Routledge
Taylor & Francis Group

Economic Research-Ekonomska Istraživanja

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rero20

"I" vs "me": the urbanization of "post-80s" and "post-90s" Chinese migrant workers

Siyu Xu, Jun He & Noshaba Aziz

To cite this article: Siyu Xu, Jun He & Noshaba Aziz (2023) "I" vs "me": the urbanization of "post-80s" and "post-90s" Chinese migrant workers, Economic Research-Ekonomska Istraživanja, 36:1, 2179514, DOI: 10.1080/1331677X.2023.2179514

To link to this article: https://doi.org/10.1080/1331677X.2023.2179514

Published online: 20 Apr 2023.

ECONOMIC RESEARCH-EKONOMSKA ISTRAŽIVANJA
2023, VOL. 36, NO. 1, 2179514
https://doi.org/10.1080/1331677X.2023.2179514

"I" vs "me": the urbanization of "post-80s" and "post-90s" Chinese migrant workers

Siyu Xu[a], Jun He[a] and Noshaba Aziz[b]

[a]College of Economics and Management, Nanjing Agricultural University, Nanjing, China;
[b]School of Economics, Shandong University of Technology, Zibo, China

ABSTRACT

The difference in self-identity among migrant workers of the new generation leads them towards different desires regarding urbanization. In this regard, it is imperative to explore the influence of self-identity on the migrant workers' willingness to stay. To explore the phenomenon empirically, the current study used data sourced from the China Migrants Dynamics Survey (CMDS), during the year, 2017. The study employed the Heckman two-stage selection model to explore the study objective. Further, the study also employed the machine learning methods for robustness check. The outcome showed that the "I" identity has a more significant impact on the urbanization by migrant workers belonging to the "post-90s". In comparison, the identity of "Me" has a more significant impact on the urbanization by migrant workers belonging to the era of the 1980s. And it is clear that if "post-80s" and "post-90s" migrant workers are uniformly divided into the union of new generation, the differences and characteristics within them may conceal. The overall findings proposes that based on the differences in migrant workers' self-identity, both born in the 1980s and 1990s, there is a need to formulate related policies to promote their residence and boost urbanization.

ARTICLE HISTORY
Received 7 July 2022
Accepted 6 February 2023

KEYWORDS
New generation of migrant workers; self-identification; urbanization; Heckman two-stage selection model; machine learning

JEL CODES
O15; R23

1. Introduction

According to the report of International Labor Organization, approximately 164 million migrant workers are found globally in 2020 (IOM 2020). According to the researchers of the International Organization of Migration, it is shown that about 35-40 million people migrate globally every 5 years (Abel & Sander, 2014). Asia is recorded as one of the region with extensive population flow; especially in the context of China, the employment growth rate of migrant workers during the years 2016 to 2020 was found at 1.5%, 1.7%, 0.6%, 0.8%, and −1.8%, respectively (National Bureau of Statistics, 2021). In the recent years, the growth rate of the total number of

CONTACT Jun He hejun@njau.edu.cn

migrant workers has begun to decline. The two-way flow between them has become the primary choice for the migrant workers. Moreover, the Chinese urbanization strategy has changed from the initial quantitative expansion to a new type of urbanization oriented by quality improvement. The key to realize the quality improvement of urbanization depends on how migrant workers settle in cities by engaging in long term employment. The share of labor force in the cities, long-term residence, and occupation helps in improving the level of urbanization.

Since the new generation of migrant workers (born after the 1980s) came for work in cities, many literature have focused on the intergenerational differences between the new generation of migrant workers and the older generation of migrant workers (born in the 1960s or 1970s). In this regard, Zhao et al. (2018) conducted study by comparing the differences between these two generations and found that the new generation migrants are more educated and skilled and are likely to work in manufacturing and service industries than in the construction industry. Moreover, they tend to allocate more time to non-farm activities and have more months away from home per year in urban areas. Chen and Liu (2016) found that the socio-cultural conditions of the first-generational migrants drive their settlement intention, so economic incentives are found as more important aspects for the new-generation rural migrants.

Compared with the first generation of migrant workers, the new generation of migrant workers, mainly those born in the 1980s and 1990s, are not strongly dependent on the countryside. They do not have a strong social identity for the cities. The new generation of migrant workers lives in cities with more uncertainty and a lack of social belonging. Most of the existing literature takes the work of the new generation of farmers as the overall research object (Hu et al., 2011; Cheng et al., 2014; Chen & Liu, 2016; Zhilin Wang & He, 2019). However, with the gradual return of the first generation of migrant workers, the "post-80s" and "post-90s" are the mainstream of the migrant worker groups. The demographic structure of migrant workers has changed. These years, the "post-90s" have gradually become the backbone of rural labor mobility. This growth signals the coming of a new era and requires a better understanding of these young people (Lee & Qi, 2021). So, there arises questions such as is there any difference between "post-80s" migrant workers and "post-90s" migrant workers? Does this difference affect their urbanization choices?

In the existing literature, couple of studies have been conducted on the "post-80s" and "post-90s" migrant workers. In this regard, Han (2010) counted the "post-90s" migrant workers as the third generation reflecting the point that "post-90s" migrant workers want to improve their living standard, and enter to the cities to find good jobs and prefer to stay in cities instead of going back to the countryside. Likewise, Han pointed out that the "post-90s" migrant workers have the most distinctive characteristics than the previous two generations of migrant workers, i.e., they never cultivated land, and never showed same affection for the land as their fathers did. Therefore, migrant workers born after 1990 have been given consideration in the prevailing literature . Contrary to the above debate, Wang (2013) holds different view and argues that there is no "gap" between the "post-90s" and the "post-80s" migrant workers like the new generation of migrant workers and the older generation of migrant workers. Therefore, it is inappropriate to refer to the "post-90s" migrant workers as "the third

generation of migrant workers," as it is easy to exaggerate the difference between "post-90s" and "post-80s" migrant workers. "Post-80s" and "post-90s" migrant workers still belong to the category of "new generation migrant workers" and have similar group characteristics. The existing literature focusing on the differences between the "post-80s" and "post-90s" migrant workers are mostly based on empirical evidence, which only focuses on part of the differences between the two groups and lacks discussions on the reasons for the differences or profound theoretical insights.

Inkeles (1969) pointed out that the gradual transition from a traditional personality to a modernized is the only way for a country to realize modernization truly. People in radically changed social environments also experience significant changes in the basic nature of their personalities. The process of urbanization is not only a process of population transfer from rural to urban areas but also a function of modern personality formation. And the appearance of personality needs to rely on the embedding of social networks. The "post-80s" migrant workers are deeply influenced by traditional culture, and their values have been formed in their early years. The "post-90s" migrant workers are a generation who grew up with the Internet, and they can receive almost the same information as those born in cities. Therefore, the "post-90s" migrant workers have nearly the same personality as those born in cities. So how to explain the differences between the "post-80s" and "post-90s" migrant workers? To answer this question, the current explores the personality formation of "post-80s" and "post-90s" migrant workers by following the self theory of Mead's (1962) and also also attempts to explain their differences based on urbanization choices.

