

# **Predicting Migrant Workers' Intention to Acquire Local Citizenship (Hukou) in Emerging Cities Using Machine Learning Models: A Case Study of Five Dongguan Factories**

**Feng-Shuo Chang**

Ph.D., Associate Professor, Guangdong University of Petrochemical Technology  
(Maoming, Guangdong Province, China)

E-mail: [feng.shuo.chang@gmail.com](mailto:feng.shuo.chang@gmail.com)

<https://orcid.org/0000-0002-4315-0194>

**Elisa Impara**

Ph.D., Senior Research Associate, King's College London, Institute of Psychiatry,  
Psychology & Neuroscience (London, United Kingdom)

E-mail: [elisa.impara@googlemail.com](mailto:elisa.impara@googlemail.com)

<https://orcid.org/0000-0002-0853-8425>

Chang, Feng-Shuo and Elisa Impara (2022) Predicting Migrant Workers' Intention to Acquire Local Citizenship (Hukou) in Emerging Cities Using Machine Learning Models: A Case Study of Five Dongguan Factories. *Future Human Image*, Volume 18, 10-27. <https://doi.org/10.29202/fhi/18/2>

*This paper examines the data from a period when hukou migration was severely restricted, and aims to provide a reproducible analytical methodology with machine-learning methods to model migrants' willingness to change their hukou. The data in this work is based on questionnaires and interviews collected in five factories in Dongguan from 2005 to 2007. After performing stepwise variable selection, our best-performing model is based on three main predictors: having children, education level, and household income. Analysis of interviews suggests that citizenship is regarded as an individual's legal status and is used in exchange for other types of capital.*

*Keywords: logistic regression, hukou migration, migrant workers, Dongguan, hukou points system*

Received: 28 October 2022 / Accepted: 16 November 2022 / Published: 15 December 2022

---

© Chang, Feng-Shuo, 2022

© Impara, Elisa, 2022

## Introduction

In the 2000s, China's coastal cities realized they needed more migrants to obtain a local household registration due to labor shortages. According to the Chinese household registration policy at the time, although these floating populations were Chinese citizens, they did not have a local hukou in their place of residence (their household was not registered as living there). As they did not have local citizenship, they could not enjoy complete social welfare and rights. Some scholars believe that migrant workers, who cannot obtain local hukou, let employers circumvent social welfare duties: this allows China to provide cheap productivity, attract foreign investment to set up factories, and promote the development of manufacturing: this is particularly visible in coastal zones with special economic status (Chan, 2010; Knight et al., 2011; Lai, 2007).

After China joined the World Trade Organization (WTO) in 2001, the demand for labor in coastal cities with manufacturing industries increased sharply. In addition, since 2004, the central government has formulated a plan to develop the inland economy to coordinate regional development. The simultaneous development of coastal and inland areas led to labor shortages, especially in 2004-2006 and 2009 (Chan, 2010). When coastal areas are no longer the only choice for migrant workers, the original policy that restricts migrant workers from applying for household registration in the local area must be re-adjusted to avoid the loss of labor due to the strict settlement policy. In hukou policy, the main way to obtain citizen status in other cities (that is, to switch to hukou status) is to inherit a family member's hukou, achieve higher education, or be given a job. However, with the demand for talent and with funds for local development, a hukou is also beginning to be available through investment, real estate purchases, and the fulfilment of specific talent needs (Chan, 2010; Cui & Cohen, 2015; Wu, 2013). After experiencing labor shortages and various settlement policies, Zhongshan in 2009 and Dongguan in 2010 began to adopt a points system to determine eligibility for hukou migration: these places have become popular as a result of manufacturing industries that attracted large numbers of migrant workers since the 1990s.

Nevertheless, many scholars question whether the accumulation of points in the points system favors skilled workers over unskilled workers. Coupled with the central allocation policy in this period (the quota for hukou migration), the local government policy for attracting technical personnel was only old wine in a new bottle, and empirical research shows that the hukou reform at that time did not achieve the expected results (Cui & Cho, 2020; Wu, 2013). However, under such policy obstacles, people who are still considering hukou migration are likely to have good reasons for doing so. If we can understand the motives of migrant workers who wanted to change their hukou at that time, this can be helpful in terms of adjusting the hukou policy so to attract laborers, thereby increasing urban competitiveness and balancing regional development.

Change of residence and change of household registration are two related terms, yet they hold different meanings. Most studies have focused on the physical migration of migrant workers or their willingness to live in the city where they work (Fan, 2011; Huang & Zhang, 2013; Yue et al., 2010; Zhu, 2007). Some researchers have also analyzed the characteristics of migrant workers who have successfully changed their hukou (Lu, 2008; Treiman & Zhang, 2011). However, there are few studies on migrant workers' rationale for changing their household registration during this period (Cai & Wang, 2008; Tang & Feng, 2012). To better understand what both change of residence and change of household registration really entail,

it is helpful to explore these concepts separately from physical migration. Defining citizenship as a 'legal status conversion' and link it to available resources might be equally helpful. Bauder (2008) reinterpreted the ideas of 'migrant' and 'citizen' from the perspective of capital. Bauder believed that citizenship has the two functions: 'distinction' and 'subordination'; outsiders who acquire citizenship can cross or reduce the boundaries between distinction/subordination, achieve economic progress or a desired social status. Bauder gave the examples of Chinese investments in Hong Kong and Taiwan; immigration to Canada, which allowed family wealth to survive and expand; and the example of Lebanese people using their immigration status in Canada to gain the ability to travel to and from Lebanon with ease. Here, Bauder emphasized that the role of 'citizen status' allows for the conversion of 'capital' into different social advantages.

