban and Regional Development in Asia held in Seoul in July 2016, the managing editor, and an anonymous referee for helpful comments and suggestions. We are also grateful to Zhu Mengbing for excellent research assistance. In this paper, the Asian Development Bank recognizes "China" as the People's Republic of China. The usual disclaimer applies.

concentration of economic activity is a highly topical issue. If one expects the gains to outweigh the costs, then urbanization should be promoted, for instance, by facilitating rural-to-urban migration. If costs dominate, appropriate policy must be implemented in order to refrain urban expansion.

The People's Republic of China (PRC) offers the best example of a large increase in rural–urban migration, having undergone such a process that has fed the development of cities for more than 15 years. In an early contribution, Au and Henderson (2006) document the productivity gains that could emerge from larger cities in the PRC. They conclude that, apart from the largest cities, large productivity gains might also emerge from a larger number of medium-sized cities. Using microdata from the mid-2000s, Combes, Démurger, and Li (2015) estimate the specific externality that migrants exert on native urban workers—beyond the impact of standard agglomeration variables such as total employment density and city land area—and find evidence of a large positive effect of the city's migrant share on urban residents' wages. This paper complements the analysis over time by evaluating the evolution of productivity gains in cities in the PRC between 2002 and 2013. Measuring possible changes has important policy implications because the success of urban policies ultimately depends on the nature of the gains, which may have changed over time. Another contribution of this paper consists of estimating urban gains not only in terms of total earnings but also earnings per hour, which is a closer measure of labor productivity, as well as for both skilled and unskilled workers and by gender. In doing so, we combine nationally representative household income surveys conducted in 2002, 2007, and 2013 with city-level data.

We find that city characteristics have a significant impact on labor productivity and that the characteristics that matter have changed over time. In 2002, the migrant externality was very strong. To a lesser extent, access to foreign markets through proximity to a seaport also played a positive role. In 2007 and, to a greater extent, in 2013, agglomeration variables that are standard in the literature and mostly reflect city size (e.g., employment density and land area) had become crucial determinants of a city's productivity. By contrast, market access, whether internal or external, played no significant direct role. As for rural migrants, they still enhanced urban residents' earnings, but the effect was less than half what it had been in 2002. These findings show that the PRC moved from urban gains shaped by specific features (e.g., large disparities across cities in terms of access to the sea and migrant shares) in 2002 to a situation in 2007 and 2013 where variables more typically affecting urban workers' earnings are playing a more influential role. What matters for cities in the PRC now is their capacity to host migrants as well as their overall size in terms of both density and land area. Urban gains and their evolution over time are very similar on a total and per hour earnings basis. Surprisingly, we find little difference in earnings gains between skilled and unskilled workers, yet the city effects are slightly larger for skilled workers. This may be due to the fact that the substitution and externality effects from migrants are difficult to disentangle from

one another with the data sets we use. Last, female workers seem to gain slightly more from cities than male workers do.

Section II briefly recalls the theoretical background and the associated empirical strategy, and presents the data we use. Section III is devoted to measuring the impact of city characteristics for all workers taken together on both a total and per hour earnings basis. Section IV evaluates agglomeration and migration gains separately for skilled and unskilled workers, and for male and female workers. Section V concludes.

## II. Mechanisms, Empirical Specification, and Data

### A. Background and Methodology

Marshall (1890) suggested long ago that gains from cities arise from various local external economies of scale. The underlying mechanisms, which are now well understood, were classified by Duranton and Puga (2004) as sharing (specialized inputs, diversity, or risk); matching (quantity and quality of matches in local labor markets); and learning (creating, diffusing, and accumulating knowledge). These mechanisms either directly impact household utility or increase their productivity, which in turn leads to higher earnings. The largest part of the empirical literature augments standard wage equations with city variables to evaluate the determinants of the nominal gains that result from larger cities. Similar conclusions are usually reached when one works on total factor productivity measures. Comparing nominal gains to the higher cost of living in cities, and therefore assessing real income gains in cities, is the focus of another strand of the literature.<sup>1</sup>

We use a standard two-step procedure for the estimation of agglomeration effects discussed in Combes and Gobillon (2015) and extended by Combes, Démurger, and Li (2015) to properly account for the role of migrants. The procedure consists in estimating the following specification:

$$\log w_{it} = X_{it}\theta_t + L_{cst}\gamma_t + \delta_{ct} + \varepsilon_{it} \quad (1)$$

$$\delta_{ct} = U_{ct}\beta_t + \alpha_t + v_{ct} \quad (2)$$

The first-step estimation of equation (1) evaluates the impact on individual  $i$ 's wage at date  $t$ ,  $w_{it}$ , of city-time fixed effects,  $\delta_{ct}$  for city  $c$  where worker  $i$  is employed at date  $t$  and of city  $c$ 's characteristics for sector  $s$  where  $i$  is employed,  $L_{cst}$ , net of the role of individual characteristics  $X_{it}$ . We estimate equation (1) for each year separately, which makes all estimated parameters year specific. Controlling for individual characteristics is crucial to remove the bias arising from a

