



**HAL**  
open science

# Productivity gains from agglomeration and migration in Chinese cities over 2002-2013

Pierre-Philippe Combes, Sylvie Démurger, Shi Li

► **To cite this version:**

Pierre-Philippe Combes, Sylvie Démurger, Shi Li. Productivity gains from agglomeration and migration in Chinese cities over 2002-2013. 2017. halshs-01468602

**HAL Id: halshs-01468602**

**<https://shs.hal.science/halshs-01468602v1>**

Preprint submitted on 15 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 1709 – February 2017

## **Productivity gains from agglomeration and migration in Chinese cities over 2002-2013**

Pierre-Philippe Combes, Sylvie Démurger, Shi Li

### **Abstract:**

We evaluate the evolution of productivity gains from Chinese cities over time, from 2002 to 2013. In 2002, rural migrants were exerting a strong positive externality on natives' earnings, which were also higher when access to foreign markets through access to sea was higher. In 2007 and then further in 2013, city size (employment density but also land area) has become the crucial determinant of productivity whereas market access, internal or external, plays no direct role. Rural migrants still enhance natives' earnings, though the effect is more than half lower than in 2002. Urban gains, and their evolution over time, are very similar on total and per hour earnings. Skilled workers and females seem to gain slightly more from cities than unskilled workers and males.

### **Keywords:**

urban development; agglomeration economies; wage disparities; migration; China

### **JEL codes:**

O18, R12, R23, J31, O53

# Productivity gains from agglomeration and migration in Chinese cities over 2002-2013\*

Pierre-Philippe Combes<sup>†</sup>      Sylvie Démurger<sup>‡</sup>      Shi Li<sup>§</sup>

February 14, 2017

## Abstract

We evaluate the evolution of productivity gains from Chinese cities over time, from 2002 to 2013. In 2002, rural migrants were exerting a strong positive externality on natives' earnings, which were also higher when access to foreign markets through access to sea was higher. In 2007 and then further in 2013, city size (employment density but also land area) has become the crucial determinant of productivity whereas market access, internal or external, plays no direct role. Rural migrants still enhance natives' earnings, though the effect is more than half lower than in 2002. Urban gains, and their evolution over time, are very similar on total and per hour earnings. Skilled workers and females seem to gain slightly more from cities than unskilled workers and males.

**JEL Codes:** O18, R12, R23, J31, O53.

**Keywords:** urban development; agglomeration economies; wage disparities; migration; China.

---

\*This paper was prepared for the Asian Development Bank conference on *Urban and Regional Development in Asia* held in Seoul, July 1-2, 2016. We are grateful to an anonymous reviewer for careful reading of our paper and constructive comments, as well as to the conference participants. We also thank Zhu Mengbing for excellent research assistance.

<sup>†</sup>Univ Lyon, CNRS, GATE L-SE UMR 5824, 93 chemin des Mouilles, F-69130 Ecully, France; Sciences Po, Department of Economics, 28, Rue des Saints-Pères, 75007 Paris, France. Also research fellow at the CEPR. Email: ppcombes@gmail.com.

<sup>‡</sup>Univ Lyon, CNRS, GATE L-SE UMR 5824, 93 chemin des Mouilles, F-69130 Ecully, France. Also research fellow at IZA, Bonn, Germany. Email: demurger@gate.cnrs.fr.

<sup>§</sup>Business School, Beijing Normal University, China. Also research fellow at IZA, Bonn, Germany. Email: lishi@bnu.edu.cn.

## Introduction

Although the empirical literature on productivity gains from larger cities has matured a lot over the past decade (Combes and Gobillon, 2015), evidence about the characteristics and the role of cities in developing countries is still largely incomplete (Chauvin, Glaeser, Ma and Tobio, 2016). This stands in sharp contrast with urbanisation rates that are much lower but increasing rapidly in many of them compared to developed countries where urbanisation is fairly stable. Quantifying gains and costs from such a rapidly evolving spatial concentration of economic activity is therefore a highly topical issue. If one expects gains to be larger than costs, urbanisation should be further promoted, for instance by facilitating rural-to-urban migration. If costs dominate, appropriate policy must be implemented in order to refrain urban expansion.

