Prepared for the 2016 ASSA-AREUEA Conference

Do Rural Migrants Benefit from Urban Labor Market Agglomeration Economies? Evidence from Chinese Cities

(Preliminary draft)

Guangliang Yang
School of Economics
Fudan University
E-mail: yangguangliang08@gmail.com

Lixing Li
China Center for Economic Research, National School of Development
Peking University
E-mail: lilixing@nsd.pku.edu.cn

Shihe Fu
Research Institute of Economics and Management
Southwestern University of Finance and Economics
E-mail: fush@swufe.edu.cn

Abstract: We combine the 2005 China Inter-Census Population Survey data and the 2004 China Manufacturing Census to test whether workers, particularly rural migrants, benefit from labor market Marshallian externalities in manufacturing industries in Chinese cities. We find that workers in general, and rural migrants in particular, benefit from labor market pooling effect (measured by total employment in a city-industry cell) and human capital externalities (measured by share of workers with a college degree or above in a city-industry cell). These findings are robust to various sorting bias tests. However, rural migrants benefit much less compared with local or urban workers, possibly because rural migrants suffer from lack of social network and from double "discrimination" in terms of being both "rural" and "migrants." Our findings have policy implications on how Chinese cities can attract skilled workers during the rapid urbanization process coupled with global competition.

Key words: Rural migrants; labor market agglomeration economies; Marshallian externalities; labor market pooling; human capital externalities

JEL Code: J30; J61; J71; O18; R23

1. Introduction

A firm can benefit from the concentration of firms in a city; such benefit is called business agglomeration economies. Business agglomeration economies are generally classified into two types: localization economies resulting from the concentration of same-industry firms in a city and urbanization economies resulting from the concentration of different-industry firms in a city. In the dynamic context, the counterparts of localization and urbanization economies are dubbed Marshallian externalities and Jacobs externalities (Glaeser, et al., 1992; Henderson, 2003; Rosenthal and Strange, 2004). Likewise, a worker can benefit from the concentration of workers in the same workplace; such benefit is called labor market agglomeration economies. Similarly, benefit from the concentration of same-industry workers is called labor market Marshallian externalities, while benefit from the concentration of industry diversity and general city scale is called labor market Jacobs externalities (Fu, 2007; Andini, et al, 2012; Groot, et al., 2014). These labor market agglomeration economies occur in employment clusters through labor market pooling, information exchange and knowledge spillovers, and networking (Andini, et al., 2012; Moretti, 2004a). Such benefit can enhance workers' productivity and therefore wages if workers are paid by their marginal product revenue.

Most of existing empirical evidence focuses on labor market Jacobs externalities, particularly, city size wage premium. Yannow (2007) finds that large city size is associated with positive wage premium and tests various channels of these agglomeration economies. Glaeser and Maré (2007) find that workers receive higher wages in cities possibly because cities help accumulate human capital through learning and knowledge spillovers. Fu and Ross (2013) find that even taking into account sorting of high ability workers urban wage premium still exists. Rosenthal and Strange (2006) and Moretti (2004a, 2004b) also have similar findings. Only a few studies test the labor market Marshallian externalities.[1] Fu (2007) finds that workers benefit from the concentration of same-industry or same-occupation workers in the Boston metropolitan area. Groot, et al. (2014) find labor market Marshallian externalities in the Netherlands.

Most of the empirical studies on labor market agglomeration economies use data from developed countries, such as USA, UK (Melo and Graham, 2009), Italy (Andini, 2012). Studies on labor market agglomeration economies in developing countries, including China, are very rare.[2] [3] China is having a rapid urbanization and many cities have been

[1] Many studies test Marshallian externalities in business sectors (Henderson, 2003; Rosenthal and Strange, 2001; Ellison, et al. 2010).

[2] There are quite a few studies on business agglomeration in developing countries, for a review, see Combes and Gobillon (2014).

[3] Maria Bonomi Barufi (2014) find that high employment density is associated with high wage in the Brazilian formal labor market. Zhang and Zhao (2014) estimate the income-distance elasticity of rural migrants in China and find that migrants are willing to pay for living in big cities.

growing persistently since their transition to a market economy in 1980s. Each year a large number of migrants move from countryside to cities to work and some of them have become permanent urban residents. Given that there are still institutional barriers to migration, such as the residential registration (*hukou*) system preventing farmers from moving into cities freely and a high degree of local government intervention to cities, it is natural to ask such questions: do labor market agglomeration economies exist in Chinese cities or in general in cities of developing countries? If so, how large is the effect? Do rural migrants also benefit from labor market agglomeration economies given the massive rural-urban migration and rapid urbanization?

Using data from the 2005 China Inter-Census Population Survey and the 2004 Manufacturing Census and following the standard wage model specifications in the literature, we find that in general workers benefit from the labor market Marshallian externalities in Chinese cities: the labor market pooling effect, measured by total employment in a two-digit industry in a city, and human capital externalities, measured by the share of workers who have a college degree or above in a two-digit industry in a city.