In China, many migrants do not arrive in cities seeking permanent residence. Smart and Smart's (2001) fieldwork in a small-town factory in the Pearl River Delta found that the primary objective for migrant workers is the accumulation of wealth. If migrant workers need to decide whether to migrate permanently, buy a house locally, or relocate their hukou, the primary factor in their decision is whether the action would be helpful for the accumulation of wealth. In addition to economic considerations, which vary in different phases of life, the process of migrant workers seeking either hukou or physical migration involves individual sense of belonging or the meaning of hukou to them (Wang & Fan, 2012). However, these relatively subjective considerations are not easy to clarify. By defining hukou migration as a 'legal status conversion', we can separate it from the broader sense of migration and distinguish it from physical migration. Tang and Feng (2012) divided settlement intention into two models: 'hukou conversion' and 'not involving hukou conversion' (just living in the city until they want to leave). They found that for migrants with higher resources and who already own real estate, physical relocation was sufficient, and household registration was not necessary. This finding suggests that capital accumulation is the primary goal of migrant workers.

## **Materials and Methods**

### **Migrant Workers' Hukou Migration Intentions in a Migrant City**

The migration pattern in China's emerging cities is different from that in traditional large cities like Beijing, Shanghai, Guangzhou and in all provincial capital cities. In February 2018, Dongguan cancelled the point-based household registration system that had been in place since 2010. In April of the same year, Beijing officially started a trial of the points system. In terms of the hukou migration quota, Guangzhou, Guangdong's largest city, allowed only 3,000 hukou migrations per year from 2010 to 2018. By contrast, Dongguan, an emerging city known as 'the world's factory', allowed 12,370 hukou migrations in 2010 and lifted quota limits on the number of hukou migrants in 2018. An empirical study in the Pearl River Delta in 2008 found that the hukou migration willingness of migrant workers in Shenzhen and Zhuhai was higher than that of such workers in Guangzhou (Cai & Wang, 2008). This was due to the differences between emerging cities and traditional cities in the migrant and local population structure, which led to differences in hukou policies and migration patterns. Traditional cities have proportionally larger locally registered populations, while many migrant or emerging cities were originally small rural townships.

Previous studies on local citizenship (or local hukou) implied a positive correlation between hukou migration and physical migration: they focused on traditional large cities, and

relatively few studies consider the migration of migrant workers in emerging cities (Cai & Wang, 2008; Cui et al., 2015; Fan, 2002; Huang & Zhang, 2013; Yue et al., 2010; Zhu, 2007). This article aims to reinterpret the hukou relocation intentions of migrants through the perspective of citizenship as a form of capital. It focuses on the meaning of local citizenship to migrant workers in emerging cities. This research follows up on the studies that have been conducted in Dongguan since 1989 and analyses the migration intentions of migrants in migrant cities before 2010 (Li, 1997; Smart & Lin, 2007; Smart & Smart, 2001). Especially as the local hukou policy has been transformed from restricting the flow of people to helping attract specific human resources, it is important to re-examine the intentions behind migration. Finally, given the importance of this subject area, emphasis is placed on the reproducibility of research methods: we achieved this via predictive analytics that feature use of a wide range of procedures (e.g. stepwise variable selection, cross-validation, bootstrapping).

### **The Study Subject Matter and Sampling Design**

This research consists of two parts (Chang, 2009): semi-structured interviews, and a survey, which took place from 2005 to 2007. All data were collected in five factories in Dalang Town, Dongguan. Two hours north of Hong Kong by bus, the coastal city of Dongguan contained 4.6 million migrants and 1.8 million local citizens in 2009, in contrast to historically large cities, such as Shanghai, where there are 14 million local citizens – nearly three times the number of migrants in the city (five million) (Statistics Bureau of Dongguan, 2010a; Statistics Bureau of Shanghai, 2010). This concentration of migrant workers is a result of the export manufacturing economy that has, since 1978, been driven by the neighboring city of Shenzhen. Shenzhen became one of the first special economic zones (SEZs) in China, which meant that it and the surrounding cities, such as Dongguan, offered reduced taxes and land to attract foreign investors. According to the Statistics Bureau of Dongguan (2010b), more than 85% of the migrant workers there from 1996 to 2000 worked in the manufacturing sector. Previous studies have also shown that female and young workers make up the majority of migrants (Chan, 2010; Roberts, 1997).

The dramatic success of economic development in Dongguan was supported not only by cooperation among local elites and foreign investors, but also by the presence of migrant workers from outside Dongguan. However, due to their lack of local household registration status, these temporary migrants (Table 1), were excluded from the social security system (Smart & Lin, 2007). Li (1997) compared the income determination of migrants in Dongguan with that of migrants in Meizhou (a city in a more remote location) and found that high education and skill levels had a significant and positive effect on migrants' income in Meizhou. In Dongguan, however, a high education level resulted in high income among only professional and technical workers. Moreover, local governments applied household conversion measures that favored professional investors and house buyers, which created a new privileged group of permanent migrants with advanced education or special skills. Migrant workers with fewer skills and less education were exploited as inexpensive labor by elites (Chan, 2010; Solinger, 1999).

Household registration place and current residency	Three types of citizens
<b>Local status:</b> Place of residency is the same as place of household registration.	<b>Native:</b> Local residents since birth  <b>Permanent migrant:</b> Migrants whose household registration has been converted from non-local to local status (transfer place of household registration)
<b>Nonlocal status:</b> Place of residency is different from place of household registration.	<b>Temporary migrant</b>

**Table 1.** Definitions of Native, Permanent Migrant, and Temporary Migrant

The development background of Dalang Township in this study is similar to that of Henggang Township, which was studied by Smart and Lin (2007). Both are small towns that attract a large number of migrants due to their manufacturing sector, and there are thousands of migrant workers in factories in both towns. Normally, the factories provide migrant workers three meals per day and accommodations. The dormitories are 4-12-person bedrooms designed for single migrants; they are usually located next to factories and have curfew and controlled access. Limited en-suite rooms for employees with families are provided to managerial employees. Married migrant workers or couples are more likely to rent accommodations away from the factory.