<sup>1</sup>See, for instance, Moretti (2013) and Combes, Duranton, and Gobillon (2016).

possible nonrandom sorting of individuals across cities depending on their abilities. The sorting of individuals has been documented to be large in some economies, although this does not seem to be the case in the PRC as shown by Combes, Démurger, and Li (2015). Introducing city-sector variables, typically here only the logarithm of the share of sector  $s$  in city  $c$  employment, allows identifying the role of the characteristics of the sector within the city beyond the role of other nonsector-specific city characteristics captured by  $\delta_{ct}$ .<sup>2</sup>

The second step disentangles the characteristics of the city (vector  $U_{ct}$ ) that impact city productivity  $\delta_{ct}$ . We pool the 3 years of data together and compare estimates when some (or all) estimated parameters are year specific. The two main variables of interest are employment per square kilometer (density) and the share of internal migrants. The literature has shown the former to be the main driver of urban productivity. The role of the latter has been highlighted in the case of the PRC by Combes, Démurger, and Li (2015). These authors discuss the functional forms to be used in order to properly interpret the estimated coefficients. The logarithm of the employment density of urban residents identifies the standard impact of density that can be compared to the one obtained in the literature for other economies. Then, the impact of the logarithm of the inverse of 1 minus the share of migrants in the city's total employment corresponds to the sum of the impact of density—for a given number of urban residents, migrant inflows increase density—and the separate migrant effect net of such density gains. The latter effect results from a complementarity (if positive) or substitution (if negative) effect between urban residents and migrants in the production function and a possible further migrant externality (either positive or negative). Unfortunately, these last two effects cannot be separately identified without appealing to extended data and a more complex approach, as surveyed by Lewis and Peri (2015).

We augment the second-step specification by a number of city characteristics that can shape productivity in cities in the PRC. Two margins shape city size: density and land area. Density corresponds to the role of the intensive margin: that is, a larger number of workers per square kilometer within the city boundary. Land area captures the role of the extensive margin, the city being made larger by extending its boundaries at a given density: that is, under a simultaneous proportional increase in total employment.

Since trade takes place between locations and workers are mobile, the literature has also emphasized the role of the access to distant markets that can generate imported external economies in cities. We use two variables to disentangle the roles of internal and international markets. The first variable is a within-PRC

---

<sup>2</sup> $L_{cst}$  is centered with respect to its city mean,  $L_{c,t}$ , so that it does not capture the role of the overall distribution of city-sector shares in the city, a city characteristic we want to be captured by  $\delta_{ct}$ . Conversely, individual characteristics are not centered with respect to their city mean because the possible correlation between city and individual characteristics is interpreted as a pure composition effect, which does not arise from agglomeration externalities and thus needs to be controlled for so that  $\delta_{ct}$  encompasses city externalities only.

market potential based on Harris (1954), which is the sum of the employment density in other cities in the PRC divided by the distance to the city being considered. It assesses how far the city is from other large cities. As a large share of the PRC's exports is coursed through coastal ports, the second variable corresponds to the city's distance to the closest seaport.<sup>3</sup>

Lastly, and even if this variable has never proved to be very influential, we also consider the role of the city's industrial diversity measured by the inverse of a Herfindhal index computed on the shares of each industry in the city's total employment.

Although we pool together different years, the impact of local variables is mostly identified over the cross section of cities because time variations are much smaller than across cities and also because few cities are sampled for the 3 years that we consider. The literature emphasizes two types of possible biases. The larger one is due to the spatial distribution of workers, which could depend on unobserved skills that largely affect their productivity, and does not necessarily result from city externalities. This can make the estimated impact of local variables twice as large as it would be for the same individual who would locate in different cities in different periods. Since we do not have access to an individual panel where workers are followed across years, we can only control for observable individual skills. For example, a large spatial sorting according to skills has been observed in France and the United States. In the case of France, Combes, Duranton, and Gobillon (2008) show that controlling for observable skills is not strong enough to remove the bias. By contrast, no sorting according to unobserved heterogeneity seems to be present in the United States according to Baum-Snow and Pavan (2013). As for the PRC, Combes, Démurger, and Li (2015) show that no spatial sorting occurs according to observable skills. Therefore, only a large sorting of unobserved skills, which should not be correlated to observed skills, could significantly bias the estimates. We see this as unrealistic, at least until many years of intensive migration and possibly unequal access to higher education have further shaped the spatial distribution of individual skills.

The second possible bias arises from reverse causality. Local characteristics could be driven by local productivity, and not the reverse, if the location choices of firms and workers are endogenous. The local density variable is the most suspicious one under this perspective. This is why we use the specification proposed by Combes, Démurger, and Li (2015), which decomposes total employment density into the impacts of urban employees and the share of migrants. The urban residents' density, which is shown to have the same marginal impact on local productivity as total density, should be much less prone to reverse causality as urban residents settled in the city prior to the review period.