China offers the best example of such a large increase of rural-urban migration that has been feeding the development of cities for more than fifteen years. In an early contribution, Au and Henderson (2006) document the gains that could emerge from larger cities in China. They conclude that, apart from the largest cities, large productivity gains could emerge from a larger number of medium-sized cities. Using micro data from the mid-2000s, Combes, Démurger and Li (2015) estimate the specific externality that migrants exert on native 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 migrants' share on urban residents' wages. The present paper aims at complementing the analysis over time by evaluating the evolution of productivity gains from Chinese cities between 2002 and 2013. Measuring possible changes has important policy implication because the success of urban policies ultimately depends on the nature of the gains, which may have changed over time. Another contribution of the paper consists in estimating city gains not only on total earnings but also on earnings per hour, a closer measure of labour productivity, as well as for both skilled and unskilled workers separately, and by gender. For doing so, we combine nationally representative household income surveys conducted for the years 2002, 2007 and 2013 with city-level data.

We find that city characteristics have a significant impact on labour productivity and that the characteristics that matter have changed over time. In 2002, the migrant externality was very strong, and to a lower extent access to foreign markets through an access to the sea was also playing a positive role. In 2007 and then further in 2013, agglomeration variables that are more standard in the literature and that are mostly reflecting city size (such as employment density and land area) have become crucial determinants of city productivity. By contrast, market access, internal or external, plays no significant direct role. As for rural migrants, they still enhance urban resident workers' earnings, but the effect is more than half lower than in 2002. These findings show that China has moved from urban gains shaped by some of its specific features (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 earnings are playing a more influential role. What matters for Chinese cities 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 total and per hour earnings. Surprisingly we find only little difference in urban gains for skilled and unskilled workers, yet city effects are slightly larger for skilled workers. This may come in particular from the fact that the substitution effect and the externality effect from migrants are somewhat difficult to disentangle with the datasets we use. Last, female workers seem to gain slightly more from cities than male workers do.

Section 1 briefly recalls the theoretical background and the associated empirical strategy, and presents the data we use. Section 2 is devoted to measuring the impact of city characteristics for all workers taken together, on both total and per hour earnings. As a complement, Section 3 evaluates agglomeration and migration gains separately for skilled and unskilled workers, and for male and female workers. Section 4 concludes.

## 1 Mechanisms, empirical specification, and data

### 1.1 Background and methodology

Marshall (1890) suggested long ago that gains from cities arise from a number of local external economies of scale. The underlying mechanisms are now well understood and classified by Duranton and Puga (2004) in sharing (specialised inputs, diversity, or risk), matching (quantity and quality of matches on local labour markets) and learning (creating, diffusing and accumulating knowledge). These mechanisms either directly impact households' utility or increase their productivity, which in turns leads to higher earnings. The largest part of the empirical literature augments standard wage equations by city variables in order 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 (TFP) measures. Comparing nominal gains to the higher costs of living in cities, and therefore assessing real income gains in cities, is the purpose of another literature<sup>1</sup>.

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

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

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

The first-step estimation of specification (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 city  $c$ 's

---

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

characteristics for sector  $s$  where  $i$  is employed,  $L_{cst}$ , net of the role of individual characteristics  $X_{it}$ . We estimate equation (1) on each year separately, which makes all estimated parameters year-specific. Controlling for individual characteristics is crucial to remove the bias arising from a possible non-random sorting of individuals across cities depending on their abilities. The sorting of individuals has been documented to be large for some countries although it does not seem to be the case in China as shown by Combes et al. (2015). Introducing city-sector variables, typically the logarithm of the share of sector  $s$  in city  $c$  employment,  $L_{cst}$ , allows identifying the role of the characteristics of the sector within the city beyond the role of other non-sector 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 three years of data together and compare estimates when some of, or all, estimated parameters are year-specific. The main two variables of interest are employment per square kilometre (density) and the share of internal migrants. The former has been shown by the literature to be the main driver of urban productivity. The role of the latter has been highlighted in the case of China by Combes et al. (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 countries. Then the impact of the logarithm of the inverse of one minus the share of migrants in the city total employment corresponds to the sum of the impact of density (for a given number of urban residents, migrants inflow increases density) and the separate migrant effect net of such density gains. The latter effect itself results from a complementarity (if positive) or a 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 Chinese cities. 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 kilometre within the city boundary. Land area captures the role of the extensive margin, the city being made larger by extending its boundaries at given density, that is, under a simultaneous proportional increase of total employment.