To explore the phenomenon empirically, the current study uses the data of the China Migrants Dynamic Survey (CMDS) during the period 2017 and explores the impact of self-identity on the urbanization choices in the case of "post-80s" and "post-90s" migrant workers. The study initially compares with the "post-80s" migrant workers, the "post-90s" migrant workers' "I" identity significantly influences urbanization. Secondly, suppose the study only divides the migrant workers into the new and old ages without considering the internal heterogeneity of the new generation of migrant workers. In that case, some factors that affect the urbanization of migrant workers may be ignored. The possible marginal contributions of this paper are that, first, it is different from previous studies on migrant workers' urbanization, which emphasize economic factors (Stark & Bloom, 1985; Johnson, 2003) and identity (Yang & Curdt-Christiansen, 2021), this paper establishes a theoretical framework of migrant workers' self-identification, and analyzes the urbanization degree of migrant workers in the "post-80s" and "post-90s" from the perspective of "I" and "Me" identity. Second, unlike most literature that supposed the "post-80s" and "post-90s" migrant workers as a whole of the new generation of migrant workers in their studies, the current study considers that the "post-90s" migrant workers who grew up in the Internet age, are different than the growth environment of "post-80s". If they are uniformly divided into the union of new generation, the differences and characteristics within the "post-80s" and "post-90s" migrant workers may conceal. The third is to use the methods of bagging and random forest in machine learning. The robustness test of the relevant influencing factors shows that the identity of "I" has a greater impact on the urbanization of "post-90s" migrant workers, and the identity of "Me" has a greater impact on the urbanization of "post-80s" migrant workers.

The rest of this paper is structured as follows: the second part is the research conceptual framework and hypotheses. The third part is the data and variable selection. The fourth part is the empirical approach, and the last part is the conclusion.

2. Conceptual framework and hypotheses

According to the study of Mead (1962), it is believed that the self is a process rather than an entity, and is not innate but occurs through various processes of social experience and activities. In the process of one's connection and mutual influence with others, Mead brought the notion of "I" and "Me" into the individual's action by dividing the self, where "I" portrays ones' attitudes and "Me" expresses others' attitudes. "I" may occur when he/she defends their rights and demands to get rid of customs and established laws, while the "Me" is to adopt the attitude of others to ensure the recognition of his rights or to have the dignity of being a member of the community. The formation of "I" and "Me" can also be explained by the embeddedness theory of Granovetter's (1985), which divides embeddedness into relational and structural embeddedness. Relational embeddedness shows that an individual's economic behavior is influenced by long-term interactive bilateral relationships such as parents and friends. A mere association with a group of people modifies the individual's behavior. The influence of the overall structure of a person's embedded network on the individual is regarded as structural embeddedness. The two embeddedness methods make the socialization degree of the new generation of migrant workers of "post-80s" and "post-90s" different. Specifically, migrant workers, who born in the 1990s have a higher level of education on average than those who born in the 1980s. In addition, they reside in modern era with heavily reliant on TV and mobile phones. They have a better understanding of the outside world and already have a simple concept of equality and democracy (Han, 2010). The "post-90s" were born with the wave of globalization, and their group characteristics are mostly based on personal and personality development as the pursuit and downplay of authority (Wei, 2018). The professional use of the Internet and the same sources of information as their peers born in the surrounding cities manifest their social network as "connected individualism" (Wellman, 2001). They are prone to "low socialization" characteristics–more care about personal feelings and interests, which is "I" identity. Although there is no extreme phenomenon of "over socialization," "post-80s" are more socialized than the "post-90s" migrant workers, and "post-80s" follow the social culture to which they belong and use this as a code of conduct, so they pay more attention to the "Me" identity.

"I" identity is reflected in the psychological integration of migrant workers. To realize the long-term residence of migrant workers in the place of migration, they must have a particular psychological integration with the area of migration (Gu et al., 2022). Tan et al. (2022) found the correlation of the social cohesion policy with the regional development policy in local residence for migrants could create the necessary synergy that would contribute to the integration of migrants in the local community. Psychological integration is manifested in recognizing the local social culture and the belief that no one is different from the local residents. Vroome and Tubergen (2014) also found that, compared with economic factors, the social and cultural integration of the floating

population has a more significant impact on the willingness to reside. For the "post-90s", their lives are more better than previous generations. With the rapid spread of information in the Internet age, they disagree with the gap between urban and rural areas, so it is more difficult for them to accept that they are migrant workers and "being discriminated against" in the place where they move. They hope to get rid of the established prejudice and enjoy various rights as equals as the locals. Therefore, compared with the "post-80s", the "post-90s" pay more attention to the sense of psychological integration, that is, "I" identity. The "Me" identity is manifested in social prestige, while social prestige is mainly occupational prestige (Li, 2005). Zhou (2005) argues that occupational prestige ranks are a kind of "collective cognition," indicating that the intersubjective process of social judgment produces prestige. Inkeles (1969) believes that modern people should choose occupations based on personal interests and talents rather than engaging in fields respected by tradition. Constant and Massey (2003) found that higher occupational prestige would reduce the return rate of migrants. Liu and Wang (2019) found that individuals' occupational type and human capital have become the most significant factors that currently determine or affect the urban residency intention of the floating population. Most of the "post-1980s" migrant workers have formed families. When living in the influx, they hope to gain a kind of recognition of their own identity and recognition as a local member through their occupation. When the "post-90s" began to migrate, the inflow areas provided essential public services for the floating population, that gradually resulted in positive outcomes (Xiao et al., 2019). And this has reduced their need for social relations, or the recognition relied on occupations. Youngers' motivation of getting a high rank of occupational prestige is weaker than that of the "post-80s". Therefore, "post-80s" pay more attention to their occupational prestige, i.e. "Me" identity.

Based on the above discussion, the current study proposes the following hypothesis: Compared with the "post-80s" migrant workers, the "I" identity of the "post-90s" migrant workers has a more significant impact on their urbanization, which means the "post-90s" migrant workers are more concerned about the psychological integration than the "post-80s" migrant workers. Compared with "post-90s" migrant workers, "post-80s" migrant workers' "Me" identity has a more significant impact on their urbanization. Which means "post-80s" migrant workers are more concerned about their occupational prestige than "post-90s" migrant workers (Figures 1 and 2).

3. Methods

3.1. Data sources and participants

The current study used the data gathered from China Migrants Dynamic Survey (CMDS) database during the year 2017. The data is collected stratifically from surveying migrants of 31 provinces of China by the National Health Commission. It is a large-scale follow-up survey data, including basic population information, years of migration, employment characteristics, and social integration. As the research object of the current study is to explore migrant workers' new generation, so the samples with urban household registration and the age of more than 37 years are excluded. There are 46,953 valid samples for the "post-80s" generation and 29,821 valid samples for the "post-90s" generation.

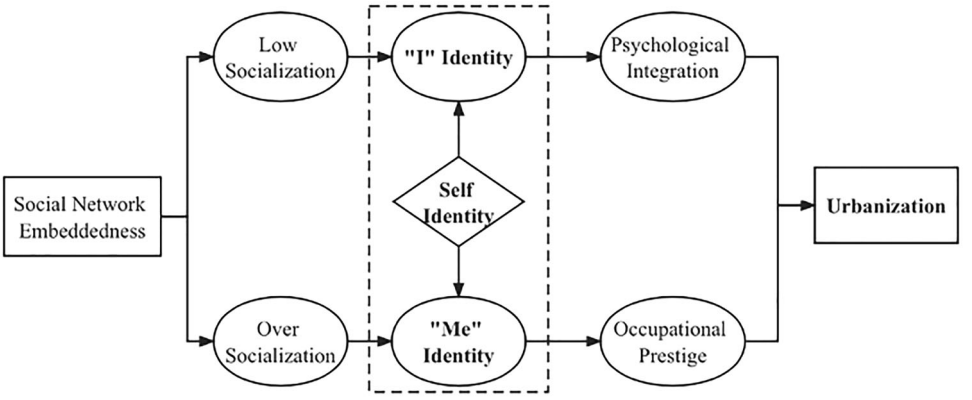

Figure 1. Conceptual framework operationalized in the current study.
Source: authors' own estimations.

