

The Effect of the Hukou System on Mobility Choice
and Housing Expenditure in China

A thesis submitted by

Ziqi Zhou

In partial fulfillment of requirements for the degree of

Master in Science

in

Economics

Tufts University

May 2023

© 2023, Ziqi Zhou

Advisor: Yannis M. Ioannides

Abstract

This thesis explores the effect of the hukou system on individuals' residential mobility behaviors and housing expenditure, and specifically the influences of hukou type and hukou registered location. The results suggest that agricultural hukou holders are more likely to move across provinces with a transfer of hukou registered location, but less likely to move within provinces. Inter-province migrants may face varying frictions in different origin-destination combinations, and this effect may differ by the hukou type. The following tests verify that the influences of the agricultural hukou type is statistically different on moving choices. A two-stage Heckman selection model, supports a potential deterring effect of agricultural hukou in inter-province moving choices with a transfer of hukou registered location. Furthermore, this study reveals a significant negative selection bias on household's housing expenditure. My empirical investigations are based on data from the Chinese Family Panel Study.

Acknowledgement

Without the contributions and support of many people, I would not be able to complete my thesis, and for that, I am sincerely grateful. I want to extend my deepest thanks to Professor Ioannides for sharing his vast expertise and unwavering commitment to academic rigor in economic research. I could not have accomplished this without his guidance and support. I would also like to express my gratitude to Dr. Hardman for her meticulous review and attention to detail, which helped to improve the quality of my work.

I would like to extend a special thanks to Qi Peng for his invaluable insights and discussions on my thesis. I am also grateful to my dear friends Chuancheng Li, Tiantian Ji, Ziwei Wang, Hanzhen Wang, Sudarshan Bashyal, and Kenneth Koenen for their support and encouragement throughout the process.

Finally, I want to express my heartfelt appreciation to my parents and my large family for their unconditional love and support. Their encouragement and belief in me have been a constant source of motivation and inspiration throughout my academic journey.

Table of Contents

1. Introduction	1
2. Data.....	9
3. Econometric Model	16
3.1 Binary Probit Models	16
3.2 Multinomial Logit Model.....	17
3.3 Intention of Moving Hukou Registered Location	18
3.4 Hukou Type Changes	20
3.5 Housing Expenditure.....	20
4. Empirical Results.....	21
4.1 Binary Probit Models	21
4.2 Multinomial Logit Model.....	22
4.3 Intention of Moving Hukou Registered Location	26
4.4 Hukou Type Changes	27
4.5 Housing Expenditure.....	28
5. Hypothesis Tests and Robustness Tests	30
5.1 Effect Equality in Probit Models.....	30
5.2 Equality of Effect in Multinomial Logit Model.....	31
5.3 Placebo Test in Multinomial Logit Model.....	32
6. Conclusion.....	33
Reference	35
Appendix.....	38

List of Appendices

Appendix I. Tables of Survey Information	38
Appendix II. Supplementary Explanation on Construction of Some Variables	43
Appendix III. Variable Definition.....	44
Appendix IV. Variables Availabilities by Year	46
Appendix V. Value Explanation of Categorical Variables	47
Appendix VI. Estimates for Probit Model of Binary Moving Choice	49
Appendix VII. Estimates for Binomial Moving Choice Models with Interaction Term	53
Appendix VIII. Heckman Selection Estimates for Moving Intention	60
Appendix IX. Hypothesis Tests Conducted in Section 5.....	61

List of Tables

Table 1 – Channels of Agricultural to Urban Hukou Conversion

Table 2 – Summary Statistics of Key Variables in Finalized Data

Table 3 – Summary of Hukou Registered Location Information

Table 4 – Urban Hukou by Urban Residents

Table 5 – Community Change by Province Change

Table 6 – Hukou Change by Province Change

Table 7 – Ratio of Different Types of Moving by Survey Wave

Table 8 – Estimates Results for Logit Models

Table 9 – Estimated Relative Risk Ratio Results for Logit Models

Table 10 – Estimates Results for Logit Models with Current Income

Table 11 – Estimated Relative Risk Ratio Results for Hukou Type Changes

Table 12 – Housing Expenditure Estimations

Table 13 – Z-test Results of Lagged Agricultural Hukou Indicator

Table 14 – Placebo Test Results IN Multinomial Logit Model

1. Introduction

Hukou—translated as “household registration system”—is a unique component of the Chinese institutional system that records household information and ensures the social benefit of the family. The system records people as permanent residents of an area and collects information including date of birth, place of birth, spouse, children, moving records, etc. Furthermore, it establishes the eligibility for the social resources (Lu, 2008) at one’s hukou location. These resources include health services, eligibilities for housing purchases, education for children, etc. At the same time, especially in large populated regions, the local governments impose restrictions regarding the transfer of the hukou into their prefectures.

Researchers and policy makers have long been discussing reform in this system and its profound influence in Chinese society. This thesis empirically examines the role of the hukou system, specifically the aspects of hukou type and hukou registered location, in the residential mobility behavior and housing expenditure. Since the hukou registered location could be different from the current residence location, both residence location and hukou registered location can be studied separately. This thesis studies the role of hukou system in moving choices and housing expenditures and relates mobility with households’ housing expenditures choices.

The hukou system has been embedded in the institutional system of China since the Shang Dynasty (1600 to 1046 B.C.). Past literature has summarized the development of the system and its major changes (Chan and Zhang, 1999; Lu, 2008; Chan, 2013; Cai, 2022, Tang and Wu, 2022). The modern hukou system was established in 1951 when the Ministry of Public Security promulgated a regulation called *Interim Provisions for the Administration of Urban Residence* in urban regions in China. In 1954, most rural regions had established the hukou system based on the

First National Population Census in 1953. In this stage, the hukou system mainly functioned to record information and did not limit migration within the country.

From 1958 to 1979, a strict dichotomous classification in the hukou system was established, which divides hukou into agricultural and non-agricultural (urban). In 1958, *The People's Republic of China Hukou Registration Regulation* was implemented, which clarified seven ways of converting the hukou type from agricultural to urban hukou. They are listed in the following table together with the channels of converting agricultural to urban hukou.

Table 1. Channels of Agricultural to Urban Hukou Conversion

(1) recruited as permanent employees by a state-owned enterprise;
(2) displaced due to state-initiated land expropriation;
(3) recruited for enrollment in an institution of higher education;
(4) promoted to administrative positions;
(5) relocated because of family crises (such as moving to a city to live with and look after a sick parent);
(6) joining the army and demobilized to cities;
(7) deemed to belong to special categories (either recipients of compensation for past policy mistakes or people who had endured personal sacrifices and hardships because of their work for the state).

Source: Chan, Kam Wing. "The Chinese Hukou System at 50." *Eurasian Geography and Economics*, vol. 50, no. 2, 2009, pp. 197–221

People who committed certain crimes could be stripped of their non-agricultural hukou status (Chan, 2009). At this stage, restrictions on migration were started and have imposed long-lasting influences on Chinese society till now. In the literature, Wang *et al.* (2021) focus on a subgroup of urban hukou holders, who migrated to the city and converted from agricultural to urban hukou holders. These migrants have another name: Non-Zhuan-Fei (NZF). Wang *et al.* research consumption habits and find that there is no significant difference between the total expenditures of NZF and urban households, but NZF households put more weight on conspicuous consumption than practical consumption.

The hukou system at this stage played a very important role in the rationing of resources in the planned economy. Only urban hukou holders were eligible to work for non-agricultural employers,

and were assigned rations of grains and food. The agricultural hukou holders worked in agricultural activities and were expected to supply their own food.

The third stage of the hukou system involves the relaxation of the dichotomous restrictions from 1979 to 2001. These economic reforms (also known as “Reform and Open Up”) that started in 1978 allowed for the de-collectivization of agriculture activities, the opening up of foreign investment, and the starting of businesses by entrepreneurs. The regulations and policies increased the size of free agricultural labor in rural areas and increased the labor supply in non-agricultural sectors. In 1984, a transitional type of hukou called “hukou with self-supplied food grain” appeared for a short time period. It allowed migrants from rural regions to move to small towns. In 1985, the government began to issue temporary residence permits and citizen ID cards that further relaxed the restriction on longer-term migration. The temporary residence permits were issued to people who would live in a place that is different from their hukou registered locations for more than three months.

In 1997, the Ministry of Public Security published “*Opinions on Promoting the Reform of Residence Registration Management in Small Towns*” and announced some follow-up policies, which allowed agricultural hukou holders to convert their hukou to small towns or medium-size cities and become urban hukou holders. This was a breakthrough in relaxing restrictions on hukou type conversion, but the high barrier of conversions to urban hukou holders in big cities remained (Liang and Ma, 2004; Tang and Wu, 2022). In this stage, the moving of residence location and hukou location could be separate, which meant that one could move to a new city but still be registered as a hukou holder in the original place.

The fourth stage started in 2002 and continues into the present as some provinces are relaxing this restriction and eliminating the distinctions between agricultural and urban hukou holders. In

2002, Jiangsu province started to remove the dichotomous hukou types in the hukou system and rename all hukou as resident hukou. About 20 out of 34 provinces have eliminated the dichotomous hukou type. In 2014, the Chinese State Council published *Opinions on Further Promoting the Reform of the Household Registration System* and proposed to eliminate the restriction of hukou moving in small cities, relax the restriction of hukou moving in medium cities, and maintain certain restrictions on big cities (those with population of three to five million) and megacities (those with over five million people). The proposal also called for the abolition of the dichotomous hukou type in the hukou system. In 2022, the government eliminated the restriction for cities with a population of less than three million.

As a matter of fact, population mobility in most Chinese cities is not restricted by the hukou system and most inter-city migrants are moving unimpeded (Liang and Ma, 2004; Tang and Wu, 2022). The exceptions are movers to mega cities. The reform in 2014 also proposed an experimental plan to implement customized plans for around 60 cities. Among the new policies, a point-based system was planned for Beijing, Ningbo, Dalian, Qingdao, Dongguan, Guangzhou, Shenzhen, and Jinjiang. This system would provide a new channel to apply to be a permanent resident in these big cities and each applicant would be assessed based on a set of criteria. The methods of evaluation differ by city, but the aim is the relaxation of the restriction of population flows and encouragement of well-educated people to work in these cities.

Beijing, for example, is one of the two largest cities in China. It imposes relatively strict restrictions on possessing a hukou. There are nine criteria, including employment, residency, education, innovation and entrepreneurship, taxation, age, honor and rewards, and criminal record. An applicant or potential permanent resident is to be evaluated and get a score: the “points.” Applicants who score above a certain level of “points” can gain permanent residency. The point-

based hukou application is not the only path to gaining permanent residency in a mega city; other ways include marriage, employment, etc. According to the statistics released by the Beijing Municipal Human Resources and Social Security Bureau, there are 124,657 people who applied to be hukou holders in Beijing in 2018, with eventually 6,019 applicants gaining the hukou. The total number of applicants in 2022 is 130,111, with 6006 of them “admitted.” The yearly amount of those gaining the hukou in Beijing has been around 6,000 since 2018.

A key feature embedded in the hukou system, its influence as a migration barrier, is broadly discussed in literature. Some study the restriction of the hukou system (Xing and Zhang, 2017) while other research pay attention to returning migrants (Chunyu *et al.*, 2013; Xu *et al.*, 2017; Zhang *et al.*, 2020). Returning migrants who used to work in the city but eventually returned to their registered locations. Zhang *et al.* (2020) conclude that a higher hukou barrier leads to a higher probability of returning among low-skilled, rural, and inter-provincial migrants. Xing and Zhang (2017) conjecture and empirically show that rural-urban migrants strongly prefer cities with larger populations. Based on hukou policy documents between 2000 and 2016, Zhang *et al.* (2018) construct a hukou registration stringency index of local hukou qualification at the city level, that provides a vital perspective for comparing the hukou restriction.

Note that the city as an administrative level in China is a component of a province and consists of districts. Each district includes villages (rural areas) and communities (urban areas). Figure 1 as follows illustrate the prefectures of a small city, Zibo in Shandong province. The left figure shows the provincial boundary and each grid is a province. The highlighted province is Shandong province. In the middle is the prefectures in Shandong and boundary at the district level. The highlighted region is Zibo City, consisting of eight districts in the central area of the province. The figure on the right displays the prefectures of Zibo City with the community level boundary. The

community level refers to community in the urban area and village in the rural area. This level is the smallest administrative unit in this thesis.