Since trade takes place between locations and workers are mobile, the literature has also emphasised the role of the access to distant markets that can generate “imported external economies”

---

<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.

in cities. We use two variables to disentangle the role of internal and international markets respectively. The first variable is a within-China market potential à la Harris (1954), which is the sum of the employment density in other Chinese cities divided by the distance to the city considered. It assesses how far the city is from other large cities. As a large share of China exports is operated through coastal ports, the second variable corresponds to the city 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 industrial diversity measured by the inverse of a Herfindhal index computed on the shares of each industry in the city total employment.

Although we pool together different years, the impact of local variables is mostly identified over the cross-section of Chinese cities because time-variations are much lower than across cities but also because few cities are sampled for all the three years we consider. The literature emphasises two types of possible biases. The largest one is due to the spatial distribution of workers depending on unobserved skills that largely affect their productivity without resulting from city externalities. This can make the estimated impact of local variables twice larger than what it is for the same individual who would locate in different cities. Since we do not have access to an individual panel where workers are followed across years, we can only control for observable individual skills. A large spatial sorting according to skills is observed in France or the US. 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 US according to Baum-Snow and Pavan (2013). As for China, Combes et al. (2015) show that no spatial sorting occurs according to observable skills. Therefore, only a large sorting on unobserved skills, which should not be correlated to observed skills, could significantly bias the estimates. We see this as rather unrealistic, at least as long as many years of intensive migration and a possible unequal access of higher university degrees have not 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 location choices of firms and workers are endogenous. The local density variable is the most suspicious one under this perspective. This is the reason why we use here the specification proposed by Combes et al. (2015), which decomposes total employment density in the impact of native employees and of the share of migrants respectively. 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 have settled long ago in the city.

By contrast, the migration variable is likely to be affected by reverse causality. Migrants choose

---

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

locations where employment conditions, and earnings in particular, are favourable. By appealing to specifications considering different sets of control variables and by using various instrumental variable strategies, Combes et al. (2015) find evidence of biases in China very similar in magnitude to those usually documented for other countries. At around 20% at maximum, this is not very large. Addressing reverse causality is beyond the scope of the present paper but it would deserve further investigation in the future. This would require more sophisticated strategies, which could be possibly difficult to implement with available data as emphasised by Lewis and Peri (2015).

Other local variables could be endogenous too, and this 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, one should note that the problem could be more severe here for the land area variable, because city boundaries are regularly reshaped by the Chinese administration in order to match recent local demographic evolution. Again, the causality would go the other way round. We have checked that very close magnitudes were obtained when we use a land area definition that is constant over time (using the 2007 value) but this again would deserve further investigation provided that more comprehensive data is available.

## 1.2 Data

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

Individual databases for the years 2002 and 2013 are part of the China Household Income Project (CHIP), whereas the 2007 dataset is a sub-sample 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 three years fairly comparable. Yet, because working time is documented only in the CHIP-2002 and CHIP-2013 data, we compute hourly wages for these two years only. Another feature to highlight is that the surveyed prefecture-level cities are not all the same across the three years, and vary substantially. 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 the three surveys, and 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

---

<sup>4</sup>See Combes et al. (2015) for a discussion of city-level data sources.