The key identification issue is that workers may sort into different cities and industries based on unobserved city, industry, and individual attributes. We define agglomeration variable at the city-industry cell and are able to control for city and industry fixed effects. To deal with the issue of workers' sorting based on unobserved individual ability, we follow the literature on using occupation attributes as proxy for skills (Bacolod, Bernardo, and William, 2009) and add worker occupation dummies as a proxy for unobserved worker ability. In addition, we split the sample by local residents who never moved and migrants, by young and old workers, to test the robustness of our estimates. Reassuringly, for subsamples of workers who are less likely to sort across cities and industries, our agglomeration estimates are very robust, particularly, for human capital externalities.

Furthermore, we find that rural migrants also benefit from labor market agglomeration economies but benefit much less compared with workers with an urban *hukou* or local workers who never moved. This is not because most of rural migrants are low skilled preventing them from reaping fully the benefit from agglomeration, neither because most of rural migrants work in informal sectors which generate little spillovers. We find that even among high-skilled workers sample, rural migrants receive much less benefit from agglomeration. We conjecture that this may be because rural migrants lack local social network or they suffer from "double discrimination" in urban labor markets for being "rural" and being "migrants."

That rural migrants benefit from labor market agglomeration but benefit much less than do urban workers and local workers has important policy implications. The growth of urban population in China is mainly driven by rural-urban migration. Migrants provide massive labor force for growing manufacturing industries. But the low labor cost advantage of Chinese manufacturing industries is waning and the new trend of manufacturing industry development requires skilled labor. In addition, for cities to grow persistently, skilled workers are important (Glaeser, 2005; Simon and Nardinelli, 2002). However, majority of rural-urban migrants are low skilled. How can Chinese cities and manufacturing industries attract skilled workers? Skills can be acquired by formal school education and by informal social interactions such as information exchange, imitating, learning by doing, etc. Although on-the-job training and adult continuation education are feasible, they are very limited in scale; employment agglomeration provides a feasible channel for social interactions occurring on daily bases to promote knowledge spillovers and learning. If agglomeration benefit is important, this may be the way through which Chinese cities become more skilled. Local governments therefore may design labor and urban development policies to help rural migrants benefit more from urban labor market agglomeration to improve skills and accumulate human capital.

The rest of the paper is organized as follows: Section 2 introduces the data; Section 3 discusses the model specification and identifications; Section 4 presents the empirical results and Section 5 concludes.

2. Data

We use two datasets to test labor market agglomeration economies in Chinese cities. The first dataset is based on the 2005 Inter-Census (1%) Population Survey conducted by the State Bureau of Statistics of China. We obtained a random sample of 0.5% sample size. The population census data has a structure similar to the U.S. decennial census, including individual person's characteristics, household information, and labor market performance information. The second dataset is the 2004 manufacturing census, which surveyed all firms in manufacturing industries by the State Bureau of Statistics of China. It includes firm location, total employment by education, accounting and financial, and other firm characteristics variables and enables us to calculate precise measure of total employment by industry and education. We merge the two datasets by city-industry cell, where industry is defined at the two-digit level. We then construct a set of agglomeration variables at the city-industry level for each worker based on where and in which industry a worker was working. Our approach is very similar to Moretti (2004b) where he merges the U.S. firm census data with the U.S. decennial population census data by metropolitan area-industry cell to estimate the effect of human capital externalities on firm productivity because education information is not in the firm census data but in the population census data. Using city-industry cell link can also enable us to test labor market Marshallian externalities without dealing with worker sorting across cities and across industries since we can control for city and industry fixed effects and our identification comes from variations across city-industry cells.

Wage is defined by annual labor income or salary divided by months worked, so our wage variable is monthly wage. To remove the influence of outliers, we winsorize the wage data at the top and bottom 0.25 percentile. we select only workers of primary working age (between 18 and 60). To ensure there are enough number of workers in each city-industry cell, we require at least 20 workers in a city-industry cell. Increasing the cutoff improves our estimation since this reduces measurement errors of our agglomeration variables. After dropping observations with missing values of required variables, we finally obtain a sample of 172,002 workers. This sample contains 35 industries, 71 occupations, 345 cities, 7,832 city-industry cells. On average there are 33 two-digit manufacturing industries in a Chinese city.

We construct two agglomeration variables for each city-industry cell. The total employment in a city-industry cell measures intra-industry labor market pooling effect in a city, or how a worker can benefit from the concentration of the same industry workers in a city. College share is calculated by the number of employees with an associate degree or above in a city-industry cell divided by the total employment in that city-industry cell, measuring intra-industry human capital externalities or knowledge spillover effects in a city.[4] Table 1 provides the summary statistics for our key variables.

3. Econometric Model Specification and Identification

Following the literature on labor market agglomeration economies, we specify our baseline model as follows:

$$\log W_{ijk} = \alpha_k + \beta_1 X_i + \beta_2 A_{jk} + \beta_3 H_{jk} + \varepsilon_{ijk}, \qquad (1)$$

where W_{ijk} is the monthly income (wage) of work *i* working in industry *j* in city *k*. Independent variables are defined as follows.