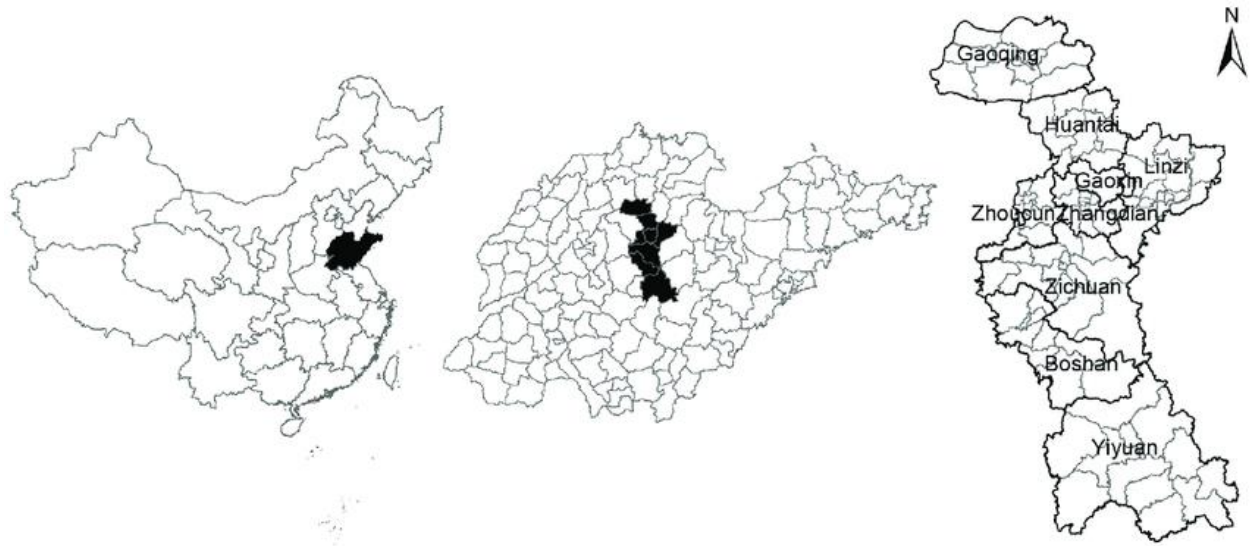


Figure 1. Prefectures of Zibo, Shandong Province. Source: Wang, Tao. *Location of the study area, Zibo City in Shandong Province, China*. Map. ResearchGate. https://www.researchgate.net/figure/Location-of-the-study-area-Zibo-City-in-Shandong-Province-China_fig7_248385803

A growing literature is discussing the hukou system's influence on the housing market (Chen *et al.*, 2019; Liao and Zhang, 2020; Lyu, 2022). Lyu (2022) focuses on the influence of car-restricted policy on the housing market, which is also a consequence of the hukou system. Similar to the restrictions on moving of the hukou location, house transactions and car sales are influenced by the hukou system, especially in big cities like Beijing. Chen *et al.* (2019) base the estimation of urban hukou value on housing prices in Jinan City and find that the value in the immigrant-dominated housing markets and top primary school districts are higher. It provides an empirical grounding for the studies of the effect of different hukou types. Researchers have conducted simulations to study the reform in the hukou system (Hertel and Zhai, 2006; Ito, 2008; Song, 2021; Whalley and Zhang, 2007). Song (2021) conducts a cost-benefit analysis and simulation of the hukou system reform. He concludes that the cost of successful reform of the hukou policy could be offset by the benefits of increasing GDP and reducing income inequality. His paper provides

an optimistic perspective of conducting a reform to improve the social benefit and economic growth in China. Hertel and Zhai (2006) also use simulation results to show that the reforms in the hukou system and increased off-farm labor mobility would reduce the inequality across urban-rural incomes. Whalley and Zhang (2007) investigate the sources of inequality in China through numerical simulation methods and found that the impacts of removing Hukou restrictions on migration are smaller but still significant.

The housing market in China is an important topic in the study of the Chinese economy. It is a source of wealth inequality and of the impediments faced by rural-to-urban migrants wanting to own their housing. Zheng *et al.* (2006) look at the violations of Pareto-efficient spatial sorting conditions in an ordered-location-choice model. Chen (2016) investigates the market failure and assesses housing policy changes in 2010 in 19 major cities in China. He illustrates that a high transaction cost will cause a high probability for market failure. Chen concludes that the new policies in 2010 have improved the efficiency of the housing market, preventing housing prices from soaring and maintaining stable housing prices. Li and Huang (2006) focus on housing behavior and land policy. They qualitatively expressed concern with increasing social inequality as a result of the hukou system and a two-class society stratification between urban residents and migrants. Jiang *et al.* (2017) and Hao and Tang (2015) survey the willingness to settle down in cities and found that surprisingly few rural migrants chose to do so. Using US data from Annual Housing Survey (AHS), Henderson and Ioannides (1986) model the joint decision of tenure choice and housing demand and estimated the parameters in a maximum likelihood framework.

According to the US Census, a person in the US is expected to move 11.7 times in their lifetime based on the data from American Community Survey (ACS). This estimate is larger than that in my dataset, on average 0.1 time per person. However, it is natural to believe the estimates of

Chinese people mobility frequency is underestimated. This is because moving times is a constructed variable, which reach the upper limit when the interviewer moves to a new place in every wave of the survey. In the literature on residential mobility, Kan (2007) uses data from the Panel Study of Income Dynamics (PSID) in the US to investigate the effect of local social capital, proxied by whether the family received help from the community. He concluded that there was a significant deterring effect of local social capital on moving, especially long-distance moving. Using the 2013 China Household Finance Survey, Liao and Zhang (2020) study how hukou type affects housing tenure choice and wealth accumulation in urban China. They illustrate that agricultural hukou is a source of wealth inequality and explain the difficulties faced by rural-to-urban migrants in owning housing. Migrants with agricultural hukou tend to face more uncertainty, having limited access to mortgage finance, and are less eligible for low-cost housing when compared to urban hukou holders. Both studies above modeled both the moving choice and tenure choice with probit models and adopted instrumental variables in dealing with the endogeneity between the key independent variables and the response variable. Christensen and Timmins (2022) focus on the influence of housing on neighborhood choice based on the Housing Discrimination Study (HUD) data from 28 US cities. They found that minorities in the US are more likely to move to neighborhoods with fewer opportunities and more crime, which worsens social inequality. Deng and Gustafsson (2007) study migration behavior in China and conclude that the probability of becoming a permanent migrant from rural to urban areas is positively related to parental education, to belonging to the ethnic majority, and to the parent's membership in the Communist Party.

My research investigates the impact of the hukou system on the household moving decision and its influence on the housing market in China. The impact of moving can be divided into different components such as the influence of hukou type (agricultural or urban) and location of

the hukou. More specifically, the moving choices can be further identified as local moving within the province (intra-province) or long-distance moving across provinces (inter-province). Both types of moving will be considered in the models.

The rest of the thesis proceeds as follows. Section 2 describes my data and provides details on the key variables, constructed variables, and related explanation. Section 3 outlines the econometrics models. Section 4 presents and discusses the empirical results from the estimations. Section 5 reports the tests including hypothesis tests and robustness tests. Section 6 concludes with the findings.

2. Data

The main data I used is from China Family Panel Studies (CFPS), funded by the 985 Program of Peking University and carried out by the Institute of Social Science Survey of Peking University.

CFPS is a nationally representative, biennial longitudinal general social survey in China, which collected data every two years from 2010. The dataset provides rich information of life habits, profiles, and perceptions on some social norms of individuals, families, and the communities involved in this survey. It consists of an adult questionnaire, a children questionnaire, a family roster questionnaire, and a family questionnaire in each survey collection.

14,960 households from 25 provinces were successfully interviewed in the 2010 baseline survey. All family members with blood, marital, or adoptive ties with the household, and newly born or adopted children were defined to be the CFPS gene members, who were tracked and interviewed in the follow-up surveys. This excludes Hong Kong SAR, Macau SAR, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan. Since the population in the 25 provinces account for 95% of the population in China, the CFPS dataset is still a national representative sample (Xie *et al.*, 2017).

In the years 2010 and 2014, community questionnaires were included during the investigation, which provide general information about the community. The content covered by the community questionnaire is displayed in Table A1 in the appendix.

In CFPS, individuals who are younger than 16 are defined as children, according to the user guides of CFPS dataset. The adult questionnaire contains information on demography, education, marriage, family relations, work, personal income, health, etc. Details of the main content of adult questionnaire in 2010 are included in Table A3. Table A4 shows the added questions in the following years of adult questionnaires.

The questionnaire for children is designed differently; it was divided into two sections: the child section completed by the children themselves and the parent section completed by the guardians (Xie *et al.*, 2017). CFPS includes different questions for each age group: 0-1 (1 excluded), 1-2, 3-5, and 6-15. Only children age 10 and above need to finish the child section themselves.

The family questionnaire recorded geographical information, living conditions, social interactions, income and expenditure, and financial assets and possessions of the family. Details of content covered by the family questionnaire is displayed by Table A2. The family roster questionnaire recorded information of relationships among household members and their demographic information.

However, since the CFPS dataset only discloses province information. Specific cities or town names were recoded as an ID number for privacy purposes and therefore many policy-related cities cannot be identified from the dataset. The CFPS dataset provides rich information about the Chinese society, economy, population, education, and health.

After cleaning and selecting variables in the large main datasets of five survey waves, the final dataset contains 168,283 observations in total from 2010 to 2018. Since each wave of interviews were not finished in one year, parts of the interviews were conducted in the year after the start of this wave. Thus, even in the same wave of the survey, the values of yearly variables may differ such as the GDP per capita of the province. The summary statistics are displayed in Table 2. The definitions of variables are included in Appendix III.

Table 2. Summary Statistics of Key Variables in Finalized Data

Variable	N	Mean	SD	Skewness	Min	p25	p50	p75	Max
female	168271.00	0.50	0.50	-0.02	0.00	0.00	1.00	1.00	1.00
urban	164833.00	0.47	0.50	0.11	0.00	0.00	0.00	1.00	1.00
hukou_change	134745.00	0.02	0.13	7.55	0.00	0.00	0.00	0.00	1.00
type_change	105985.00	0.06	0.32	5.39	0.00	0.00	0.00	0.00	2.00
income	106180.00	13.73	37.63	125.11	0.00	0.00	4.99	20.00	8739.25
eduy	162683.00	7.35	4.84	-0.19	0.00	5.00	9.00	10.00	23.00
house_debts	165563.00	17.66	188.44	289.99	0.00	0.00	0.00	0.00	68408.06
resivalue	166177.00	26.32	59.42	10.98	0.00	3.00	10.00	25.79	4085.25
size	139586.00	137.34	115.49	6.48	1.00	72.00	102.00	176.00	3000.00
house_exp	163331.00	5.69	20.45	13.51	0.00	0.76	1.94	3.93	980.02
house_cost	101236.00	-6.06	99.84	-11.69	-8048.61	-6.05	0.00	1.41	8520.28
mortgage	165656.00	1.64	18.50	38.37	0.00	0.00	0.00	0.00	1351.61
trust	126499.00	6.58	2.18	-0.38	0.00	5.00	7.00	8.00	10.00
age	168271.00	45.51	16.94	0.16	-8.00	31.00	46.00	58.00	110.00
age2	168271.00	23.58	16.17	0.81	0.64	9.61	21.16	33.64	121.00
finc_pc	160906.00	12.35	19.40	23.03	0.00	3.74	8.15	15.20	1556.53
period	168271.00	2.95	1.39	0.05	1.00	2.00	3.00	4.00	5.00
agr_hukou	168271.00	0.73	0.44	-1.03	0.00	0.00	1.00	1.00	1.00
married	168271.00	0.78	0.42	-1.34	0.00	1.00	1.00	1.00	1.00
own	168271.00	0.86	0.35	-2.02	0.00	1.00	1.00	1.00	1.00
comm_change	134745.00	0.09	0.28	2.96	0.00	0.00	0.00	0.00	1.00
prov_change	134745.00	0.01	0.11	8.72	0.00	0.00	0.00	0.00	1.00
local_move	134745.00	0.08	0.27	3.13	0.00	0.00	0.00	0.00	1.00
long_move	134745.00	0.01	0.08	11.68	0.00	0.00	0.00	0.00	1.00
phy_move	134745.00	0.00	0.04	25.77	0.00	0.00	0.00	0.00	1.00
hukou_move	134745.00	0.01	0.08	13.17	0.00	0.00	0.00	0.00	1.00
bi_move	134745.00	0.11	0.31	2.57	0.00	0.00	0.00	0.00	1.00
rationed_move	966.00	0.79	0.41	-1.43	0.00	1.00	1.00	1.00	1.00
pop	168271.00	5958.22	3094.06	0.45	349.00	3479.00	4653.00	9405.47	12489.00
gdp	167278.00	24087.26	17546.25	1.34	1096.20	10697.34	19495.69	30463.68	81748.98
gdppc	167278.00	40039.39	20201.50	1.63	12989.37	26490.61	34446.65	46368.99	122355.66
diff_gdppc	105285.00	5009.18	5805.52	4.35	-86686.58	3077.04	4275.63	6200.29	99113.83

Data Source: Chinese Family Panel Study (CFPS) and China Data Online

The availability of variable in each year are summarized in Appendix IV. Since some categorical variables are encoded as double type in the dataset, I also included the explanation of each value of these variables in Appendix V.