The qualitative sample (n = 30), consisting of local residents and representatives of institutions or factories, was selected through contact with the local plant managers. Thematic content analysis was the primary analytical tool.<sup>1</sup> Samples of successful household migration cases are not easily available due to privacy concerns: in order to achieve sample heterogeneity, the study included interviewees who bought a house in Dongguan and interviewees who were locals. 18 residents, including 6 locals, were interviewed. 5 of the other 12 non-locals had bought houses there. Interviews focused on three topics:

1. migration (motivation and future planning);
2. residence (status and future purchase of a house in Dongguan);
3. relocation intentions (current household registration status and relocation plans).

In addition to interviewing factory managers and retired government officials, this study also included: agency and company representatives visiting the HQs of architectural firms; and representatives of real estate associations. A total of 12 people were interviewed about the migration of workers and the status of migrant housing.

The survey research was mainly completed in 2007. Fixed sampling and convenience sampling were adopted for data collection. The first stage of sampling involved snowball sampling to identify five local Taiwanese factories. In the second stage, questionnaires were distributed according to the number of participants the managers of these factories could spare and the ratio of the factory's workers living in non-factory accommodations versus

---

<sup>1</sup> See Chang (2009) for further details regarding qualitative analysis of these interviews.

those staying in the dormitory. The managers at each factory collected 60-100 questionnaires, and 383 surveys were collected in total. Finally, after removing relocated households, locally registered households, those with income deviations, and invalid surveys, the final sample consisted of 267 data-points. To analyze this data, we constructed a binary logistic regression model to predict intention status. Stepwise variable selection was used to select the best-performing model, whereas cross-validation and bootstrapping techniques were used to ensure reproducibility and validity of the estimates. Analysis was conducted using R Studio.

Table 2 shows the official census data for Dongguan from 1990 and 2005. These data are the closest and most readily available for comparison to Li's (1997) survey data in Dongguan in 1992, as well as to our survey data in 2007 (official data regarding permanent migrants are not available). Both the data from Li's fieldwork in Dongguan and our data showed that hukou migrants (permanent migrants) represented a small number of cases (Table 3). According to the census data, the largest increase was among the temporary migrants, and this is evident in the data changes in the two surveys. These two datasets indicate that even with the strong influx of migrants, Dongguan residents who became permanent migrants were not common in either of the survey samples.

<b>Household registration status</b>	1990	2005
<b>Nonlocal</b>	25.20%	78%
<b>Local</b>	73.80%	21%
<b>Total population (millions)</b>	1.7 million	7.5 million

**Table 2.** Population Census Data of Dongguan in 1990 and 2005. Source: Li (1997).

<b>Household registration status</b>	<b>Identity</b>	<b>Li's survey in Dongguan, 1992</b>	<b>Chang's survey in Dongguan, 2007</b>
Nonlocal	Temporary migrants	18%	78%
Local	Hukou migrants (permanent migrants)	7%	6%
	Natives	75%	16%
Total number of respondents		2,191	383

**Table 3.** Si-ming Li's Field Survey Data and the Data from This Survey. Source: Chang's field survey (2009).

## Variables and Measurements

### Measuring the Intentions of Migrant Workers

Respondents who wanted to change their place of hukou registration to Dongguan were asked to provide their motivation via a multiple-choice question: 1) to avoid excessive paperwork, 2) for their children's education, 3) to seek better job opportunities, 4) to access

local welfare, or 5) other. These options were based on the interview data, and the respondents who answered <other> were asked to write down the reason.

### **Measurement of Migrants' Characteristics**

There has been little research on the migration intentions of migrant workers. Neither studies by Cai and Wang (2008) or Tang and Feng (2012) included child(ren) or household type variables, but the variable of settlement intention is relatively informative and can be used in measurements. Gender and marital status capture important latent information regarding society expectations. Traditionally, females of the age of consent are expected to return to their hometown for marriage (Fan, 2003). Previous studies have shown that the majority of migrant workers in Dongguan are young women (Chan, 2010; Roberts, 1997). However, in the two studies mentioned above, the gender variable showed differences in hukou migration between men and women. Cai and Wang's study (2008) in the PRD found that men were more likely to change their place of hukou registration, but Tang and Feng's study (2012) found this to be the case for women instead. Married respondents are expected to behave differently than single respondents because of different family responsibilities. Yang (2000) and Roberts (2002) found that married migrants are more likely to migrate with their family. Moreover, married migrants are more likely than single migrants to be return to their hometown (Wang & Fan, 2006; Zhao, 2002). When migrating with children, factors such as permission for children to attend local public schools and an increase in family size directly affect migrants' financial situation and migration decisions (Montgomery, 2012; Wang & Fan, 2006; Zhao, 2002; Zhu, 2001). Married people also showed stronger intention to change their hukou than unmarried people in the Jiangsu household migration intention study (Tang & Feng, 2012).

Age is a very informative measure when it comes to people's intention to migrate or return to their hometown. Studies show that older migrants are more likely to return (Zhao, 2002), whereas younger people intend to migrate (Yang, 2000). In Tang and Feng's study (2012), younger migrant workers were found to be more likely to change their hukou registration.

