This is a repository copy of *The complex relationship between neighbourhood types and migrants' socio-economic integration: the case of urban China*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/145444/

Version: Accepted Version

This is a post-peer-review, pre-copyedit version of an article published in Journal of Housing and the Built Environment. The final authenticated version is available online at: https://doi.org/10.1007/s10901-019-09670-2

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The Complex Relationship between Neighbourhood Types and Migrants' Socio-economic Integration: The Case of Urban China

Jing Zou, Yu Chen, Jie Chen[1]

Abstract: Migrants are important parts of China's urban population, but still limited knowledge is known about the contextual determinants of their socio-economic integration in Chinese cities. Based on a large national micro-level data extracted from the 2014 China Migrant Dynamic Survey, we extend the literature on migrants' socio-economic integration by examining the relationships between neighbourhood types and different dimensions of socio-economic integration. Our multi-dimensional index of socio-economic integration covers economic, socio-cultural and identity perspectives. The results show that migrants demonstrate significantly higher levels of overall socio-economic integration when living in formal neighbourhoods (composed of commercial properties, work unit and affordable housing), compared with those residing in informal neighbourhoods (e.g. urban villages). Moreover, migrants living in commercial property neighbourhoods are best integrated economically, and those living in affordable and work unit housing neighbourhoods are more socio-culturally integrated than others. Inner city old housing neighbourhoods are positively associated with socio-cultural integration,

compared with urban villages, while both neighbourhood types are negatively correlated with identity integration. Our findings remain robust after controlling for potential endogeneity bias caused by migrants' self-selection into different neighbourhood types. We conclude that neighbourhood types do matter for migrants' socio-economic integration but there is large heterogeneity of the correlations across different migrant groups.

Keywords: neighbourhood, socio-economic integration, migrants, self-selection, urban, China

The Complex Relationship between Neighbourhood Types and Migrants' Socio-economic Integration: The Case of Urban China

Abstract: Migrants are important parts of China's urban population, but still limited knowledge is known about the contextual determinants of their socio-economic integration in Chinese cities. Based on a large national micro-level data extracted from the 2014 China Migrant Dynamic Survey, we extend the literature on migrants' socio-economic integration by examining the relationships between neighbourhood types and different dimensions of socio-economic integration. Our multi-dimensional index of socio-economic integration covers economic, socio-cultural and identity perspectives. The results show that migrants demonstrate significantly higher levels of overall socio-economic integration when living in formal neighbourhoods (composed of commercial properties, work unit and affordable housing), compared with those residing in informal neighbourhoods (e.g. urban villages). Moreover, migrants living in commercial property neighbourhoods are best integrated economically, and those living in affordable and work unit housing neighbourhoods are more socio-culturally integrated than others. Inner city old housing neighbourhoods are positively associated with socio-cultural integration, compared with urban villages, while both neighbourhood types are negatively correlated with identity integration. Our findings remain robust after controlling for potential endogeneity bias caused by migrants' self-selection into different neighbourhood types. We conclude that neighbourhood types do matter for migrants' socio-economic integration but there is large heterogeneity of the correlations across different migrant groups.

Keywords: neighbourhood, socio-economic integration, migrants, self-selection, urban, China

1. Introduction

Migrants' socio-economic integration is important for the society, as it can provide economic and cultural benefits and ensure the stability of the society (Chen and Wang 2015). Numerous

studies have been conducted in different contexts to explore the influencing factors of socio-economic integration, such as individual characteristics, family composition, institutional arrangements, local labor market and social capital (Schwarzweller 2006; Wang and Fan 2012; Shubin and Dickey 2013). However, only a limited number of studies have paid due attention to the heterogeneity of neighbourhood types as well as its roles in influencing migrants' integration in the city. Neighbourhood represents an important place for residents to participate into the wider urban society through information exchange, social interaction and access to services and facilities. It is therefore important to explore the association between neighbourhood types and migrants' socio-economic integration.

Over the last four decades, Chinese cities have experienced large-scale rural-to-urban migration and massive neighbourhood changes. According to the National Bureau of Statistics, over 245 million residents in Chinese cities were migrants at the end of 2016 (PFPC 2017). Constrained by *hukou* and other institutional barriers, migrants are marginalized as they are denied access to many local benefits and services (Démurger et al. 2009). Most of them rent housing in peri-urban areas with limited integration into the local urban society (Wang and Fan 2012; Huang and Tao 2015). Since the beginning of 2014, China has developed a new people-oriented urbanization initiative aimed at facilitating migrants' integration in the city and further promoting sustainable urban development (State Council 2014). Therefore, migrants' socio-economic integration has become an important policy concern to both central and local governments.

Only in recent years has the literature on China begun to examine the role of neighbourhood in influencing migrants' socio-economic integration. Studies have primarily focused on neighbourhood characteristics and social interaction between migrants and local residents, using data from a single city, such as Guangzhou and Nanjing (Forrest and Yip 2007; Zhu et al. 2012; Wang et al. 2016). However, little is known about the relationship between neighbourhood types and migrants' socio-economic integration in different dimensions, such as economic, social, cultural and identity integration. Furthermore, migrants are new-comers in the city, and tend to choose to live in a neighbourhood contingent on their choices and constraints in the housing market. Their self-selection into different neighbourhood types is likely to influence their socio-economic integration. However, very few studies address this significant methodological

challenge when examining the relationship between neighbourhood types and integration.

Our goal in this paper is to fill in the above gaps in the literature by examining the association between neighbourhood types and migrants' socio-economic integration in Chinese cities. The specific advantage of our study is that the Chinese urban setting, assisted by our unique dataset, enables us to explore in detail the complex relationships between a rich number of neighborhood types and various dimensions of socio-economic integration. Upon this, we attempt to answer the following two questions: to what extent have migrants living in different neighbourhood types integrated in the city differently in terms of economic, socio-cultural and identity perspectives? How does migrants' self-selection into different neighbourhood types influence the association between neighbourhood types and socio-economic integration? Drawing on data from the 2014 China Migrant Dynamic Survey, we first construct a multi-dimensional index of socio-economic integration using factor analysis. Five different types of neighbourhoods are identified according to the dominant housing tenure types and built environment, i.e. commercial properties, affordable housing, work-unit housing, inner-city old housing awaiting regeneration, and urban villages. Following previous studies (Liu et al. 2013), housing in the first three categories is regarded as formal, whilst the latter two are characterized with informal housing[2]. Formal and informal neighbourhoods are used thereafter for convenience. We then investigate the associations between neighbourhood types and migrants' socio-economic integration through a variety of model specifications.

Our work contributes to the literature in two key respects. First, different from most previous studies that focused on neighbourhood characteristics and social interaction between migrants and local residents, we explore how migrants' different dimensions of socio-economic integration vary in different neighbourhoods. Second, we employ the Propensity Score Matching (PSM) method to control for potential bias caused by migrants' self-selection into different neighbourhood types. Our paper provides evidence showing that different neighbourhood types matter for migrants'

---

[2] According to previous studies, informal housing settlements are characterized with substandard housing units, over-crowding, insecure housing tenure and insufficient service provision (Zheng et al. 2009; Wang et al. 2010; Wang 2011; Liu et al. 2013; Ouyang et al. 2017). In our paper, inner-city old housing awaiting regeneration has the above features including obsolete houses with minimal living space, traffic congestion and overcrowding. Housing tenure is not secure due to demolition and regeneration in the near future. Therefore, it belongs to informal housing. It does not overlay with inner-city work-unit housing that is still in good building conditions.

socio-economic integration in Chinese cities. The study would not only shed light on China's urban development policy making but also give useful reference for other countries.

The remainder of the paper is arranged as follows. Section 2 reviews existing studies on socio-economic integration and provides contextual information on migrants' neighbourhood choices and integration in Chinese cities. This is followed by the discussions of methods and data. Our empirical findings and robustness checks are presented in Section 4. Section 5 provides further discussion and section 6 concludes with a brief summary and policy implications.

2. Literature review

2.1 Previous studies on socio-economic integration

With massive international migration in Europe and US and rapid rural-to-urban internal migration in developing countries such as India and China, the socio-economic integration of migrants in receiving societies has been a significant policy challenge all over the world (Goldlust and Richmond 2006; Robinson 2010; Hainmueller et al. 2016). Socio-economic integration is an important part of a wider concept of integration. Earlier literature equates integration with assimilation which refers to the processes in which migrants gradually abandon their original customs and adapt to the mainstream society, especially over generations (Park 1928; Gordon 1964). Portes and Zhou (1993) challenge the traditional assimilation theory by pointing out three different trajectories of migrants' integration into the main society, i.e. upward mobility, downward mobility and economic integration without social integration. They raise the concept of segmented assimilation by arguing that some migrants may integrate well economically but keep their traditional culture and custom. A major critique about the assimilation theory is that it regards integration as a one-way process, i.e. ethnic minorities adopt the culture and customs of the main society and achieve similar opportunities to the ethnic majority. Recent studies regard integration as a two-way process where migrants and the receiving society adapt to each other to achieve both social inclusion and cultural diversity (Shubin and Dickey 2013). Various policies following the two-way concept are implemented in countries such as UK and Netherland.