From the mean values of the years of education and non-agricultural hukou population, we can tell that there are some potential biases in the sample of the dataset, since these values are slightly smaller than Chinese census record of in 2010 and 2020. This is also explained by the Guide of CFPS dataset, that is, family members with higher educational degrees and higher incomes are prone to be busy and do not stay at home and finish the survey.

For the information of individual income, I use net family income per capita to replace the total personal income due to the high ratio of 0 values despite the high availability. The reason behind the high ratio of 0 values is unknown. It is thus not appropriate to employ in the following analysis. The house expenditure is constructed by CFPS and it includes expenses on paid rent, house maintenance, property management, and utility bills. In this way, the variable is useful and meaningful for both renters and owners of their current residences.

Hukou registered location information is available only in the baseline survey of 2010. In the following survey waves, the corresponding variables have higher ratios of missing values, and most are those who have different hukou registered locations from their current residency province. Thus, to obtain hukou registered location information for the vast amount of missing or “not applicable” values, I used variables of whether the location of hukou is in the same province of current residence to infer the location information of hukou for the wave of 2012, 2014, and 2016. Two indicator variables of whether hukou is in the province showed up in the datasets, one from the adult dataset and the other from the family economics surveys. Because the hukou registered location information is also from the adult survey, which is answered by individuals themselves in

the interview, the indicator variable from the adult questionnaire was used as a main reference. The variable in the family economic surveys was used as a complement for missing values in the main reference variables. For the wave of 2018, the indicator variable of whether hukou is in the province are not recorded, then I use the variable of whether the individual changed the hukou registered location instead. Individuals who did not move their hukou were still using the same hukou registered location in the 2016 wave. Note that the location of hukou is at provincial level, so attentions is only paid to the provincial information. Therefore, the hukou stringency index constructed by Zheng *et al.* (2018) cannot be utilized. Instead of focusing on return migrants, this paper focuses on a broader range of migrants who either return or leave their registered place. I am able to capture the hukou location changes during the migration in my data.

While this strategy makes up for missing information, there are some logically inconsistent observations which conflict with the information in the dataset. For example, some people answered that their hukou were in the same province of their residency, but the hukou variable already recorded that the location of hukou was in another province. Because of this, I keep the hukou registered location information recorded by the data for now, since this is the direct information in the dataset. This does not occur in 2018 since the mechanism to identify the hukou registered location avoided such confusions. The portion of such inconsistent hukou registered location information and availability of hukou information including the inconsistent ones is summarized in Table 3.

Table 3: Summary of Hukou Registered Location Information

	Total	Availability	#Inconsistent
2010	33,526	33519	-
2012	35,351	35328	158
2014	35,663	35646	96
2016	33,194	33192	56
2018	30,549	28294	-

In cleaning of the variable of agricultural hukou (*agr_hukou*), I dropped the observations that do not have a hukou, and foreigners since they are not classified as agricultural or non-agricultural hukou type. Note that agricultural hukou does not indicate residence in rural area. The distribution of agricultural or urban hukou who live in urban or rural area are summarized in Table 4.

Table 4. Urban Hukou by Urban Residents

	Urban Hukou		Total	
	No	Yes		
Urban Resident	No	80,832	6,014	86,846
	Yes	38,923	39,064	77,987
	Total	119,755	45,078	168,271

I constructed a set of moving variables that includes two pairs (four types) of moving behaviors: local moves and long-distance moves, as well as moves to non-hukou locations and moves to hukou locations. Specifically, a local move refers to a change in community but not province, while a long-distance move involves a change in both community and province. A move to a non-hukou location refers to a long-distance move that does not change the individual's hukou location, while a move to a hukou location refers to a long-distance move that changes the individual's hukou location. More details on this construction can be found in Appendix II. It's worth noting that this construction does not account for moves within the same community, which are difficult to trace given the design of the CFPS.

For the sake of further analysis, I created a binary variable indicating whether any of the aforementioned types of moves occurred for each individual in a given year. Table 5 and Table 6 display the number of community changes by province change and the number of province changes by hukou change, respectively. Table 7 includes the number of moves and the corresponding ratio in the total sample for each survey wave.

Table 5. Community Change by Province Change

	Community Change		Total	
	No	Yes		
Province Change	No	119,831	738	120,569
	Yes	13,210	966	14,176
	Total	123,197	11,548	134,745

Table 6. Hukou Change by Province Change

	Hukou Change		Total	
	No	Yes		
Province Change	No	132,223	818	133,041
	Yes	276	1,428	1,704
	Total	132,499	2,246	134,745

Table 7. Ratio of Different Types of Moving by Survey Wave

	local_move	long_move	phy_move	hukou_move
2012	3,576	42	7	35
Ratio	0.101	0.001	0	0.001
2014	4,005	270	130	140
Ratio	0.112	0.008	0.004	0.004
2016	3,089	283	35	248
Ratio	0.093	0.009	0.001	0.007
2018	2,540	371	30	341
Ratio	0.083	0.012	0.001	0.011
Total	13,210	966	202	764

However, there is a small portion of data that is inconsistent with the information provided in the dataset. For example, as shown in Table 5, it is illogical for an individual to remain in the same community but change provinces, unless there has been a change in provincial borders, which is unlikely but may require further investigation.

3. Econometric Model

3.1 Binary Probit Models

Ioannides and Zanella (2008) estimate a switching regression to model the selection of neighborhood when considering the neighborhood effect in the formation of children's human capital and the process of cultural transmission. Similarly, in this thesis, the first step is to model the moving choices of the five types using probit models. Based on individual, family, and community characteristics, a binary probit model is estimated for each type of moving behavior as follows:

$$\mathbf{M}_{it} = \beta_0 + \beta_1 hukou_type_{t-1} + \beta_2 incpc_{t-1} + \mathbf{X}\boldsymbol{\beta} + \gamma_1 Prov_t.FE + \gamma_2 Prov_{t-1}.FE + \epsilon$$

Vector \mathbf{M}_{it} is the general moving indicator of individual i in period t , where the moving includes five types of moves: local moving, inter-province moving, non-hukou location moving, hukou location moving, and binary moving. The matrix \mathbf{X} include age, square term of age, lagged term of whether solely own the residence, lagged term of whether live in urban area, lagged term of years of education, lagged term of total expenditure on housing, lagged term of total mortgage, the difference in GDP per capita between the consecutive year. Inspired by Kan (2007), the model also includes local social capital, the family's cohesion to the community, as a regressor. The interaction with neighbors in the community has an impact on the family's cohesion to the community, and therefore influences the decision to move. The model includes two types of fixed effects. The first is the fixed effect of the current residence province, which controls for the difficulties of moving within the province. The second is the fixed effect of the previous residence province, which controls for the challenges of moving out of the province. For local moving, where the residence province in the current period and previous period is the same, only one fixed effect is included in the model. The results are summarized in Appendix VI.

3.2 Multinomial Logit Model

Next, I created a multinomial choice variable *move*, which includes four levels: not move (*move* = 0), local intra-province move (*move* = 1), inter-province move without change in hukou (*move* = 2) and inter-province move with change in hukou (*move* = 3). Note that the long-distance inter-province move is indeed the sum of the latter two types of moving, thus long-distance moving is not a level in the new variable. Based on this, I fitted a multinomial logit model as follows. It should be noted that the long-distance inter-province move is the sum of the two previous types of moving; therefore, the long-distance moving is not included as a level in the new variable. Based on this, a multinomial logit model is fitted as follows. To simplify the analysis and emphasize the focus on the moving type with a change in hukou registered location, levels 2 and 3 of the four-level moving variable are merged into a three-level moving choices variable. This variable will also be analyzed in section 4.

$$l_{ijt} = \log\left(\frac{p_{ij}}{p_{i0}}\right) = \beta_0 + \beta_1 \text{agri}_{hukou_{t-1}} + \mathbf{X}\boldsymbol{\beta} + \epsilon_{ijt}.$$

l_{ijt} is the log odds of individual i choosing choice j at time period t , $j = \{1,2,3\}$ in the case of not moving is the base level. Since the models with provincial fixed effect did not converges, I did not consider the fixed effect for now. The matrix \mathbf{X} includes all information covered in the probit model and the average family net income; $\text{agri}_{hukou_{t-1}}$ is the indicator variable of whether the individual possesses an agricultural hukou in the previous period. Follow the structure proposed by Greene (2018), and the probability p_{ij} is thus, without loss of generality, the conditional probability of individual i choosing j as

$$p_{ij} = \begin{cases} \frac{1}{1 + \sum_{k=1}^3 \exp(\mathbf{z}_i \boldsymbol{\gamma}_k)}, & \text{if } i = 0; \\ \frac{\exp(\mathbf{z}_i \boldsymbol{\gamma}_j)}{1 + \sum_{k=1}^3 \exp(\mathbf{z}_i \boldsymbol{\gamma}_k)}, & \text{if } i \geq 1, \end{cases}$$

where $\mathbf{z}_i\boldsymbol{\gamma}_k = \beta_0 + \beta_1 agri_hukou_{t-1} + \mathbf{X}\boldsymbol{\beta} + \epsilon_{ijt}$.

To better understand the differences between moving with a change in hukou registered location and those without, the base level was changed to hukou move ($move=3$), and the comparison of estimations is included in Table 8.

3.3 Intention of Moving Hukou Registered Location

In addition to looking at the moving choice, I am also investigating the intention of moving hukou registered location to another province and, implicitly, whether the moving of hukou registered location is influenced by the system. First, I am constructing a binary variable *rationed_move*, with a value of 1 for people who move to another province with hukou ($hukou_move = 1$), a value of 0 for people who move to another province but without a change in hukou ($non_hukou_move = 1$), and a missing value for people who do not move. I am utilizing the two-stage estimation method – Heckman selection model (Greene, 2018; Heckman, 1979) – to capture the correct potential correlation between the moving choices and moving intention. The equations to estimate are:

$$rationed_move = \mathbf{X}\boldsymbol{\beta} + \mu_{1i};$$

$$select = z_i\boldsymbol{\gamma} = \beta_0 + \beta_1 house_exp_t + \beta_2 finc_pc + \beta_3 agr_hukou_t + \beta_4 own_t + \mu_{2i}.$$

The model for rationed move is conditional on the same information set used in section 3.1 when modeling for the binary moving choices. The selection variable is not observed here and represents the intention to move with the transfer of hukou registered location. The log likelihood for observation i is

$$\begin{cases} w_i \ln \phi \left\{ z_i \boldsymbol{\gamma} + \frac{(y_i - x_i) \rho}{\sigma} \right\} - \frac{w_j}{2} \left(\frac{y_i - \mathbf{X}\boldsymbol{\beta}}{\sigma} \right)^2 - w_i \ln(\sqrt{2\pi}\sigma), & \text{if } y_i \text{ is observed;} \\ w_i \ln \phi(-z_i \boldsymbol{\gamma}), & \text{if } y_i \text{ is observed,} \end{cases}$$

where $\Phi(\cdot)$ is the standard cumulative normal distribution function. The variables that explain the intention includes the current period's housing expenditure, family income per person, whether the person holds an agricultural hukou, and whether they own a house. The key difference from the information set in $\mathbf{X}\boldsymbol{\beta}$ is that the information explaining the intention are from the current period instead of the previous periods. It is important to note that the Heckman selection model estimates directly the inverse hyperbolic tangent of the correlation, rather than the correlation $\rho = \text{corr}(\mu_{1i}, \mu_{2i})$. The inverse hyperbolic tangent is defined as:

$$\text{atanh } \rho = \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho}$$

Then a χ^2 test is conducted to test significance of $\rho \neq 0$. If the null hypothesis of $\rho = 0$ is rejected, it is concluded that the selection bias exists in the moving choices. The standard error of the residual in the *rationed_move* equation is not directly estimated. Instead, the logarithm of it, $\ln\sigma$, will be estimated for numerical stability. The product of the correlation and the standard error, $\lambda = \rho\sigma$, is a commonly used estimates for selectivity effect.