α_k : city fixed effect, used to control for unobserved city attributes based on which workers may sort across cities.

X_i : individual attributes, including standard variables in a wage equation such as age, age squared, gender, marital status, years of migration, education attainment (high school, associate, college, and master degree or above), minority identity, and variables of

[4] An associate degree refers to graduation from a two or three year college.

Chinese characteristics that may affect individual wage, including whether a worker has an urban *hukou*, types of employers, and types of labor contract.[5]

A_{jk}: the logarithm of total employment in manufacturing industry *j* in city *k*. This measures labor market pooling effect within an industry in a city: how a worker benefits from the concentration of the same industry workers in a city. [6]

H_{jk}: the share of employees with an associate degree or above in industry *j* in city *k*. It measures human capital externalities or knowledge spillover effects within an industry in a city.

ε_{ijk}: error term, may not be independent and identically distributed.

A few identification issues arise in the estimation of the coefficients of the two agglomeration variables: total employment and college share in a city-industry cell. The key concern is that workers may sort into different workplaces or industries based on some unobserved factors. We discuss them in turn.

First, workers may sort into different cities based on unobservable city attributes. Also cities may have different productive and consumption amenities that affect workers' productivity and residential location choices. We include city fixed effects to control for this.

Second, agglomeration economies may be specific to industries. For example, high-tech industries tend to generate stronger knowledge spillover effect (Henderson, 2003) while informal sectors are less likely to generate spillovers. We include industry fixed effects to control for this.

Third, after controlling for city and industry fixed effects, sorting across industries in a given city is less likely simply because there are not many industries available in a given city. This reasoning is similar to Bayer et al. (2008) in that residents are less likely to sort across residential locations at the block level simply because housing markets at the block level is thin. However, such sorting could still occur due to unobserved individual ability. For example, workers with better local social network are more likely to work in industries

[5] Minority dummy is set to 0 if a worker belongs to *Han* and 1 if a worker belongs to non-*Han* ethnicity. Employer types include social organizations and public sector, state-owned enterprises, collectively-owned enterprises, proprietary, private enterprises, and others. Labor contract types include fixed-term contract, long term contract, and no contract.

[6] Alternatively, we can use total employment. But if we assume the production function is of Cobb-Douglas type and wage equals the marginal product of labor, then the logarithm of total employment is a preferred specification.

where firm performance is more stable and employee fringe benefit is generous. Existing studies have used different approaches to deal with this issue. For example, Glaeser and Maré (2006) use individual panel data and individual fixed effects to control for unobserved persona attributes; Rosenthal and Strange (2006) use geographic features as instrumental variables (IV) for agglomeration variables to break the correlation between unobserved individual attributes and agglomeration; Moretti (2004c) uses a city's historical demographic structure and the presence of a land-grant college as IVs. Fu and Ross (2013) used residential location at the census tract level as proxy for unobserved ability.[7] Due to data constraint, we cannot employ any of them. However, a growing literature uses occupation attributes to proxy for skills (Bacolod, et al., 2009; Bacolod, et al., 2010). We follow this literature and use a worker's occupation to proxy for the worker's unobserved ability. A rule of thumb to test this method is to see if coefficients of education category variables become attenuated after inclusion of occupation fixed effects since in general unobserved ability should be positively correlated with observed ability such as education attainment.[8]

Fourth, even if we include city, industry, and occupation fixed effects, these controls may not be perfect and it is still possible that workers sort into different city-industry cells based on unobservables. Since workers with a high degree of mobility are more likely to sort across cities and industries than those with a low degree of mobility, we split the sample into migrants and local residents who never moved. If both types of workers benefit similarly from agglomeration, then sorting bias is not a serious problem.

Fifth, China has transitioned from a planned economy to a market economy. Older workers who experienced the centrally planned economy and the transition period face more stringent mobility constraints: they are more likely to be affiliated with state-owned enterprises, having more family dependents, and with different human capital and skills that may not be easily transferable or adapted to market economy. That is, old workers tend to have low mobility, while young workers tend to have high mobility. Sorting bias should be stronger in the younger sample. We also estimate the models for young and old workers sub-samples separately to gauge the seriousness of sorting bias issue.

Because our wage data is in 2005 and the agglomeration variables are measured in 2004 from the manufacturing census data, it is the lagged labor market agglomeration that generates current wage premium. This data structure also mitigates the endogeneity issue that agglomeration and worker wage are simultaneously determined.

Model 1 is estimated for the full sample but also for subsample of rural migrants. Complementary to the sample splits, we also interact rural and migrant dummies with

[7] Recent studies have used randomized experiments (Afridi, et al., 2015).
[8] Arguably occupation choice is endogenous in a wage model but we can do no better here.

agglomeration variables to check the robustness of results. The next section reports and discusses all empirical results.