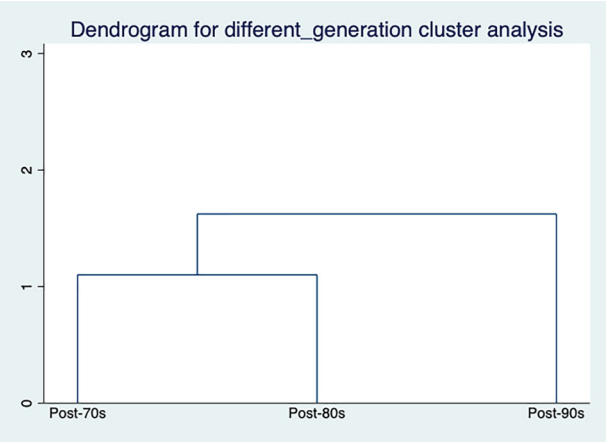

Figure 2. Cluster analysis of migrant workers of differen generation.
Source: authors' own estimations.

3.2. Variables selection

3.2.1. Explained variables

In the first stage, the explained variable is whether the current status is working. If the migrant worker is already employed, it will be assigned a value of 1. And if it is temporarily unemployed, it will be given a value of 0. In the second stage, the explained variable is the choice of urbanization, which is expressed by the residence preference of migrant workers to stay in cities . According to naire question "Are you going to stay in the cities where you work in the future?" the explained variable is assigned a value of 1 if the answer is "yes"; otherwise, it is assigned a value of 0.

3.2.2. Explanatory variables

The core explanatory variable is self-identity and by following the study of Mead (1962), self-identity is divided into "I" identity and "Me" identity. This paper used psychological integration into the city to represent "I" identity and occupational prestige to represent

"Me" identity. The five questions in the questionnaire were selected for psychological integration. They are "I like the city/place I live in now", "I pay attention to the changes in the city/place I live in now", "I am very willing to integrate into the local people and become a member of them", "I feel that the locals are willing to accept me as one of them" and "I feel that I am already a local". For any of the above questions, if one chooses "totally disagree," it will get 1 point, "disagree" will earn 2 points, "basically agree" will get 3 points, and "completely agree" will get 4 points. Finally, add up the scores for the above questions. Occupational prestige matches the current main occupation of migrant workers with Li's (2005) score on Chinese occupational prestige measurement, which the lowest score being 0 and the highest being 72.41.

3.2.3. Control variables

Referring to the existing literature, the study also included other variables such as individual characteristics (logarithm of monthly income, education, gender, marriage, health, first migration is independent and the whole migration time), family characteristics (parental migration experience, the logarithm of monthly house rent and have a house/apartment), urban features (system of medical insurance for urban workers and personal social security card) as control variables. The summary characteristics of variables employed in the current study are illustrated below (see Table 1).

3.3. Empirical method

At the time of surveying, it is found that not all migrant workers participated in the work. However, most current studies have ignored this part of migrant workers. If these samples are eliminated or ignored, the estimation results may be biased. The reason is that when migrant workers are working, their willingness to stay in the cities is more stronger. Therefore, if this part of the sample is ignored, the model of migrant workers is a self-select rather than a random sample, and this non-random sample selection will lead to biased estimates. So, to eliminate the problem of sample selection bias, Heckman (1979) constructed a two-stage selection model to solve this problem. Therefore, this study uses Heckman two-stage selection model to divide migrant workers' urbanization into two stages, whether to work and the willingness to stay in cities. The first stage is the probit model to observe whether the functional status of migrant workers is affected. The second stage is the revised urbanization model, which further follows the influence of migrant workers' urbanization.

The model is:

$$\Pr(work = 1) = \Phi\left(\sum\nolimits_n \alpha^n Z_j^n\right) \tag{1}$$

$$\text{Urbanization}_i = \sum\nolimits_n \alpha^n Z_j^n + \beta\gamma_i + \epsilon_i \tag{2}$$

Equation (1) is the Heckman first-stage job selection model, $Pr(work = 1)$ which indicates that migrant workers have participated in work; $\Phi\left(\sum\nolimits_n a^n Z_j^n\right)$ indicates the

Table 1. Descriptive statistics.

		"post-80s"		"post-90s"	
Variable	Value	Mean	Std.	Mean	Std.
Residence preference	1 = Yes; 0 = No	0.847	0.360	0.806	0.396
Work	1 = Yes; 0 = No	0.872	0.335	0.776	0.417
Psychological integration	5 to 29 means the lowest level to the highest level	16.183	2.422	15.896	2.422
Occupational prestige	0 to 72.41means the lowest level to the highest level	39.813	12.407	39.644	13.356
Income	Yuan/month(log)	8.180	1.000	8.033	0.907
Education	1 = illiteracy; 2 = primary school; 3 = junior high school; 4 = senior high school; 5= junior college; 6 = university; 7 = Master or above	3.485	1.018	3.736	0.994
Gender	1 = male; 0 = female	0.520	0.500	0.410	0.492
Marriage	1 = Yes; 0 = No	0.921	0.269	0.503	0.500
Health	1 to 4 means the lowest level to the highest level	3.862	0.369	3.909	0.302
Self-mobility	1 = Yes; 0 = No	0.452	0.498	0.472	0.500
Mobility time	Year(s)	5.500	4.621	3.535	4.146
Family experience	1 = Yes; 0 = No	0.248	0.432	0.461	0.498
Rent (for house)	Yuan/month (log)	5.156	2.870	4.593	3.012
House	1 = Yes; 0 = No	0.257	0.437	0.188	0.391
SMIUW	1 = Yes; 0 = No	0.198	0.399	0.187	0.390
Social security card	1 = Yes; 0 = No	0.478	0.500	0.437	0.496

Source: authors' own estimations.

probability distribution function of standard normal distribution. Z means the factors that affect migrant workers' urbanization.

Equation (2) is the modified urbanization model for the second stage of Heckman. Equation (2) γ_i is the inverse Mill's ratio, which is used to overcome sample selection bias. The inverse Mill's ratio is calculated based on the estimated results of the first stage, and the specific calculation formula is as follows:

$$\gamma_i = \frac{\varphi\left(\sum_n \alpha^n Z_j^n\right)}{\Phi\left(\sum_n \alpha^n Z_j^n\right)} \tag{3}$$

Among them, $\varphi(\cdot)$ and $\Phi(\cdot)$ represent the standard normal distribution's probability density and distribution function, respectively. If γ_i is significantly non-zero, it means that there is a problem of sample self-selection and the Heckman selection model is effective.

For robustness check, this study also employed Bagging and random forest tests. The bootstrap aggregating (bagging) method is proposed by Breiman (1996) to estimate the test error, and machine learning is used to set the bootstrap sample B = 500. Due to the resampling with replacement, each tree has new observations. Average the predictions of these decision trees (without X_i), and record as the out-of-bag prediction value $\hat{y}_{i,OOB}$ for the i observation. The out-of-bag prediction value of all observations is $\{\hat{y}_{i,OOB}\}_{i=1}^{n}$, and compare the out-of-bag prediction value $\hat{y}_{i,OOB}$ with the actual observation value y_i. The out-of-bag error can be obtained as OOB mean square error:

$$MSE_{OOB} \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_{i,OOB} - y_i\right)^2 \tag{4}$$

Based on the bagging method, the random forest can measure the importance of variables. For each decision tree in the random forest, measure the decrease in the residual sum of squares caused by each variable. According to the magnitude of the decline, the average for each decision tree measures the importance of this variable. %IncMSE represents the percentage of the out-of-bag error caused by removing a variable from the model. IncNodePurity represents the variable importance calculated based on the training samples.