<sup>3</sup>The 12 major coastal seaports in the PRC in terms of the volume of freight handled are, in decreasing order, Shanghai, Ningbo, Guangzhou, Tianjin, Qingdao, Qinhuangdao, Dalian, Rizhao, Yingkou, Yantai, Lianyungang, and Zhanjiang.

By contrast, the migration variable is likely to be affected by reverse causality. Migrants choose locations where employment conditions and earnings, in particular, are favorable. By appealing to specifications considering different sets of control variables and by using various instrumental variable strategies, Combes, Démurger, and Li (2015) find evidence of biases in the PRC very similar in magnitude to those usually documented for other economies. At around 20% at most, this is not very large. Addressing reverse causality is beyond the scope of the present paper, though it deserves further investigation. As emphasized by Lewis and Peri (2015), this would require more sophisticated strategies, which might be difficult to implement with the data available.

Other local variables could be endogenous too, which is why we present estimates with and without controlling for them. Again, the literature that has made some attempts to instrument these variables never reported any large bias. Yet, the problem could be more severe here for the land area variable because city boundaries are regularly reshaped by the authorities to match local demographic evolution. Again, the causality would go the other way round. We have checked that similar magnitudes were obtained when we use a land area definition that is constant over time (using the 2007 value) but this again deserves further investigation provided that more comprehensive data are available.

## B. Data

The data used in this paper come from various sources. Equation (1) is estimated using three different nationally representative household income surveys for the years 2002, 2007, and 2013. Equation (2) requires city-level data compiled from the China City Statistical Yearbook edited by the National Bureau of Statistics (NBS). For the city-level share of migrants, we use the 2000 population census, the 1% 2005 population survey, and the 2010 population census for the years 2002, 2007, and 2013, respectively.<sup>4</sup>

Individual databases for 2002 and 2013 are part of the China Household Income Project (CHIP), while the 2007 data set is a subsample of the NBS Urban Household Survey.<sup>5</sup> The CHIP survey is closely related to the NBS household survey, especially on the income section, which makes all 3 years fairly comparable. Yet, because working time is documented only in the CHIP 2002 and CHIP 2013 data, we compute hourly wages for these 2 years only. Furthermore, the surveyed prefecture-level cities vary substantially across all 3 years. The 2002 survey covers 53 cities in 12 provinces, the 2007 survey covers 87 cities in 16 provinces, and the 2013 survey covers 99 cities in 14 provinces. There are only 20 cities common to all

<sup>4</sup>See Combes, Démurger, and Li (2015) for a discussion of city-level data sources.

<sup>5</sup>For 2002, a detailed description of the survey can be found in Li et al. (2008). The NBS 2007 data set is described in Combes, Démurger, and Li (2015).

three surveys, while there are 38 cities common to the CHIP 2002 and CHIP 2013 surveys.

A major feature of the three individual databases is that they cover registered urban residents only (the urban *hukou* holders).<sup>6</sup> This means that an important segment of the urban labor market, comprising migrants not officially registered in cities, is excluded. Representing about 17% of the PRC's 1.3 billion people, migrants constitute an important social grouping, yet they are clearly marginalized (Cai, Park, and Zhao, 2008; Démurger et al. 2009). The exclusion of migrants whose *hukou* is not in the same location as their place of employment has an important implication for our estimates: our data set refers to urban residents (natives) only and one cannot infer that earnings determinants across the cities under review would equally apply to migrant workers. Nevertheless, we provide an estimate of the impact that migrants exert on local urban workers simultaneously with agglomeration effects.

All explanatory variables are defined consistently across all 3 years for both estimation steps. The only exception is for enterprise ownership in equation (1), which is more detailed in 2002 and 2013 in comparison with 2007.

### III. Agglomeration Gains in 2002, 2007, and 2013

#### A. Total Earnings

Table 1 displays the results obtained for all workers' earnings for the estimation of equation (2). Rows with the variable's name without vintage (e.g., "Density") correspond with the effect common to all years. This is also the total effect for year 2013. The effects for 2007 or 2002 are obtained by adding to this number the coefficient reported in the corresponding line (e.g., "Density 2007" or "Density 2002"). The significance test denotes whether the effect for 2007 (or 2002) is significantly different from the 2013 effect. The first five columns provide estimates based on total earnings and the last two columns give estimates based on earnings per hour. As indicated above, the number of hours worked is not available in the microdata for the year 2007, which implies that these per hour earnings estimates are based on 2002 and 2013 only.

In column (1), density only is introduced in the specification. We obtain an elasticity of total earnings with respect to density equal to 0.097 for 2013, as reported in the first line of the table. The value estimated for 2007 is not significantly different from the one for 2013. The elasticity for these 2 years is close to the values reported by Combes, Démurger, and Li (2015) and is almost three times larger

<sup>6</sup>The *hukou* (household registration) system is a distinctive institutional feature of the PRC that divides the population based on occupation (agricultural or nonagricultural) and place of residence. By entitling access to social benefits to local *hukou* holders only, the system has limited population mobility for decades. Most (rural) migrants still hold their hometown *hukou*, which prevents them from permanently settling in cities.