4. Empirical Results

4.1 Existence of labor market agglomeration economies

To test whether labor market agglomeration economies exist in Chinese cities, we estimate different versions of Model (1). Table 2 presents the estimation results. Column 1 reports the result of a simple wage model with individual attributes, agglomeration variables, and city fixed effects. The coefficients of the individual attributes variables have expected signs and reasonable magnitudes. Both the coefficients of total employment and college share in a city-industry cell are positive and statistically significant, suggesting the existence of Marshallian externalities in urban labor markets. But this result may be driven by industry specific attributes or individual sorting by unobserved ability.

We add industry fixed effects to Column 1 and the coefficients of individual attributes variables remain almost identical (Column 2), suggesting that there is little sorting across industries based on observed individual attributes. However, the coefficient of logarithmic total employment attenuates from 0.0052 to 0.0015 and becomes insignificant, suggesting that agglomeration benefit from same-industry peers in a city may be mainly industry-specific. The human capital externalities effect remains highly significant albeit attenuated by about 30% (from 0.5044 to 0.3431). These results suggest that after controlling for industry-specific factors, substantial agglomeration economies still exist.

To test whether workers may sort across industries in a given city based on unobserved ability biasing the estimates of agglomeration variables, we add occupation fixed effects to Column 1, aiming to capture unobserved worker ability. Column 3 presents the result. As discussed in the previous section, if occupation fixed effects can absorb part of unobserved ability and since observed and unobserved ability should be positively correlated, then, the coefficients of education variables should attenuate significantly. Column 3 indeed shows that the coefficients of education variables do attenuate by between 15% and 24%. For example, the coefficient of college degree dummy decreases by 20%, from 0.7560 to 0.6032. This pattern is consistent with the findings in Fu and Ross (2013) where residential fixed effects are used to proxy for unobserved worker ability. After controlling for unobserved ability, human capital externalities effect still exists.

Column 4 further adds both industry and occupation fixed effects. This baseline model specification is preferred since we have controlled for city- and industry-specific attributes and unobserved ability. The result suggests that although not statistically significant, doubling the employment size of an industry in a city increases the wage of a worker in that industry by 0.29%. This is a very conservative estimate since industry employment is very likely to have measurement errors and individual sorting may not be perfectly controlled for. [9] On the other hand, human capital externalities effect remains important and significant: a one percentage point increase in college share of an industry-city cell raises a worker's wage by about 0.36% (or a one standard deviation increase in college share in an industry-city cell raises wage by about 3.06%). (to be added: compare estimates with existing studies from US.)

To check the robustness of our preferred specification, we estimate the baseline model by a set of sample splits: local residents who never moved versus migrants, migrants with below or above median migrating years (2.5 years), and workers with below or above median ages (33 years old). Compared with migrants, inexperienced migrants, and young workers, local residents, experienced migrants, and old workers should have less sorting due to stronger social network and more attachment to family and housing. The results presented in Table 3 confirm this conjecture. Column 1 replicates the baseline model result of Column 4 in Table 2. Columns 2 and 3 show that compared with migrants, local workers benefit more from labor market pooling and human capital externalities, possibly due to their stronger localized social network. Columns 4 and 5 show that compared with new migrants, experienced migrants benefit more from both labor market pooling and human capital externalities, suggesting that working in cities longer may help develop social network and accumulate human capital through learning from peers. Similar patterns hold for young versus old workers as indicated in Columns 6 and 7.

Taking together the results from Tables 2 and 3, we conclude that there exist economically important labor market Marshallian externalities in manufacturing industries in Chinese cities. Specifically, there is weak evidence for labor market pooling effect but strong and robust evidence for human capital externalities effect in Chinese urban labor market.

4.2 Rural migrants benefit from labor market agglomeration economies

Given the massive migration from rural areas to cities during the past decades and many of the rural migrants work in manufacturing industries, it is natural and important to test whether rural migrants benefit from agglomeration economies in urban labor markets in China. We estimate the baseline model for rural migrant sample. Column 1 of Table 4 reports the result and shows that rural migrants benefit significantly from labor market Marshallian externalities. Specifically, Column 1 shows that for rural migrants working in a two-digit manufacturing industry in a city, doubling the industry employment size in that city increases the wage by 1.23%; a ten percentage points increase in college share in that

[9] When we select city-industry employment greater than 300 workers to reduce measurement errors, the coefficient of log(Employment) is 0.0068 and significant at the 10% level, and the coefficient of college share remains similar (0.3969 significant at the 1% level).

industry located in that city increases the wage by 2.1%. These findings are robust to subsamples of rural migrants with less or more migrating experience (Columns 2 and 3) and to subsamples of young or old workers (Columns 4 and 5), except that human capital externalities effects are not statistically significant for experienced rural migrants. To summarize, rural migrants in general also benefit from labor market agglomeration economies in Chinese cities.

A recent study by Yu, et al. (2015) using the 2007 China Household Income Project data finds that rural migrants who have worked in other provinces are more likely to become entrepreneurs when they return home, suggesting that rural migrants have accumulated human capital in cities. This is consistent with our findings.