There are different measures of integration. Gordon (1964) believes that integration includes seven aspects: cultural exchanges, intermarriage, structural assimilation, ethnic identity, value and

power conflicts, discrimination and prejudice. Forrest and Kearns (2001) claim that integration covers social network and social capital, social order and social control, place attachment and identity. Kearns and Whitley (2015) examine integration from three dimensions: trust reliance and safety, social relations and sense of community. Ager and Strang (2008, p. 166) provide a comprehensive framework of integration which includes 'achievement and access across the sectors of employment, housing, education and health; assumptions and practice regarding citizenship and rights; processes of social connection within and between groups within the community; and structural barriers to such connection related to language, culture and the local environment'. Among them, socio-economic integration is an important domain which is crucial for migrants' life at destination.

Empirical studies have identified numerous factors that affect migrants' socio-economic integration, such as socio-demographic characteristics and institutional systems. For instance, Intermarriage has a positive association with integration (McCaa 1989). Functional factors such as educational attainment, employment, ability to speak local dialect have positive effects on integration (Robinson 2010). Schneeweis (2011) finds that years of schooling and early education are positively correlated with the integration of students with migrantion background. Piller and Takahashi (2011) indicate that linguistic assimilation facilitates integration of migrants and contributes to social cohesion. Furthermore, social interaction and trust among different groups can promote their social and economic integration (Letki 2008; Laurence 2011). However, close ties with family members at hometown can hinder the integration of young rural migrants at destination (Schwarzweller 2006).

2.2 Migrants' integration in urban China

Different from international migration discussed above, migrants in China are defined as internal migrants who do not possess local household registration (hukou) status at destination. Most of the migrants share the same ethnicity as local urban residents. However, due to strict hukou constraints, migrants cannot obtain local hukou status automatically after they migrated from one place to the other. Without local hukou status, they are excluded from access to local social benefits and services, and discriminated against in terms of occupational attainment,wages, and application of subsidized housing (Démurger et al. 2009; Chen 2011). Thus, the integration

theories discussed in Section 2.1 are relevant to China because Chinese migrants are still confronted with many difficulties when integrating into the urban society as a result of insititutional discrimination.

Wang et al. (2015) identify four dimensions most relevant to Chinese migrants as economic, social relation, cultural and psychological integration. Similarly, Lin et al. (2017) discuss integration from the perspectives of acculturation, integration willingness, social insurance, economy and social communication. Chen and Wang (2015) show that education, labour market outcomes, social interactions and migration distance significantly influence migrants' social integration in Shanghai. Lu et al. (2013) reported that social capital are crucial for economic integration of migrants in urban China. Based on data from the rural-urban migrants survey in Fujian Province in 2009, Yue et al. (2013) further indicate that social networks with local residents significantly falitate migrants' acculturation, socio-economic and psychological integration. In addition, Wei and Gao (2017) find that the use of social media can promote social integration of new urban migrants by developing their social networks and participating in the society.

Besides these factors, many studies agree that the hukou institution remains an important barrier to migrants' integration in cities. For example, Wang and Fan (2012) utilize the data of a questionnaire survey of 1100 migrants in Wuhan in 2008 and find that the household registration system is a persistent obstacle to migrants' socio-economic integration. The results suggest that local household registration significantly increases indidviduals' income and the probability of identity integration. Rural and urban migrants may have lower trust in the local government than local residents, which is related to the social exclusion of migrants due to the hukou institution (Niu and Zhao 2018).

2.3 Migrants' neighbourhood choices and integration in urban China

The role of neighbourhood in influencing migrants integration represents an emerging interesting area for research, because the residential environment of a neighbourhood, both physical and social, tends to influence migrants' participation into the wider urban society. A neighbourhood in the Chinese context refers to the community blocks consisting of residential buildings with similar design and open space, with or without a gate. There have been enormous neighbourhood changes as a result of the post-1980 housing reforms, inner-city regeneration and

rapid urban expansion into suburban areas. Evidence of residential segregation between migrants and local residents has been reported (Wu et al. 2014; Shen 2017). The majority of local residents live in work unit housing and commercial property neighbourhoods. Work unit housing, as the main housing type before 1980, accommodates residents who may share some attributes in common because of current/previous affiliation to work units. Commercial properties emerged after the housing reforms, and have high building standards and good access to amenities and services (Ma et al. 2018). As many migrants conduct low-paid jobs, most commercial properties are beyond their affordability. They are also excluded from purchasing affordable housing without local hukou status. Most of them rent properties in the private market with both formal and informal housing. In particular, urban villages have become popular migration destination because of its cheap housing and convenient location after urban expansion. Urban villages are usually characterized with over-crowding, inadequate facilities and insufficient service provision (Liu et al. 2013; Ouyang et al. 2017).

Neighbourhoods with different social and physical environment may play an important role in influencing migrants' socio-economic integration in the city. Interaction between different social groups might help migrants develop social networks which enable them to obtain social and economic opportunities within and beyond their neighbourhoods (Logan and Spitze 2002). For example, based on a survey of three areas in Nanjing, Wu and He (2005) find that community-based social interaction is particularly useful for marginalized residents to survive in the city. Using survey data of rural migrants in Fujian in 2009, Yue et al. (2013) reveal that migrant-resident ties facilitate migrants' socio-economic, acculturation and psychological integration. Liu et al. (2017) find that neighbourhood built environment and social support significantly influence migrants' life satisfaction.

Studies have highlighted the impacts of neighbourhood characteristics on social interaction. Using data from three neighbourhoods in Guangzhou, Forrest and Yip (2007) state that social interaction and mutual assistance are reduced when people move from older communities to commercial property neighbourhoods where residents value privacy, safety and comfort. Similar conclusions are reached in Wang et al. (2017b) which indicates that older and poorer courtyard housing and temporary shelters have higher social interaction in Nanjing, because of open space

and residents' similar socio-economic status. However, Zhu et al. (2012) find that residents in commercial property neighbourhoods have stronger neighbourhood attachment than those in traditional neighbourhoods in Guangzhou. A recent study indicates that residential segregation, e.g. high percentage of migrants in a neighbourhood, has negative effects on the perceptions of migrants' social integration (Liu et al. 2018).

In summary, studies so far have paid due attention to the relationship between neighbourhood types and social interaction. Relatively little is known about the association between neighbourhood types and migrants' overall socio-economic integration and its different dimentions. Moreover, most studies focus on one city, such as Guangzhou, Nanjing or Shanghai, and do not control for migrants' self-selection into different neighbourhood types. In this paper we will extend the literature by exploring the association between neighbourhood types and integration in different dimensions, using a national survey covering different cities. Furthermore, we use PSM estimation to control for sample selection bias, which acts as a robustness check to compensate for the deficiency of the OLS method.

3. Research Design

3.1. Methodology

Following previous studies (Yue et al. 2013; Wang et al. 2015; Liu et al. 2017), migrants' socio-economic integration in Chinese cities covers economic, socio-cultrual and identity perspectives. Our first stage is to measure socio-economic integration using factor analysis. A variety of indicators are identified to measure these different dimensions, and they are then weighted to calculate migrants' overall socio-economic integration in the host city. At the second stage, we run multivariable linear regression models to examine the determinants of migrants' overall socio-economic integration as well as the different domains. In particular, we focus on the roles of neighbourhood characteristics to see whether migrants living in different types of neighbourhoods are integrated into the urban society differently. Our benchmark OLS model is specified as follows,

$$\text{Integration} = \alpha_0 + \alpha_1 \text{Neighbourhood} + \alpha_2 X_i + \varepsilon \qquad (1)$$

The dependent variables are migrants' socio-economic integration and its different dimensions. Neighbourhood type is the core explanatory variable. X_i are control variables including socio-demographic characteristics. Following previous studies (e.g. Wang and Fan 2012; Wang et al. 2016), gender, age, education, household registration, household composition, homeownership are included into the models, as these variables are likely to influence individuals' integration experiences. Mobility patterns, such as inter-provincial or intra-provincial migration, may affect integration, as inter-provincial migrants tend to encounter more difficulties in getting familiar with local culture and customs (Chen and Wang 2015). Moreover, city dummy variables are added into the model to control for city-level variations. α_0 is the constant term; α_1 and α_2 are correlation coefficients, and ε is the error term.