Table 1. City Determinants of Total and Per Hour Earnings

	Total Earnings					Hourly Earnings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.097*** (0.025)	0.059*** (0.022)	0.088*** (0.022)	0.092*** (0.024)	0.096*** (0.025)	0.070*** (0.025)	0.086*** (0.027)
Density 2007	-0.005 (0.033)	-0.017 (0.028)	-0.030 (0.026)	-0.028 (0.030)	-0.030 (0.034)		
Density 2002	-0.069 (0.043)	-0.067* (0.035)	-0.053 (0.033)	-0.082* (0.041)	-0.091** (0.043)	-0.044 (0.036)	-0.085* (0.047)
Migrants		0.267*** (0.061)	0.210*** (0.060)	0.205*** (0.062)	0.207*** (0.062)	0.194*** (0.066)	0.185*** (0.067)
Migrants 2007		0.106 (0.075)	0.096 (0.071)	0.092 (0.076)	0.086 (0.077)		
Migrants 2002		0.491*** (0.120)	0.390*** (0.114)	0.466*** (0.131)	0.412*** (0.133)	0.393*** (0.127)	0.412*** (0.145)
Land area			0.080*** (0.017)	0.089*** (0.027)	0.092*** (0.028)	0.072*** (0.023)	0.097*** (0.030)
Land area 2007				0.008 (0.039)	0.011 (0.041)		
Land area 2002				-0.046 (0.043)	-0.059 (0.044)		-0.058 (0.048)
Market potential			0.093* (0.055)	0.088 (0.055)	0.099 (0.084)	0.057 (0.075)	0.070 (0.091)
Market potential 2007					0.048 (0.125)		
Market potential 2002					-0.137 (0.148)		-0.070 (0.162)
Distance to seaport			-0.020** (0.009)	-0.019** (0.009)	-0.009 (0.014)	-0.029** (0.012)	-0.010 (0.015)
Distance to seaport 2007					0.002 (0.019)		
Distance to seaport 2002					-0.052** (0.023)		-0.056** (0.026)
Diversity			-0.109** (0.048)	-0.115** (0.048)	-0.111** (0.056)	-0.097* (0.056)	-0.095 (0.061)
Diversity 2007					0.006 (0.145)		
Diversity 2002					-0.104 (0.146)		-0.112 (0.160)
Observations	239	239	239	239	239	152	152
R <sup>2</sup>	0.67	0.80	0.82	0.83	0.83	0.84	0.85

Notes: The panel of cities is not balanced (99 cities for 2013, 87 cities for 2007, and 53 cities for 2002.) Time dummies are included in all specifications. Standard errors in brackets. \*\*\* = p < 0.01, \*\* = p < 0.05, \* = p < 0.10. Source: Authors' calculations.

than usual estimates for developed economies. Large density economies seem to prevail in the PRC, at least from the mid-2000s onward. By contrast, the effect appears much lower in 2002, even though the estimated year-specific effect is not significantly different from zero. We will return to this issue once other variables have been introduced.



Column (2) adds the role of migrants, which is found to be largely positive albeit to a decreasing degree between 2002 and 2013. In this augmented specification, the effect of density is lower. This can be due to the fact that if denser cities host more migrants, then estimates in column (1) capture both effects. The correlation of density with other city characteristics is also high; in particular, it is positive with market access and negative with distance to seaports and land area. Meaningful estimates are therefore obtained in column (3), where all variables are introduced together, and even more in columns (4) and (5), our preferred specification, which progressively considers more year-specific effects.

Most control effects compensate each other with regard to density, and its impact, at 0.096 in column (5), is very close to that of column (1). While the density impact is 33% lower in 2007, though not significantly because of the imprecision of the estimation, it is significantly lower in 2002 to the point that it is fully canceled out. Hence, while there were no density gains present in the PRC in 2002, by 2013 these gains were three times larger than in developed economies. Moreover, while the extensive margin of city size is mostly found to have no impact in developed economies, it generates further gains in the PRC from 2007 onward. As with density, the impact of land area is significantly different from zero in 2007, while it is not significant though slightly positive in 2002.<sup>7</sup>

By contrast, the impact of migrants was three times larger in 2002 than what it was in 2013, with 2007 presenting an intermediate value that is somewhat closer to the 2013 estimate. The 2002 impact is very large. Typically, moving from the first to the last quartile of the migrant share variable increases natives' earnings by 17.5% in 2002, but by only 8.5% in 2013.<sup>8</sup> Distance to a seaport also presents a significant negative effect in 2002, which vanishes in the latter years. This finding suggests that proximity to the sea (and thus access to international markets) was a factor of productivity gains in 2002 that did not seem to hold in subsequent years. This could be related to the changing role of international trade in the PRC's economic growth over the past decade. Finally, a city's industrial diversity was found to be only marginally significant in all 3 years under review, as usually found in the literature.