4. Results and discussion

4.1. Cluster analysis of migrant workers

Cluster analysis is a very useful technique, and it attemps obtain that partition which minimizes the within-cluster scatter or maximizes the between-cluster scatter (Jain et al., 2000). In this paper, the index of cluster analysis is self-identity–"I" identity and "Me" identity and devides migrant workers in three groups based on their ages. Ward method in clustering method is selected, and Euclidean Distance is selected as the measurement standard, as well as the original data is standardized. The result after clustering is shown in Figure 3. According to the result of cluster analysis and the similarity of migrant workers' self-identity are theoretically described as two groups. One group includes migrant workers who born in the 1970s and 1980s and the other includes migrant workers who born in the 1990s.

4.2. Correlation between self-identity and the urbanization of "post-80s" and "post-90s" migrant workers

Table 2 reports the results of correlation between self-identity on the urbanization of "post-80s" and "post-90s" migrant workers. The logistic regressions imply that there is a positive correlation between self-identity and the urbanization of migrant workers, which is parallel to the finding of the recent study of Liu and Wang (2019) and Gu et al. (2022). The results also provide a basic test for the further Heckman two-stage selection model (Table 2).

4.3. Influence of self-identity on the urbanization of "post-80s" and "post-90s" migrant workers

By using Heckman two-stage selection model, models (1) and (3) in Table 3 are about the impact of "I" identity on the urbanization of migrant workers. The regression results show that for "post-80s" and "post-90s" migrant workers, the psychological integration is significant at 1% level, indicating that the psychological integration has a significant positive impact on the urbanization of "post-80s" and "post-90s" migrant workers, which is parallel to the finding of the recent study of Gu et al. (2022). But comparing the marginal effects, the "post-90s" migrant workers'

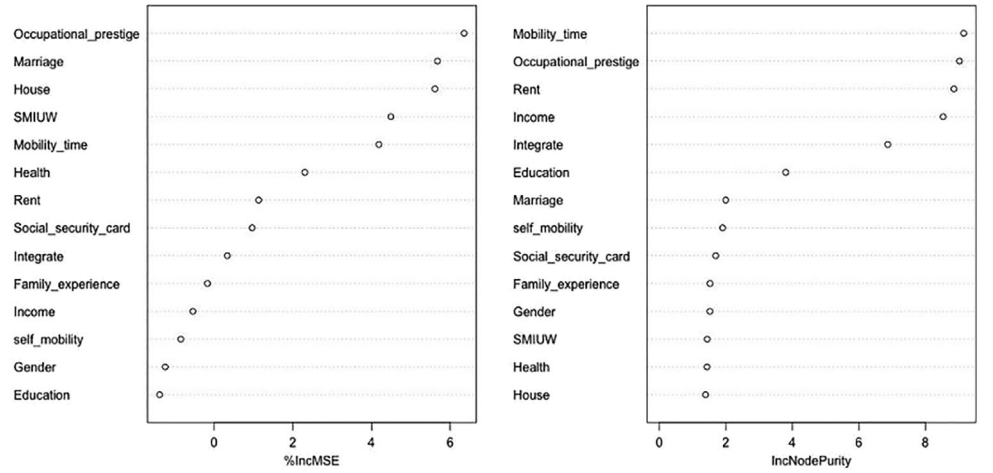

Figure 3. Percentage of "post-80s" migrant workers' urbanization out-of-bag error and the importance of variables.
Source: authors' own estimations.

Table 2. Results of correlation between self-identity on the urbanization of "post-80s" and "post-90s" migrant workers.

Variable	"post-80s" Urbanization	"post-90s" Urbanization
Psychological integration	1.234***	1.211***
	(0.01)	(0.01)
Occupational prestige	1.004***	1.001
	(0.00)	(0.00)
Income	1.069***	1.104***
	(0.01)	(0.02)
Education	1.195***	1.088***
	(0.02)	(0.02)
Gender	0.928**	1.020
	(0.03)	(0.03)
Marriage	1.767***	1.563***
	(0.08)	(0.06)
Health	1.163***	1.199***
	(0.05)	(0.07)
Self-mobility	1.134***	1.065*
	(0.03)	(0.04)
Mobility time	1.432***	1.372***
	(0.05)	(0.05)
Family experience	1.021***	1.023***
	(0.01)	(0.01)
Rent (for house)	1.728***	2.124***
	(0.08)	(0.15)
House/apartment	1.031***	1.026***
	(0.00)	(0.01)
Urban employee medical insurance	1.190***	1.161***
	(0.04)	(0.04)
N	39767	24037

Note. Standard errors are in parentheses. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.
Source: authors' own estimations.

psychological integration increases by 1, and the probability of choosing to reside in cities will increase by 2.85%. While in the case of "post-80s" migrant workers' psychological integration increases by 1, the likelihood of choosing to live in cities will increase by 2.64%. It is concluded that "post-90s" migrant workers are more

Table 3. Results of the effect of self-identity on the urbanization of "post-80s" and "post-90s" migrant workers.

	"post-80s"				"post-90s"			
	(1)		(2)		(3)		(4)	
Variable	Work	Urbanization	Work	Urbanization	Work	Urbanization	Work	Urbanization
Psychological integration	−0.0219***	0.0264***			−0.0343***	0.0285***		
	(0.005)	(0.001)			(0.006)	(0.001)		
Occupational prestige			0.0049***	0.0003*			0.0009	0.0002
			(0.001)	(0.000)			(0.001)	(0.000)
Income	0.0837***	0.0074***	0.0813***	0.0083***	0.1608***	0.0210***	0.1627***	0.0194***
	(0.009)	(0.002)	(0.009)	(0.002)	(0.012)	(0.004)	(0.012)	(0.004)
Education	−0.0736***	0.0171***	−0.0880***	0.0197***	−0.0676***	0.0084***	−0.0758***	0.0146***
	(0.012)	(0.002)	(0.012)	(0.002)	(0.015)	(0.003)	(0.015)	(0.003)
Gender	0.6685***	−0.0167**	0.6590***	−0.0118*	0.4100***	0.0075	0.4093***	0.0079
	(0.023)	(0.007)	(0.023)	(0.007)	(0.028)	(0.006)	(0.028)	(0.006)
Marriage	0.0288	0.0838***	0.0162	0.0875***	0.3315***	0.0702***	0.3305***	0.0728***
	(0.046)	(0.007)	(0.046)	(0.007)	(0.028)	(0.007)	(0.028)	(0.007)
Health	0.1000***	0.0142***	0.0883***	0.0273***	0.0148	0.0333***	−0.0050	0.0498***
	(0.029)	(0.005)	(0.029)	(0.005)	(0.044)	(0.010)	(0.044)	(0.010)
Self-mobility	0.1112***	0.0110***	0.1062***	0.0132***	0.4903***	0.0095	0.4853***	0.0153**
	(0.023)	(0.004)	(0.023)	(0.004)	(0.028)	(0.007)	(0.028)	(0.007)
Mobility time	−0.0425***	0.0042***	−0.0434***	0.0049***	−0.1580***	0.0017	−0.1599***	0.0031
	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.003)	(0.002)
Family experience	−0.1400***	0.0433***	−0.1385***	0.0395***	−0.2300***	0.0476***	−0.2263***	0.0421***
	(0.024)	(0.004)	(0.024)	(0.004)	(0.027)	(0.006)	(0.027)	(0.006)
Rent (for house)	0.0341***	0.0022***	0.0339***	0.0034***	−0.0126***	0.0041***	−0.0127***	0.0051***
	(0.004)	(0.001)	(0.004)	(0.001)	(0.005)	(0.001)	(0.005)	(0.001)
House/apartment	−0.6205***	0.0565***	−0.6426***	0.0753***	−0.9457***	0.0470***	−0.9871***	0.0772***
	(0.025)	(0.008)	(0.025)	(0.008)	(0.033)	(0.014)	(0.033)	(0.015)
Urban employee	0.1367***	0.0470***	0.1249***	0.0478***	0.3282***	0.0614***	0.3232***	0.0616***
medical insurance	(0.031)	(0.005)	(0.031)	(0.005)	(0.039)	(0.008)	(0.039)	(0.008)
Social security card	−0.0420*	0.0045	−0.0454*	0.0062	−0.0666**	0.0010	−0.0715**	0.0033
	(0.024)	(0.004)	(0.024)	(0.004)	(0.029)	(0.006)	(0.029)	(0.006)
Intercept	1.0607***	0.1109***	0.6633***	0.4285***	1.1573***	−0.0993*	0.6869***	0.2477***
	(0.152)	(0.031)	(0.139)	(0.032)	(0.213)	(0.051)	(0.200)	(0.051)
Inverse Mill's Ratio	−0.0851*		−0.0329		0.0650*		0.0778**	
	0.044)		(0.045)		(0.035)		(0.035)	
N	39767		39767		24037		24037	
Wald	2206.03***		1139.64***		1204.55***		625.15***	