One important methodological challenge concerns migrant's self-section into different neighbourhood types, which is likely to influence the relationship between neighbourhood characteristics and socio-economic integration. To control for the sample selection bias, we employ the propensity score matching (PSM) method developed by Rosenbaum and Rubin (1983). The main idea of this method is to build a counterfactual control group, in order to examine the impact of changes on key outcomes. Specifically, the control group and the experimental group are built through constructing migrants who have similar characteristics such as gender, age, education and *hukou*, but only differ in terms of neighbourhood types. Then the association between neighbourhood types and integration is analyzed by comparing the outcomes of the above two groups. The details of the PSM method are available in Appendix I. In brief, we first calculate the propensity score which is defined as 'the conditional probability of receiving a particular treatment given observed covariates' (Rosenbaum and Rubin 1983, p. 41), using logit models. We then employ three matching approaches, i.e. nearest-neighbor matching, radius matching and kernel matching, and examine the average effect of the treatment on the treated (ATT).

3.2. Data and variables

Our data come from the 2014 China Migrant Dynamic Survey, conducted by the National Population and Family Planning Committee (PFPC 2015). The survey targeted migrants aged 15-59 who had resided in the host city for over one month and did not have local hukou status.

The probability proportionate to size (PPS) sampling method was employed to select respondents in eight cities with different development levels and population sizes in different parts of the country, i.e. Chengdu, Jiaxing, Qingdao, Xiamen, Shenzhen, Beijing, Zhengzhou, Zhongshan. Among them, Shenzhen and Beijing belong to first-tier mega-cities; Chengdu, Qingdao, Xiamen and Zhengzhou are classified as second-tier cities; Jiaxing and Zhongshan are third-tier cities. The survey records information on migrants' demographic characteristics, labour market outcomes, housing and integration experience in the local society. The total sample size is 15,866 individuals.

Dependent variable

As socio-economic integration is measured from economic, socio-cultural and identity perspectives, we identify 14 variables in the survey relevant to these dimensions for the factor analysis. They are occupation type, monthly household income, access to pension and medical insurance, the number of social organizations migrants participated in, the number of social activities attended, types of neighbours, getting along with the locals, views and attitudes, familiarity with local dialects, belonging and self-identification. Descriptive statistics of these variables are displayed in Table A1 in Appendix II. After the factor analysis, the overall integration and the different dimensions are calculated, as shown in Section 4.1.

Explanatory variables

At the second stage we examine the determinants of migrants' socio-economic integration in the multivariable linear regression models. Table 1 displays a summary of the explanatory variables. About 55% of the migrants are male. They are generally young, with those under the age of 35 accounting for 64.72%. Migrants have relatively low educational attainment; 59.83% had junior high school or below, and less than 15% had been to colleges or universities. The majority hold rural *hukou* status, with only 14.02% having urban *hukou* status. About 70% of the migrants live with their spouses, and 48.39% are accompanied by their children in the city. This demonstrates that family migration has become an important feature of China's internal migration. Another feature associated with migration is that many people move across provincial boundaries, accounting for 54.67% of the sample. These migrants travelled a long distance and might encounter difference in terms of culture and customs. Consistent with previous studies, many migrants rent housing due to financial constraints and institutional obstacles, with those owning

properties accounting for only 9.93%. About 26.31% of the respondents live in neighbourhoods with formal housing, including neighbourhoods of commercial properties (17.27%), affordable housing (4.29%), and work unit housing (4.75%). However, the majority of the migrants find accommodation in informal neighbourhoods with inferior residential environment, in particular, 58.33% live in urban villages.

4. Results and findings

4.1. Measurement of integration

We conduct the Kaiser-Meyer-Olkin (KMO) and Bartlett test of sphericity to check the adaptability and appropriateness of using the 14 indicators discussed in Section 4.2. to measure migrants' socio-economic integration in Chinese cities. The results show that the KMO is 0.643 and the P value of the Bartlett test of sphericity is 0.000. Therefore, factor analysis is suitable for our analysis. In addition, Cronbach α is 0.632, and the communalities of variables are all over 0.4, except the variables of getting along with the locals (0.363) and views and attitudes (0.301). Therefore, the scale which consists of 14 indexes has good reliability and validity. All data are standardized by extremum method, and the varimax method is chosen in factor rotation. Six components are extracted whose eigenvalues are greater than 1, and they explain 63.11% of the total variance. The variance contribution rate and rotated component matrix are showed in Table A2 in Appendix II.

As is clear in Table A2, three dimensions can be extracted according to the loadings of the six components. Occupation type (X1), monthly household income (X2), access to pension (X3) and medical insurance (X4) constitute the first dimension which is related to "economic integration". Number of organizations participated (X5), number of activities attended (X6), types of neighbours (X7), getting along with the locals (X8), views and attitudes (X9), and familiarity with local dialects (X10) belong to socio-cultural integration. The third dimension is about identity integration, which is composed of belonging to the city (X11), a member of the city (X12), thinking of yourself as a hometown member (X13) and thinking of yourself as a native (X14).

According to the results of the factor analysis, the overall and different dimensions of socio-economic integration are calculated. In order to visualize the patterns, we translate the

measurements into values between 1 and 100 based on standardized scores[3] (He 2010). Note the original measurements are used for the regression analysis in Section 4.2. Migrants' average overall socio-economic integration is not very high (39.53). As to different dimensions, identity integration is the highest (54.32), followed by socio-cultural integration (50.01) and economic integration (16.92). As most migrants regard themselves as part of the city, this pushes up the measurement of identity integration, though only a small proportion of migrants identify themselves as native residents. The results are in line with Wang et al. (2015) who reported that migrants' overall integration is low, and economic integration is the lowest. It suggests that migrants adapt themselves better in socio-cultural aspects compared with economic ones. This may be explained by the fact that migrants in Chinese cities are of the same ethnicity to local residents and are not confronted with the huge cultural difference experienced by immigrants in the US and Europe.

The histograms of socio-economic integration for migrants living in formal/informal and five different types of neighbourhoods are displayed in Figures 1 and 2, respectively. Compared with migrants living in informal neighbourhoods, those residing in formal ones are significantly associated with higher levels of socio-economic integration (Fig. 1). As shown in Figure 2, migrants living in commercial property neighbourhoods have the highest levels of overall integration, economic and identity integration, whilst those in urban villages have the lowest level of socio-economic integration. Those living in work unit neighbourhoods have the highest level of socio-cultural integration. The correlations between neighbourhood types and integration are statistically significant.

4.2. Regression analysis

OLS regression results are displayed in Tables 2 and 3, with formal/informal and five different types of neighbourhoods, respectively. After we control for socio-demographic characteristics, mobility patterns and city dummies, migrants living in formal neighbourhoods show significantly higher levels of overall integration, socio-cultural and identity integration, compared with those in informal neighbourhoods. Those who live in formal neighbourhoods are

---

[3] Standardized formula: factor vale after conversion=(factor value +B) *A
A=99/(MAX(factor value)-MIN((factor value))
B=1/A- MIN((factor value))

likely to have more opportunities to interact with local residents, which may facilitate their integration in cities. When we add five neighbourhood types in the model, different patterns occur. Compared with migrants living in inner city old housing (the default category), those in commodity properties, affordable and work unit housing neighbourhoods show significantly higher levels of overall integration, whilst those in urban villages are least well integrated. In terms of economic integration, migrants in commodity properties are best integrated, followed by those living in urban villages. One explanation may concern migrants' self-selection, i.e. those who are well integrated economically, e.g. with high income and professional jobs, are likely to live in commercial properties. It is also possible that those living in commercial property neighbourhoods get more information and economic resources by interacting with neighbours of similar socio-economic status and participating in community organizations. Many migrants are engaged in business or self-employed in urban villages. Their economic integration seems not significantly lower than those living in affordable and work unit housing neighbourhoods. As to socio-cultural integration, migrants in affordable and work unit housing are more integrated than those in commercial properties. Inner city old housing is positively associated with socio-cultural integration relative to urban villages. Migrants in inner city old housing may have more opportunities to communicate with locals than those in urban villages due to their locations in the city center. Regarding identity, migrants feel more integrated in the three formal types of neighbourhood. There is no significant difference between migrants living in the two informal neighbourhood types. Living in formal neighbourhoods may serve as a sign of social status which is relevant to identity integration (Pow 2007).

In terms of individual characteristics, females are less integrated than males, especially in terms of economic and socio-cultural integration. However, their identity integration seems higher than that of male migrants, all else being equal. Age has a non-linear relationship with integration. Migrants under the age of 25 are associated with the lowest integration levels, especially regarding economic integration. Migrants aged 25-35 have higher socio-cultural integration than those under 25. As to identity integration, no significant difference exists at all ages. Consistent with previous studies, education has a significantly positive effect on migrants' integration, especially in economic and socio-cultural integration, as education provides more opportunities for economic

and social participation. Urban *hukou* status is significantly positively related to overall integration. As to household composition, both partner and children present in the host city facilitate migrants' integration. In particular, living with partners enhances migrants' economic integration, as partners usually work and contribute to household income. Living with children in the city facilitates migrants' identity integration, since these migrants are more likely to regard the destination city as their new home. Regarding mobility patterns, long-distance move as a result of inter-provincial migration has a significantly negative effect on migrants' integration in all dimensions. This is consistent with previous studies as inter-provincial migrants may encounter more difficulties in adapting themselves to local customs than those originating from the same province. Migrants who stay longer in the host city may be able to pick up local knowledge of the labour market and develop a wider social network and are therefore more economically and socially integrated. However, the effects are not linear; with the increase of duration in the city, migrants' integration tends to increase first and then decrease, especially in terms of economic integration. This corresponds with the conventional inverted U-shaped relationship between duration and economic integration. Finally, it is not surprising to find that homeownership is significantly and positively correlated with integration in all dimensions. Homeowners are more likely to invest time and develop social networks in their neighbourhoods. They are also likely to have a stronger sense of belonging than tenants (Liu et al. 2018).