The housing maintenance cost is rent for renter, and annualized financial cost for owners, constructed following Henderson and Ioannides (1986) and King (1980). Specifically, the housing maintenance cost for owner i in province k in year t is

$$\text{house_cost}_{ikt} = w_i(\bar{r}_{kt} + m^{\circ}_k) - \Delta w_i,$$

where w_i denotes individual i 's residence value; r_k is the average nominal rate of mortgage interest for five-year and longer periods; m°_k denotes the rate of maintenance that account for the rate of cost to maintain the residential place at a certain level of services. It is constructed by the property depreciation length for residential house in China, which is 70 years. Hence, I set the

maintenance rate is 1.429% ($= \frac{100\%}{70}$). Nevertheless, it should be noted that the property right for the residential place is automatically extended at the end of the 70-year period.

3.4 Hukou Type Changes

Then to investigate how hukou type changed as people moved, I fit another multinomial logit model as follows:

$$change_{ijt} = \log\left(\frac{p_{ij}}{p_{ij_0}}\right) = \beta_0 + \beta_1 agri_{hukou_{i,t-1}} + \beta_2 move_{FixEffect_{it}} + \beta X + \epsilon_{ijt},$$

where $change_{ijt}$ is the log odds for individual i to change the type of hukou or not (j) at time period t , j represents the three possible types of change: 0 for no change, 1 for changing from urban to agricultural, and 2 for changing from agricultural to urban. To facilitate comparison, the base level has been set to the three types of change, allowing for a clearer understanding of the differences between each choice.

3.5 Housing Expenditure

In the next step, I study the housing expenditure behavior and take into account the influence of the hukou system. There is a possibility that the moving choices may impact housing expenditures due to factors such as ownership of a new house or additional costs induced by the policy in the new place. To account for the correlation between the moving choices and the housing expenditures. I apply the Heckman selection model (Greene, 2018; Heckman, 1979), which is mentioned in section 3.3. However, in this case, the selection is fully observed and is, in fact, the moving choice. The specifications are:

$$house_exp_i = \mathbf{x}_i\boldsymbol{\beta} + \delta_{1i};$$

$$bi_move = z_i\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta} + \delta_{2i},$$

where $\mathbf{x}_i\boldsymbol{\beta} = \beta_0 + \beta_1 house_cost + \beta_2 cpi + \beta_3 rinc_pc + \beta_4 size + \beta_5 urban + \beta_6 own + FE$

and the second equation uses the same set of control variables in section 3.2, in addition to provincial fixed effect, but excludes the housing expenditure variable. Note that the selection variable does not refer to a specific type of moving choice, but to a broader concept of binary moving choice, that is, 1, if individual i move ($local_move = 1$ or $hukou_move = 1$ or $non_hukou_move = 1$), and 0, if not. Similarly, testings on the significance of correlation between μ_{1i} and μ_{2i} will determine if there is selection bias on the housing expenditure.

4. Empirical Results

4.1 Binary Probit Models

In the probit models, some levels of the province variable had to be omitted due to collinearity and perfect prediction issues. However, the full results are presented in Appendix VI. The results show that individuals living in urban areas are generally more likely to move. This could be due to the fact that urban areas have more opportunities for individuals to move, particularly in terms of housing markets. The results also show that the agricultural hukou indicator is not significant, suggesting that the type of hukou does not have a significant impact on the likelihood of moving.

Results with the control variables are as follows: age and owning a property in the city have a negative effect on moving choices, while years of education, average net family income, and trust in the neighborhood have a positive effect. People with more education and higher income should have a greater capacity to invest in new properties. Thus, they may be more incentivized to move.

The housing expenditure and total housing mortgage have negative effect since people with higher financial burden may be more risk-averse and wish to avoid risk of default. Significant at 90% confidence level, the negative effect of trust in neighbors means that the more one has higher cohesion with the neighborhood, the less likely one would move to a new place. Influences of GDP per capita, difference in GDP per capita, and lagged difference in GDP per capita are significant

at 99% confidence level but the estimation value is very small (smaller than 0.0001). The base level of province fixed effect is Beijing. The coefficients of current province fixed effects are mostly negative and that of the previous province are mostly positive. This makes sense: people are facing higher restrictions to move to Beijing and less willing to leave Beijing given the resource of the city. This result is consistent with the conclusion of Xing and Zhang (2017) about the attractiveness of big cities.

I included an interaction term between the agricultural hukou indicator and lagged residence province fixed effects due to the potential interaction between hukou type and the residence province. This is because people may face varying frictions in different origin-destination combinations, and this effect may differ by the hukou type. The results of the models with interaction terms are displayed in Appendix VII. The signs of estimates that appear in the models with interaction terms are similar to those in the models without the interaction. The significant interactions suggest the existence of moving frictions. However, it is challenging to determine whether individuals with agricultural hukou in the past have higher or lower incentives to move and how this varies across provinces.

4.2 Multinomial Logit Model

In the multinomial logit model with not moving as base level, the hukou type is the key variable of my interests. The estimation results are presented in Table 8. Column (1) to (3) show the estimates of the original logit model, while column (4) to (6) display the estimates of relative risk ratio (RRR), which is defined as $RRR_m = \exp(\beta_m)$, the exponential of corresponding coefficient in logit model. The RRR value represents how many times more likely on average is that the individual will make the current choice over the base level choice, as the variable of interest

changes by one unit. Hence, when RRR is greater than one, individuals are more likely to choose the current choice. Otherwise, individuals would rather choose the base level choice.

Table 8. Estimates Results for Logit Models

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Logit Local Move	Logit Non-Hukou Move	Logit Hukou Move	RRR Local Move	RRR Non-Hukou Move	RRR Hukou Move
age	-0.0187*** (0.00501)	-0.0496 (0.0574)	-0.0453** (0.0214)	0.981*** (0.00492)	0.952 (0.0546)	0.956** (0.0205)
age2	0.00924* (0.00501)	-0.0125 (0.0694)	0.0167 (0.0232)	1.009* (0.00505)	0.988 (0.0686)	1.017 (0.0236)
L.female	0.00337 (0.0269)	0.117 (0.254)	-0.344*** (0.114)	1.003 (0.0269)	1.124 (0.285)	0.709*** (0.0810)
L.own	-0.342*** (0.0370)	-0.380 (0.332)	-0.403*** (0.152)	0.710*** (0.0263)	0.684 (0.227)	0.668*** (0.101)
L.urban	0.0943*** (0.0312)	0.187 (0.301)	0.120 (0.130)	1.099*** (0.0342)	1.206 (0.363)	1.127 (0.147)
L.agr_hukou	-0.0637* (0.0356)	-0.113 (0.324)	0.287* (0.156)	0.938* (0.0334)	0.893 (0.289)	1.333* (0.209)
L.eduy	0.00751** (0.00350)	0.0416 (0.0364)	0.0151 (0.0155)	1.008** (0.00353)	1.042 (0.0380)	1.015 (0.0157)
L.finc_pc	0.00207*** (0.000776)	0.00388 (0.00649)	0.00406** (0.00188)	1.002*** (0.000778)	1.004 (0.00652)	1.004** (0.00189)
L.house_exp	0.000381 (0.000596)	0.00323 (0.00333)	0.000334 (0.00252)	1.000 (0.000596)	1.003 (0.00334)	1.000 (0.00252)
L.mortgage	0.000848* (0.000456)	-783.9 (885.5)	-0.0107 (0.0107)	1.001* (0.000456)	0 (0)	0.989 (0.0106)
L.trust	-0.0199*** (0.00601)	-0.0538 (0.0591)	0.00497 (0.0266)	0.980*** (0.00589)	0.948 (0.0560)	1.005 (0.0267)
L.gdppc	-4.95e-06*** (9.33e-07)	-4.39e-05*** (1.19e-05)	-2.60e-05*** (4.16e-06)	1.000*** (9.33e-07)	1.000*** (1.19e-05)	1.000*** (4.16e-06)
diff_gdppc	8.01e-06*** (2.61e-06)	6.15e-05*** (5.88e-06)	7.37e-05*** (2.86e-06)	1.000*** (2.61e-06)	1.000*** (5.88e-06)	1.000*** (2.86e-06)
L.diff_gdppc	2.62e-05*** (3.46e-06)	0.000115*** (1.35e-05)	9.74e-05*** (6.11e-06)	1.000*** (3.46e-06)	1.000*** (1.35e-05)	1.000*** (6.11e-06)
Constant	-0.992*** (0.147)	-3.500** (1.388)	-3.275*** (0.596)	0.371*** (0.0544)	0.0302** (0.0419)	0.0378*** (0.0225)
Observations	58,215	58,215	58,215	58,215	58,215	58,215

Standard errors in parentheses for column (1)-(3)

seEform in parentheses for column (4)-(6)

*** p<0.01, ** p<0.05, * p<0.1

The result indicates that people with agricultural hukou are 1.333 times more likely to move with hukou location compared to staying in the same place. Control variables such as age, gender, house ownership, urban residence, years of education, and net family income per person are

statistically significant. Females are less likely to move with the change in hukou location, while homeownership acts as a barrier to all types of moves. Meanwhile, higher education levels and family net incomes encourage people to move to new places.

To further investigate how the decision to move across provinces with a change in hukou differs from other types of moves, Table 9 includes the results of the multinomial logit model with the base level set as a hukou move ($move=3$). Column (1) to (3) show the same RRR results from Table 8, while columns (4) to (6) compare the decision to move hukou registered location in inter-province moving to three other types of mobility choices. Additionally, columns (7) and (8) display the models of nested moving choices with the exact same set of explanatory variables.