Note. Standard errors are in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$.
Source: authors' own estimations.

concerned about psychological integration than "post-1980s" migrant workers. For "post-90s" migrant workers, "I" identity greatly impacts urbanization.

Models (2) and (4) in Table 3 are about the impact of "Me" identity on urbanization. The regression results show that occupational prestige has a significant positive impact on the "post-80s" migrant workers' urbanization, while occupational prestige has no significant impact for the "post-90s" migrant workers' urbanization. It means that when faced with whether or not to live in cities, occupational prestige has a more significant impact on the choice of "post-80s" migrant workers. For "post-80s" migrant workers, "Me" identity greatly impacts migrant workers' urbanization that supports the hypothesis. From the perspective of individual characteristics, monthly income has a significant positive impact on the urbanization for "post-80s" and "post-90s" migrant workers. Still, compared with "post-80s" migrant workers, monthly income has a significant impact on the urbanization of "post-90s". The reason may be that for the "post-90s" migrant workers, working is the way to earn wages and to survive in the cities, while the "post-80s" migrant workers are more concerned about their development in the cities. So compared with the "post-90s" migrant workers, monthly income has less impact on the urbanization of "post-80s" migrant workers. Education is also found to significantly and positively influence urbanization of migrant workers born in the 1980s and 1990s, which is parallel to the finding of the recent study of Cheng (2021). However, the impact of education on the urbanization of "post-80s" migrant workers is greater than that of "post-90s" migrant workers that infers that "post-80s" migrant workers with lower education levels cannot find satisfactory jobs in cities and choose to return home to find employment. The migration time has a significant positive impact on the urbanization of "post-80s" migrant workers, while it has no significant effect on the "post-90s". It reflects the point that the average migration time of the "post-80s" is longer, and their new social capital is higher than that of the "post-90s". From the perspective of family characteristics, parents' mobility experience has a significant positive impact on the urbanization of "post-80s" and "post-90s" migrant workers. Relevant intergenerational studies have found that the mobility behavior of parents will affect the migration behavior of offspring (Fang & Shi, 2018). Having a house/apartment is also found to significantly and positively influence the urbanization of "post-80s" and "post-90s" migrant workers; that is, owning a house/apartment in the city will significantly prompt migrant workers to choose urbanization. From the perspective of urban characteristics, participating in urban employee medical insurance has a significant positive impact on the urbanization of "post-80s" and "post-90s" migrant workers. But compared with "post-80s" migrant workers, participating in urban employee medical insurance has a greater impact on the "post-90s" urbanization, indicating that "post-90s" migrant workers have higher demand for public goods than "post-80s".

4.4. Impact of self-identity on the urbanization of migrant workers under the traditional division of new and old generations of migrant workers

Based on most scholars' research, this paper also divides the migrant workers into the old generation and the new generation to investigate the rationality of dividing the

new generation of migrant workers into "post-80s" and "post-90s" migrant workers for analysis. Models (5) and (7) in Table 4 are about the effect of "I" identity on urbanization. The regression results show that for the old and new generations of migrant workers, psychological integration has a significant positive effect on the urbanization of migrant workers. And comparing the marginal impact of the two, the psychological integration of the older generation of migrant workers and the new generation of migrant workers increases by 1, and the probability of urbanization will increase by 2.77% and 2.73%, respectively. The impact of migrant workers' urbanization is smaller between the new and old generations of migrant workers than that between the "post-80s" and "post-90s" migrant workers. Further, models (6) and (8) in Table 3 are about the impact of "Me" identity on the urbanization of the new and old generations of migrant workers. The regression results show that occupational prestige has a significant positive impact on the urbanization of the old and new generations of migrant workers. And the effect on the older generation of migrant workers is slightly more significant than that of the new generation of migrant workers. Judging from the results, if migrant workers are only divided into new and old ages, the heterogeneity between "post-80s" and "post-90s" migrant workers may be ignored, ignoring some factors may affect migrant workers' urbanization factors.

4.5. Heterogeneity results

Previous analysis has confirmed the impact of self-identity on the urbanization of the new generation of migrant workers. However, the above conclusion is only the average effect of the whole sample and does not consider the group heterogeneity of "post-80s" and "post-90s" migrant workers. Next, the current study grouped the samples based on three dimensions such as region, marital status, and education level to obtain more detailed research conclusions. Areas are divided into east, middle, and west, marital status divided into married and unmarried, and education is divided into less than nine years, 9-12 years, and more than 12 years. Tables 5–10 shows the estimation results based on the logistic model. Overall, the three dimensions of the regional endowment effect, family endowment effect, and human capital effect all show corresponding heterogeneity.

4.5.1. Regional endowment effect

Cities in the eastern region have an advantage in providing employment opportunities and income for migrant workers compared with cities in the central and western areas. From Tables 5 and 6, the urbanization and economic development is significantly different between east and west, which is parallel to the finding of the recent study of Gu et al. (2022). The "post-80s" migrant workers who flow into the region of the west each increase by one unit; psychological integration will make the ratio of migrant workers choosing urbanization to be 1.25 times higher than that of the "post-80s" migrant workers in the eastern and middle regions. For "post-90s" migrant workers, when their inflows are in the eastern part, compared with the central and western areas, each additional unit of psychological integration will increase the rate of migrant workers choosing to urbanize. The occupational prestige of the "post-80s"

Table 4. Results of the effect of self-identity on the urbanization of the old generation and the new generation migrant workers.