4.3. PSM estimation and results

Given the potential bias due to endogeneity between neighbourhood type and integration as a result of migrants' self-selection into different neighbourhoods, we use the propensity score matching method to check potential bias. Our analysis is limited to two types of neighbourhoods, formal and informal, rather than five different types, due to computing complexity.

4.3.1. Estimated propensity score

The first step of the PSM method is to estimate the propensity score (PS) using a logit model, following previous studies (Dehejia and Wahba 2002; Becker and Ichino 2002). The dependent variable is whether to live in a formal neighbourhood or not. It is related to migrants' personal characteristics, mobility patterns and city dummies. The results, as shown in Table A3 in Appendix II, are consistent with previous studies. For example, age, education and urban *hukou* have

positive and significant effects on living in formal neighbourhoods. However, living with a partner and inter-provincial migration have significantly negative impactss. In addition, duration in the city has a U-shaped relationship.

4.3.2. Sampling Matching Results

In order to test the rationality of the matching variables and the matching process, we conduct both the balanced hypothesis test and the common support hypothesis test. Three approaches are used to estimate the ATT, i.e. the nearest neighbor matching, radius matching and kernel matching approaches. Due to space constraints, only the results of the nearest neighbor matching test are reported here, as the test is most widely used. First, the PS value distribution of the experimental group and the control group is basically the same in a common support hypothesis, as shown in Figure A1 in Appendix II. Before matching, the PS value distribution of the experimental group is higher than that of the control group. After matching, the kernel density function of the experimental group is close to that of the control group, indicating that the socio-economic characteristics of the two groups are not very different. Second, the two groups have no significant difference in the balanced hypothesis, as shown in Table A4 in Appendix II. After matching, the absolute value of the standard deviation of each variable is less than 20%, which is consistent with the hypothesis, i.e. there is no significant difference between the experimental group and the control group after matching in terms of socio-economic characteristics. For example, before matching, the values of the treatment group and the control group with urban *hukou* status are 0.240 and 0.105, respectively, and the difference is statistically significant. However, after matching, the values of the two groups are 0.240 and 0.220, respectively, with no significant difference.

4.3.3. PSM results and robust checkness

The results of PSM in Table 4 show, whether it is before or after matching, integration of migrants in the experimental group is significantly higher than that in the control group. This indicates that migrants living in formal neighbourhoods are better integrated than those residing in informal neighbourhoods, even after controlling for their self-selection into different neighbourhood types. According to the nearest neighbor matching results, the integration of the treatment and control groups are 0.1664 and 0.0654 after PS matching, respectively. The

difference is statistically significant. That is to say, the integration of migrants living in formal neighbourhoods increases 154.4% (0.1010/0.0654), compared with that of migrants living in informal neighbourhoods. When using the methods of Radius matching and Kernel matching, the increase rates of integration of migrants living in formal neighbourhoods are 131.8% (0.0944/0.0716) and 145.5% (0.0985/0.0677), respectively. This shows the robustness of our results.

Meanwhile, ATT for different dimensions of migrants' integration are also analyzed (Table 5). Compared with migrants living in informal neighbourhoods, those residing in formal neighbourhoods are associated with significantly higher degrees of integration in economic, socio-cultural and identity dimensions. The results of PSM are slightly different from those of OLS; after we control for migrants' self-selection into formal and informal neighbourhoods, we find those in formal neighbourhoods have higher levels of economic integration than those in informal neighbourhoods, and the result is statistically significant.

5. Discussion

Our data have demonstrated that neighbourhood types are significantly correlated with migrants' socio-economic integration measured from the economic, socio-cultural and identity perspectives. The PSM method confirms that migrants living in formal neighbourhoods are positively associated with socio-economic integration, after alleviating the endogenous bias caused by the incomparability between the experimental and control groups. Migrants living in urban villages are least integrated into the urban society. One explanation concerns the nature of social interaction and cohesion in different neighbourhood types. Urban villages are usually characterized with overcrowding, inferious housing conditions and high turnover rate. Previous studies have discussed the positive roles of urban villages in providing low-cost housing for in-coming migrants. Such roles are particularly important because many local urban governments fail to provide affordable housing for migrants despite their significant contribution to the local urban development. Urban villages gradually beome enclaves where migrants outnumber local residents. However, the concentration of migrants does not reduce social interactions within urban villages. As shown in previous studies, residents in urban villages and temporary shelters have

more neighbouring activities than those in formal neighbourhoods, especially commercial property neighbourhoods (Wang et al. 2017b). This could partly be explained by the open space available in these informal settlements. A more important explanation is that migrants as vulnerable social groups are more likely to rely on social networks to survive in the city, because they are confronted with institutional discrimination resulting from the hukou system. Nevertheless, the social connections of migrants living in urban villages are often confined to small and homogeneous social networks composed of migrants. Their social interaction with local residents, especially those with high socio-economic status, are limited. Indeed, the lack of intergroup relations between migrants and local residents is often regarded as the downside of migrant enclaves (Wang et al. 2017a). As shown in various studies (e.g. Yue et al. 2013), it is social ties with local residents that are crucial for migrants' economic and social integration in the city. The residential segregation between migrants and local residents is likely to exert negative consequences on integration for migrants living in urban villages. In addition, urban villages are governed by local village collectives. Most migrants are tenants who are excluded from the decision-making of local governance and do not have incentive to participate in community organizations and activities (Wang et al. 2017a).

On the contrary, migrants living in formal neighbourhoods tend to have more opportunities of interaction with the locals than those living in urban villages. This might help them obtain social and economic opportunities within and beyond their neighbourhoods, which is conducive to their socio-economic integration (Wu and He 2005; Liu et al. 2013). In the meanwhile, this study finds that the level of socio-cultural integration is much higher in affordable and work unit housing compared with that in commercial property neighbourhoods. This is consistent with previous studies which suggest that residents in commercial property neighbourhoods have less social interaction than those in traditional ones, because of the preference for privacy (Forrest and Yip 2007; Zhu et al. 2012). However, the levels of social cohesion and social solidarity are still high in commercial property neighbourhoods with a more organized governance system (Wang et al. 2017a), and residents have strong neighbourhood attachment (Pow 2007).

This study enriches the literature of integration by highlighting the strong correlation between neighbourhood type and migrants' socio-economic integration. It has limitations. First, the PSM

method is not exempt from the bias of unobserved covariates which cannot be adjusted by the propensity score (Rosenbaum and Rubin 1983; Joffe and Rosenbaum 1999). Although we have controlled for variables such as gender, age, education and hukou status, migrants in the control and experimental groups might also differ in their length of stay in the neighbourhood, workplace characteristics, the number and socio-economic status of their social ties, which are likely to influence socio-economic integration but unavailable in the survey. Therefore, we are unable to include these variables into the models. Second, our research is based upon cross-sectional data which do not allow us to explore the causal effects of neighbourhood types on socio-economic integration. Although we take measures to control for migrants' self-selection into different neighbourhood types by using the PSM method, we do not argue that the relationship is causal because of potential bias resulting from reverse causation. The PSM method aims to control for the differences and make the treatment and control groups more comparable. It is useful in dealing with self-selection problems, but it does not provide the final assurance to establish a causal relationship (Bai 2011). The Instrumental Variable (IV) method provides a potential solution to the problem of reverse causality. However, due to data limitation, we are unable to identify appropriate instrumental variables which only affect an individual's neighborhood choice but do not influence their socio-economic integration. Alternatively, longitudinal data would be useful in exploring the temporal relations between neighbourhood characteristics and integration. Nonetheless, we propose a credible establishment of statistical correlation between neighbourhood types and migrants' socio-economic integration, which is useful for both academic analysis and policy implication. Moreover, the study provides a base for future research on this important topic.