The majority of migrant workers in China are from agricultural households. They are farmers from villages who have no or only some skills in non-agricultural activities. The agricultural household category is used in comparison with non-agricultural or urban households, the other hukou category in the hukou system. Migrants in non-agricultural households tend to have more years of education and work experience than those in agricultural households (Fan, 2002). Knight et al. (1999), Huang (2001), Guo and Iredale (2004), and Liang (2004) found that migrants in agricultural hukou categories were less likely to acquire specific occupations and that migrants in urban households found greater opportunities in non-agricultural jobs, such as administrative, professional, and technical positions. However, the hukou type was not used by the Cai and Wang (2008) in the PRD study or by Tang and Feng (2012) in the Jiangsu study.

Education was measured by the respondents' years of education. Previous studies have shown that more years in education have a positive influence on both return migrants and continuing migrants (Wang & Fan, 2006; Zhao, 2002). This result seems to confirm that those with longer time in education, not only have an advantage in the job market, but also benefit from the household migration policy. However, in the Jiangsu study by Tang and Feng (2012), respondents with higher education levels showed higher hukou migration intention; yet there was no statistically significant correlation between education and hukou migration intention in the PRD study by Cai and Wang (2008).

Occupations are divided into skilled versus unskilled jobs. Migrant workers in administrative and managerial positions and in jobs requiring specific techniques and certificates are categorized as skilled workers. Several studies have found that administrative and professional occupations are positively correlated with higher incomes (Bian & Logan, 1996; Cao & Nee, 2005; Meng & Zhang, 2001). Occupational classification is also significant in Tang and Feng's Jiangsu study (2012). This work found that professional and technical personnel were less likely to consider hukou migration than those engaged in sales and services.

Income was measured by monthly household income, and the respondents were asked to choose from a range. The ordinal measure was based on the tax-free personal allowance in 2007, and the responses to the preliminary questionnaire and interviews suggested that such use could have avoided inaccurate or missing values due to the sensitive nature of the topic. Concern about net income is one of the fundamental determinants of migrants' destination decisions. Income was significant in the PRD study (Cai & Wang, 2008), which asked for the average monthly income. The Jiangsu study divided income variables into intervals and did not show significant results (Tang & Feng, 2012).

## Results

The section presents the findings from both statistical analysis and the semi-structured interviews about their views on hukou migration.

Table 4 shows that our sample consisted of more women than men and that most participants were under 30 years old. Almost 60% were married, but only 61, accounting for 23% of the total, actually lived in Dongguan with their child(ren). Most respondents had less than 12 years of education, and most of the respondents were non-professional or technical workers. The households were concentrated in the low- and middle-income groups. 51 respondents (19%) expressed willingness to relocate hukou.

	Categories	Frequency	Percentage
Demographic controls			
Gender	Female	147	0.55
	Male	120	0.45
Age	≤ 25	102	0.38
	26–30	88	0.33
	31	77	0.29
Marital status	Married	158	0.59
	Other	109	0.41
Housing status	Living with children	61	0.23
	Not living with children	206	0.77
Hukou type	Nonagricultural hukou	100	0.37
	Agricultural hukou	167	0.63
Education (levels)	0	1	0.004

	6	10	0.037
	9	43	0.161
	12	145	0.543
	18	67	0.251
	20	1	0.004
Recode Education	Under 10 years	54	0.202
	Above 10 years	213	0.798
Skilled worker	Skilled	99	0.37
	Nonskilled	168	0.63
Family income (yuan)	< 3601	127	0.48
	3601–26600	125	0.47
	>26600	15	0.06
Recode Income	Low	127	0.476
	Mid	125	0.468
	High	15	0.056
Intention to change hukou registration location	Yes	51	0.19
	No	216	0.81

**Table 4.** List of Profiles

### **Determinants of Migrant Workers' Local Hukou Intentions**

This section presents the results of binary logistic regression, where hukou migration intention was modelled as a function of migrants' socio-demographic factors. Variable selection was performed using stepwise procedures (Table 5), which allowed us to select the best performing model. Backward-stepwise selection starts with all the predictors and sequentially removes them until the best fit is found (Hastie et al., 2009: 58). The metric involved in this process is the Akaike information criterion (AIC): Table 5 shows the difference in AIC score between the best model and the full model. This part also discusses the findings from the multiple-choice questionnaire and the semi-structured interviews.

	<b>Model</b>	<b>AIC</b>
<b>Full (Model 1)</b>	intention = gender + age + married + Nonagri + withchild + recode_edu + skilled + recode_income	166.29
<b>Model 2</b>	intention = gender + age + married + Nonagri + withchild + recode_edu + recode_income	164.44
<b>Model 3</b>	intention = age + married + Nonagri + withchild + recode_edu + recode_income	162.58
<b>Model 4</b>	intention = age + Nonagri + withchild + recode_edu + recode_income	160.85
<b>Model 5</b>	intention = age + withchild + recode_edu + recode_income	159.21
<b>Reduced (Model 6)</b>	intention = withchild + recode_edu + recode_income	158.51

**Table 5.** Forward Stepwise Selection Regression. The tables show the six iterations performed by the algorithm: model 6 is the best performing model, with the lowest AIC score.

### **Binary Logistic Regression Models on the Determinants of Migrant Workers' Local Hukou Intentions**

The results of the binary logistic regression on the determinants of migrant workers' local hukou intentions are summarized in Table 6. The assumptions for this classifier have been fulfilled: Assumption 1 (dichotomous response variable), Assumption 2 (continuous or categorical explanatory variables) and Assumption 3 (independence of observations), and Assumption 4 (no multicollinearity, Table 7). Assumption 5 (linearity assumption) is not applicable to our study because no continuous predictors were used.