Table 9. Estimated Relative Risk Ratio Results for Logit Models

VARIABLES	(1) Base = Not Move	(2) Base = Not Move	(3) Base = Not Move	(4) Base = Hukou Move	(5) Base = Hukou Move	(6) Base = Hukou Move	(7) Base = Not Move	(8) Base = Not Move
age	0.981*** (0.00492)	0.952 (0.0546)	0.956** (0.0205)	1.046** (0.0224)	1.027 (0.0224)	0.996 (0.0602)	0.980*** (0.00488)	0.959** (0.0206)
age2	1.009* (0.00505)	0.988 (0.0686)	1.017 (0.0236)	0.983 (0.0228)	0.993 (0.0234)	0.971 (0.0704)	1.010** (0.00503)	1.014 (0.0236)
L.female	1.003 (0.0269)	1.124 (0.285)	0.709*** (0.0810)	1.411*** (0.161)	1.416*** (0.165)	1.586* (0.431)	1.006 (0.0269)	0.704*** (0.0805)
L.own	0.710*** (0.0263)	0.684 (0.227)	0.668*** (0.101)	1.496*** (0.227)	1.063 (0.164)	1.023 (0.364)	0.709*** (0.0261)	0.668*** (0.101)
L.urban	1.099*** (0.0342)	1.206 (0.363)	1.127 (0.147)	0.887 (0.115)	0.975 (0.129)	1.070 (0.344)	1.097*** (0.0341)	1.132 (0.147)
L.agr_hukou	0.938* (0.0334)	0.893 (0.289)	1.333* (0.209)	0.750* (0.117)	0.704** (0.112)	0.670 (0.233)	0.936* (0.0332)	1.353* (0.212)
L.eduy	1.008** (0.00353)	1.042 (0.0380)	1.015 (0.0157)	0.985 (0.0153)	0.992 (0.0156)	1.027 (0.0399)	1.008** (0.00352)	1.015 (0.0157)
L.finc_pc	1.002*** (0.000778)	1.004 (0.00652)	1.004** (0.00189)	0.996** (0.00188)	0.998 (0.00189)	1.000 (0.00666)	1.002*** (0.000776)	1.004** (0.00187)
L.house_exp	1.000 (0.000596)	1.003 (0.00334)	1.000 (0.00252)	1.000 (0.00252)	1.000 (0.00257)	1.003 (0.00413)	1.000 (0.000591)	1.000 (0.00251)
L.mortgage	1.001* (0.000456)	0 (0)	0.989 (0.0106)	1.011 (0.0108)	1.012 (0.0108)	0 (0)	1.001* (0.000457)	0.990 (0.0103)
L.trust	0.980*** (0.00589)	0.948 (0.0560)	1.005 (0.0267)	0.995 (0.0264)	0.975 (0.0264)	0.943 (0.0597)	0.980*** (0.00587)	1.006 (0.0267)
L.gdppc	1.000*** (9.33e-07)	1.000*** (1.19e-05)	1.000*** (4.16e-06)	1.000*** (4.16e-06)	1.000*** (4.22e-06)	1.000 (1.25e-05)	1.000*** (9.18e-07)	1.000*** (4.16e-06)
diff_gdppc	1.000*** (2.61e-06)	1.000*** (5.88e-06)	1.000*** (2.86e-06)	1.000*** (2.86e-06)	1.000*** (3.48e-06)	1.000** (5.87e-06)	1.000*** (2.47e-06)	1.000*** (2.82e-06)
L.diff_gdppc	1.000*** (3.46e-06)	1.000*** (1.35e-05)	1.000*** (6.11e-06)	1.000*** (6.11e-06)	1.000*** (6.25e-06)	1.000 (1.38e-05)	1.000*** (3.29e-06)	1.000*** (5.98e-06)

Constant	0.371*** (0.0544)	0.0302** (0.0419)	0.0378*** (0.0225)	26.45*** (15.76)	9.807*** (5.962)	0.798 (1.181)	0.388*** (0.0566)	0.0351*** (0.0210)
Observations	58,215	58,215	58,215	58,215	58,215	58,215	58,215	58,215

seEform in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The results suggest that net family income per person has a significant effect on moving choices, with higher-income individuals being more likely to choose to move to a different province with a hukou location change, and less likely to choose to remain in the same place. The estimated effects for the difference of GDP per capita in the residence province are close for the model of four- and three-level moving choices variables.

To investigate the potential impact of net family income per person on moving choices, I have included this variable and re-estimated the model presented in section 3.2. The results are shown in Table 10, which displays the RRR estimates for the multinomial logit model with not moving as the base level. As in Table 8, column (1) to (3) present the estimates of the original logit model, and column (4) to (6) show the estimates of the RRR.

Table 10. Estimates Results for Logit Models with Current Income

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Logit Local Move	Logit Non-Hukou Location Move	Logit Hukou Location Move	RRR Local Move	RRR Non-Hukou Location Move	RRR Hukou Location Move
age	-0.0168*** (0.0050)	-0.0507 (0.0547)	-0.0464** (0.0202)	0.9834*** (0.0050)	0.9506 (0.0520)	0.9546** (0.0193)
age2	0.0077 (0.0050)	-0.0116 (0.0669)	0.0176 (0.0212)	1.0078 (0.0051)	0.9885 (0.0661)	1.0177 (0.0216)
L.female	0.0020 (0.0272)	0.0868 (0.2604)	-0.3771*** (0.1195)	1.0020 (0.0273)	1.0906 (0.2840)	0.6859*** (0.0820)
L.own	-0.3336*** (0.0379)	-0.3065 (0.3601)	-0.3661* (0.1877)	0.7163*** (0.0271)	0.7361 (0.2651)	0.6934* (0.1301)
L.urban	0.1012*** (0.0318)	0.1494 (0.3311)	0.1127 (0.1323)	1.1065*** (0.0352)	1.1611 (0.3844)	1.1193 (0.1480)
L.agr_hukou	-0.0491 (0.0377)	-0.1534 (0.3456)	0.2582 (0.1688)	0.9521 (0.0359)	0.8578 (0.2965)	1.2946 (0.2185)
L.eduy	0.0064* (0.0036)	0.0349 (0.0331)	0.0130 (0.0166)	1.0064* (0.0036)	1.0355 (0.0343)	1.0131 (0.0168)
finc_pc	0.0020*** (0.0006)	0.0047*** (0.0014)	0.0032*** (0.0012)	1.0020*** (0.0006)	1.0047*** (0.0014)	1.0032*** (0.0012)
L.finc_pc	0.0015** (0.0007)	0.0032 (0.0039)	0.0035*** (0.0013)	1.0015** (0.0007)	1.0032 (0.0039)	1.0035*** (0.0013)

L.house_exp	0.0004 (0.0006)	0.0028 (0.0035)	0.0003 (0.0022)	1.0004 (0.0006)	1.0029 (0.0035)	1.0003 (0.0022)
L.mortgage	0.0007* (0.0004)	-785.1912 (597.8170)	-0.0098* (0.0056)	1.0007* (0.0004)	0.0000 (0.0000)	0.9902* (0.0055)
L.trust	-0.0215*** (0.0061)	-0.0604 (0.0636)	-0.0071 (0.0280)	0.9787*** (0.0060)	0.9414 (0.0599)	0.9929 (0.0278)
L.gdppc	-0.0000*** (0.0000)	-0.0000*** (0.0000)	-0.0000*** (0.0000)	1.0000*** (0.0000)	1.0000*** (0.0000)	1.0000*** (0.0000)
diff_gdppc	0.0000*** (0.0000)	0.0001*** (0.0000)	0.0001*** (0.0000)	1.0000*** (0.0000)	1.0001*** (0.0000)	1.0001*** (0.0000)
L.diff_gdppc	0.0000*** (0.0000)	0.0001*** (0.0000)	0.0001*** (0.0000)	1.0000*** (0.0000)	1.0001*** (0.0000)	1.0001*** (0.0000)
Constant	-1.0515*** (0.1484)	-3.3514** (1.3065)	-3.1192*** (0.5810)	0.3494*** (0.0518)	0.0350** (0.0458)	0.0442*** (0.0257)
Observations	56,692	56,692	56,692	56,692	56,692	56,692

Standard errors in parentheses for column (1)-(3)

seEform in parentheses for column (4)-(6)

*** p<0.01, ** p<0.05, * p<0.1

The results suggest that net family income per person is a significant factor in explaining mobility choices at a 95% confidence level. Interestingly, the significance of the agricultural hukou indicator diminishes once the income variable is included in the control variables. This may be because the variation in mobility choices explained by the agricultural hukou indicator was already explained by the income variable. To verify this conjecture, I conduct a χ^2 test to compare the effects of the agricultural hukou indicator in Table 8 and Table 10. The result suggests that the effects in two estimations are not significantly different. Statistically speaking, the effect of agricultural hukou indicator does not change significantly. Therefore, the lack of significance of the agricultural hukou indicator in Table 10 may be due to the income variable providing a better explanation of the variation in mobility choices.

4.3 Intention of Moving Hukou Registered Location

The estimates of the model are displayed in Appendix VIII. At the confidence level of 95%, the result indicates that there exists a selection bias ($\chi^2 = 7.56, p = 0.006$) given the current-period personal and household characteristics, such as hukou type, net family income per person, homeownership, and housing cost. The correlation coefficient is estimated at $\rho = 0.5$. It is worth

noticing that the intention to move with transfer of hukou registered location is 17.2 percentage higher if the individual holds an agricultural hukou. Based on the result from the Heckman selection model, there are 424 out of 98,177 observations being identified as selected. This means that there are 220 more individuals who intended to move across provinces and transfer the hukou registered location but not doing so, which provides potential evidence on the influence of hukou system. To further examine whether the hukou system leave an impact, especially a deterring one, on the moving choice across province, it might be necessary to identify the moving direction, that is, whether the move is from a small city to a big city or to leave a big city for a small city. This is hard to achieve due to the limitation of the dataset I am employing.

4.4 Hukou Type Changes

In the following table, estimations of hukou type changes are summarized. Each choice is set as a base level once to comparison with the other two.

Table 11. Estimated Relative Risk Ratio Results for Hukou Type Changes

VARIABLES	(1) Changed to Agri. Base = 0	(2) Changed to Urban Base = 0	(3) Not Changed Base = 1	(4) Changed to Urban Base = 1	(5) Not Changed Base = 2	(6) Changed to Agri. Base = 2
age	0.940*** (0.0108)	0.942*** (0.00995)	1.064*** (0.0122)	1.003 (0.0154)	1.061*** (0.0112)	0.997 (0.0153)
age2	1.065*** (0.0118)	1.055*** (0.0110)	0.939*** (0.0104)	0.990 (0.0148)	0.948*** (0.00989)	1.010 (0.0151)
L.female	0.926 (0.0628)	0.949 (0.0585)	1.080 (0.0733)	1.025 (0.0930)	1.054 (0.0651)	0.976 (0.0885)
L.own	0.931 (0.0882)	0.643*** (0.0502)	1.075 (0.102)	0.691*** (0.0838)	1.555*** (0.121)	1.447*** (0.175)
L.urban	1.575*** (0.116)	1.528*** (0.103)	0.635*** (0.0469)	0.970 (0.0961)	0.655*** (0.0441)	1.031 (0.102)
L.eduy	1.009 (0.00864)	1.008 (0.00794)	0.991 (0.00849)	0.999 (0.0115)	0.992 (0.00781)	1.001 (0.0115)
L.finc_pc	0.995* (0.00295)	0.995* (0.00266)	1.005* (0.00298)	1.000 (0.00395)	1.005* (0.00268)	1.000 (0.00395)
L.house_exp	1.000 (0.00162)	1.002 (0.00112)	1.000 (0.00162)	1.002 (0.00195)	0.998 (0.00112)	0.998 (0.00195)
L.mortage	0.999 (0.00191)	1.001 (0.000869)	1.001 (0.00191)	1.002 (0.00209)	0.999 (0.000868)	0.998 (0.00208)
L.trust	1.002 (0.0154)	1.003 (0.0141)	0.998 (0.0153)	1.001 (0.0206)	0.997 (0.0140)	0.999 (0.0206)
L.gdppc	1.000*** (2.09e-06)	1.000*** (1.97e-06)	1.000*** (2.09e-06)	1.000 (2.84e-06)	1.000*** (1.97e-06)	1.000 (2.84e-06)

diff_gdppc	1.000 (5.94e-06)	1.000 (5.66e-06)	1.000 (5.94e-06)	1.000 (8.11e-06)	1.000 (5.66e-06)	1.000 (8.11e-06)
L.diff_gdppc	1.000 (7.51e-06)	1.000 (7.27e-06)	1.000 (7.51e-06)	1.000 (1.03e-05)	1.000 (7.27e-06)	1.000 (1.03e-05)
Move:	1.054	1.204**	0.948	1.142	0.830**	0.876
Local Move	(0.107)	(0.105)	(0.0966)	(0.152)	(0.0727)	(0.116)
Move:	1.494	8.91e-07	0.670	5.96e-07	1.123e+06	1.677e+06
Non-Hukou Move	(1.089)	(0.000738)	(0.488)	(0.000494)	(9.300e+08)	(1.389e+09)
Move:	0.677	0.946	1.478	1.398	1.057	0.715
Hukou Move	(0.318)	(0.351)	(0.694)	(0.829)	(0.392)	(0.424)
Constant	0.0359*** (0.0120)	0.0791*** (0.0237)	27.86*** (9.303)	2.204* (0.976)	12.64*** (3.790)	0.454* (0.201)
Observations	58,215	58,215	58,215	58,215	58,215	58,215

seEform in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The result show that, the fixed effect of local move is significant at 95% confidence level in the choices of changing hukou type to urban area. This means that movers who move within the province are more likely to change their hukou type from agricultural hukou to urban hukou compared to those non-movers, which makes sense but the important result here is that the lack of significance of other two types of movers. This finding might imply that the majority of people who change their agricultural hukou to urban hukou are local movers within the province, while inter-province movers are less motivated to change their hukou type. In addition, people living in the urban area are more likely to change their hukou type in both directions. It is less surprising to the find of the change from agricultural hukou to urban hukou. Further research could help shed light on the reasons behind the other findings.