Overall, we find that the two standard city-size variables, density and land area, as well as the presence of migrants, generate higher earnings in cities in 2007 and 2013, with other characteristics playing no significant direct role.<sup>9</sup> The impact of migrants in 2007 and 2013 is, however, less than half of what it was in 2002. At that time, density and land area had no effect, access to the sea being the only other city characteristic significantly impacting earnings. These findings highlight

<sup>7</sup>These comments are based on additional regressions where a dummy normalization is done so that significance tests for the total effects for 2002 and 2007 are obtained. They are available upon request.

<sup>8</sup>Note that the effect is divided by two despite the fact that the interquartile gap in the migrant share variable increased from 2 to 2.6 between 2002 and 2013.

<sup>9</sup>Some variables (market potential, for instance) could drive density or the presence of migrants and thus have an indirect effect. Yet, the strategy developed here does not allow this to be assessed.

the PRC's move from urban gains shaped by some of its specific features in the early 2000s—large disparities across cities in terms of access to the sea and the presence of significant numbers of rural migrants—to a situation where variables that typically affect urban earnings are now playing an influential role. What matters for cities in the PRC now is their capacity to host migrants as well as their overall size, in terms of both density and land area.

First-step estimates for equation (1) from which the city fixed effects used in the second step are obtained are provided in Appendix Table A.1. We do not further comment on the impact of individual characteristics, which are consistent with usual findings for the PRC (Démurger, Li, and Yang 2012). The first-step specification also includes the role of a city–industry-specific variable, specialization, which is measured by the logarithm of the share in the local economy of employment in the firm's industry. Being specialized in an industry increases earnings in a city, which is an additional agglomeration effect typically found in the literature. The effect seems to have reinforced over time: it is significant on earnings per hour only in 2002, while the magnitude of the impact in 2013 is larger than that usually obtained.

As mentioned above, the cities surveyed for each wave differ across time. The regressions provided in Table 1 are performed on the resulting nonbalanced panel. We checked that the distribution of city characteristics does not differ too much between waves. As a further robustness check, we run the same regressions on the subsample of cities that are common in all three waves of the survey. Results are presented in Table A.2 in the Appendix. Our main conclusions remain unchanged, even though some estimates are less precise because of the much smaller sample size.

## B. Earnings per Hour

The agglomeration literature mostly focuses on productivity externalities, which are better tested empirically by earnings per hour than by total earnings. The 2002 and 2013 surveys allow us to compute for each worker a measure of earnings per hour, which should thus be closer to labor productivity. If the number of hours worked varies significantly across cities, as documented by Rosenthal and Strange (2008) for the United States, the city determinants of total and per hour earnings can differ. As shown by the last two columns in Table 1, this appears not to be the case in the PRC, whether only density and migrants are interacted with year (column [6]) or all city variables (column [7]).<sup>10</sup> The estimated impact of density is only very slightly lower on earnings per hour than it is on total earnings. Importantly, both

<sup>10</sup>This is supported by the fact that the impact of city characteristics on hours worked is not significant. Indeed, when using the same specification for hours as for earnings, market potential is the only local characteristic to have a significant (negative) impact (in 2002 only). Results are available upon request.

density and land area effects still totally offset in 2002. Neither the intensive nor the extensive margins of city size increase labor productivity at that date. By contrast, the larger presence of migrants does significantly increase earnings per hour, with an impact equal to the one estimated on total earnings. Again, the effect is less than half in 2013, but is still significant. Overall, both total and per hour earnings are influenced by the same city characteristics and to the same extent.

#### IV. Gender and Skills Heterogeneity

The literature shows that both the returns to individual characteristics and the returns to city characteristics, typically gains from larger city size, should and do differ across skills. High-skilled workers tend to benefit more from agglomeration, in particular when local externalities result from technological spillovers (Bacolod, Blum, and Strange 2009). The comparison of agglomeration effects between genders is less frequent; Phimister (2005) is one of the rare exceptions. The data we use allow us to test whether city characteristics have different impacts on skilled and unskilled workers, and on male and female workers. The results are reported in Table 2.

Consistent with what is often observed for developed economies, skilled workers gain slightly more from larger city size (both higher density and larger land area) than unskilled workers. This finding holds for both total and per hour earnings, suggesting that at least part of city-size gains arise from technological spillovers rather than from market effects. Yet, the skills gap is not large since point estimates for high-skilled workers are only around 10% higher than for their unskilled counterparts. Because of the lack of the estimates' precision, the gap is not significantly different from zero most of the time.

A similar conclusion is reached for the impact of the city's migrant share for the year 2013. The estimated effects are very similar for skilled and unskilled workers, and even slightly larger for unskilled workers on total earnings. When the migrant effect is twice as large in 2002, the gap in favor of skilled workers is also larger, with the marginal impact being around 25% larger, though it is not significant. A larger positive impact of migrants on skilled urban residents was expected as migrants are more likely to substitute for unskilled workers in production functions.