		<b>Reduced Model</b>	
	<i>Variable</i>	<i>Coeff (S.E)</i>	<i>OR (95% CI)</i>
(Intercept)		-19.5701 (993.5538)	0.00
Migration condition	Living with children	1.1463 (0.4278) **	3.15 (1.36, 7.28)
Cultural/Human capital	Education Level: above 10	17.2345 (93.5537)	30538588.49 (0, Inf)
Economic capital	Household income level: mid	1.2885 (0.4634) **	3.63 (1.46, 9)
	Household income level: high	1.2118 (0.7578)	3.36 (0.76, 14.84)
Null deviance	183.63 on 187 degrees of freedom		
Residual deviance	148.51 on 183 degrees of freedom		
AIC	158.51		
Number of Fisher scoring iterations:	17		

**Table 6.** Best Performing Binary Logistic Regression Model on the Determinants of Migrant Workers' Local Hukou Intentions.

\* Significant at  $p < 0.05$ ; \*\*  $p < 0.001$

Residual deviance is lower than null model.

ORs measure the strength of association between response and predictors: OR >1 means greater odds of association; OR =1 means no association; OR <1 means lower odds of association.

	GVIF	Df	GVIF^(1/(2*Df))
withchild	1.019965	1	1.009933
recode_edu	1.000000	1	1.000000
recode_income	1.019965	2	1.004954

**Table 7.** Diagnostics: Checking for Multicollinearity via Variance Inflation Factor (VIF). Assumption of multicollinearity has not been violated.

In the first model (full model, Table 5), the initiated stepwise selection contained the basic demographic characteristics of gender, age, marital status, education and income levels, skilled worker status and household registration type as predictors to identify the migrants' hukou migration intentions: the AIC score for this model was 166.2. The best performing model found after 6 iterations can be specified as:

$$\text{logit}[P(\text{intention}=1)] = a + b1 \text{ housing status} + b2 \text{ education} + \text{INCOME } T\beta$$

where intention is the categorical binary response variable (with  $j = 0$ , no intention of acquiring local citizenship and  $j = 1$ , intending to acquire local citizenship);  $a$ ,  $b$  are scalar coefficients; **housing status** and **education** are binary predictors;  $\beta$  is a vector of coefficients; **INCOME** is a vector of multiclass predictor.

The AIC score for this (reduced) model was 158.51, a considerable reduction from the full model. Information on the agricultural background of individuals (skilled/unskilled variable) was not selected by the stepwise selection, but this has proved extremely important in terms of theoretical models on migration in China and larger scale studies across multiple industries (e.g., Guo & Iredale, 2004; Huang, 2001; Knight et al., 1999; Liang, 2004; Tang & Feng, 2012). However, the selection of the final model in this work (and the exclusion of agricultural status as a meaningful predictor) is due to our small sample size and its focus on manufacturing.

Table 6 shows how coefficients quantify the impact of changes in various predictors on the probability of migrant workers' intention to obtain local citizenship (hukou migration intent) in terms of both odd ratios and log odds. The results show that the odds of obtaining local citizenship for migrant workers are, respectively, 3.15 for individuals living with their child(ren), 3.63 for mid-income respondents and 3.36 for high-income respondents compared to the baseline. Living with children and being at a mid-income level both record statistical significance: this result is consistent with previous arguments that affordable education for their children is a key factor in migrants' migration and return intentions (Wang & Fan, 2006; Zhao, 2002; Zhu, 2001). Education level, despite being selected by the stepwise algorithm, registered a large standard error, suggesting its interpretation may not truly represent the actual population. We decided to not remove this term in our model because it makes sense at theoretical level (Li, 2010) and because of the overall fit.

The estimate for income is consistent with the results of the Cai & Wang study (2008) in that high-income earners show a relatively high willingness to migrate. In terms of education, Tang and Feng's (2012) revealed the same result: higher education was related to a relatively high hukou migration intention. The statistically significant result of the variable "Whether living with child(ren) in Dongguan" supports the emphasis on the importance of the family in Chinese migration studies and validates the obstacles migrants' children face in accessing education in places where their hukou is not registered, which is one of the main problems migrant workers face (Montgomery, 2012; Wang & Fan, 2006; Zhao, 2002; Zhu, 2001).

Basic demographic variables, such as gender, age, and household type, carried no predictive power, even though they were statistically significant in both Pearl River Delta and Jiangsu studies. The effect of age on household migration was not important, which was consistent with findings from the Pearl River Delta study. When it comes to occupation, we included only whether respondents had certain skills: this variable was not selected by stepwise procedure. This differed from the Jiangsu study of Tang and Feng (2012), which found that professional and technical people were less likely to consider hukou migration. However, it is important to stress that, unlike previous work, we focused on predictive performance. Despite our efforts, the resulting reduced model might have been affected by homogeneity in the sample. We sampled employees in five factories, all in the manufacturing industry, and most of the factories had low skill thresholds, had relatively few highly skilled or senior executives and employed mostly foreigners or Taiwanese, who were not included in the sample. However, we encourage a replication of our approach (using stepwise selection with the same full model specification) in a more heterogeneous sample.

### **Evaluation of the Best Performing Model**

In this work, two approaches have been implemented to ensure testing accuracy of the selected models: cross-validation and bootstrapping.

Cross-validation is a class of methods that can be used to evaluate the ability of statistical models to generalize for out-of-sample data. As a method of model evaluation, cross-validation is more robust than residual analysis; this is because, unlike residual evaluation, cross-validation provides an indication of how a classifier performs when making predictions on unseen data. The dataset is divided such that some data points are used to train the classifier (in the case of this work, binary logistic regression) and some are used to test its classification accuracy. This process avoids overfitting, which is a modelling error where the model effectively memorizes the data, picking up random fluctuations of noise in the data. For the purpose of this work, the sample has been randomly partitioned into a train set (70% of the sample) and a test set (30% of the sample), in line with current practices (Liu & Cocea, 2017).