	Old generation				new generation			
	(5)		(6)		(7)		(8)	
Variable	Work	Urbanization	Work	Urbanization	Work	Urbanization	Work	Urbanization
Psychological integration	−0.0273***	0.0277***			−0.0289***	0.0273***		
	(0.005)	(0.001)			(0.003)	(0.001)		
Occupational prestige			0.0093***	0.0004*			0.0032***	0.0003**
			(0.001)	(0.000)			(0.001)	(0.000)
Income	0.0852***	0.0182***	0.0823***	0.0171***	0.1181***	0.0100***	0.1176***	0.0102***
	(0.007)	(0.002)	(0.007)	(0.002)	(0.007)	(0.002)	(0.007)	(0.002)
Education	0.0345**	0.0168***	0.0169	0.0203***	−0.0838***	0.0139***	−0.0960***	0.0178***
	(0.014)	(0.002)	(0.014)	(0.003)	(0.009)	(0.002)	(0.009)	(0.002)
Gender	0.6244***	0.0354***	0.6196***	0.0423***	0.5600***	−0.0063	0.5544***	−0.0044
	(0.024)	(0.008)	(0.024)	(0.008)	(0.017)	(0.004)	(0.017)	(0.004)
Marriage	−0.1065	0.0227	−0.1206	0.0257*	0.5018***	0.0714***	0.4956***	0.0783***
	(0.120)	(0.015)	(0.119)	(0.016)	(0.020)	(0.005)	(0.020)	(0.005)
Health	0.1082***	0.0312***	0.0925***	0.0375***	0.0355	0.0203***	0.0203	0.0337***
	(0.020)	(0.004)	(0.020)	(0.004)	(0.024)	(0.005)	(0.024)	(0.005)
Self-mobility	0.0120	0.0124***	0.0081	0.0117***	0.3629***	0.0059	0.3589***	0.0100***
	(0.026)	(0.004)	(0.026)	(0.004)	(0.017)	(0.004)	(0.017)	(0.004)
Migration time	−0.0055***	0.0026***	−0.0065***	0.0035***	−0.0749***	0.0043***	−0.0762***	0.0052***
	(0.002)	(0.000)	(0.002)	(0.000)	(0.002)	(0.001)	(0.002)	(0.001)
Family experience	−0.0438	0.0370***	−0.0461	0.0375***	−0.2734***	0.0477***	−0.2711***	0.0420***
	(0.043)	(0.008)	(0.043)	(0.008)	(0.017)	(0.004)	(0.017)	(0.004)
Rent (for house)	0.0142***	0.0026***	0.0124***	0.0036***	0.0108***	0.0033***	0.0107***	0.0043***
	(0.004)	(0.001)	(0.004)	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)
House/apartment	−0.5553***	0.0308***	−0.5878***	0.0565***	−0.7534***	0.0578***	−0.7836***	0.0814***
	(0.027)	(0.009)	(0.027)	(0.009)	(0.019)	(0.007)	(0.019)	(0.007)
Urban employee	0.0568	0.0712***	0.0474	0.0741***	0.2323***	0.0505***	0.2233***	0.0511***
medical insurance	(0.038)	(0.006)	(0.038)	(0.007)	(0.023)	(0.005)	(0.023)	(0.005)
Social security card	−0.0193	−0.0022	−0.0291	0.0000	−0.0455**	0.0037	−0.0497***	0.0058
	(0.024)	(0.004)	(0.024)	(0.004)	(0.018)	(0.003)	(0.018)	(0.004)
Intercept	0.9787***	−0.0858**	0.3696**	0.2942***	0.9778***	0.0574**	0.5188***	0.3906***
	(0.167)	(0.034)	(0.152)	(0.036)	(0.118)	(0.026)	(0.109)	(0.026)
Inverse Mill's Ratio	0.2876***		0.3473***		−0.0384		−0.0129	
	(0.067)		(0.065)		(0.026)		(0.027)	
N	45135		45135		63804		63804	
Wald	3772.96***		2100.38***		2437.41***		1160.00***	

Note. Standard errors are in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$.
Source: authors' own estimations.

Table 5. Results of the regional endowment effect: The "I" identity estimation results of samples by regions.

	"post-80s"						"post-90s"					
	West		Middle		East		West		Middle		East	
Variable	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)
Psychological integration	0.223***	1.250	0.214***	1.239	0.208***	1.232	0.195***	1.215	0.185***	1.204	0.204***	1.227
Control variable	Y		Y		Y		Y		Y		Y	
N	8816		7409		16938		5932		4117		10632	
Pseudo R2	0.0764		0.0899		0.0871		0.0679		0.0916		0.0729	

Note. Standard errors are in parentheses. ***p<.01, **p<.05, *p<.1.
Source: authors' own estimations.

Table 6. Results of the regional endowment effect: The "Me" identity estimation results of samples by regions.

	"post-80s"						"post-90s"					
	West		Middle		East		West		Middle		East	
Variable	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)
Occupational prestige	0.003	1.003	−0.001	0.999	0.006***	1.006	0.004	1.004	−0.003	0.997	−0.001	0.999
Control variable	Y		Y		Y		Y		Y		Y	
N	8816		7409		16938		5932		4117		10632	
Pseudo R2	0.0378		0.0553		0.0570		0.0764		0.0899		0.0871	

Note. Standard errors are in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$.
Source: authors' own estimations.

Table 7. The family endowment effect: The "I" identity estimation results of samples by marriage.

	"post-80s"				"post-90s"			
	Unmarried		Married		Unmarried		Married	
Variable	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)
Psychological integration	0.199***	1.220	0.213***	1.237	0.190***	1.209	0.196***	1.217
Control variable	Y		Y		Y		Y	
N	3420		36347		13346		10691	
Pseudo R2	0.0540		0.0782		0.0559		0.0735	

Note. Standard errors are in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$.
Source: authors' own estimations.

Table 8. The family endowment effect: The "Me" identity estimation results of samples by marriage.

	"post-80s"				"post-90s"			
	Unmarried		Married		Unmarried		Married	
Variable	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)
Occupational prestige	0.002	1.002	0.003**	1.003	0.001	1.001	0.001	1.001
Control variable	Y		Y		Y		Y	
N	3420		36347		13346		10691	
Pseudo R2	0.0228		0.0439		0.0278		0.0440	

Note. Standard errors are in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$.
Source: authors' own estimations.

migrant workers in the eastern region has a significant impact on the residence preference. In contrast, the occupational prestige of the "post-90s" migrant workers in the different areas has no significant effect on the willingness to stay in cities. This shows that in terms of urbanization for migrant workers, the "post-90s" migrant

Table 9. The human capital effect: The "I" identity estimation results of samples by education.

	"post-80s"						"post-90s"					
	≤9 Years		9-12 Years		>12 Years		≤9 Years		9-12 Years		>12 Years	
Variable	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)
Psychological integration	0.199***	1.220	0.230***	1.259	0.258***	1.294	0.173***	1.189	0.226***	1.253	0.193***	1.213
Control variable	Y		Y		Y		Y		Y		Y	
N	23670		9677		6420		11286		7466		5285	
Pseudo R2	0.0610		0.0867		0.1058		0.0570		0.0769		0.0858	

Note. Standard errors are in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$.
Source: authors' own estimations.

Table 10. The human capital effect: The "Me" identity estimation results of samples by education.

	"post-80s"						"post-90s"					
	≤9 Years		9-12 Years		>12 Years		≤9 Years		9-12 Years		>12 Years	
Variable	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)	B	Exp(B)
Occupational prestige	0.005***	1.005	0.001	1.001	0.002	1.002	0.002	1.002	0.000	1.000	−0.001	0.999
Control variable	Y		Y		Y		Y		Y		Y	
N	23670		9677		6420		11286		7466		5285	
Pseudo R2	0.0292		0.0494		0.0625		0.0325		0.0396		0.0600	

Note. Standard errors are in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$.
Source: authors' own estimations.

workers in the eastern region and the "post-80s" migrant workers in the western area care more about the identity of "I". In contrast, the "post-80s" migrant workers in the eastern part care more about "Me" identity. One of the reasons is that the "post-90s" migrant workers who grew up in the Internet era are willing to stay in the more economically developed eastern region where they are more adaptable to emerging things than the "post-80s", so they have a more stronger psychological integration in the eastern region. Most "post-80s" migrant workers are married and have established their businesses, and they are relatively clear about their future development. As for whether to stay in the eastern region to continue working and living, they pay less attention to their subjective feelings and pay more attention to the attitudes of others towards them.