To the best of our knowledge, a gender gap for urban returns has never been documented in the PRC. We do find small differences in favor of females, even though the size of the sample makes it difficult to obtain significant gaps. Gaps are also slightly larger for earnings per hour than for total earnings. Female workers are found to gain slightly more from city size (both higher density and larger land area). They also gain slightly more from the presence of migrants. A possible explanation for the larger benefits for females in cities may relate to household-level choices of which neighborhood to live in. If the location decision is made by men,

Table 2. City Determinants of Total and Per Hour Earnings

	Total		Hourly			Total			Hourly	
	Skilled	Unskilled	Skilled	Unskilled	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(7)	(8)
Density	0.098*** (0.029)	0.089*** (0.031)	0.088*** (0.033)	0.078** (0.033)	0.088*** (0.026)	0.104*** (0.029)	0.074** (0.029)	0.098*** (0.033)		
Density 2007	-0.033 (0.040)	-0.024 (0.043)			-0.013 (0.036)	-0.044 (0.041)				
Density 2002	-0.095* (0.050)	-0.096* (0.053)	-0.092 (0.057)	-0.079 (0.057)	-0.066 (0.045)	-0.121** (0.051)	-0.060 (0.049)	-0.115** (0.057)		
Migrants	0.193*** (0.072)	0.212*** (0.077)	0.202** (0.082)	0.196** (0.083)	0.187*** (0.065)	0.240*** (0.073)	0.180** (0.071)	0.188** (0.081)		
Migrants 2007	0.111 (0.090)	0.072 (0.096)			0.072 (0.081)	0.084 (0.091)				
Migrants 2002	0.443*** (0.156)	0.365** (0.166)	0.415** (0.178)	0.324*** (0.178)	0.412*** (0.141)	0.413*** (0.157)	0.399** (0.153)	0.445** (0.176)		
Land area	0.129*** (0.032)	0.072*** (0.034)	0.129*** (0.037)	0.086** (0.037)	0.081*** (0.029)	0.101*** (0.033)	0.080** (0.032)	0.115*** (0.036)		
Land area 2007	-0.014 (0.048)	0.027 (0.051)			0.026 (0.043)	0.006 (0.048)				
Land area 2002	-0.095* (0.052)	-0.054 (0.055)	-0.092 (0.059)	-0.051 (0.059)	-0.045 (0.046)	-0.067 (0.052)	-0.036 (0.050)	-0.077 (0.058)		
Market potential	0.043 (0.098)	0.156 (0.104)	0.002 (0.112)	0.161 (0.112)	0.116 (0.088)	0.065 (0.099)	0.058 (0.096)	0.077 (0.111)		
Market potential 2007	0.082 (0.147)	0.068 (0.156)			0.048 (0.132)	0.056 (0.148)				
Market potential 2002	-0.080 (0.174)	-0.124 (0.184)	-0.018 (0.198)	-0.043 (0.199)	-0.133 (0.157)	-0.111 (0.175)	-0.035 (0.170)	-0.091 (0.196)		

Continued.

Table 2. *Continued.*

	Total		Hourly		Total		Hourly	
	Skilled (1)	Unskilled (2)	Skilled (3)	Unskilled (4)	Male (5)	Female (6)	Male (7)	Female (8)
Distance to seaport	-0.022 (0.016)	-0.001 (0.017)	-0.025 (0.018)	0.001 (0.019)	-0.004 (0.015)	-0.017 (0.016)	-0.002 (0.016)	-0.021 (0.018)
Distance to seaport 2007	0.008 (0.023)	0.005 (0.024)			-0.015 (0.020)	0.022 (0.023)		
Distance to seaport 2002	-0.043 (0.027)	-0.049* (0.029)	-0.047 (0.031)	-0.049 (0.031)	-0.055*** (0.025)	-0.048* (0.028)	-0.059** (0.027)	-0.050 (0.031)
Diversity	-0.133*** (0.066)	-0.196*** (0.070)	-0.171** (0.075)	-0.155** (0.075)	-0.099* (0.059)	-0.110* (0.066)	-0.082 (0.064)	-0.098 (0.074)
Diversity 2007	-0.004 (0.170)	0.027 (0.180)			0.029 (0.153)	-0.033 (0.170)		
Diversity 2002	-0.064 (0.172)	-0.083 (0.182)	-0.010 (0.196)	-0.159 (0.196)	-0.030 (0.155)	-0.212 (0.173)	-0.045 (0.168)	-0.215 (0.193)
Observations	239	239	152	152	239	239	152	152
R <sup>2</sup>	0.73	0.83	0.68	0.86	0.83	0.80	0.84	0.82

Notes: The panel of cities is not balanced (99 cities for 2013, 87 cities for 2007, and 53 cities for 2002). Time dummies are included in all specifications. Standard errors in brackets. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , \* =  $p < 0.10$ .  
 Source: Authors' calculations.

either because of cultural and social norms or because male earnings are higher, females have to adapt to the nonfreely chosen location with respect to their own job opportunities. Better job matches are more easily found in dense and large areas, and externalities are more important there than in less dense and smaller locations where job opportunities are less abundant. As a result, the gender gap should be lower in cities benefiting from characteristics such as larger size and a larger pool of migrants.