4.5 Housing Expenditure

The estimates for housing expenditure are reported in the following table. Column (1) is the OLS estimates, column (2) reports the OLS estimates with household fixed effect, and column (3) reports the estimates of Heckman selection model, which considers the correlation between the housing expenditure and moving choices.

Table 12. Housing Expenditure Estimations

VARIABLES	(1) OLS	(2) OLS with Household Fixed Effect	(3) Error Correction with Binary Move
house_exp			
house_cost	-0.0054*** (0.0009)	-0.004*** (0.001)	-0.0804*** (0.0099)
cpi_infla	7.0864** (3.4544)	42.254*** -1.562	-40.6946*** (13.3316)
finc_pc	0.0721*** (0.0061)	0.020*** -0.006	0.1155*** (0.0201)
size	0.0126*** (0.0009)	0.017*** -0.002	0.0158*** (0.0043)
agr_hukou	-0.5275* (0.3005)	-0.223 -0.36	-2.5400** (1.0139)
urban	1.7991*** (0.2640)	1.286*** -0.468	2.7642*** (0.9898)
own	-0.4564 (0.3285)	1.017*** -0.306	-0.8538 (1.0316)
Constant	1.7745 (4.2648)	-36.029* (19.798)	60.3337*** (16.1368)
Province FE	Yes	Yes	Yes
Household FE	No	Yes	No
athrho			-0.1105** (0.0445)
lnsigma			3.4638*** (0.0098)
Observations	49,930	84,503	58,126

Note: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The significant effect of agricultural hukou type vanishes when the household fixed effect is included. This may indicate that there exists some unobserved information that is correlated with the agricultural hukou variable that is not considered in the simple OLS estimation. The χ^2 test following the Heckman selection model indicates that there exists significant selection bias ($\chi^2 = 4.87, p = 0.0274$) and the correlation coefficient is $\rho = -0.110004$. The estimated standard error σ is 31.93772 and thus the selectivity effect λ is -3.513278. That is to say, the housing expenditure changes as the moving choices changes. Taking the selection bias into account, the influence of housing maintenance cost and agricultural hukou are significant. The housing expenditure decreases 2540 yuan on average for agricultural hukou holders, ceteris paribus, which

is 2012.5 yuan more compared to the OLS estimates. This result reveals the connection between the moving choices and housing expenditures under the influence of the hukou system.

5. Hypothesis Tests and Robustness Tests

After obtaining the estimations for moving choices and housing prices based on the set of information from individual, family, and province, I am interested in whether the variables, especially those related to the hukou system, affect the moving choices in the same way statistically. In this section, I will conduct several hypothesis testings in order to examine the effect of variables of my interest in the moving choices.

5.1 Effect Equality in Probit Models

A joint test for whether the effect of a variable in all four types of moving choices are equal in a χ^2 test, with an exclusion of the binary moving choices variable. The hypothesis setup is as follows. The null hypothesis is that the effect of x on four types of moves are the same; alternative hypothesis is that at least one effect of variable x is not equal to the effect in other three estimations:

$$H_0: \beta_{xy_1} = \beta_{xy_2} = \beta_{xy_3} = \beta_{xy_4}$$

$$H_1: \text{not}(\beta_{xy_1} = \beta_{xy_2} = \beta_{xy_3} = \beta_{xy_4})$$

The results of the statistical tests indicate that the influence of the agricultural hukou type on different types of moving choices is not homogeneous, as the test statistics is 6.34 and the p-value is 0.0963, allowing for rejection of the null hypothesis at a 90% confidence level. Further pairwise comparisons are conducted to investigate whether the influence on a subset of the moving choices is different. Similarly, the test statistics for the urban area indicator is 8.83, and the p-value is 0.0316, allowing for rejection of the null hypothesis that the effects are the same at a 90% confidence level. A summary of all the test results is presented in Table X.1 in Appendix X.

5.2 Equality of Effect in Multinomial Logit Model

Considering multinomial logit model of moving choices in section 4.2, I teste again for the effect equality of the variable on different types of moving choices, including the base level choice – not moved. The difference between this test and the test in section 5.1 is the exclusion of controlling for province fixed effects, and the interaction effect between the province and hukou type. I conduct a pairwise comparison in the effect for two types of moves. The setup for the Z-test is as follows, where $i \neq j$ and $i, j = 1, 2, 3, 4$:

$$H_0: \beta_{xy_i} = \beta_{xy_j};$$

$$H_1: \beta_{xy_i} \neq \beta_{xy_j}.$$

Table 13 below displays the effect comparison of hukou type variable. The first column of comparison between move choice y_1 versus move choice y_2 indicates the estimated coefficient in the multinomial logit model with y_2 set as the base level of moving choice. Column “e^b” refers to the RRR estimates mentioned in section 4. As the results indicated, I reject the null hypothesis at 90% confidence level that the effects of the agricultural hukou are the same between the choice of not moving, local move, and inter-province move with change in hukou location. The test suggests that the difference between no move and inter-province moves without hukou location change is not significant. Consequently, the effects of the agricultural hukou for four types of moving choices are not jointly same. Similar patterns are observed in the test for the three-level moving choices model. All effects are significantly different from zero at 90% confidence level.

Table 13. Z-test Results of Lagged Agricultural Hukou Indicator

Comparisons			b	z	P>z	e^b	e^bStdX
Panel A. Four-Level Moving Choices							
Not Move	vs	Local Move	0.0637	1.792	0.073	1.066	1.029
Not Move	vs	Non-Hukou Move	0.1130	0.349	0.727	1.120	1.052
Not Move	vs	Hukou Long Move	-0.2872	-1.835	0.066	0.750	0.879
Local Move	vs	Not Move	-0.0637	-1.792	0.073	0.938	0.972
Local Move	vs	Non-Hukou Move	0.0493	0.152	0.880	1.050	1.022
Local Move	vs	Hukou Long Move	-0.3509	-2.205	0.027	0.704	0.854
Non-Hukou Move	vs	Not Move	-0.1130	-0.349	0.727	0.893	0.951

Non-Hukou Move	vs	Local Move	-0.0493	-0.152	0.880	0.952	0.978
Non-Hukou Move	vs	Hukou Long Move	-0.4002	-1.150	0.250	0.67	0.836
Hukou Long Move	vs	Not Move	0.2872	1.835	0.066	1.333	1.138
Hukou Long Move	vs	Local Move	0.3509	2.205	0.027	1.420	1.171
Hukou Long Move	vs	Non-Hukou Move	0.4002	1.150	0.250	1.492	1.197
Panel B. Three-Level Moving Choices							
Not Moved	vs	Hukou Not Moved	0.0656	1.853	0.064	1.068	1.030
Not Moved	vs	Hukou Moved	-0.3022	-1.928	0.054	0.739	0.873
Hukou Not Moved	vs	Not Moved	-0.0656	-1.853	0.064	0.936	0.971
Hukou Not Moved	vs	Hukou Moved	-0.3678	-2.310	0.021	0.692	0.848
Hukou Moved	vs	Not Moved	0.3022	1.928	0.054	1.353	1.145
Hukou Moved	vs	Hukou Not Moved	0.3678	2.310	0.021	1.445	1.180

Note: b is the raw coefficient, z is the z -score for test of $b=0$, $P>|z|$ is the p -value for z -test, $e^{b} = \exp(b)$ is the factor change in odds for unit increase in X , and $e^{bStdX} = \exp(b*SD \text{ of } X)$ is the change in odds for SD increase in X .

It is worth mentioning that although the result for GDP per capita and its first difference are significantly different from zero, the raw estimate in the multinomial logit model is extremely close to zero. Thus, the effect of GDP per capita and its first difference appear to have the same impact on the moving choices.

5.3 Placebo Test in Multinomial Logit Model

In order to assess the robustness of the model results, particularly with respect to the impact of agricultural hukou indicator, I conducted a placebo test using a randomly generated variable distributed uniformly on the interval $[0,1]$. I anticipate that the effect of this placebo variable will be insignificant, thereby indicating that the distribution of hukou type is not random. Specifically, observations with values greater than or equal to 0.5 are assigned to hold the random agricultural hukou, while the remaining observations are assigned to hold the urban hukou.

Table 14. Placebo Test Results IN Multinomial Logit Model

VARIABLES	(1)	(2)	(3)
	Local Move	Non-Hukou Location Move	Hukou Location Move
age	0.982*** (0.00492)	0.952 (0.0547)	0.955** (0.0204)
age2	1.009* (0.00505)	0.988 (0.0687)	1.016 (0.0235)
L.female	1.005 (0.0270)	1.125 (0.286)	0.704*** (0.0804)
L.own	0.709*** (0.0262)	0.675 (0.223)	0.677*** (0.103)
L.urban	1.123***	1.246	1.039

	(0.0323)	(0.354)	(0.128)
L.rand_agr_hukou	1.022	1.082	1.034
	(0.0268)	(0.272)	(0.116)
L.eduy	1.009***	1.046	1.007
	(0.00338)	(0.0363)	(0.0148)
L.finc_pc	1.002***	1.004	1.004**
	(0.000771)	(0.00623)	(0.00201)
L.house_exp	1.000	1.003	1.000
	(0.000601)	(0.00339)	(0.00254)
L.mortgage	1.001*	0	0.989
	(0.000456)	(0)	(0.0108)
L.trust	0.980***	0.945	1.007
	(0.00588)	(0.0557)	(0.0267)
L.gdppc	1.000***	1.000***	1.000***
	(9.29e-07)	(1.20e-05)	(4.10e-06)
diff_gdppc	1.000***	1.000***	1.000***
	(2.61e-06)	(5.87e-06)	(2.85e-06)
L.diff_gdppc	1.000***	1.000***	1.000***
	(3.45e-06)	(1.35e-05)	(6.07e-06)
Constant	0.336***	0.0258***	0.0540***
	(0.0468)	(0.0346)	(0.0302)
Observations	58,215	58,215	58,215

seEform in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14 presents the result from the placebo test. The lack of significance of the placebo implies the and the conclusions drawn in residential mobility models are robust. To avoid the influence of seeds in Stata’s random number generator, I repeat the process and confirm the results are consistent.

6. Conclusion

This thesis examines the impact of the hukou system on residential mobility and housing expenditure. Specifically, it focuses on the influence of hukou type and registered location on these outcomes using data from the China Family Panel Study. Moving choice models are constructed to analyze changes in residence and hukou registration. They reveal that agricultural hukou holders are more likely to move across provinces with changing their hukou registered location, but less likely to move within the same province. Inter-province migrants may face varying frictions in different origin-destination combinations, and this effect may differ by the hukou type. Further

tests demonstrate that the effects of agricultural hukou type are not significantly different across various types of moving choices.

I employ a Heckman selection model in order to correct for the selection bias in two-stage estimations when modeling intention of moving and housing expenditures. In studying intention of moving choices, I found potential deterring impact of agricultural hukou in the inter-province moving choices with transfer of hukou registered location. In studying the housing expenditures of households, the expenditure is found to bias negatively on the moving choices. Moreover, the result for the changes in hukou type implies that the majority of those who change their agricultural hukou to urban hukou are local movers within the province.

Regarding recommendations for future research in this topic, two directions could be considered. The first is to have richer data regarding the city. In this case, the classifications of cities (Liao and Zhang, 2020) can be used to specify the direction of moving and to focus on the transfer in and out of the big cities with higher-quality social benefits. A second is to introduce new variables and information to capture the influence of hukou. Hukou type and registered location are of vital importance in the hukou system but richer information is needed to depict the how hukou functions in Chinese economy.