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



only (the urban *Hukou* holders<sup>6</sup>). This means that an important segment of the urban labour market, composed of migrants not officially registered in cities, is excluded. Representing about 17% of China’s 1.3 billion people, migrants constitute an important group to be reckoned with, yet clearly marginalized (Cai, Park and Zhao, 2008; Démurger, Gurgand, Li and Yue, 2009). The exclusion of migrants whose *Hukou* is not where they work has an important implication for our estimates: our data set refers to urban residents (‘natives’) only and one cannot infer that earnings determinants across Chinese cities we identify here 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 years for both estimation steps. The only exception is for enterprise ownership in specification (1), which is more detailed in 2002 and 2013 compared to 2007 (see Appendix A).

## 2 Agglomeration gains in 2002, 2007 and 2013

### 2.1 Total earnings

Table 1 displays the results obtained on all workers’ earnings for the estimation of specification (2). Rows with the variable’s name without vintage, ‘Density’ for instance, correspond to the effect common to all years. This is the total effect for year 2013 too. The effect for 2007 (2002 respectively) is obtained by adding to this number the coefficient reported in the corresponding line, ‘Density 2007’ for instance (‘Density 2002’ respectively). The significance test denotes whether the effect for 2007 (2002 respectively) is significantly different from the 2013 effect. The first five columns provide estimates based on total earnings while the last two columns give estimates based on earnings per hour. As indicated above, the number of hours worked is not available in the micro data 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 non-significantly different from the one for 2013. The elasticity for these two years is close to the values reported by Combes et al. (2015) and this is almost three times larger than usual estimates for developed countries. Large density economies seem to prevail in China, at least from the mid 2000s onwards. By contrast, the effect appears much lower in 2002, even though the estimated year-specific effect is not significantly different from zero here. We will come back to this issue once other variables are introduced.

---

<sup>6</sup>The *Hukou* or household registration system is a distinctive institutional feature of China, which divides the Chinese population according to occupation (agricultural and non-agricultural) and place of residence. By entitling access to social benefits to local *Hukou* holders only, the system has strongly limited population mobility towards cities for decades. Until today, 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 <sup>a</sup> (0.025)	0.059 <sup>a</sup> (0.022)	0.088 <sup>a</sup> (0.022)	0.092 <sup>a</sup> (0.024)	0.096 <sup>a</sup> (0.025)	0.070 <sup>a</sup> (0.025)	0.086 <sup>a</sup> (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 <sup>c</sup> (0.035)	-0.053 (0.033)	-0.082 <sup>c</sup> (0.041)	-0.091 <sup>b</sup> (0.043)	-0.044 (0.036)	-0.085 <sup>c</sup> (0.047)
Migrants		0.267 <sup>a</sup> (0.061)	0.210 <sup>a</sup> (0.060)	0.205 <sup>a</sup> (0.062)	0.207 <sup>a</sup> (0.062)	0.194 <sup>a</sup> (0.066)	0.185 <sup>a</sup> (0.067)
Migrants 2007		0.106 (0.075)	0.096 (0.071)	0.092 (0.076)	0.086 (0.077)		
Migrants 2002		0.491 <sup>a</sup> (0.120)	0.390 <sup>a</sup> (0.114)	0.466 <sup>a</sup> (0.131)	0.412 <sup>a</sup> (0.133)	0.393 <sup>a</sup> (0.127)	0.412 <sup>a</sup> (0.145)
Land area			0.080 <sup>a</sup> (0.017)	0.089 <sup>a</sup> (0.027)	0.092 <sup>a</sup> (0.028)	0.072 <sup>a</sup> (0.023)	0.097 <sup>a</sup> (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 <sup>c</sup> (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 <sup>b</sup> (0.009)	-0.019 <sup>b</sup> (0.009)	-0.009 (0.014)	-0.029 <sup>b</sup> (0.012)	-0.010 (0.015)
Distance to seaport 2007					0.002 (0.019)		
Distance to seaport 2002					-0.052 <sup>b</sup> (0.023)		-0.056 <sup>b</sup> (0.026)
Diversity			-0.109 <sup>b</sup> (0.048)	-0.115 <sup>b</sup> (0.048)	-0.111 <sup>b</sup> (0.056)	-0.097 <sup>c</sup> (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: Non-balanced panel of cities: 99 for 2013, 87 for 2007 and 53 for 2002. Time dummies are included in all specifications. Standard errors in brackets. <sup>a</sup>: p<0.01, <sup>b</sup>: p<0.05, <sup>c</sup>: p<0.10.