4.3 Rural migrants benefit less from labor market agglomeration economies

A careful comparison reveals that although rural migrants benefit from labor market agglomeration economies in Chinese cities, they benefit much less compared with workers with an urban *hukou* or workers who are local. Table 5 presents a set of such results. To facilitate comparison, Columns 1 and 2 replicate the baseline results for the full sample (same result of Column 2 of Table 2) and for the rural migrants sample (same result of Column 1 of Table 4). It is striking that rural migrants benefit 42% less than the full sample in terms of human capital externalities. Such under-compensation persists compared with workers with an urban *hukou* (Column 3), workers who are local residents in a city regardless of *hukou* (Column 4), workers who are local urban residents (Column 5), and urban workers moving across cities (Column 6). These results hint that rural workers benefit much less from human capital externalities in two dimensions—being "rural" and being "migrants".

In terms of benefit from labor market pooling, although it is not informative to compare the rural migrants sample with the full sample, it is straightforward to see that rural migrants benefit 30% to 40% less compared with workers with an urban *hukou* (Column 3), workers who are local urban residents (Column 5), and urban workers moving across cities (Column 6). Again, this pattern hints that rural workers benefit much less from labor market pooling in an industry in a city in two dimensions—being "rural" and being "migrants".

To make use of the full sample information, we create four dummy variables for these four worker categories： urban migrants, rural migrants, local urban, and local rural workers. We choose local urban workers as the default category and interact the other three category dummies with the two agglomeration variables while keeping their main effects. Column 7 presents the results. For labor market pooling effects, local urban workers benefit significantly: doubling industrial employment size increases their wage by 1.18%; workers with an urban *hukou* and moving across cities (urban migrants) enjoy similar benefit (0.9%); however, local rural workers suffers a penalty of 1.38% less compared with local urban workers; more strikingly, rural migrants suffers a penalty of 2.40% less compared with local urban workers. A similar pattern holds for human capital externalities: a ten percentage points increase in college share in a city-industry cell increases the wage of local urban workers by 5.5%; urban migrants enjoy similar benefit; however, both local rural workers and rural migrants suffer 6% less compared with local urban workers. All these results again suggest rural migrants receive an under-compensation from labor market agglomeration economies because of their being "non-urban" and "non-local."

4.4 Why do rural migrants benefit less from labor market agglomeration economies?

Why do rural migrants benefit much less from labor market agglomeration economies compared with local, urban residents? We cannot fully answer this question in this study but it seems there are at least three possible explanations. First, most of rural migrants are low skilled with less education attainment (98.56% of rural migrants finished high school or less education), which may prevent them from learning from other workers in the same industry. [10]Second, rural migrants lack social network in cities preventing them finding better jobs and learning from spillovers. Third, rural migrants are discriminated in the urban labor markets because they are non-urban and non-local. We are unable to test the second and the third but can rule out the first.

To test whether low education attainment hinders rural migrants reaping fully the benefit from labor market agglomeration economies, we estimate the baseline model for low-skilled and high-skilled worker samples where high skilled is defined as having an associate degree or above and low skilled is defined as having a high school diploma or below. Table 6 reports the results. Column 1 shows that although the benefit from labor market pooling is not statistically significant, low-skilled workers do benefit significantly from human capital externalities (coefficient is 0.3112), more so than rural migrants (coefficient is 0.2095). Column 2 shows that for low-skilled worker sample, being "rural" deprives almost all the benefit from both types of agglomeration economies. Column 3 further shows that being "migrant" (being "non-local") wipes out all benefit from labor market pooling, possibly due to lack of local social network. Columns 4-6 presents the same set of results for high-skilled workers. Although for high-skilled workers, being "migrant" does not affect their ability to reap agglomeration benefit, being "rural" is very harmful, making the benefit from agglomeration economies completely disappear. Interactive models using local urban workers as default reveal the same pattern that rural migrants, whether they are low-skilled or high-skilled, receive much less benefit from agglomeration economies (Columns 2 and 3 in Table 7). Taking together, we can infer that it is not education disparity, but the disparity in terms of rural and urban *hukou*, nonlocal

[10] Many rural migrants work in informal sectors and there sectors generally generate little knowledge spillovers. Our model specifications have included industry fixed effects and can rule out this interpretation.

and local residence that prevent rural migrants reaping the full benefit from urban labor market agglomeration economies. This is consistent with the literature on discrimination against rural *hukou* migrants (Afridi, et al., 2015; need to add more citations).

Rural migrants generally lack local social network compared with local urban residents; they may be discriminated by urban residents and employers. [11]Either or both can explain why rural migrants benefit much less from labor market agglomeration economies than do local urban workers. We cannot distinguish or test these two hypotheses in the current study but anecdotal evidence tends to support the discrimination hypothesis. (cite reference, due to culture and value difference in rural and urban area…)

5. Conclusion

Labor market agglomeration economies have attracted much attention in developed countries. This paper offers complementary evidence for labor market agglomeration economies in Chinese cities. We find that in general workers benefit from Marshallian externalities, including intra-industry labor market pooling effect and human capital externalities in cities. This finding is robust to various sorting bias tests. We also find that rural migrants also benefit from labor market agglomeration economies, but they benefit much less than do workers who have an "urban *hukou*" or who are local. We provide evidence to show that this is not because rural migrants are generally low-skilled preventing them from reaping fully the benefit from labor market agglomeration economies. The two alternative interpretations are that rural migrants lack local social network and that they are discriminated by local urban residents and employers. Testing these two hypotheses warrants future studies.