6. Conclusions

We extend the literature on migrant's socio-economic integration by examining the correlations between different types of neighbourhoods and migrants' different dimensions of integration by using a national survey of migrants in China in 2014. Based on factor analysis, we construct a comprehensive index of socio-economic integration, which covers economic, socio-cultural and identity perspectives. Our results show that Chinese migrants' socio-economic integration in their host cities is limited, which is consistent with previous studies (Wang and Fan

2012; Chen and Wang 2015; Wang et al. 2015). Then, we use econometric models to empirically examine the relationship between neighbourhood types and migrants' socio-economic integration, and controll for migrants' self-selection into different neighbourhoods through the PSM method. The results indicate that migrants living in formal neighbourhoods show a significantly higher level of overall integration, compared with those residing in informal neighbourhoods. The finding remains robust after we control for potential self-selection bias. In terms of the three dimensions, migrants living in commercial property neighborhoods have the best economic integration, followed by those living in urban villages. In contrast, migrants in affordable and work unit housing neighbourhood are more socio-culturally integrated than those in commercial properties, inner city old housing and urban villages. Inner city old housing is positively associated with socio-cultural integration relative to urban villages. As to identity, migrants feel more integrated in the three formal types of neighbourhood, with no significant difference between those living in inner city old housing and urban villages.

With further urbanization in China, more and more migrants will settle down in cities (Xie and Chen 2018). Migrants' integration into the local urban society becomes a critical challenge for sustainable urban development. Findings in this paper thus carry far-reaching policy implications. It would be beneficial for the government to promote migrants' socio-economic integration from the neighbourhood perspective. Living in formal neighbourhoods goes hand in hand with migrants' enhanced socio-economic integration in the city, even after controlling for migrants' self-selection into different neighbourhood types. Decent living environment, access to services and faculties, and social mix with local residents in formal neighbourhoods are likely to be positively correlated with integration. It is important for policies to provide such neighbourhood environment to migrants. As to the design of commercial property neighbourhoods, it would be useful to include public space and shared facilities that can increase the likelihood of social interaction. Meanwhile, more public services and infrastructure facilities should be provided in informal neighbourhoods, especially in urban villages, to improve migrants' participation in the wider urban society. Migrants living in affordable housing seem better integrated than those in informal housing. Institutional barriers to migrants' access to affordable housing should be removed to provide them with more opportunities of accessing decent residential environment.

References

Ager, A., & Strang, A. (2008). Understanding integration. A conceptual framework. *Journal of Refugee Studies*, 21(2), 166–191.

Bai, H. (2011). Using Propensity Score Analysis for Making Causal Claims in Research Articles. *Educational Psychology Review, 23,* 273-278

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores Estimation of. *The Stata Journal2*, *2*(4), 358–377.

Chen, Y. (2011). Occupational Attainment of Migrants and Local Workers: Findings from a Survey in Shanghai's Manufacturing Sector. *Urban Studies*, 48(1), 3-21

Chen, Y., & Wang, J. (2015). Social integration of new-generation migrants in Shanghai China. *Habitat International*, *49*, 419–425.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, *84*(1), 151-161.

Démurger, S., Gurgand, M., Li, S., & Yue, X. (2009). Migrants as second-class workers in urban China? A decomposition analysis. *Journal of Comparative Economics*, *37*(4), 610–628.

Forrest, R., & Kearns, A. (2001). Social cohesion, social capital and the neighbourhood. *Urban Studies*, *38*(12), 2125–2143.

Forrest, R., & Yip, N. M. (2007). Neighbourhood and neighbouring in contemporary Guangzhou. *Journal of Contemporary China*, *16*(50), 47–64.

Goldlust, J., & Richmond, A. H. (2006). A Multivariate Model of Immigrant Adaptation. *International Migration Review*, *8*(2), 193.

Gordon, M. M. (1964). The Nature of Assimilation. *Assimilation in American Life: The Role of Race, Religion and National Origins*, 60–83.

Hainmueller, J., Hangartner, D., & Pietrantuono, G. (2016). *Catalyst or Crown: Does Naturalization Promote the Long-Term Social Integration of Immigrants? SSRN.*

He, X. (2010). Applied multivariate statistical analysis [M]. *China Statistics Press* (*in Chinese*).

Huang, Y., & Tao, R. (2015). Housing migrants in Chinese cities: current status and policy design. *Environment and Planning C: Government and Policy*, *33*(3), 640–660.

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: Propensity scores. *American Journal of Epidemiology, 150*(4), 327-333.

Kearns, A., & Whitley, E. (2015). Getting There? The Effects of Functional Factors, Time and Place on the Social Integration of Migrants. *Journal of Ethnic and Migration Studies*, *41*(13), 2105–2129.

Kleit, R. G. (2001). The role of neighborhood social networks in scattered-site public housing residents' search for jobs. *Housing Policy Debate*, *12*(3), 541–573.

Laurence, J. (2011). The Effect of Ethnic Diversity and Community Disadvantage on Social Cohesion. *European Sociological Review*, *27*(1), 70–89.

Letki, N. (2008). Does diversity erode social cohesion? Social capital and race in british neighbourhoods. *Political Studies*, *56*(1), 99–126.

Lin, Y., Zhang, Q., Chen, W., & Ling, L. (2017). The social income inequality, social integration and health status of internal migrants in China. *International Journal for Equity in Health*, *16*(1), 139.

Liu, L., Huang, Y., & Zhang, W. (2018). Residential segregation and perceptions of social integration in Shanghai, China. *Urban Studies*, *55*(7), 1484–1503.

Liu, Y., Zhang, F., Wu, F., Liu, Y., & Li, Z. (2017). The subjective wellbeing of migrants in Guangzhou, China:

The impacts of the social and physical environment. *Cities*, *60*, 333–342.

Liu, Z., Wang, Y., & Tao, R. (2013). Social Capital and Migrant Housing Experiences in Urban China: A Structural Equation Modeling Analysis. *Housing Studies*, *28*(8), 1155–1174.

Logan, J. R., & Spitze, G. D. (2002). Family Neighbors. *American Journal of Sociology*, *100*(2), 453-476.

Lu, Y., Ruan, D., & Lai, G. (2013). Social capital and economic integration of migrants in urban China. *Social Networks*. *35* (3), 357-369.

Ma, J., Dong, G., Chen, Y., & Zhang, W. (2018). Does satisfactory neighbourhood environment lead to a satisfying life? An investigation of the association between neighbourhood environment and life satisfaction in Beijing. *Cities*, *74*, 229–239.

McCaa, R. (1989). Isolation or assimilation? A log linear interpretation of Australian marriages, 1947-60, 1975, and 1986. *Population Studies, 43*(1), 155-162.

National Population and Family Planning Commission of P.R. China (PFPC). (2015). Report on China's Migrant Population Development 2015. Beijing, China: *China Population Press.*

National Population and Family Planning Commission of P.R. China (PFPC). (2017). Report on China's Migrant Population Development 2017. Beijing, China: *China Population Press.*

Niu, G., & Zhao, G. (2018). Identity and trust in government: A comparison of locals and migrants in urban China. *Cities*, *83*, 54–60.

Ouyang, W., Wang, B., Tian, L., & Niu, X. (2017). Spatial deprivation of urban public services in migrant enclaves under the context of a rapidly urbanizing China: An evaluation based on suburban Shanghai. *Cities*, *60*, 436–445.

Park, R. E. (1928). Human Migration and the Marginal Man. *American Journal of Sociology*, *33*(6), 881–893.

Piller, I., & Takahashi, K. (2011). Linguistic diversity and social inclusion. *International Journal of Bilingual Education and Bilingualism*, *14*(4), 371-381.

Pinkster, F. (2009). Neighborhood-based networks, social resources, and labor market participation in two Dutch neighborhoods. *Journal of Urban Affairs*, *31*(2), 213–231.

Portes, A., & Zhou, M. (1993). The New Second Generation: Segmented Assimilation and its Variants. *The ANNALS of the American Academy of Political and Social Science*, *530*(1), 74–96.

Pow, C. P. (2007). Securing the "civilised" enclaves: Gated communities and the moral geographies of exclusion in (Post-)socialist Shanghai. *Urban Studies*, *44*(8), 1539–1558.

Robinson, D. (2010). The neighbourhood effects of new immigration. *Environment and Planning A*, *42*(10), 2451–2466.

Rosenbaum, P. R., & Rubin, D. B. (1983). Biometrika Trust The Central Role of the Propensity Score in Observational Studies for Causal Effects The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Scharf, T., Phillipson, C., & Smith, A. E. (2005). Social exclusion of older people in deprived urban communities of England. *European Journal of Ageing*, *2*(2), 76-87.

Schneeweis, N. (2011). Educational institutions and the integration of migrants. *Journal of Population Economics*, *24*(4), 1281-1308.

Schwarzweller, H. K. (2006). Parental Family Ties and Social Integration of Rural to Urban Migrants. *Journal of Marriage and the Family*, *26*(4), 410.

Shen, J. (2017). Stuck in the suburbs? Socio-spatial exclusion of migrants in Shanghai. *Cities*, *60*, 428–435.

Shubin, S., & Dickey, H. (2013). Integration and mobility of Eastern European migrants in Scotland. *Environment and Planning A*, *45*(12), 2959–2979.

Stafford, M., Bartley, M., Sacker, A., Marmot, M., Wilkinson, R., Boreham, R., & Thomas, R. (2003). Measuring

the social environment: Social cohesion and material deprivation in English and Scottish neighbourhoods. *Environment and Planning A*, *35*(8), 1459–1475.