Column (2) adds the role of migrants, which is found to be largely positive, and decreasing 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 is also large with other city characteristics: 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 as regards density, and its impact, at 0.096 in column (5) is very close to the one in column (1). Importantly, 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, down to fully canceling out. Hence, no density gains were present in China in 2002, while they are three times larger than in developed countries nowadays. Moreover, whereas the extensive margin of city size is most of the time found to have no impact in these countries, it generates further gains in China from 2007 onwards. Indeed, as for density, the impact of land area turns significantly different from 0 in 2007 but 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 is in 2013, 2007 presenting an intermediate value closer to the 2013 estimate though. 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, and by only 8.5% in 2013<sup>8</sup>. Distance to seaports also presents a significant negative effect in 2002, which vanishes afterwards. This finding suggests that proximity to the sea, and thus access to international markets, was a factor of productivity increase in 2002 that does not seem to be so important in more recent years. This could be related to the changing role of international trade in Chinese economic growth over the past decade. Last, diversity is only very marginally significant, 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, other characteristics playing no significant direct role.<sup>9</sup> The impact of migrants is, however, lower than what it was in 2002, by more than half. 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 China'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 large presence of rural migrants– to a situation where variables that typically affect urban earnings are now playing an influential role. What matters for

---

<sup>7</sup>These comments are based on additional regressions where a dummy normalisation 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 of them, as for instance market potential, could drive density or the presence of migrants and thus have an indirect effect. Yet, the strategy developed here does not allow to assess that.

Chinese cities 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 specification (1) from which the city fixed effects used in the second step are obtained are provided in Appendix A. We do not further comment the impact of individual characteristics, which are consistent with usual findings for China (Démurger, Li and Yang, 2012). The first-step specification also includes the role of a city-industry specific variable, specialisation, which is measured by the logarithm of the share in the local economy of employment in the firm's industry. Being specialised in an industry increases earnings in this city, a further agglomeration effect typically found in the literature. Here again, the effect seems to have reinforced over time: it is significant on earnings per hour only in 2002, and the magnitudes for 2013 are larger than those 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 non-balanced panel. We checked that the distribution of city characteristics does not differ too much between waves. As a further robustness check, we also run the same regressions on the sub-sample of cities that are common to all three waves of the survey. Results are presented in Table 4 in Appendix B. Our main conclusions remain unchanged, even though some estimates are less precise because of the much smaller sample size.

## 2.2 Earnings per hour

In the agglomeration economies literature, the theory mostly focuses on productivity externalities, which are better tested empirically by earnings per hour than by total earnings. Interestingly, the 2002 and 2013 surveys allow us to compute for each worker a measure of earnings per hour, which should thus be closer to labour productivity. Would the number of hours worked vary significantly across cities, as documented by Rosenthal and Strange (2008) for the US, the city determinants of total and per hour earnings could differ. As shown by the last two columns in Table 1, this appears not to be the case in China<sup>10</sup>, whether only density and migrants are interacted with year, column (6), or all city variables, column (7). The estimated impact of density is only very slightly lower on earnings per hour than it is on total earnings. Importantly, both density and land area effects still totally offset in 2002. Neither the intensive nor the extensive margins of city size increase labour 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 more than twice lower in 2013 but still significant. Overall, both total and per hour earnings are influenced by the same city characteristics, and to the same extent.

---

<sup>10</sup>This is supported by non-significant impacts of city characteristics on hours worked. 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.

### 3 Gender and skills heterogeneity

The literature shows that both the returns to individual characteristics and the returns to city characteristics, typically gains from larger 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 gender is less frequent, Phimister (2005) being one of the rare exceptions. The data we use allow us to test whether city characteristics impact differently skilled and unskilled workers, on the one hand, and male and female workers, on the other hand. Results are reported in Table 2.