That rural migrants cannot benefit fully from urban labor market agglomeration economies has important policy implications. The growth of population and employment in Chinese cities is mainly driven by massive rural-urban migration, implying that the skill intensity in many Chinese cities becomes diluted (cite statistics). However, globalization and outsourcing of manufacturing firms from China to other developing countries suggests that Chinese cities must become "skilled" to gain competitive advantage. It is unlikely to send most rural migrants back to school to receive formal education; however, social interactions in cities, especially in concentrated urban labor markets, provide another channel of learning and human capital accumulation. Designing policies to help rural migrants gain fully the benefit from urban labor market agglomeration to improve their productivity is a pressing task. An immediate policy suggestion would be to remove the *hukou* barrier and allow rural migrants to enjoy the same employee benefits and public

[11] Chen, et al. (2015) find that rural migrants in China receive much lower wage if they search jobs through social network.

services as do local urban residents. This would facilitate rural migrants to settle down in cities and to be integrated into urban culture and life. Fortunately, the Chinese government policies are moving toward this direction.

Reference

Afridi, F., Li, S., Ren, Y., 2015, Social identity and inequality: the impact of China's *hukou* system, *Journal of Public Economics* 123, 17-29.

Andini, M., Guido de Blasio, Gilles Duranton, William C. Strange, 2012, Marshallian labor market pooling: evidence from Italy, Working Paper.

Bacolod, M., Blume, B., Strange, W., 2009, Skills in the city. *Journal of Urban Economics* 65, 136-153.

Bacolod, M., Blum B., Strange, W., 2010, Elements of skill: traits, intelligences, education, and agglomeration, *Journal of Regional Science* 50, 245-280.

Bayer, P., Ross, S., Topa, G., 2008, Place of work and place of residence: informal hiring networks and labor market outcomes, *Journal of Political Economy* 116, 1150-1196.

Combes, P.P., Gobillon, L., 2014, The empirics of agglomeration economies, Working Paper.

Ellison, G., Glaeser, E., Kerr, W., 2010, What causes industry agglomeration? Evidence from coagglomeration patterns, American Economic Review 100, 1195–1213.

Fu, S., 2007, Smart cafe ́ cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics* 61, 86–111.

Fu, S., Ross, S., 2013, Wage premia in employment clusters: How important is worker heterogeneity? *Journal of Labor Economics* 31, 271-304.

Glaeser, Edward, HediKallal, Jose Scheinkman, and Andrei Shleifer. 1992. Growth in Cities, *Journal of Political Economy*, 100, 1126–1152.

Glaeser, Edward L., and David C. Mare ́, 2001, Cities and skills. *Journal of Labor Economics* 19, 316–42.

Glaeser, E., 2005, Reinventing Boston: 1630-2003, *Journal of Economic Geography* 5, 119-153.

Groot, S., de Groot, H., Smit, M., 2014, Regional wage differences in the Netherlands: micro evidence on agglomeration externalities, *Journal of Regional Science* 54, 503-523.

Henderson, V., 2003. "Marshall's Scale Economies," *Journal of Urban Economics*, 53, 1–28.

Maria, A., Barufi, B., 2014, Agglomeration economies in the Brazilian formal labor market, Working Paper, The University of São Paulo.

Melo, P.C., Graham, D.J., 2009, Agglomeration Economies and Labour Productivity: Evidence from Longitudinal Worker Data for GB's Travel-to-Work Areas, Spatial Economics Research Centre Discussion Paper No.31, London School of Economics.

Moretti, Enrico, 2004a, Human capital externalities in cities. In Handbook of urban and regional economics, vol. 4, ed. J. Vernon Henderson and Jacques-Francois Thisse, 2243–91. Amsterdam: Elsevier.

Moretti, E., 2004b, Workers'Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions, *American Economic Review* 94, 656-690.

Moretti, E., 2004c, Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data, *Journal of Econometrics* 121, 175-212.

Rosenthal, Stuart andWilliam Strange. 2001. "The Determinants of Agglomeration," *Journal of Urban Economics*, 50, 191–229.

Rosenthal, S., Strange, W., 2004, Evidence on the Nature and Source of Agglomeration Economies, in J.V. Henderson, and J. Thisse (eds.), *Handbook of Regional and Urban Economics*, Volume 4. North-Holland: Elsevier, pp. 2119–2171.

Rosenthal, S., Strange, W., 2006, The attenuation of human capital spillovers: A Manhattan skyline approach, *Journal of Urban Economics* 64, 373–89.