Reference:

1. All China Marketing Research LLC. (2020, April 12). China Data Online. <https://www-china-data-online-com.ezp2.lib.umn.edu/asp/contact.asp>.
2. Cai, Jifei. "Thoughts on the Reform Direction and Legislation of the Household Registration System in the New Development Stage in China." *China Development Observation*, 2022, pp. 82-85.
3. Chan, Kam Wing. "The Chinese Hukou System at 50." *Eurasian Geography and Economics*, vol. 50, no. 2, 2009, pp. 197–221.
4. Chan, Kam Wing and Li Zhang. "The Hukou System and Rural-Urban Migration in China: Processes and Changes." *China Quarterly*, vol. 160, no. 160, 1999, pp. 818-55.
5. Chen, W. D. "Policy Failure or Success? Detecting Market Failure in China's Housing Market." *Economic Modelling*, vol. 56, 2016, pp. 109–21.
6. Chen, Yu, *et al.* "Valuing the Urban Hukou in China: Evidence from a Regression Discontinuity Design for Housing Prices." *Journal of Development Economics*, vol. 141, Elsevier B.V, 2019, p. 102381–.
7. *China Family Panel Study*. Peking University, [2010-], 2010.
8. Chunyu, Miao David *et al.* "Interprovincial Return Migration in China: Individual and Contextual Determinants in Sichuan Province in the 1990s." *Environment and planning. A*, vol. 45, no. 12, 2013, pp. 2939-58.
9. Christensen, Peter, and Christopher Timmins. "Sorting or Steering: The Effects of Housing Discrimination on Neighborhood Choice." *The Journal of Political Economy*, vol. 130, no. 8, 2022, pp. 2110–63.
10. Deng, Quheng; Gustafsson, Bjorn. "China's Lesser Known Migrants. (in Chinese. With English Summary.)." *Jingji Yanjiu/Economic Research Journal*, vol. 42, no. 4, 2007, p. 137.
11. Greene, W. H. 2018. *Econometric Analysis*. 8th edition. New York: Pearson.
12. Hao, Pu and Shuangshuang Tang. "Floating or Settling Down: The Effect of Rural Landholdings on the Settlement Intention of Rural Migrants in Urban China." *Environment and planning. A*, vol. 47, no. 9, 2015, pp. 1979-99.
13. Henderson, J. Vernon, and Yannis M. Ioannides. "Tenure Choice and the Demand for Housing." *Economica* 53, no. 210, 1986, pp. 231–46.
14. Hertel, Thomas and Fan Zhai. "Labor Market Distortions, Rural–Urban Inequality and the Opening of China's Economy." *Economic Modelling*, vol. 23, no. 1, 2006, pp. 76–109.

15. Ioannides, Yannis M., and Giulio Zanella. "Searching for the Best Neighborhood: Mobility and Social Interactions", 2008.
16. Ito, Junichi. "The Removal of Institutional Impediments to Migration and Its Impact on Employment, Production and Income Distribution in China." *Economic Change and Restructuring*, vol. 41, no. 3, 2008, pp. 239-65, ABI/INFORM Collection; Social Science Premium Collection.
17. Jiang, Ying *et al.* "Rural Migrant Workers' Intentions to Permanently Reside in Cities and Future Energy Consumption Preference in the Changing Context of Urban China." Transportation research. Part D, *Transport and Environment*, vol. 52, 2017, pp. 600-18.
18. Kan, Kamhon. "Residential Mobility and Social Capital." *Journal of Urban Economics*, vol. 61, no. 3, 2007, pp. 436–57.
19. Li, Si-Ming and Youqin Huang. "Urban Housing in China: Market Transition, Housing Mobility and Neighbourhood Change." *Housing studies*, vol. 21, no. 5, 2006, pp. 613-23.
20. Liang, Zai and Zhongdong Ma. "China's Floating Population: New Evidence from the 2000 Census." *Population and development review*, vol. 30, no. 3, 2004, pp. 467-88.
21. Liao, Yu, and Junfu Zhang. "Hukou Status, Housing Tenure Choice and Wealth Accumulation in Urban China." *IDEAS Working Paper Series* from RePEc, 2020.
22. Lu, Yilong. "Does Hukou Still Matter? The Household Registration System and Its Impact on Social Stratification and Mobility in China." *Social Sciences in China*, vol. 29, no. 2, 2008, pp. 56–75.
23. Lyu, Xueying. "Car Restriction Policies and Housing Markets." *Journal of Development Economics*, vol. 156, Elsevier B.V, 2022.
24. "Point-based Hukou System"(Translated) *Baidu Baike*, 2022, <https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%E7%A7%AF%E5%88%86%E8%90%BD%E6%88%B7%E5%88%B6/13344255>.
25. Song, Yang. "Cost-Benefit Analysis of the Hukou Reform: Simulation Evidence from a Theoretical Labor Market Model." *China Economic Quarterly International*, vol. 1, no. 2, Elsevier B.V, 2021, pp. 109–19.
26. Tang, Qiong, and Jing Wu. "Change and Reform of Household Registration System Since the Founding of New China and Enlightenment from Past Experience." *Construction of Old Area (Laoqujianshe)*, vol. 3, 2022, pp. 25–31.
27. Wang, Junhui, Shuang Ai, and Mian Huang. "Migration History, Hukou Status, and Urban Household Consumption." *Economic modelling*, 97 (2021): 437–448.

28. Wu, Weiping. "Sources of Migrant Housing Disadvantage in Urban China." *Environment and Planning A: Economy and Space*, vol. 36, no. 7, 2004, pp. 1285-304.
29. Xie, Yu *et al.*, "China Family Panel Studies User's Manual", Version 3, 2017.
30. Xing, Chunbing, and Junfu Zhang. "The Preference for Larger Cities in China: Evidence from Rural-Urban Migrants." *China economic review*, 43 (2017): 72–90.
31. Xu, Zhihua *et al.* "Return Migration in China: A Case Study of Zhumadian in Henan Province." *Eurasian Geography and Economics*, vol. 58, no. 1, 2017, pp. 114-40.
32. Zhang, Jipeng *et al.* "Return Migration and Hukou Registration Constraints in Chinese Cities." *China Economic Review*, vol. 63, 2020, p. 101498.
33. Zheng, Siqi, *et al.* "Housing-Choice Hindrances and Urban Spatial Structure: Evidence from Matched Location and Location-Preference Data in Chinese Cities." *Journal of Urban Economics*, vol. 60, no. 3, 2006, pp. 535–57.
34. Zhou, Min and Wei Guo. "Housing Characteristics and Health in Urban China: A Comparative Study of Rural Migrants and Urban Locals." *Population Space and Place*, vol. 29, no. 1, 2023.

Appendix I. Tables of Survey Information

Table A1. Main Content of Community questionnaire of CFPS 2010 and CFPS 2014

Module	Questionnaire Content
Infrastructure	type of village/urban neighborhood, respondent's position, facilities, bulletin, land borders, administrative area, water sources, fuels
Population Structure	total number of households, population, registered residents, permanent residents, migrant population, age structure, fertility and mortality, ethnic groups
Economy	basic living allowances, price level
Administration	type of administrative staff, working conditions, neighborhood transportation
History & Politics	historical changes, whether it is a tourist area, whether there are high-pollution enterprises, latest election of village/urban neighborhood's committee
Housing Price	highest price in history, highest last month, average price last month
Environment & Transportation	transportation hours to the closest county, town, provincial capital, mineral resources, natural disasters, land resources
Labor Economy	labor force structure, gross output of agriculture & non-agriculture sectors, net personal income per capita, price of assisting workers, distribution of the "big family names"
HealthCare& Fertility	area of medical station, number of medical stall, progress of rural cooperative medical services, one-child policy
Finance	Collectively-owned enterprises and their output total fiscal revenue and its resources, total financial expenditure and the items
Respondent's information	gender, age, political status, level of education of the respondent, information of the director of the village/urban neighborhood, names of professions of other respondents
Interviewer' Observations	economic status, tidiness of roads, mental state of villagers/neighborhood, the homogeneity of villagers/neighborhood, architecture pattern, congestion, type of village/neighborhood, topography, features of respondents

Note: replicated from Table 4 of user guide of China Family Penal Study (CFPS)

Table A2. Main Content of the Family Questionnaire of CFPS 2010

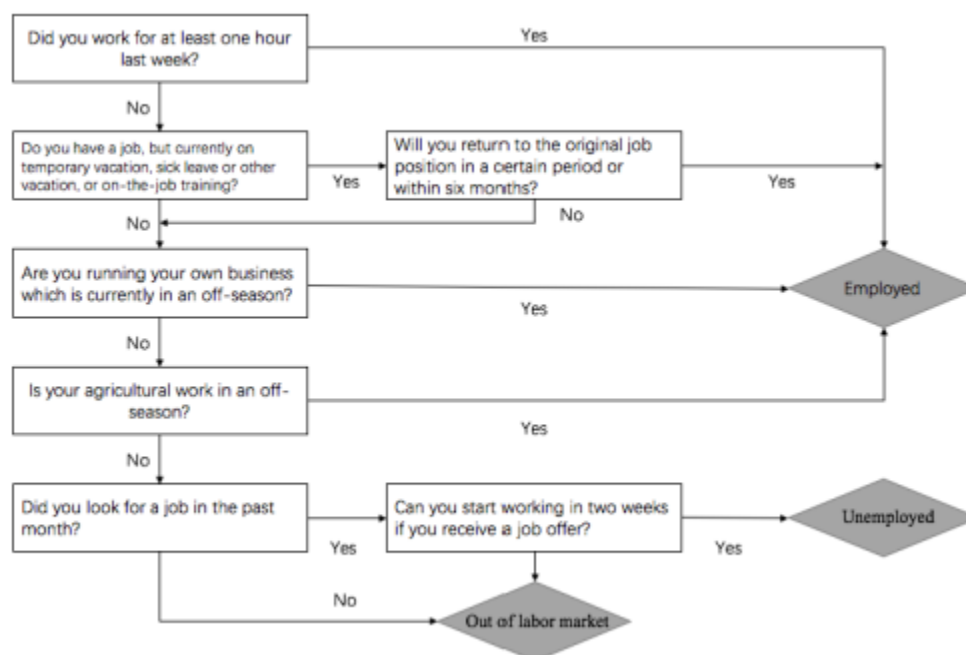
Module	Content
Geography and transportation	Nearest public transportation, medical stations, high schools, commercial centers
Living conditions	Use of water, fuel resources, electricity, bathroom conditions, garbage disposal, employment of housemaids
Social interactions	Spring Festival visits, gift giving, family-lineage, ancestor worship, neighborhood interactions, communications with relatives
Housing	House ownership, self-built houses or commercial apartments, houses for rent, house area, living time, market value and rent of the houses, the apartment structures, other estates, difficulty in housing
Management conditions	<u>U module for working outside</u> (people who are outside, working address, time devoted, whether they go home during vacations, the transfer payment, whether their family has increased/decreased assisting workers because they are working outside) , the governments support, reason for poverty. <u>V module for non-agricultural management</u> (type of non-agricultural industry, number of participants, total assets, shares held by family members, number of employees, turnover, after-tax profits), houses for rent, land and other means of production for rent property for sell, demolition of the houses, land acquisition
Family income	Savings, financial products, pensions/social security/basic living allowances, salaries/rewards/allowances/bonuses, non-salary/agricultural income, value of gifts
Family assets	Insurance indemnity, others' debts, value of collections, present value of other assets
Family expenditure	Highest expenditure, loans, daily expenditure (food, travel, communications, etc.) .special expenditure(family appliances, medical care, education, commercial insurances) , donations, total expenditure
Durable goods	Cars, motorcycles, tractors, televisions
Agricultural production	Land type, land areas, revenue and expenditure, types of corps, output, sales, income, types of domestic animals and fishing, output, sales, income, raising conditions of domestic animals
Interviewer Observations	Respondent's housing situation, tidiness of the family, mental state of their family members, relations of family members, relations between elderly and the young, relations between genders, personal characteristics of the respondent

Note: replicated from Table 5 of user guide of China Family Penal Study (CFPS)