State Council. (2014). National new urbanization planning. Retrieved from http://www.gov.cn/xinwen/2014-03/16/content-2639841.htm

Wang, M., Cheng, H., & Ning, Y. (2015). Social integration of migrants in Shanghai's urban villages. *Acta Geographica Sinica*, *70*(8), 1243–1255.

Wang, W. W., & Fan, C. C. (2012). Migrant Workers' Integration in Urban China: Experiences in Employment, Social Adaptation, and Self-Identity. *Eurasian Geography and Economics*, *53*(6), 731–749.

Wang, Z., Zhang, F., & Wu, F. (2016). Intergroup neighbouring in urban China: Implications for the social integration of migrants. *Urban Studies*, *53*(4), 651–668.

Wang, Z., Zhang, F., & Wu, F. (2017a). Neighbourhood Cohesion under the Influx of Migrants in Shanghai. *Environment and Planning A Economy and Space*, 49(2), 407-425.

Wang, Z., Zhang, F., & Wu, F. (2017b). Social Trust Between Rural Migrants and Urban Locals in China – Exploring the Effects of Residential Diversity and Neighbourhood Deprivation. *Population, Space and Place*, *23*(1), 1-15.

Wei, L., & Gao, F. (2017). Social media, social integration and subjective well-being among new urban migrants in China. *Telematics and Informatics*, 34, 786-796.

Wu, F., & He, S. (2005). Changes in traditional urban areas and impacts of urban redevelopment: A case study of three neighbourhoods in Nanjing, China. *Tijdschrift voor Economische en Sociale Geografie*, *96*(1), 75–95.

Wu, Q., Cheng, J., Chen, G., Hammel, D. J., & Wu, X. (2014). Socio-spatial differentiation and residential segregation in the Chinese city based on the 2000 community-level census data: A case study of the inner city of Nanjing. *Cities*, *39*, 109–119.

Xie, S., & Chen, J. (2018). Beyond homeownership: Housing conditions, housing support and rural migrant urban settlement intentions in China. *Cities*, *78*, 76–86.

Yue, Z., Li, S., Jin, X., & Feldman, M. W. (2013). The Role of Social Networks in the Integration of Chinese Rural-Urban Migrants: A Migrant-Resident Tie Perspective. *Urban Studies*, *50*(9), 1704–1723.

Zhu, Y., Breitung, W., & Li, S. ming. (2012). The Changing Meaning of Neighbourhood Attachment in Chinese Commodity Housing Estates: Evidence from Guangzhou. *Urban Studies*, *49*(11), 2439–2457.

Appendix I

The PSM methods used in the study

Scholars define the propensity score as the conditional probability of a given set of observation covariates assigned to the treatment group (Rosenbaum and Rubin 1983).

$$p(X) = \Pr[B = 1|X] = E[B|X] \tag{I1}$$

In formula (I1), X is the multidimensional eigenvector of the control group, and B is the core dependent variable, which equals 1 if migrants live in formal neighbourhoods and 0 otherwise. If we can get a propensity score, we can estimate the ATT by comparing the potential differences between the experimental and control groups as follows (Becker and Ichino 2002),

$$
\begin{aligned}
\text{ATT} &= E[Y_{1i} - Y_{0i}|B_i = 1] \\
&= E\{E[Y_{1i} - Y_{0i}|B_i = 1, p(X_i)]\} \\
&= E\{E[Y_{1i}|B_i = 1, p(X_i)] - E[Y_{0i}|B_i = 0, p(X_i)]|B_i = 1\}
\end{aligned}
\tag{I2}
$$

Here Y_{1i} and Y_{0i} respectively indicate the potential results of the experimental group and the control group. Following previous studies (Becker and Ichino 2002; Dehejia and Wahba 2002), we use the following Logit model to get the propensity score.

$$p(X_i) = \Pr(B_i = 1|X_i) = \frac{\exp(\alpha X_i)}{1+\exp(\alpha X_i)} \tag{I3}$$

Here, X is the same as above. It is assumed to affect migrants' tendency of living in different neighbourhoods. α is a coefficient vector. Then, we use three matching methods to estimate the ATT, i.e. Nearest neighbor matching, Radius matching and Kernel matching. The Nearest neighbor matching method is the most commonly used matching method. It finds the individual with the smallest difference of PS value in the control group with the experimental group as its comparison object, which can be represented as follows,

$$C(i) = min_j \|p_i - p_j\| \tag{I4}$$

However, the Radius matching method is to set the radius beforehand, and to find all the control samples in the unit circle within the set radius range. The radius is positive. As the radius decreases, the matching requirements become more stringent. The formula is shown as follows,

$$C(i) = \{p_j | \|p_i - p_j\| < r\} \tag{I5}$$

In the above two formulas, C refers to the control group. We let T represent the experimental group, Y_i^T and Y_j^C respectively represent the observation results of the experimental group and the control group. Meanwhile, let the control group's number matched with observation i∈T by N_i^C and define the weights $w_{ij} = \frac{1}{N_i^C}$ if j∈C(i), otherwise, $w_{ij}=0$. In addition, suppose that the experimental group has N^T observations. We can estimate the ATT following (Becker and Ichino 2002).

$$\tau^M = \frac{1}{N^T}\sum_{i\in T} Y_i^T - \frac{1}{N^T}\sum_{j\in C} w_j Y_j^C \tag{I6}$$

Then, we use M as the matching method, $w_j = \sum_i w_{ij}$. Assuming that the weights remain unchanged, the validity of the neighborhood is independent. We can estimated the variance of τ^M as the following,

$$Var(\tau^M) = \frac{1}{N^T}Var(Y_i^T) + \frac{1}{(N^T)^2}\sum_{j\in C}(w_j)^2 Var(Y_i^C) \tag{I7}$$

The third method (the Kernel matching method) is slightly different from the first two. It is to construct a virtual object to match the treatment group. The principle is to average the weight of the existing control variables, and the value of the weight is inversely related to the PS value difference between the treatment group and the control group. The ATT of this method is estimated as following:

$$\tau^K = \frac{1}{N^T}\sum_{i\in T}\left\{Y_i^T - \frac{\sum_{j\in C} Y_j^C G\left((p_j - p_i)/h_n\right)}{\sum_{k\in C} G((p_k - p_i)/h_n)}\right\} \tag{I8}$$

In formula (I8), G(.) means the Gaussian kernel function, and bandwidth parameter is h_n.

$$\frac{\sum_{j\in C} Y_j^C G\left((p_j - p_i)/h_n\right)}{\sum_{k\in C} G((p_k - p_i)/h_n)}$$

It is a consistent estimate of Y_{0i} with counterfactual results.

Table 1 Descriptive statistics of explanatory variables

Variables		Percentage（%）
Neighbourhood type	Commodity housing	17.27
	Affordable housing	4.29
	Work unit housing	4.75
	Inner city old housing	15.37
	Village in the city	58.33
Gender	Male	54.98
	Female	45.02
Age	Aged below 25	22.91
	Aged between 25 and 35	41.81
	Aged between 35 and 45	26.33
	Aged over 45	8.95
Education	Junior high school and below	59.83
	High school	25.38
	College or above	14.79
Hukou	Agricultural	85.98
	Non-agricultural	14.02
Parter_present	Not live in the local residences	30.11
	Live in the local residences	69.89
Child_present	Not live in the local residences	51.61
	Live in the local residences	48.39
Flow scope	Intra-province mobility	54.67
	Inter-province mobility	45.33
Homeowner	Homeowner	9.93
	Tenant	90.07
Total		15,866

Table 2 OLS results on integration of migrants in formal/informal neighbourhoods

Variables	(1) Integration	(2) Economic integration	(3) Socio-cultural integration	(4) Self-identity
NeighbourhoodFormal	0.098***	0.019	0.155***	0.113***
	(0.007)	(0.013)	(0.012)	(0.014)
Female	-0.031***	-0.107***	-0.019*	0.021*
	(0.006)	(0.011)	(0.010)	(0.011)
Age				
Aged below 25 (ref.)				
Aged between 25 and 35	0.059***	0.150***	0.032**	0.009
	(0.008)	(0.016)	(0.014)	(0.016)
Aged between 35 and 45	0.050***	0.156***	0.001	0.005
	(0.010)	(0.018)	(0.016)	(0.019)
Aged over 45	0.025**	0.144***	-0.047**	-0.010
	(0.012)	(0.023)	(0.021)	(0.024)
Education				
Junior high school and below (ref.)				
High school	0.070***	0.056***	0.116***	0.041***
	(0.007)	(0.014)	(0.012)	(0.014)
College or above	0.185***	0.323***	0.233***	0.029
	(0.010)	(0.018)	(0.016)	(0.019)
Nonfarmer	0.070***	-0.012	0.091***	0.118***
	(0.009)	(0.018)	(0.016)	(0.018)
Parter_present	0.042***	0.154***	-0.023	0.008
	(0.009)	(0.017)	(0.015)	(0.017)
Child_present	0.029***	-0.006	-0.005	0.086***
	(0.007)	(0.014)	(0.012)	(0.014)
Inter-provincial-move	-0.103***	-0.045***	-0.227***	-0.040***
	(0.008)	(0.015)	(0.013)	(0.015)
Moveyears	0.010***	0.019***	0.007**	0.006*
	(0.002)	(0.003)	(0.003)	(0.003)
Moveyearssq	-0.000*	-0.001***	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Homeowner	0.180***	0.199***	0.045**	0.284***
	(0.011)	(0.020)	(0.018)	(0.021)
City dummies	YES	YES	YES	YES
Constant	0.054***	-0.272***	0.358***	0.051**
	(0.011)	(0.021)	(0.019)	(0.022)
Observations	15,866	15,866	15,866	15,866
R-squared	0.274	0.105	0.268	0.086

Note: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% level, respectively.