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

A similar conclusion is reached for the impact of the city 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 larger, in 2002, the gap in favour of skilled workers is also larger, the marginal impact being around 25% larger, but, again, non significantly. A larger positive impact of migrants on skilled natives was expected as migrants are probably more substitutable to unskilled workers than to skilled workers in the production function.

To the best of our knowledge, the gender gap in urban returns has never been documented for China. We do find small differences, in favour of females, even though again 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, as regards the role of both density and 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 the choice, within households, of location, be the city or the neighbourhood within the city where they live. If the location decision is made by men more than by women, either because of cultural and social norms or because male earnings are larger within the couple, females have to adapt to this non-freely chosen location with respect to their own job opportunities. Better job matches are easier in dense and large areas, and externalities are more important there than in smaller size locations where job opportunities are less abundant. As a result, the gender gap should be lower in cities benefiting from 'good characteristics', a larger size and a larger pool of migrants.

Table 2: City determinants of total and per hour earnings

	Total		Hourly		Total		Hourly	
	Skilled (1)	Unskilled (2)	Skilled (3)	Unskilled (4)	Male (5)	Female (6)	Male (7)	Female (8)
Density	0.098 <sup>a</sup> (0.029)	0.089 <sup>a</sup> (0.031)	0.088 <sup>a</sup> (0.033)	0.078 <sup>b</sup> (0.033)	0.088 <sup>a</sup> (0.026)	0.104 <sup>a</sup> (0.029)	0.074 <sup>b</sup> (0.029)	0.098 <sup>a</sup> (0.033)
Density 2007	-0.033 (0.040)	-0.024 (0.043)			-0.013 (0.036)	-0.044 (0.041)		
Density 2002	-0.095 <sup>c</sup> (0.050)	-0.096 <sup>c</sup> (0.053)	-0.092 (0.057)	-0.079 (0.057)	-0.066 (0.045)	-0.121 <sup>b</sup> (0.051)	-0.060 (0.049)	-0.115 <sup>b</sup> (0.057)
Migrants	0.193 <sup>a</sup> (0.072)	0.212 <sup>a</sup> (0.077)	0.202 <sup>b</sup> (0.082)	0.196 <sup>b</sup> (0.083)	0.187 <sup>a</sup> (0.065)	0.240 <sup>a</sup> (0.073)	0.180 <sup>b</sup> (0.071)	0.188 <sup>b</sup> (0.081)
Migrants 2007	0.111 (0.090)	0.072 (0.096)			0.072 (0.081)	0.084 (0.091)		
Migrants 2002	0.443 <sup>a</sup> (0.156)	0.365 <sup>b</sup> (0.166)	0.415 <sup>b</sup> (0.178)	0.324 <sup>c</sup> (0.178)	0.412 <sup>a</sup> (0.141)	0.413 <sup>a</sup> (0.157)	0.399 <sup>b</sup> (0.153)	0.445 <sup>b</sup> (0.176)
Land area	0.129 <sup>a</sup> (0.032)	0.072 <sup>b</sup> (0.034)	0.129 <sup>a</sup> (0.037)	0.086 <sup>b</sup> (0.037)	0.081 <sup>a</sup> (0.029)	0.101 <sup>a</sup> (0.033)	0.080 <sup>b</sup> (0.032)	0.115 <sup>a</sup> (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 <sup>c</sup> (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)
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 <sup>c</sup> (0.029)	-0.047 (0.031)	-0.049 (0.031)	-0.055 <sup>b</sup> (0.025)	-0.048 <sup>c</sup> (0.028)	-0.059 <sup>b</sup> (0.027)	-0.050 (0.031)
Diversity	-0.133 <sup>b</sup> (0.066)	-0.196 <sup>a</sup> (0.070)	-0.171 <sup>b</sup> (0.075)	-0.155 <sup>b</sup> (0.075)	-0.099 <sup>c</sup> (0.059)	-0.110 <sup>c</sup> (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
$R^2$	0.73	0.83	0.68	0.86	0.83	0.80	0.84	0.82

Notes: See Table 1.