## V. Conclusion

Gains from urbanization are evident in the PRC, where they seem to be larger than in developed economies. The nature of these gains has evolved rapidly and the city characteristics that shaped such gains in 2002 are different from those in effect in 2013. The presence of rural migrants and access to a seaport were the largest productivity drivers in cities in the PRC in 2002. City size, which is measured by both the intensive margin (employment density) and the extensive margin (land area), became the most important determinant over the next decade even as the migrant externality remained strong. In 2013, workers in cities benefited from more migrants, increased density, and a larger land area. This contrasts with developed economies, where density has the largest effect (and with much lower elasticity). Internal market access, which is almost the only other important local characteristic in developed countries, does not seem to play a major role in the PRC.

In this paper, we do not assess possible biases due to missing city variables or reverse causality. In earlier studies on developed economies, reverse causality was found to affect estimates by no more than 20%. Yet, this should be carefully evaluated in future studies on the PRC because high migration rates could make such biases larger. Furthermore, the land area of cities is also regularly updated by authorities in the PRC based on past development, which could affect the estimated impact. Similarly, we do not find evidence of any strong spatial sorting according to observed individual skills, but this issue could emerge progressively through the endogenous location choices of an increasingly heterogeneous population with regard to skills.

Lastly, the migrant externality we document is largely a black box that needs to be better understood. The impact of migrants is a mix of externality and substitution effects. Identifying these effects separately is needed to design consistent local policies. Identifying these channels separately requires working on larger data sets because some differences (for instance, between skilled and unskilled workers) seem to emerge from our results. Investigating in more detail the role of workers' allocation between industries and occupations, which sharply differs between local residents and migrants, is necessary too. Finally, as migrants themselves become a larger share of a city's population, measuring the impact of city characteristics on their own earnings would help to complete the analysis of

urbanization gains in the PRC. Data limitations in the surveys used in this study did not allow us to investigate these additional issues.

## References

- Au, Chun-Chung, and J. Vernon Henderson. 2006. "Are Chinese Cities Too Small?" *Review of Economic Studies* 73 (3): 549–76.
- Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009. "Skills in the City." *Journal of Urban Economics* 65 (2): 136–53.
- Baum-Snow, Nathaniel, and Ronni Pavan. 2013. "Inequality and City Size." *Review of Economics and Statistics* 93 (5): 1535–48.
- Cai, Fang, Albert Park, and Yaohui Zhao. 2008. "The Chinese Labor Market in the Reform Era." In *China's Great Economic Transformation*, edited by Loren Brandt and Thomas Rawski, 167–214. Cambridge: Cambridge University Press.
- Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. "What is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States." *Journal of Urban Economics* 98:17–49.
- Combes, Pierre-Philippe, Sylvie Démurger, and Shi Li. 2015. "Migration Externalities in China." *European Economic Review* 76: 152–67.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63 (2): 723–42.
- \_\_\_\_\_. 2016. "The Costs of Agglomeration: Land and House Prices in French Cities." Mimeograph.
- Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Urban and Regional Economics, Volume 5A*, edited by Gilles Duranton, J. Vernon Henderson, and Will Strange. Amsterdam: North-Holland.
- Démurger, Sylvie, Marc Gurgand, Shi Li, and Ximing Yue. 2009. "Migrants as Second-Class Workers in Urban China? A Decomposition Analysis." *Journal of Comparative Economics* 37 (4): 610–28.
- Démurger, Sylvie, Shi Li, and Juan Yang. 2012. "Earnings Differentials between the Public and Private Sectors in China: Exploring Changes for Urban Local Residents in the 2000s." *China Economic Review* 23 (1): 138–53.
- Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics, Volume 4*, edited by J. Vernon Henderson and Jacques-François Thisse, 2063–117. Amsterdam: North-Holland.
- Harris, Chauncy. 1954. "The Market as a Factor in the Localization of Industry in the United States." *Annals of the Association of American Geographers* 44 (4): 315–48.
- Lewis, Ethan, and Giovanni Peri. 2015. "Immigration and the Economy of Cities and Regions." In *Handbook of Urban and Regional Economics, Volume 5A*, edited by Gilles Duranton, J. Vernon Henderson, and Will Strange. Amsterdam: North-Holland.
- Li, Shi, Chuliang Luo, Zhong Wei, and Ximing Yue. 2008. "The 1995 and 2002 Household Surveys: Sampling Methods and Data Description." In *Inequality and Public Policy in China*, edited by Bjorn Gustafsson, Shi Li, and Terry Sicular, 337–53. Cambridge: Cambridge University Press.
- Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.
- Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (1): 65–103.