Table 3 OLS results on integration of migrants in five different types of neighbourhoods

Variables	(1) Integration	(2) Economic integration	(3) Socio-cultural integration	(4) Self-identity
Neighbourhood type				
Inner city old housing (ref.)				
Commodity housing	0.074***	0.059***	0.064***	0.094***
	(0.010)	(0.020)	(0.018)	(0.020)
Affordable housing	0.102***	0.003	0.126***	0.162***
	(0.016)	(0.030)	(0.027)	(0.030)
Unit housing	0.097***	0.044	0.131***	0.109***
	(0.015)	(0.029)	(0.026)	(0.030)
Urban village	-0.020**	0.035**	-0.089***	-0.004
	(0.008)	(0.016)	(0.014)	(0.016)
Female	-0.031***	-0.107***	-0.020**	0.021*
	(0.006)	(0.011)	(0.010)	(0.011)
Age				
Aged below 25 (ref.)				
Aged between 25 and 35	0.060***	0.148***	0.035**	0.009
	(0.008)	(0.016)	(0.014)	(0.016)
Aged between 35 and 45	0.050***	0.155***	0.003	0.006
	(0.010)	(0.018)	(0.016)	(0.019)
Aged over 45	0.025**	0.143***	-0.046**	-0.009
	(0.012)	(0.023)	(0.021)	(0.024)
Education				
Junior high school and below (ref.)				
High school	0.070***	0.056***	0.115***	0.041***
	(0.007)	(0.014)	(0.012)	(0.014)
College or above	0.184***	0.323***	0.232***	0.029
	(0.010)	(0.018)	(0.016)	(0.019)
Nonfarmer	0.069***	-0.010	0.088***	0.117***
	(0.009)	(0.018)	(0.016)	(0.018)
Parter_present	0.043***	0.153***	-0.020	0.008
	(0.009)	(0.017)	(0.015)	(0.017)
Child_present	0.029***	-0.005	-0.005	0.086***
	(0.007)	(0.014)	(0.012)	(0.014)
Inter-provincial-move	-0.102***	-0.046***	-0.224***	-0.040***
	(0.008)	(0.015)	(0.013)	(0.015)
Moveyears	0.010***	0.019***	0.006**	0.006*
	(0.002)	(0.003)	(0.003)	(0.003)
Moveyearssq	-0.000*	-0.001***	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Homeowner	0.183***	0.195***	0.053***	0.290***
	(0.011)	(0.021)	(0.018)	(0.021)
City dummies	YES	YES	YES	YES
Constant	0.067***	-0.296***	0.418***	0.054**
	(0.013)	(0.024)	(0.021)	(0.024)
Observations	15,866	15,866	15,866	15,866
R-squared	0.274	0.105	0.270	0.086

<h1>Table 4 ATT estimates of the whole sample</h1>

Methods	Sample	Treatment group	Control group	ATT	Standard Error	T-value
Nearest neighbor matching	Pre-matching	0.1664	-0.0594	0.2258	0.0073	31.06***
	Post-matching	0.1664	0.0654	0.1010	0.0122	8.28***
Radius matching	Pre-matching	0.1664	-0.0594	0.2258	0.0073	31.06***
	Post-matching	0.1660	0.0716	0.0944	0.0091	10.38***
Kernel matching	Pre-matching	0.1664	-0.0594	0.2258	0.0073	31.06***
	Post-matching	0.1662	0.0677	0.0985	0.0090	10.98***

Note: 1. "Pre-matching" refers to the samples without matching the treatment group and the control group, and "Post-matching" refers to the groups after matching.

2. ***,** and * represent significance at 1%, 5% and 10% level, respectively.

<h1>Table 5 ATT estimates for different dimensions of integration of migrants</h1>

Methods	Economic integration	Socio-cultural integration	Self-identity
Nearest neighbor matching	0.0614***	0.1058 ***	0.1292***
Radius matching	0.0418***	0.1391***	0.0980***
Kernel matching	0.0436***	0.1448***	0.1024***

Note: ***,** and * represent significance at 1%, 5% and 10% level, respectively.

Figure 1 Integration of migrants in formal and informal neighbourhoods

■ informal neighbourhood ☑ formal neigbourhood

Dimension		
integration	38.00	43.80
self-identity	53.08	57.80
socio-cultural integration	48.99	52.88
economic integration	16.68	17.60

Score (0 10 20 30 40 50 60 70)
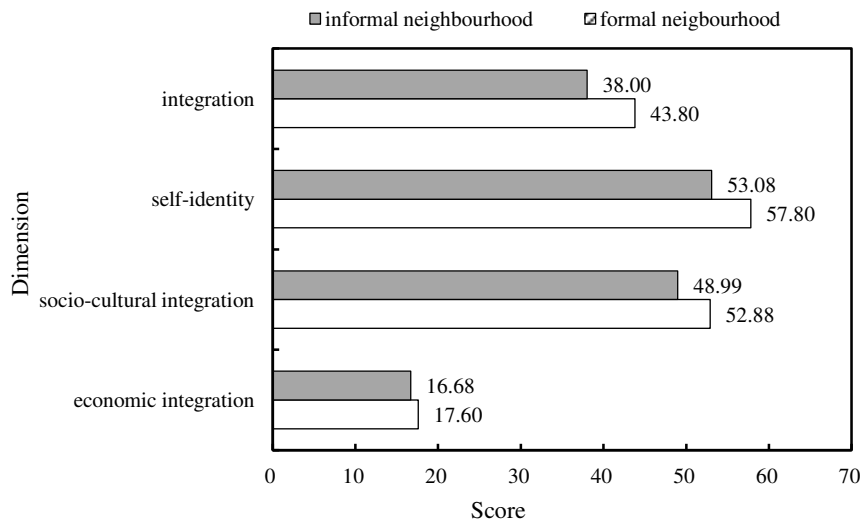
Figure 2 Integration of migrants in five different types of neighbourhoods

■ commodity housing ☑ affordable housing ⊠ work unit housing ▤ inner city old housing ◩ urban village

economic integration: 17.99, 17.02, 16.67, 16.98, 16.60
socio-cultural integration: 52.72, 52.71, 53.65, 51.69, 48.28
self-identity: 58.72, 57.61, 54.65, 54.46, 52.71
integration: 44.51, 42.95, 41.98, 40.85, 37.26