## 4 Conclusion

Gains from urbanisation are present in China, and they seem to be larger than in developed countries. However, they evolve rapidly and the city characteristics that shaped such gains in 2002 are rather different from those currently at stake. The presence of rural migrants and access to the sea were the largest productivity drivers in Chinese cities in 2002. City size, which is measured by both the intensive margin (employment density) and the extensive margin (land area), has taken over during the following decade even though the migrant externality remains strong. In 2013, workers in Chinese cities benefit from migrants, density and land area all together. This contrasts with developed countries where density has the largest effect, with a much lower elasticity. Internal market access, which is almost the only further important local characteristics there, does not seem to play a major role in China, at least for the moment.

In this paper, we do not assess possible biases due to city missing variables and reverse causality. In earlier studies on Western countries, reverse causality has been found to affect estimates by no more than 20%. Yet, this should be carefully evaluated in future contributions on China because large migration rates could make such biases larger. Moreover the land area of cities is also regularly updated by Chinese authorities based on city's past development, which could possibly affect the estimated impact. Similarly, we do not find evidence of any strong spatial sorting according to observed individual skills, but the issue could emerge progressively through the endogenous location choices of the more and more skilled-heterogeneous population.

Lastly, the migrant externality we document is largely a black box for the moment, which clearly needs to be better understood. Indeed, the impact of migrants is a mix of externality and substitution effects and identifying separately the two is of importance in order to design consistent local policies. Identifying these channels separately would require working on larger data sets also because some differences, between skilled and unskilled for instance, seem to emerge from our results. Investigating in more detail the role of the workers' allocation between industries and occupations, which sharply differs between local residents and migrants, would probably be important too. Finally, as migrants themselves become a large share of city population, measuring the impact of city characteristics on their own earnings would help completing the analysis of urbanisation gains in China. Data limitation from the surveys we use here does not allow us to investigate these additional issues.

## References

Au, C. and Henderson, V. (2006), 'Are Chinese cities too small?', *Review of Economic Studies* **73**, 549–576.

- Bacolod, M., Blum, B. S. and Strange, W. C. (2009), ‘Skills in the city’, *Journal of Urban Economics* **65**(2), 136–153.
- Baum-Snow, N. and Pavan, R. (2013), ‘Inequality and city size’, *Review of Economics and Statistics* **93**(5), 1535–1548.
- Cai, F., Park, A. and Zhao, Y. (2008), The Chinese labor market in the reform era, in L. Brandt and T. Rawski, eds, ‘China’s Great Economic Transformation’, Cambridge University Press, pp. 167–214.
- Chauvin, J. P., Glaeser, E., Ma, Y. and Tobio, K. (2016), ‘What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States’, *Journal of Urban Economics* (forthcoming).
- Combes, P.-P., Démurger, S. and Li, S. (2015), ‘Migration externalities in China’, *European Economic Review* **76**, 152–167.
- Combes, P.-P., Duranton, G. and Gobillon, L. (2008), ‘Spatial wage disparities: Sorting matters!’, *Journal of Urban Economics* **63**(2), 723–742.
- Combes, P.-P., Duranton, G. and Gobillon, L. (2016), The costs of agglomeration: Land and house prices in French cities, Mimeograph.
- Combes, P.-P. and Gobillon, L. (2015), The empirics of agglomeration economies, in G. Duranton, V. Henderson and W. Strange, eds, ‘Handbook of Urban and Regional Economics’, Vol. 5A, North-Holland, Amsterdam.
- Démurger, S., Gurgand, M., Li, S. and Yue, X. (2009), ‘Migrants as second-class workers in urban China? a decomposition analysis’, *Journal of Comparative Economics* **37**(4), 610–628.
- Démurger, S., Li, S. and Yang, J. (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 – 153.
- Duranton, G. and Puga, D. (2004), Micro-foundations of urban agglomeration economies, in V. Henderson and J.-F. Thisse, eds, ‘Handbook of Regional and Urban Economics’, Vol. 4, North-Holland, Amsterdam, pp. 2063–2117.
- Harris, C. (1954), ‘The market as a factor in the localization of industry in the United States’, *Annals of the Association of American Geographers* **44**(4), 315–348.
- Lewis, E. and Peri, G. (2015), Immigration and the economy of cities and regions, in G. Duranton, V. Henderson and W. Strange, eds, ‘Handbook of Urban and Regional Economics’, Vol. 5A, North-Holland, Amsterdam.
- Li, S., Luo, C., Wei, Z. and Yue, X. (2008), The 1995 and 2002 household surveys: Sampling methods and data description, in B. Gustafsson, S. Li and T. Sicular, eds, ‘Inequality and Public Policy in China’, Cambridge University Press, pp. 337–353.
- Marshall, A. (1890), *Principles of Economics*, Macmillan, London.
- Moretti, E. (2013), ‘Real wage inequality’, *American Economic Journal: Applied Economics* **5**(1), 65–103.
- Phimister, E. (2005), ‘Urban effects on participation and wages: Are there gender differences?’, *Journal of Urban Economics* **58**(3), 513–536.
- Rosenthal, S. S. and Strange, W. C. (2008), ‘Agglomeration and hours worked’, *Review of Economics and Statistics* **90**(105-118).