We used bootstrapping as an extra measure to estimate model performance. Classification accuracy was chosen as metrics in both cross-validation and the bootstrapping technique. Bootstrapping is a resampling method with replacement that creates several simulated datasets from a single dataset (Efron & Tibshirani, 1994). Modelling is performed on each of these simulated samples, and predictions are done by averaging results in regression problems or calculating mode values in classification (Wilmott, 2020). Unlike cross-validation, the model is tested on the entire dataset: we employed 1000 random extractions to test our model. Table 8 shows the estimates expressed in terms of original coefficients and odds ratios for the three predictors of interest: except for education level, standard errors are moderate, and bias for household status and income level is low, suggesting that these values are likely to represent true population values.

		<b>Reduced Model</b>			
<i>Variable</i>		<i>Original Coeff (S.E.)</i>		<i>Bias</i>	<i>OR</i>
(Intercept)		-3.131	(1.732)	-0.298	0.044
Migration condition	Living with children	1.006	(0.395)	0.007	2.735
Cultural/Human capital	Education level: above 10	0.971	(1.728)	0.254	2.640
Economic capital	Household income level: mid	0.845	(0.380)	0.015	2.328
	Household income level: high	1.672	(0.771)	0.004	5.322

**Table 8.** Bootstrapping (n = 1000) Results on Reduced Model

Predictive accuracy in cross validation is 77% and 80% for bootstrapping. Given the small sample size, this mild increase in accuracy during resampling is to be expected.

### **Migrant Workers' Views and Motivations Regarding Hukou Migration**

The survey identified 51 respondents who were willing to obtain local citizenship: descriptive data (Table 9) suggests that impact on children's access to schools outweighs the other reasons for acquiring hukou. Although affordable education falls under overall social welfare, the category of social welfare is secondary to children's access to education. The third factor is the amount of paperwork that migrant workers need to complete to return to the place where their hukou is registered (e.g., no criminal record, notarization of driver's license, registration of unemployment, etc.), followed by finding employment. The results are consistent Pearl River Delta studies (Cai & Wang, 2008), where respondents also cited "yearning for city life," "wanting children to stay in the city," and "feeling that having a city hukou would eliminate a lot of hassle."

<b>Motivations</b>	<b>Children's education</b>	<b>Better welfare</b>	<b>Paperwork burden</b>	<b>Employment opportunities</b>	<b>Other</b>
N	34	24	16	12	1
Percentages	39.10%	27.60%	18.40%	13.80%	1.10%

**Table 9.** Motivation of 51 Respondents Willing to Change Their Hukou (Multichoice Question Results)

Interviews with plant managers and representatives of local government agencies or businesses found that their knowledge about the hukou migration of migrant workers was rather limited. Much of the information about the willingness of migrants to change their hukou comes from direct access to migrant workers. Although the interviews are few, these and survey results show consistency in that the respondents' decisions around hukou migration

are driven by two factors: children's education and buying a property. The respondents who considered hukou migration because of their children's education pointed out that their decision was driven by the fact that their children could not attend local public schools without a local hukou and that private school fees were unaffordable. Those respondents who had already bought a home (or were considering buying a home) stated that their reason for the hukou change was the policy required them to purchase a commercial house in order to qualify for a local hukou. There were also respondents with children close to school age who were considering buying a house.

Questionnaires did not collect information on reasons for not wishing to settle in the city, but the interviews were consistent with previous studies in this regard that urban experience and a sense of belonging matter (Cai & Wang, 2008; Fan, 2011; Wang & Fan, 2012; Wu & Treiman, 2004). One interviewee explicitly mentioned that he had been a teacher in his hometown; the job paid less but was assigned as a lifetime job by the state, while his current job in the Dongguan factory was unstable, and he could be fired from it at any time.

*We do not consider this kind of job permanent... they can fire you anytime without reason.... Therefore, it is a temporary job to us.*

*Male, 33, Purchasing staff*

Some respondents mentioned that the early scrutiny of temporary residence permits made them feel that locals were not friendly towards migrant workers, which led them to define their residence and work in the city as temporary (e.g. long enough to earn sufficient money to return home). Some migrant workers interviewed also mentioned that even though they were eligible to obtain a hukou, they would be hesitant to do so because they might not be accepted by locals. The newly added local hukou implies that the benefits (e.g. shares and dividends of local enterprises) that were originally given to local people might be shared by everybody (Chan, 2010; Davies & Ramia, 2008; Deng & Smyth, 2000; Smart & Smart, 2001).

We also found that so-called hukou migration does not necessarily mean family relocation, as the term 'hukou' is clearly designed around the concept of individuals. After buying a house to obtain local hukou eligibility, it is necessary to pay additional fees based on the number of people covered by the hukou. However, one's economic situation may not allow for the whole family to obtain a new hukou as a unit. Some respondents indicated that if the whole family could not be together, they would not change their hukou. However, some respondents said that if only one person in the family had access to a local hukou, they would give priority to their children to allow them to attend local public schools. This shows that hukou, in fact, represents an individual's legal status and is used in exchange for other capital (Bauder, 2008).