Simon, C., Nardinelli, C., 2002, Human capital and the rise of American cities, 1900-1990, *Regional Science and Urban Economics* 32, 59-96.

Yu, L., Li, W., Yin, X., Zheng, X., 2015, Lose to win: entrepreneurship of returned migrants in China, Working Paper, Central University of Finance and Economics.

Zhang, J., Zhao, Z., 2011. Measuring the income–distance tradeoff for rural–urban migrants in China, IZA Discussion Paper No.7160.

Table 1: Summary Statistics of Key Variables

	Mean	Standard deviation	Min	Max
Wage (monthly)	956	707	100	10000
Male	0.544	0.498	0	1
Age	33.310	9.980	18	60
Single	0.263	0.440	0	1
Minority	0.042	0.201	0	1
High school	0.235	0.424	0	1
Associate	0.063	0.243	0	1
College	0.027	0.163	0	1
Graduate	0.002	0.047	0	1
Urban *hukou*	0.389	0.487	0	1
Migrate year	1.211	1.922	0	6
Employee (of a city-industry)	81,805	123,834	20	773,914
College share(of a city-industry)	0.113	0.085	0	1
Sample size		172,002		

Table 2: Labor market agglomeration economies

	1	2	3	4
Male	0.2189***	0.2135***	0.2080***	0.2060***
	(30.23)	(35.78)	(36.82)	(37.85)
Age	0.0124***	0.0120***	0.0117***	0.0114***
	(9.49)	(9.78)	(10.27)	(10.04)
Age squared	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-11.17)	(-11.65)	(-12.55)	(-12.29)
Single	-0.0651***	-0.0641***	-0.0555***	-0.0554***
	(-10.17)	(-10.08)	(-10.04)	(-10.12)
Minority	-0.0562***	-0.0531***	-0.0492***	-0.0477***
	(-7.31)	(-7.11)	(-6.56)	(-6.42)
High school	0.1330***	0.1322***	0.1008***	0.1004***
	(23.57)	(22.47)	(26.64)	(26.91)
Associate	0.4284***	0.4250***	0.3226***	0.3200***
	(31.01)	(31.02)	(33.12)	(33.56)
College	0.7560***	0.7522***	0.6032***	0.6005***
	(26.67)	(26.81)	(28.84)	(29.53)
Graduate	1.3198***	1.3097***	1.1195***	1.1132***
	(29.55)	(29.48)	(31.21)	(31.45)
Urban *hukou*	0.0297***	0.0375***	0.0166**	0.0217***
	(2.51)	(3.49)	(2.20)	(2.93)
Migrate year	0.0113***	0.0114***	0.0107***	0.0107***
	(8.64)	(9.51)	(8.79)	(8.91)
Log(Employment)	0.0052*	0.0015	0.0019	0.0029
	(1.72)	(0.36)	(0.66)	(0.74)
College share	0.5044***	0.3431***	0.4621***	0.3598***
	(9.97)	(5.36)	(10.03)	(6.10)
Industry fixed effects	N	Y	N	Y
Occupation fixed effects	N	N	Y	Y
Adjusted R^2	0.39	0.41	0.43	0.44

Note: City fixed effects, employee type, and work contract type dummies are included. Standard errors are clustered at the city-industry cell level. t statistics are in parentheses. "*", "**", and "***" indicate significance at the 10%, 5%, and 1% levels. Sample size: 172,002.

Table 3: Labor market agglomeration economies robustness checks

	1 Baseline	2 Local	3 Migrants	4 <2.5 years	5 ≥2.5 years	6 Below median age	7 Above median age
Log(Employment)	0.0029	0.0111**	0.0015	-0.0018	0.0147***	0.0018	0.0071
	(0.74)	(1.91)	(0.47)	(-0.51)	(3.40)	(0.50)	(1.49)
College share	0.3598***	0.3473***	0.2812***	0.2285***	0.3454***	0.3228***	0.3847***
	(6.10)	(5.06)	(3.87)	()3.05	(3.22)	(5.26)	(5.63)
Adjusted R^2	0.44	0.44	0.43	0.39	0.46	0.44	0.45
Sample size	172,002	97,478	74,524	34,975	39,549	91,426	80,576

Note: All models include the same set of independent variables as those in Column 4 of Table 2 except that Column 2 excludes migrating years. Median migrating year is 2.5. Median age is 33 in the full sample.Standard errors are clustered at the city-industry cell level. t statistics are in parentheses. "*", "**", and "***" indicate significance at the 10%, 5%, and 1% levels.