## APPENDIX

### A First-step estimates

Table 3: Individual wage disparities - OLS estimates for the first step

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

*Notes:* Estimations on individuals aged 16 to 70, 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. <sup>a</sup>: p<0.01, <sup>b</sup>: p<0.05, <sup>c</sup>: p<0.10.

## B Estimation for cities common to 2002, 2007, and 2013

Table 4: City determinants of total and per hour earnings - Cities common to 2002, 2007 and 2013

	Total earnings					Hourly earnings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.166 <sup>b</sup> (0.069)	0.112 <sup>b</sup> (0.049)	0.134 <sup>a</sup> (0.042)	0.140 <sup>a</sup> (0.044)	0.155 <sup>a</sup> (0.046)	0.130 <sup>a</sup> (0.034)	0.137 <sup>a</sup> (0.035)
Density 2007	-0.066 (0.103)	-0.099 (0.073)	-0.101 (0.061)	-0.106 (0.065)	-0.134 <sup>c</sup> (0.075)		
Density 2002	-0.194 <sup>c</sup> (0.101)	-0.141 <sup>b</sup> (0.069)	-0.158 <sup>a</sup> (0.058)	-0.172 <sup>a</sup> (0.062)	-0.210 <sup>a</sup> (0.069)	-0.118 <sup>a</sup> (0.043)	-0.147 <sup>a</sup> (0.052)
Migrants		0.326 <sup>a</sup> (0.112)	0.187 <sup>c</sup> (0.097)	0.173 <sup>c</sup> (0.100)	0.206 <sup>c</sup> (0.107)	0.119 (0.088)	0.137 (0.091)
Migrants 2007		0.366 <sup>b</sup> (0.170)	0.281 <sup>c</sup> (0.142)	0.296 <sup>c</sup> (0.156)	0.267 (0.179)		
Migrants 2002		0.390 <sup>b</sup> (0.165)	0.374 <sup>a</sup> (0.137)	0.411 <sup>a</sup> (0.147)	0.332 <sup>b</sup> (0.162)	0.443 <sup>a</sup> (0.133)	0.414 <sup>b</sup> (0.159)
Land area			0.088 <sup>a</sup> (0.024)	0.111 <sup>a</sup> (0.039)	0.121 <sup>a</sup> (0.041)	0.077 <sup>a</sup> (0.026)	0.099 <sup>a</sup> (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 <sup>c</sup> (0.077)	0.145 <sup>c</sup> (0.079)	0.077 (0.145)	0.140 (0.093)	0.236 <sup>c</sup> (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 <sup>c</sup> (0.032)
Diversity			-0.242 <sup>b</sup> (0.098)	-0.258 <sup>b</sup> (0.101)	-0.354 <sup>b</sup> (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
$R^2$	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. <sup>a</sup>: p<0.01, <sup>b</sup>: p<0.05, <sup>c</sup>: p<0.10.