4.5.2. Family endowment effect

Chinese people put their family in a more critical position than their own development, and there is a tradition of "taking care of the family" (Cai, 2003). Many previous studies have found that the migration of spouses and children increases the willingness of migrant workers to stay in cities. Comparing Tables 7 and 8, the married "post-80s" and "post-90s" migrant worker groups are more likely to choose urbanization than unmarried groups. Among married "post-80s" migrant workers, occupational prestige significantly influences their urbanization. Occupational prestige does not affect urbanization whether or not "post-90s" migrant workers are married. So "Me" identity of married "post-80s" migrant workers significantly influences their urbanization. They have long-term plans for the future, and most of them make decisions based on the family as a whole instead of only considering their own psychological integration.

4.5.3. Human capital effect

Migrant workers with different educational levels have various benefits in cities, affecting their urbanization. Higher education for "post-80s" migrant workers increases the probability of urbanization, which shows the same finding of Haryanto et al. (2021), while the "post-90s" migrant workers do not find this rule. The occupational prestige of "post-80s" migrant workers with less than nine years of education significantly impacts their urbanization, while education has no significant effect on the occupational prestige of "post-90s" migrant workers. It is not difficult to find that the "I" identity of "post-80s" migrant workers with higher education has a more significant impact on their urbanization. In comparison, the "Me" identity of "post-80s" migrant workers with lower education has a more significant effect on their urbanization. For "post-80s" migrant workers, the higher their education, the more opportunities they have to engage in higher-paying jobs and more freedom to get jobs. Therefore, psychological integration is more likely to affect "post-80s" migrant workers' urbanization.

4.6. Robustness results

Random forests are an effective tool in prediction, and they do not overfit because of the Law of Large Numbers. Injecting the right kind of randomness makes them accurate regressors. The adaptive bagging algorithm in regression was designed to reduce bias and operates effectively in regression. Forests give results competitive with boosting and adaptive bagging, yet do not progressively change the training set. Their accuracy indicates that they act to reduce bias (Breiman, 2001). For robustness check, by employing bagging and random forest test, the Figures 3 and 4 show the percentage and varying importance of the increase in out-of-bag errors in the urbanization of "post-80s" migrant workers and "post-90s" migrant workers, respectively. The impact of the selected 14 relevant urbanization variables, the psychological integration of "post-80s" migrant workers ranked ninth in the percentage of increase in out-of-bag error, while

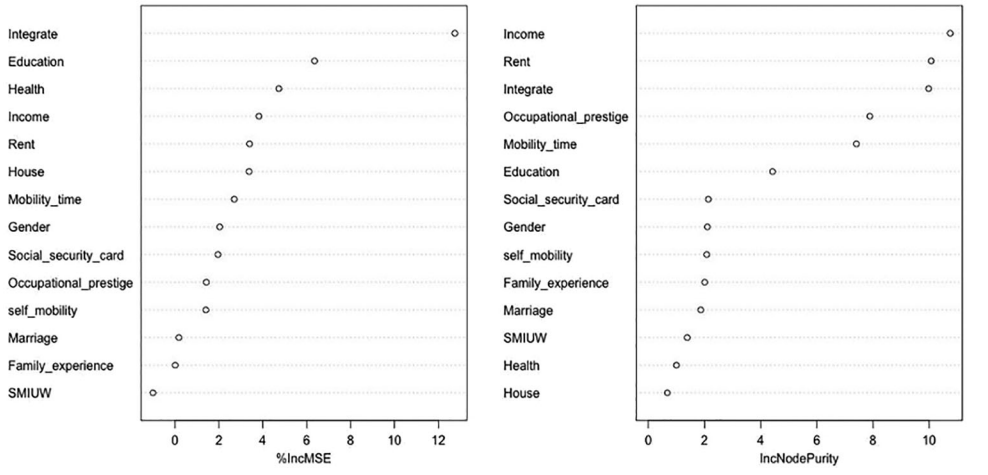

Figure 4. Percentage of "post-90s" migrant workers' urbanization out-of-bag error and the importance of variables.
Source: authors' own estimations.

that variable for "post-90s" ranks first. For the urbanization of "post-90s" migrant workers, if the variable of psychological integration is removed, the percentage of out-of-bag error will be higher than that of "post-80s" migrant workers. The importance of the variable "post-80s" migrant workers' psychological integration ranks fifth, while the variable of "post-90s" ranks third. That means compared with "post-80s" migrant workers, the psychological integration has a more significant impact on the urbanization of "post-90s" migrant workers, which means "post-90s" migrant workers pay more attention to the "I" identity. "Post-80s" migrant workers ranked first in the percentage of out-of-bag error increase in occupational prestige, while the variable "post-90s" ranked 10th. It indicates that compared with "post-90s" migrant workers, the percentage of the out-of-bag error in the variable of removing occupational prestige will increase a more significant share when discussing "post-80s" migrant workers. At the same time, the importance of the variable of "post-80s" migrant workers' occupational prestige ranks second, while the variable of "post-90s" migrant workers ranked fourth. For "post-80s" migrant workers, occupational prestige has a more significant impact on urbanization, which means when migrant workers consider urbanization, "post-80s" pay more attention to "Me" identity, consistent with the hypothesis.

5. Conclusion

Urbanization has enabled rapid economic growth and long-term social stability during China's reform and opening up, while new urbanization usually refers to people-centered urbanization. It is mainly driven by the transfer of rural labor to cities for long-term residence and employment. However, the growth rate of the profession of migrant workers is slowing down, and how to make the floating population stay in cities has become one of the main topics of current research. The "post-90s" migrant workers grew up in the era of rapid Internet development, and their thoughts and behaviors are quite different from those of the "post-80s" migrant workers. It shows that the different socialization degree of the new generation of migrant workers of "post-80s" and "post-90s". Therefore, the heterogeneity within the new generation of migrant workers cannot be ignored. This paper argues that the difference between "post-80s" and "post-90s" migrant workers are mainly reflected in their self-identity. "Post-90s" migrant workers pay more attention to personal feelings and interests, which is "I" identity, while "post-80s" pay more attention to follow the social culture to which they belong, which is "Me" identity.

Using the relevant data of the China Migrants Dynamics Survey (CMDS) in 2017, this paper empirically tests the impact of self-identity on the urbanization of "post-80s" and "post-90s" migrant workers by using Heckman selection model. The study found that the division method of the new and old generations of farmers will mask some of the factors that affect the urbanization of migrant workers. First, compared with the "post-80s" migrant workers, the "post-90s" migrant workers' "I" identity has a more significant impact on urbanization. "Post-90s" migrant workers pay more attention to psychological integration than the "post-80s" migrant workers. Second, compared with the "post-80s" migrant workers, the "post-90s" migrant workers' "Me" identity has a smaller impact on the urbanization, which means the "post-90s" migrant workers are

less concerned about their professional prestige than the "post-80s" migrant workers. The above findings are further supported by the bagging and random forest method tests in machine learning. Moreover, the heterogeneity results show that the regional endowment effect of the self-identity of migrant workers is significant. Compared with other regions, the "post-90s" migrant workers in the eastern part are more concerned about the "I" identity, and the "post-80s" migrant workers are more concerned about "Me" identity. Compared with "post-90s" migrant workers, marriage has a more significant impact on "post-80s" migrant workers' "Me" identity. Compared with the "post-90s" migrant workers, the human capital effect of the "post-80s" urbanization is more significant. This paper supposes that the differences in the self-identity of migrant workers between the "post-80s" and "post-90s" will lead to different urbanization. Therefore, the two new generations of migrant workers, the "post-80s" and "post-90s", should be studied separately. According to their different needs in the cities, make precise strategies to attract them to stay.