Phimister, Euan. 2005. "Urban Effects on Participation and Wages: Are There Gender Differences?" *Journal of Urban Economics* 58 (3): 513–36.  
 Rosenthal, Stuart S., and William C. Strange. 2008. "Agglomeration and Hours Worked." *Review of Economics and Statistics* 90 (1): 105–18.

**Appendix**

**Table A.1. Individual Wage Disparities—  
 Ordinary Least Squares Estimates for the First Step**

	2013		2007	2002	
	Total (1)	Hourly (2)	Total (3)	Total (4)	Hourly (5)
Male	0.272*** (0.015)	0.236*** (0.015)	0.246*** (0.010)	0.138*** (0.013)	0.100*** (0.014)
Years of education	0.059*** (0.003)	0.065*** (0.004)	0.065*** (0.002)	0.049*** (0.003)	0.052*** (0.003)
Experience	0.054*** (0.002)	0.044*** (0.002)	0.040*** (0.002)	0.043*** (0.003)	0.039*** (0.003)
Experience <sup>2</sup>	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Urban collective enterprises	-0.211*** (0.036)	-0.188*** (0.037)	-0.272*** (0.021)	-0.195*** (0.026)	-0.224*** (0.027)
Private enterprises	-0.190*** (0.023)	-0.207*** (0.023)	-0.218*** (0.018)	-0.136*** (0.020)	-0.203*** (0.021)
Foreign enterprises	0.047 (0.042)	0.038 (0.044)		0.184*** (0.041)	0.126*** (0.043)
Government administration	0.038 (0.029)	0.029 (0.030)		0.098*** (0.023)	0.094*** (0.024)
Other ownership			-0.220*** (0.013)		
Specialization	0.085*** (0.013)	0.075*** (0.014)	0.060*** (0.011)	0.015 (0.014)	0.032** (0.014)
Occupation dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes
City dummies	Yes	Yes	Yes	Yes	Yes
Observations	7,181	7,181	14,028	7,659	7,591
R <sup>2</sup>	0.39	0.39	0.44	0.41	0.41

Notes: Estimations for individuals aged 16–70 years old who declared working at least part of the year and earning (positive) wages. Reference category for ownership is state-owned enterprises. Standard errors in brackets. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , \* =  $p < 0.10$ .  
 Source: Authors' calculations.



Table A.2. City Determinants of Total and Per Hour Earnings—  
Cities Common to 2002, 2007, and 2013 Surveys

	Total Earnings					Hourly Earnings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.166** (0.069)	0.112** (0.049)	0.134*** (0.042)	0.140*** (0.044)	0.155*** (0.046)	0.130*** (0.034)	0.137*** (0.035)
Density 2007	-0.066 (0.103)	-0.099 (0.073)	-0.101 (0.061)	-0.106 (0.065)	-0.134* (0.075)		
Density 2002	-0.194* (0.101)	-0.141** (0.069)	-0.158*** (0.058)	-0.172*** (0.062)	-0.210*** (0.069)	-0.118*** (0.043)	-0.147*** (0.052)
Migrants		0.326*** (0.112)	0.187* (0.097)	0.173* (0.100)	0.206* (0.107)	0.119 (0.088)	0.137 (0.091)
Migrants 2007		0.366** (0.170)	0.281* (0.142)	0.296* (0.156)	0.267 (0.179)		
Migrants 2002		0.390** (0.165)	0.374*** (0.137)	0.411*** (0.147)	0.332** (0.162)	0.443*** (0.133)	0.414** (0.159)
Land area			0.088** (0.024)	0.111*** (0.039)	0.121*** (0.041)	0.077*** (0.026)	0.099*** (0.034)
Land area 2007				-0.025 (0.055)	-0.034 (0.063)		
Land area 2002				-0.045 (0.055)	-0.069 (0.059)		-0.050 (0.051)
Market potential			0.151* (0.077)	0.145* (0.079)	0.077 (0.145)	0.140 (0.093)	0.236* (0.129)
Market potential 2007					0.189 (0.202)		
Market potential 2002					0.015 (0.197)		-0.215 (0.184)
Distance to seaport			-0.018 (0.011)	-0.019 (0.012)	-0.005 (0.020)	-0.022 (0.016)	0.005 (0.022)
Distance to seaport 2007					-0.004 (0.029)		
Distance to seaport 2002					-0.040 (0.029)		-0.061* (0.032)
Diversity			-0.242** (0.098)	-0.258** (0.101)	-0.354** (0.140)	-0.022 (0.088)	0.045 (0.105)
Diversity 2007					0.202 (0.275)		
Diversity 2002					0.222 (0.268)		-0.269 (0.195)
Observations	60	60	60	60	60	76	76
R <sup>2</sup>	0.70	0.88	0.92	0.92	0.93	0.90	0.91

Notes: Balanced panel of cities (20 cities common to 2002, 2007, and 2013; 38 cities common to 2002 and 2013). Time dummies are included in all specifications. Standard errors in brackets. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , \* =  $p < 0.10$ .

Source: Authors' calculations.