## **Discussion**

In the third decade of China's reform and opening up, the rise of the inland regions' economy and two labor shortages have led to many different hukou policy reforms (e.g. housing, special talent, investment, the points system) to qualify for hukou. However, the total number of household changes suggests that hukou migration policies in the 2000s still had limits, and most were targeted for skilled workers. Since 2020, China has urged cities with a resident population of less than 3 million to lift the restrictions on hukou migration and to try to register people's hukou in their permanent place of residence rather than having them inherit

the hukou registration location of their parents. Although the details and results of shifting hukou registration to the permanent place of residence are not yet clear, it is important to study the willingness of migrant workers to change their hukou during the early period of strict hukou policies. As unskilled workers who were not favored by policies at the time still wanted to change their hukou, the reasons for this are worth documenting so to inform future policies.

The analysis of migrant workers in the manufacturing industry emerging in this paper is consistent with most of the results of previous research. The use of machine learning techniques allows us to estimate the best sampling possibilities under a limited sample size and to find the most effective way to model citizenship intention. We also attempted to provide an analytical strategy (e.g. stepwise selection, cross-validation, bootstrapping) that other researchers interested in this area could use and expand on in future research.

The number of interviews is small, and it is debatable whether the relevant phenomena found in this study can be extended to all social groups. But regardless, its content, in tandem with quantitative analysis, proves that children's education and buying a house impact on the intention of acquiring hukou. Migrant workers realized that obtaining citizenship can represent a further step into integration. Although obtaining citizenship together with family members may be the common goal, interview transcripts<sup>2</sup> also show that hukou is a personal experience that approaches a legal status; furthermore, migrant workers mainly use citizenship in exchange for the education of their children. Interviews also revealed variables that might be added to future research designs, such as:

- perception of migrant workers (welcome/not welcome);
- forecasts of the local economic outlook (for their job opportunities);
- non-agricultural work experience;
- predictions of their hometown's development prospects.

These can help us better understand the willingness of migrant workers to move their hukou or place of residence.

On the policy side, future hukou policies should consider:

- fairness of the welfare of new migrants;
- convenience of documentation procedures related to hukou migration;
- a vision statement for local development (especially education and employment opportunities).

In terms of policy advocacy, external policies should target relatively underdeveloped areas, whereas internal policies should focus on people who already have some local experience, are married and have children. The promotion of hukou migration should be accompanied by the promotion of housing (tax-free or low-interest loans), child education (qualified to local public compulsory education), and individual job opportunities (priority employment program for potential migrants). In addition, concerns exist among locals and non-locals that the influx of new migrants will dilute the dividends that locals would otherwise receive in local joint-stock enterprises. The relevant central and local authorities should address these concerns by explaining the tax and welfare policies for both locals and new migrants.

---

<sup>2</sup> Available upon request.

### **Acknowledgements**

We thank Professor Chris Pickvance (SSPSSR, University of Kent, U.K.) for his valuable suggestions on the initial version of this manuscript.

### **Funding**

This work was supported by Guangdong University of Petrochemical Technology, China (Grant No.702/519238).

### **Declaration of conflicting interests**

The authors have no conflicts of interest to declare

### **Data Availability Statement**

The data, R code and records of interviews that support the findings of this study are available from the corresponding author, Feng-shuo Chang, upon reasonable request.

## **References**

- Bauder, H. (2008) Citizenship as capital: The distinction of migrant labor. *Alternatives: Global, Local, Political* 33(3): 315–333.
- Bian, Y. and Logan, J.R. (1996) Market transition and the persistence of power: The changing stratification system in Urban China. *American Sociological Review* 61(5): 739-758.
- Cai, H. and Wang, J. (2008) Factors influencing the migration intentions of rural workers in the Pearl River Delta. *Social Sciences in China* 29(3): 157-171.
- Cao, Y. and Nee, V. (2005) Remaking inequality: Institutional change and income stratification in Urban China. *Journal of the Asia Pacific Economy* 10(4): 463-485.
- Chan, K.W. (2010) A China paradox: Migrant labor shortage amidst rural labor supply abundance. *Eurasian Geography and Economics* 51(4): 513-530.
- Chang, F.S. (2009). *Housing Inequality in Dongguan City, China: Winners and Losers in Transition* (Unpublished doctoral dissertation). University of Kent, Canterbury, United Kingdom.
- Chang, L. and Bu, Q. (2020) Review on the compulsory education status of migrant workers' children in Chinese cities. *Science Insights Education Frontiers* 7(2): 861-877.
- Cui C, Geertman S and Hooimeijer P (2015) Residential mobility of skilled migrants in Nanjing, China. *Environment and Planning A: Economy and Space* 47(3): 625-642.
- Cui H and Cho J (2020) Does the revised hukou system facilitate or restrain the short-term labor inflows into Chinese cities? *Sustainability* 12(4): 1295.
- Cui, R. and Cohen, J.H. (2015) Reform and the hukou system in China. *Migration Letters* 12(3): 327-335.
- Davies, G. and Ramia, G. (2008) Governance reform towards “Serving migrant workers”: The local implementation of central government regulations. *The China Quarterly* 193: 140-149.
- Deng, X. and Smyth, R. (2000) Non-tax levies in China: Sources, problems and suggestions for reform. *Development Policy Review* 18(4): 391-411.
- Efron, B. and Tibshirani, R.J. (1994) *An Introduction to the Bootstrap*. New York, NY: CRC Press.