Based on the empirical findings, the current study proposes the following policy suggestions. Firstly, in the process of promoting urbanization, all localities should pay more attention to the psychological integration of migrant workers, especially the "post-90s" migrant workers' sense of belonging and ownership of the city and committed to creating a good atmosphere for social integration may be more effective in attracting migrant workers to reside. Secondly, public opinion should eliminate discrimination against migrant workers' occupations, fully respect each occupational's domain, and prevent migrant workers from choosing to return because of their low occupational prestige and lack of self-identification in cities.

Acknowledgments

We are grateful to Ye Luo and two referees for very helpful comments.

Disclosure statement

No potential conflict of interest was reported by authors.

Data available statement

The data is with open access. Stata and R programs available on request.

Funding

Authors further acknowledges financial support from the Ministry of Education of China grant number 20YJA790020.

ORCID

Siyu Xu http://orcid.org/0000-0002-0398-8574
Jun He http://orcid.org/0000-0002-5751-0867
Noshaba Aziz http://orcid.org/0000-0001-5948-2807

References

Abel, G. J., & Sander, N. (2014). Quantifying global international migration fows. *Science*, *343*(6178), 1520–1522. https://doi.org/10.1126/science.1248676

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cai, Q. (2003). Migrant remittances and family ties: A case study in China. *International Journal of Population Geography*, *9*(6), 471–483. https://doi.org/10.1002/ijpg.305

Cheng, Z. (2021). Education and consumption: Evidence from migrants in Chinese cities. *Journal of Business Research*, *127*, 206–215. https://doi.org/10.1016/j.jbusres.2021.01.018

Cheng, Z., Wang, H., & Smyth, R. (2014). Happiness and job satisfaction in urban China: A comparative study of two generations of migrants and urban locals. *Urban Studies*, *51*(10), 2160–2184. https://doi.org/10.1177/0042098013506042

Chen, S., & Liu, Z. (2016). What determines the settlement intention of rural migrants in China? Economic incentives versus sociocultural conditions. *Habitat International*, *58*, 42–50. https://doi.org/10.1016/j.habitatint.2016.09.004

Constant, A., & Massey, D. (2003). Self-selection, earnings, and out-migration: A longitudinal study of immigrants to Germany. *Journal of Population Economics*, *16*(4), 631–653. https://doi.org/10.1007/s00148-003-0168-8

Fang, Y., & Shi, Z. (2018). Children of migrant parents: Migrating together or left behind. *Habitat International*, *76*, 62–68. https://doi.org/10.1016/j.habitatint.2018.05.001

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*(3), 481–510. http://www.jstor.org/ stable/2780199. https://doi.org/10.1086/228311

Gu, H., Lin, Y., & Shen, T. (2022). Do you feel accepted? Perceived acceptance and its spatially varying determinants of migrant workers among Chinese cities. *Cities*, *125*, 103626. https://doi.org/10.1016/j.cities.2022.103626

Han, C. (2010). Talking about "post-90s" migrant workers. *Country Agriculture Farmers(A)*, *2*, 22–23. in Chinese.

Haryanto, T., Erlando, A., & Utomo, Y. (2021). The relationship between urbanization, education, and GDP per capita in Indonesia. *The Journal of Asian Finance, Economics and Business*, *8*(5), 561–572. https://doi.org/10.13106/JAFEB.2021.VOL8.NO5.0561

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. https://doi.org/10.2307/1912352

Hu, F., Xu, Z., & Chen, Y. (2011). Circular migration, or permanent stay? Evidence from China's rural-urban migration. *China Economic Review*, *1*, 64–74. https://doi.org/10.1016/j.chieco.2010.09.007

Inkeles, A. (1969). Making men modern: On the causes and consequences of individual change in six developing countries. *American Journal of Sociology*, *75*(2), 208–225. https://doi.org/10.1086/224767

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4–37. https://doi.org/10.1109/34.824819

Johnson, D. G. (2003). Provincial migration in China in the 1990s. *China Economic Review*, *14*(1), 22–31. https://doi.org/10.1016/S1043-951X(02)00085-8

Lee, W. O., & Qi, J. (2021). From lost identity to identity grafting: The second generation migrant workers in Beijing. *Journal of Educational Change*, *22*(2), 191–220. https://doi.org/10.1007/s10833-020-09387-y

Li, C. (2005). Prestige stratification in the contemporary China: Occupational prestige measures and socio-economic index. *Sociological Studies*, *2*, 74–102. in Chinese.

Liu, L., & Wang, J. (2019). Rational choice of urban long-term residence willingness of floating population–empirical study based on nonlinear hierarchical model. *Population Journal*, *3*, 100–112. in Chinese.

Mead, G. H. (1962). *Mind, self and society*. The University of Chicago Press.

Stark, O., & Bloom, D. E. (1985). The new economics of labor migration. *The American Economic Review*, *75*(2), 173–178.

Tan, J., Xu, H., Yu, J., & Radu, V. (2022). Consumption and savings of migrants in China–social cohesion perspective. *Economic Research-Ekonomska Istraživanja*, *35*(1), 4029–4046. https://doi.org/10.1080/1331677X.2021.2010110

Vroome, T., & Tubergen, F. (2014). Settlement intentions of recently arrived immigrants and refugees in the Netherlands. *Journal of Immigrant & Refugee Studies*, *12*(1), 47–66.

Wang, X. (2013). On the essential characteristics of peasant workers born in 1990s -A case study of the Pearl River. *Journal of Guangxi University for Nationalities (Philosophy and Social Science Edition)*, *1*, 128–133. in Chinese.

Wang, K., & He, S. (2019). Unraveling the marginalization of new generation peasant workers in China: Cultural reproduction and symbolic construction. *Journal of Urban Affairs*, *41*(3), 282–304. https://doi.org/10.1080/07352166.2018.1451253

Wei, L. (2018). The development of modernity and post-modernity: An analysis of the "Post-90s" life values based on intergenerational comparison. *Contemporary Youth Research*, *6*, 17–23. in Chinese.

Wellman, B. (2001). Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research*, *25*(2), 227–252. https://doi.org/10.1111/1468-2427.00309

Xiao, Z., Xu, S., & Liu, J. (2019). The assessment of social integration of urban migrant population: An investigation based on 50 cities of migration destination. *Population Journal*, *5*, 96–112. in Chinese.

Yang, H., & Curdt-Christiansen, X. L. (2021). Conflicting linguistic identities: language choices of parents and their children in rural migrant workers' families. *Current Issues in Language Planning*, *22*(4), 408–426. https://doi.org/10.1080/14664208.2020.1748370

Zhao, L., Liu, S., & Zhang, W. (2018). Trends in internal migration in China: Profiles of the new-generation migrants. *China & World Economy*, *26*(1), 18–41. https://doi.org/10.1111/cwe.12227

Zhou, X. (2005). The institutional logic of occupational prestige ranking: Reconceptualization and reanalyses. *American Journal of Sociology*, *111*(1), 90–140. https://doi.org/10.1086/428687