Dimension
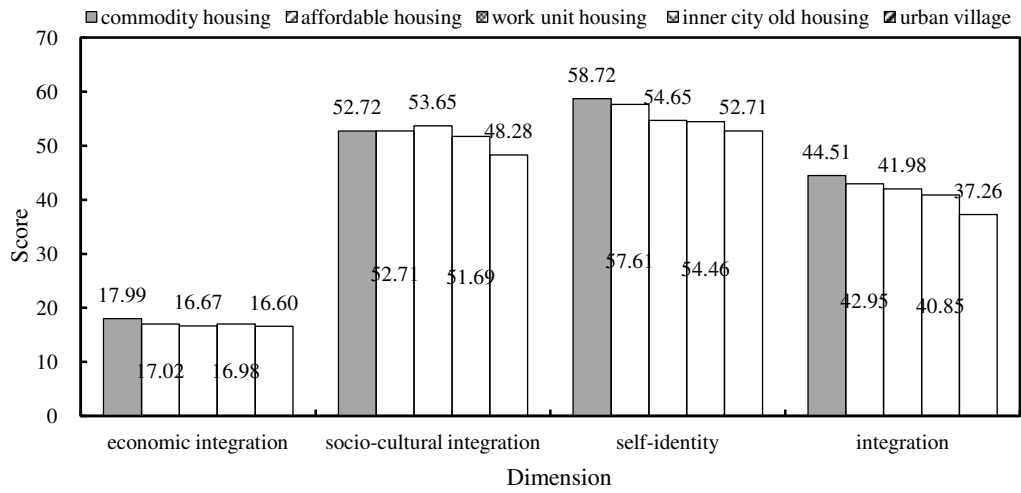
Appendix II

Table A1 Indicators of migrants' integration in urban society

Variables	Definitions	Options	% or $\bar{X} \pm s$
Occupation type	Current main occupation	Irregular employment	9.64
		Production workers and service personnel	66.89
		Businesspeople	14.66
		Managers and professionals	8.80
Monthly household income	Average monthly income of the household		6429.038±7076.009
Pension	Access to pension schemes	No	27.78
		Yes	72.22
Medical insurance	Access to social or commercial medical insurance	No	14.85
		Yes	85.15
Number of organizations (0-8)	Number of social organizations respondents participated in, such as labor union, volunteer associations, the Chinese Communist Party group of migrants/local residents, alumni association, chamber of commerce of hometown, fellow-townsman associations		0.3976±0.7725
Number of activities (0-7)	Number of social activities respondents attended, for example, community sports, social public welfare activities, election campaigns, awards events, activities organized by the homeowners' committee, management activities of residents' committees		0.6625±1.0382
Types of neighbours	Whether most of respondents' neighbours were local hukou residents or migrants	Migrants	43.59
		Not sure	6.47
		Half migrants and half locals	29.31
		Locals	20.63
Getting along with the locals	Respondents provided their views based on a five-point scale (unharmonious, quite unharmonious, so so, harmonious, quite harmonious).		3.9230±0.9390
Views and attitudes (8-40)	Respondents provided their agreement with the views on social norms and customs, based on a five-point scale (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree). 1) It's important for you to follow the customs of your hometown (such as the customs of marriage, funeral); 2) It's important for you to do things according to traditions in your hometown; 3) Your child should learn to speak native dialect; 4) Keeping your hometown lifestyle (such as eating habits) is important to you; 5) There is a big difference in health habits between you and local residents; 6) There is a big difference in dressing styles between you and local residents; 7) Your views on education or pension are quite different from those of local residents; 8) Your views on some social issues are quite different from those of local residents.		23.7940±4.0863
Familiarity with local dialects	Proficiency in the local dialect	Don't understand	14.85
		Understand some only	22.90
		Understand and speak some	22.76

		Understand and speak	39.49
Belonging to the city	You feel that you belong to the city. Respondents provided their agreement with the view based on a five-point scale (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).		3.1667±0.7013
A member of the city	You feel that you are a member of the city. Respondents provided their agreement with the view based on a five-point scale (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).		3.2091±0.6799
Think of yourself as a hometown member	Do you identify yourself as a member of folks at hometown?	No	12.93
		Yes	87.07
Think of yourself as a native	Do you identify yourself as a local resident?	No	78.15
		Yes	21.85

Table A2 Result of rotated component matrix in integration of migrants

Items	F1	F2	F3	F4	F5	F6	Communalities
X1 Occupation type	0.0450	0.1266	0.1270	0.1660	-0.0673	0.6537	0.494
X2 Monthly household income	0.0280	-0.0033	-0.0609	-0.0473	0.0958	0.8007	0.657
X3 Pension	0.0322	0.8673	0.0133	0.1068	0.0202	0.0478	0.768
X4 Medical insurance	0.0168	0.8792	0.0355	0.0228	-0.0276	0.0083	0.776
X5 Number of organizations participated in	0.0198	0.0440	-0.0025	0.8367	0.0327	0.0856	0.711
X6 Number of activities attended	0.0775	0.1031	0.0620	0.8169	0.0339	-0.0514	0.692
X7 Types of neighbours	0.0653	-0.0047	0.6964	-0.0079	0.0396	-0.0190	0.491
X8 Getting along with the locals	0.4148	0.0833	0.4183	0.0303	0.0801	0.0395	0.363
X9 Views and attitudes	0.1346	-0.0079	0.5126	0.0035	0.0878	0.1110	0.301
X10 Familiarity with local dialects	0.0504	0.0886	0.6804	0.1047	0.0674	-0.0173	0.489
X11 Belonging to the city	0.9241	0.0182	0.0570	0.0385	0.0645	0.0211	0.864
X12 A member of the city	0.9249	0.0198	0.0341	0.0353	0.0595	0.0187	0.862
X13 Think of yourself as a hometown member	0.0200	-0.0469	-0.0310	0.0253	0.8703	0.0542	0.765
X14 Think of yourself as a native	0.2468	0.0753	0.2279	0.0587	0.6947	-0.0136	0.605
Eigenvalue	1.9781	1.5774	1.4680	1.4279	1.2851	1.0992	
Variance contribution rate	0.1413	0.1127	0.1049	0.1020	0.0918	0.0785	
Cumulative variance proportion	0.1413	0.2540	0.3588	0.4608	0.5526	0.6311	

Table A3 The Logit Model of living in a formal neighbourhood

| Variables	NeighbourhoodFormal
Female	0.028 (0.040)
Age	
Aged below 25 (ref.)	
Aged between 25 and 35	0.054 (0.057)
Aged between 35 and 45	0.161**(0.068)
Aged over 45	0.268***(0.086)
Education	
Junior high school and below (ref.)	
High school	0.555***(0.048)
College or above	0.810***(0.061)
Nonfarmer	0.417***(0.058)
Parter_present	-0.321***(0.062)
Child_present	0.083(0.054)
Longmove	-0.227***(0.054)
Moveyears	-0.026**(0.012)
Moveyearssq	0.002***(0.001)
Homeowner	2.085***(0.065)
City dummies	YES
Constant	-1.191***(0.073)
Observations	15,866
R-squared	0.1355

Note: Standard errors in parentheses; ***, ** and * represent significance at 1%, 5% and 10% level, respectively.

Table A4 Test results of balanced hypothesis

Variable	Matching	Treat group	Control group	s.e.	t-value
Female	Pre-matching	0.451	0.450	0.1	0.07
	Post-matching	0.451	0.434	3.3	1.53
Aged between 25 and 35	Pre-matching	0.425	0.416	1.9	1.05
	Post-matching	0.425	0.420	1.0	0.47
Aged between 35 and 45	Pre-matching	0.258	0.265	-1.5	-0.85
	Post-matching	0.258	0.250	1.8	0.84
Aged over 45	Pre-matching	0.091	0.089	0.8	0.47
	Post-matching	0.091	0.083	2.8	1.30
High school	Pre-matching	0.313	0.233	18.1	10.29***
	Post-matching	0.313	0.297	3.6	1.59
College or above	Pre-matching	0.252	0.111	37.3	22.43***
	Post-matching	0.252	0.248	0.9	0.38
Nonfarmer	Pre-matching	0.240	0.105	36.4	21.91***
	Post-matching	0.240	0.220	5.3	2.12
Parter_present	Pre-matching	0.661	0.713	-11.2	-6.27***
	Post-matching	0.661	0.622	8.3	3.68***
Child_present	Pre-matching	0.493	0.481	2.5	1.38
	Post-matching	0.493	0.471	4.4	2.00**
Longmove	Pre-matching	0.464	0.576	-22.6	-12.55***
	Post-matching	0.464	0.408	11.3	5.19***
Moveyears	Pre-matching	4.532	4.149	8.4	4.80***
	Post-matching	4.532	4.301	5.1	2.23**
Moveyearssq	Pre-matching	43.369	35.647	9.4	5.48***
	Post-matching	43.369	40.543	3.4	1.45
Homeowner	Pre-matching	0.261	0.042	64.3	42.98***
	Post-matching	0.261	0.245	4.7	1.69*

Note: 1. "Pre-matching" refers to the samples without matching the treatment group and the control group, and "Post-matching" refers to the groups after matching.

2. ***,** and * represent significance at 1%,5% and 10% level, respectively.

Fig. A1 Kernel Density of the treatment and control Groups
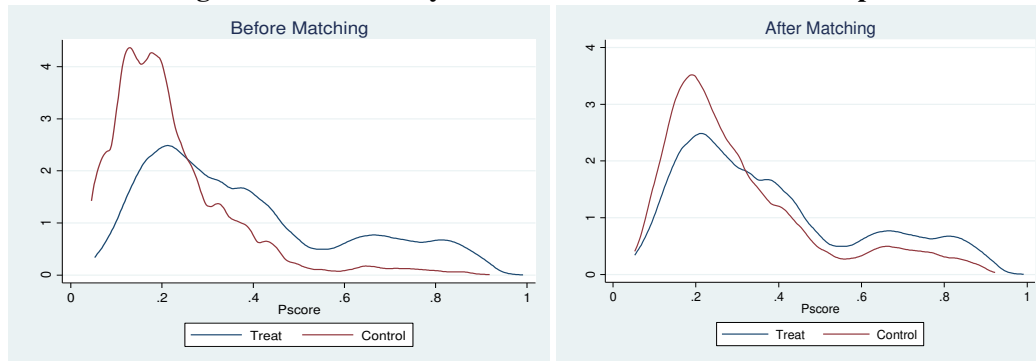
Propensity score (a) before matching   Propensity score (b) after matching