Table 4: Rural migrants benefit from labor market agglomeration economies

	1 Rural migrants sample	2 Below median migrating years	3 Above median migrating years	4 Below median age	5 Above median age
Log(Employment)	0.0123***	0.0128**	0.0115**	0.0181***	0.0104**
	(2.70)	(2.20)	(2.26)	(3.02)	(2.19)
College share	0.2095**	0.3551***	0.1119	0.2209*	0.2381**
	(1.93)	(2.64)	(0.93)	(1.67)	(2.07)
R-squared	0.30	0.25	0.33	0.26	0.35
Sample size	49,916	23,302	266,14	25,260	24,656

Note: All models include the same set of independent variables as those in Column 4 of Table 2. Columns 2 and 3 use subsamples below or above median migrating years (cutoff is 2.5 years); Columns 4 and 5 use subsamples below or above median age (cutoff is 26). Standard errors are clustered at the city-industry cell level. t statistics are in parentheses. "*", "**", and "***" indicate significance at the 10%, 5%, and 1% levels.

Table 5: Rural migrants benefit less from labor market agglomeration economies

	1 Full sample	2 Rural migrants	3 Urban	4 Local	5 Local urban	6 Urban migrants	7 Interactive model
Log(Employment)	0.0029	0.0123***	0.0178***	0.0111**	0.0197***	0.0181**	0.0118**
	(0.74)	(2.70)	(5.23)	(1.91)	(5.60)	(2.27)	(1.97)
Log(Employment)* Urban*Migrant							0.0093**
							(1.94)
Log(Employment)* Local*Rural							-0.0138***
							(-2.79)
Log(Employment)* Rural*Migrant							-0.0240***
							(-3.21)
College share	0.3598***	0.2095**	0.3483***	0.3473***	0.3287***	0.5133***	0.5539***
	(6.10)	(1.93)	(6.10)	(5.06)	(5.39)	(3.71)	(8.81)
College share* Urban*Migrant							0.0366
							(0.49)
College share* Local*Rural							-0.6150***
							(-8.03)
College share* Rural*Migrant							-0.5883***
							(-4.76)
R-squared	0.44	0.30	0.53	0.44	0.50	0.55	0.44
Sample size	172,002	49,916	66,827	97,478	57,431	9,396	172,002

Note: All models include the same set of independent variables as those in Column 4 of Table 2. Column 7 also includes dummies for urban migrant, local rural, and rural migrant categories. Standard errors are clustered at the city-industry cell level. t statistics are in parentheses. *t* statistics are in parentheses. "*", "**", and "***" indicate significance at the 10%, 5%, and 1% levels.

Table 6: Low skilled versus high skilled worker samples

	1 Low skilled	2 Low skilled	3 Low skilled	4 High skilled	5 High skilled	6 High skilled
Log(Employment)	0.0029	0.0100*	0.0052	0.0140**	0.0174***	0.0167**
	(0.72)	(1.86)	(1.18)	(2.37)	(2.89)	(2.40)
Log(Employment)* Rural		-0.0128***			-0.0382**	
		(-3.21)			(-3.45)	
Log(Employment)* Migrant			-0.0094**			-0.0102
			(-2.02)			(-0.77)
College share	0.3112***	0.5365***	0.3220***	0.5037***	0.5168***	0.4992***
	(5.48)	(8.30)	(5.37)	(6.42)	(6.66)	(6.26)
College share* Rural		-0.5555***			-0.5765***	
		(-7.89)			(-2.49)	
College share* Migrant			-0.0740			0.0353
			(-0.81)			(0.25)
Adjusted R^2	0.34	0.35	0.34	0.52	0.52	0.52
Sample size	156,034	156,034	156,034	15,968	15,968	15,968

Note: All models include individual attributes variables, city, industry, and occupation fixed effects. Standard errors are clustered at the city-industry cell level. *t* statistics are in parentheses. "*", "**", and "***" indicate significance at the 10%, 5%, and 1% levels.

Appendix: Table 7: Interactive models by skill

	1 Full sample	2 Low skilled worker sample	3 High skilled workers sample
Log(Employment)	0.0118**	0.0077	0.0178***
	(1.97)	(1.32)	(2.53)
Log(Employment)*Urban*Migrant	0.0093**	0.0002	-0.0038
	(1.94)	(0.03)	(-0.30)
Log(Employment)*Local*Rural	-0.0138***	-0.0072*	-0.0188
	(-2.79)	(-1.64)	(-1.20)
Log(Employment)*Rural*Migrant	-0.0240***	-0.0139**	-0.0458***
	(-3.21)	(-2.38)	(-2.62)
College share	0.5539***	0.5379***	0.5086***
	(8.81)	(8.26)	(6.38)
College share*Urban*Migrant	0.0366	-0.0928	0.0528
	(0.49)	(-1.12)	(0.39)
College share*Local*Rural	-0.6150***	-0.5798***	-0.5835**
	(-8.03)	(-7.55)	(-2.00)
College share*Rural*Migrant	-0.5883***	-0.4996***	-0.5028
	(-4.76)	(-4.35)	(-1.49)
Adjusted R^2	0.44	0.35	0.52
Sample size	172,002	156,034	15,968

Note: All models include individual attributes variables, dummies for urban migrant, local rural, and rural migrant categories, city, industry, and occupation fixed effects. Standard errors are clustered at the city-industry cell level. *t* statistics are in parentheses. *t* statistics are in parentheses. "*", "**", and "***" indicate significance at the 10%, 5%, and 1% levels.