- Fan, C.C. (2002) The elite, the natives, and the outsiders: Migration and labor market segmentation in Urban China. *Annals of the Association of American Geographers* 92(1): 103-124.
- Fan, C.C. (2003) Rural-urban migration and gender division of labor in transitional China. *International Journal of Urban and Regional Research* 27(1): 24-47.
- Fan, C.C. (2011) Settlement intention and split households: Findings from a survey of migrants in Beijing's Urban Villages. *China Review* 11(2): 11-41.
- Guo, F. and Iredale, R. (2004) The impact of hukou status on migrants' employment: Findings from the 1997 Beijing migrant census. *International Migration Review* 38(2): 709-731.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.
- Huang, Q. and Zhang, G. (2013) The intention of resettling in cities (towns) of China's new-generation of migrant workers and individual-level determinants. In: Wang H (ed) *Proceedings of the 2nd International Conference on Science and Social Research*. Beijing, China: Atlantis Press, pp.534-540.
- Huang, Y. (2001) Gender, hukou, and the occupational attainment of female migrants in China (1985-1990). *Environment and Planning A: Economy and Space* 33(2): 257-279.
- Knight, J., Deng, Q. and Li, S. (2011) The puzzle of migrant labour shortage and rural labour surplus in China. *China Economic Review* 22(4): 585-600.
- Knight, J., Song, L. and Huaibin, J. (1999) Chinese rural migrants in urban enterprises: Three perspectives. *Journal of Development Studies* 35(3): 73-104.
- Lai, H. (2007) Developing Central China: A new regional programme. *China: An International Journal* 5(1): 109-128.
- Li, S.M. (1997) Population migration, regional economic growth and income determination: A comparative study of Dongguan and Meizhou, China. *Urban Studies* 34(7): 999-1026.
- Li, S. (2010) The economic situation of rural migrant workers in China. *China Perspectives*. DOI: 10.4000/chinaperspectives.5332.
- Liang, Z. (2004) Patterns of migration and occupational attainment in contemporary China: 1985-1990. *Development and Society* 33(2): 251-274.
- Liu, H. and Cocca, M. (2017) Semi-random partitioning of data into training and test sets in granular computing context. *Granular Computing* 2(4): 357-386.
- Lu, Y. (2008) Does hukou still matter? The household registration system and its impact on social stratification and mobility in China. *Social Sciences in China* 29(2): 56-75.
- Meng, X. and Zhang, J. (2001) The two-tier labor market in Urban China: Occupational segregation and wage differentials between urban residents and rural migrants in Shanghai. *Journal of Comparative Economics* 29(3): 485-504.
- Montgomery, J.L. (2012) The inheritance of inequality: Hukou and related barriers to compulsory education for China's migrant children. *Pacific Rim Law and Policy Journal* 21(3): 591-622.
- Roberts, K. (2002) Female labor migrants to Shanghai: Temporary "floaters" or potential settlers? *International Migration Review* 36(2): 492-519.
- Roberts, K.D. (1997) China's "tidal wave" of migrant labor: What can we learn from Mexican undocumented migration to the United States? *International Migration Review* 31(2): 249-293.

- Smart, A. and Lin, G.C.S. (2007) Local capitalisms, local citizenship and translocality: Rescaling from below in the Pearl River Delta Region, China. *International Journal of Urban and Regional Research* 31(2): 280-302.
- Smart, A. and Smart, J. (2001) Local citizenship: Welfare reform urban/rural status, and exclusion in China. *Environment and Planning A: Economy and Space* 33(10): 1853-1869.
- Solinger, D.J. (1999) Citizenship issues in China's internal migration: Comparisons with Germany and Japan. *Political Science Quarterly* 114(3): 455-478.
- Statistics Bureau of Dongguan (2010a) Statistics reports. Available online: <http://tjj.dg.gov.cn/website/web2/main.jsp?columnId=112&parentcolumnId=114> (accessed 21 October 2010).
- Statistics Bureau of Dongguan (2010b) Statistics reports. Available online: <http://history.dg.gov.cn/DgNet/book/2001/tables/t170.htm> (accessed 21 October 2010).
- Statistics Bureau of Shanghai (2010) Statistics reports. Available online: <http://sh.gov.cn/shanghai/node2314/node24651/node26195/node26200/u21ai512618.html> (accessed 21 October 2010).
- Tang, S. and Feng, J. (2012) Understanding the settlement intentions of the floating population in the cities of Jiangsu Province, China. *Asian and Pacific Migration Journal* 21(4): 509532.
- Wang, W.W. and Fan, C.C. (2006) Success or failure: Selectivity and reasons of return migration in Sichuan and Anhui, China. *Environment and Planning A: Economy and Space* 38(5): 939-958.
- Wang, W.W. and 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.
- Wilmott, P. (2019) *Machine Learning: An Applied Mathematics Introduction*. Panda Ohana Publishing.
- Wu, L. (2013) Decentralization and hukou reforms in China. *Policy and Society* 32(1): 33-42.
- Wu, X. and Treiman, D.J. (2004) The household registration system and social stratification in China: 1955-1996. *Demography* 41(2): 363-384.
- Yang, X. (2000) Determinants of migration intentions in Hubei Province, China: Individual versus family migration. *Environment and Planning A: Economy and Space* 32(5): 769-787.
- Yue, Z., Li, S., Feldman, M.W., et al (2010) Floating choices: A generational perspective on intentions of rural-urban migrants in China. *Environment and Planning A: Economy and Space* 42(3): 545-562.
- Zhang, Z. and Treiman, D.J. (2011) Social origins, hukou conversion, and the wellbeing of urban residents in contemporary China. *Social Science Research* 42(1): 71-89.
- Zhao, Y. (2002) Causes and consequences of return migration: Recent evidence from China. *Journal of Comparative Economics* 30(2): 376-394.
- Zhu, M. (2001) The education problems of migrant children in Shanghai. *Child Welfare* 80(5): 563-569.
- Zhu, Y. (2007) China's floating population and their settlement intention in the cities: Beyond the hukou reform. *Habitat International* 31(1): 65-76.