Table A3. Main Content of the Adult Questionnaire of CFPS 2010

Module	Content	Question type
Basic information	Date of birth, birthweight, birthplace, residence. Hukou, ethnicity, family category during the Cultural Revolution, political party and organizational affiliations, time spent living with parents before age 3 and between ages 4 and 12	Baseline questions
Siblings	Number of siblings, name, date of birth, alive or not age at and cause of death, marital status, educational attainment, occupation, administrative/managerial positions, residence, living with parents, parents' cause of death	(used only in CFPS 2012)
Educational history	Educational attainment, type of schools at different educational levels, time spent at school, when did school end, name of schools, graduated or not, subject and major, educational expectation	Baseline questions
Language use	Importance of different languages, language spoken at home	Core questions
Schooling (shared)	Current level of education, type of school, major, grades, extracurricular tutoring, student activities, subjective evaluations of study and school, educational expectation	Core questions
Marriage	Current marital status (married/remarried/cohabitating/divorced/widowed) date of birth of the present/last/first spouse/cohabitation partner, time at marriage/cohabitation, pre-marital cohabitation, channel of initial contact, reason for the divorce ending the last/first marriage	Core questions
Relations with children	Evaluation of the relations with children by those aged 60 or older, intergenerational transfers	Core questions
Work	See Figure 13 (<i>attached below</i>)	Core questions
Personal income	Non-operating income, operating income, financial support from relatives and friends, government subsidies	Core questions
Time use (<i>shared module</i>)	Life, work, study and training, leisure and social activities, commuting	Core questions
Leisure	Leisure activities and frequencies, means of travel, overseas experiences	Core questions
Mobile phone and internet (shared module)	Use of mobile phones, QQ, MSN, e-mails, importance of internet, frequency and places of internet use	Core questions

Social relations	Help seeking, confiding troubles, self-reported social status	Core questions
Subjective measurements	Values, social attitudes, achievement scale, life	Core questions
Politics	Experiences of thefts or robbery, unjust treatments, media interests, evaluation of government work	Core questions
Health	Height, weight, self-reported health, discomfort, chronic diseases, experiences of hospitalization, medical expenses, coping with diseases, satisfaction with healthcare, traditional Chinese medicine, physical exercises, diet, P-ADL, smoking and drinking, sleeping, memory, depression scale, main healthcare-giver	Core questions
Mental health	K 6 Scale, CESD Scale	Rotation questions
Cognitive assessments	Literacy/vocabulary, math, word recall, number series	Rotation questions
Personal information and observations of the interviewers	Contact information, respondent, personal characteristics of respondent	Core questions



Note: replicated from Table 8 and Figure 13 of user guide of China Family Penal Study (CFPS)

Table A4. Added questions in CFPS adult questionnaire in follow-up survey

Module	Content	Question type
Added in CFPS 2012		
Information about deceased siblings	Education level and occupation of siblings who died before the baseline survey	(Only in CFPS 2012)
Pension insurance	Status of participation, fees, and benefits of various kinds of pension insurances	Extension questions
Fertility intentions	Ideal number of children	Core questions
Trust	Trust towards different types of people	Core questions
Religion	Religious beliefs and participation in religious activities	Core questions
Anchoring vignettes for health assessment and social status	Assess the health condition and social status of Core hypothetical cases	Core questions
Added in CFPS 2014		
Parent information	Parents' birth year, parents' occupation, and political affiliation when the respondent was age 14	Rotation questions
Law module	Law module	Extension questions
[EHC-RESI]	EHC Migration module	Core questions

Note: replicated from Table 9 of user guide of China Family Panel Study (CFPS)

Appendix II. Supplementary Explanation on Construction of Some Variables

For the convenience of analysis of moving choices, I created dummy variables *local_move* and *interprov_move* that identify the community change and province change of families.

Following this logic, we can distinguish three types of moving: intra-provincial moving that moved within the province (*local_move* = 1 and *interprov_move* = 0), inter-provincial moving that moved across provinces (*local_move* = 1 and *interprov_move* = 1), and those did not move (*local_move* = 0 and *interprov_move* = 0).

The variable of housing expenditure (*house_exp*) includes the rent of the house if any. But this will cause problems in the model of housing expenditure (section 3.5). Thus, I take away the rent from the housing expenditure.

Appendix III. Variable Definition

Variable Name	Definition
pid	Personal ID
fid	Family ID
cid	Community ID
provcid	Province ID
year	Year of Interview (float type)
wave	wave of survey (2010, 2012, 2014, 2016, 2018)
countyid	County ID
subpopulation	Indicator of 'large' vs 'small' province subsample
urban	Urban area (Census Bureau's definition)
hukou	Place of current Hukou
type_change	Multinomial variable of whether Hukou type changed (agricultural or urban)
to_agr	Whether changed hukou type to agricultural
to_urban	Whether changed hukou type to urban
income	Personal income (Thousand Yuan)
eduy	Years of education
house_debts	Total amount of mortgage for all houses (Thousand Yuan)
house_exp	Expenditure on housing: Adjusted (Thousand Yuan)
House_cost	Financial cost on housing (10 thousand Yuan)
mortgage	Mortgage on housing
trust	Degree of trust in neighbors (2010 missing)
age	Age
age2	Age Square/100
finc_pc	Net family income per capita (Thousand Yuan)
agr_hukou	Whether agricultural hukou type
married	Whether married
own	Whether own the current residence (by any family member)
comm_change	Whether community (cid) change
prov_change	Whether province (provcid) change
hukou_change	Whether location of hukou (hukou) change
local_move	Whether move across community but within province
long_move	Whether move across province
phy_move	Whether the inter-province move WITHOUT changing location of hukou
hukou_move	Whether the inter-province move change location of hukou
bi_move	Whether move (in any type of moving choice)
rationed_move	Whether able to move inter-province with change in hukou location
pop	Provincial population in 10 thousand (Source: China Data Online)
gdp	Provincial population in 100 million Yuan (Source: China Data Online)
gdppc	Provincial GDP per capita in 100 million Yuan (constructed from China Data Online)
diff_gdppc	Difference in GDP per capita of consecutive year in residence province
size	Building area of the house
size_consist	Whether there have been modifications on living space (2014-2018)

resivalue	Constructed variable: value of current residence (10 thousand yuan)
female	Whether gender is female
move	Multinomial Variable of Moving Choice

Appendix IV. Variables Availabilities by Year

Variable	2010	2012	2014	2016	2018
female	33526	35,342	35663	33192	30548
subpopulation	33526	35342	0	33192	30548
urban	33526	35113	33004	32965	30225
hukou	33519	35323	35646	33192	28294
hukou_change	33526	35342	35663	33192	30548
income	31896	34162	19308	9695	11666
eduy	33523	35300	35614	29584	28662
edu	33523	35300	35614	31573	30548
house_debts	33373	35196	35248	32626	30064
resivalue	33391	35287	35663	32731	30055
size	32042	33796	28477	24113	21158
house_exp	33156	34260	34563	32731	29551
mortgage	33356	34917	35663	32641	30023
trust	0	31769	31353	33129	30248
age	33526	35342	35663	33192	30548
age2	33526	35342	35663	33192	30548
finc_pc	31539	34462	33641	31978	30246
period	33526	35342	35663	33192	30548
agr_hukou	33526	35342	35663	33192	30548
married	33526	35342	35663	33192	30548
own	33526	35342	35663	33192	30548
comm_change	0	35342	35663	33192	30548
prov_change	0	35342	35663	33192	30548
type_change	0	26168	28361	26700	24756
local_move	0	35342	35663	33192	30548
long_move	0	35342	35663	33192	30548
phy_move	0	35342	35663	33192	30548
hukou_move	0	35342	35663	33192	30548
pop	33526	35342	35663	33192	30548
gdp	33526	35342	35663	33192	30548
gdppc	33526	35342	35663	33192	30548

Appendix V. Value Explanation of Categorical Variables

Table V.1. Value Explanation of *subpopulation*

Indicator of 'large' vs 'small' province subsample	Value	Freq.	Percent	Cum.
Shanghai subpopulation	1	9,233	6.96	6.96
47uangdon subpopulation	2	11,807	8.9	15.87
henan subpopulation	3	15,813	11.92	27.79
gansu subpopulation	4	16,706	12.6	40.39
47uangdong subpopulation	5	12,014	9.06	49.45
other province subpopulation	6	67,035	50.55	100

Table V.2. Value Explanation of *edu*

Stage of education	Value	Freq.	Percent	Cum.
Illiterate/Semi-literate	1	40,889	24.55	24.55
Primary school	2	34,819	20.91	45.45
Junior high school	3	48,920	29.37	74.83
Senior high school	4	24,804	14.89	89.72
2- or 3-year college	5	8,916	5.35	95.07
4-year college/Bachelor's degree	6	5,879	3.53	98.6
Master's degree	7	423	0.25	98.85
Doctoral degree	8	20	0.01	98.87
No need to go to school	9	1,888	1.13	100

Table V.3. Value Explanation of *trust*

Degree of trust in neighbors (missing in 2010)	Freq.	Percent	Cum.
0	1,593	1.26	1.26
1	1,106	0.87	2.13
2	2,242	1.77	3.91
3	4,668	3.69	7.6
4	3,999	3.16	10.76
5	33,099	26.17	36.92
6	14,306	11.31	48.23
7	15,363	12.14	60.38
8	28,039	22.17	82.54
9	6,940	5.49	88.03
10	15,144	11.97	100

Table V.4. Value Explanation of *hukou_consist*

Ownership of current residence	Value	Freq.	Percent	Cum.
Property right solely owned by the family	1	143,901	86	86
Property right partly owned by the family	2	3,642	2.18	88.18
Public house (gong fang) provided by work	3	3,660	2.19	90.37
Cheap rental house	4	1,138	0.68	91.05
Public rental house	5	1,383	0.83	91.87
Commercial house rentable in the market	6	5,525	3.3	95.17
Friends or relatives	7	4,153	2.48	97.66
Other	77	3,922	2.34	100

Table V.5. Value Explanation of *marital_status*

Current marital status	Value	Freq.	Percent	Cum.
Never married	1	23,992	14.26	14.26
Married (having a spouse)	2	131,057	77.89	92.15
Cohabitation	3	684	0.41	92.56
Divorced	4	2,694	1.6	94.16
Widowed	5	9,830	5.84	100

Table V.6. Value Explanation of *female*

Gender	Freq.	Percent	Cum.
Male	83,366	49.54	49.54
Female	84,905	50.46	100

Table V.7. Value Explanation of *move*

Multinomial Variable of Moving Choice	Freq.	Percent	Cum.
Not Move	98,136	72.83	72.83
Local Move	35,643	26.45	99.28
Non-Hukou Long Move	202	0.15	99.43
Hukou Long Move	764	0.57	100

Table V.8. Value Explanation of *type_change*

Whether Hukou type changed (agricultural or urban)	Freq.	Percent	Cum.
Not Changed	101,904	96.15	96.15
Changed to Agricultural Hukou	1,626	1.53	97.68
Changed to Urban Hukou	2,455	2.32	100

Appendix VI. Estimates for Probit Model of Binary Moving Choice

VARIABLES	(1) Local Move	(2) Long Move	(3) Non-Hukou Location Move	(4) Hukou Location Move	(5) Binary Move
age	-0.0108*** (0.0027)	-0.0144* (0.0078)	-0.0216 (0.0153)	-0.0069 (0.0086)	-0.0131*** (0.0027)
age2	0.0059** (0.0027)	0.0036 (0.0079)	0.0012 (0.0177)	-0.0019 (0.0087)	0.0075*** (0.0026)
L.own	-0.2049*** (0.0211)	-0.1270** (0.0630)	-0.1859** (0.0945)	-0.0882 (0.0695)	-0.2085*** (0.0209)
L.urban	0.0700*** (0.0173)	0.0548 (0.0475)	-0.0766 (0.0852)	0.0640 (0.0502)	0.0708*** (0.0170)
L.agr_hukou	0.0131 (0.0209)	0.0146 (0.0604)	-0.0613 (0.1051)	0.0162 (0.0662)	0.0237 (0.0206)
L.eduy	0.0058*** (0.0019)	0.0020 (0.0058)	0.0154 (0.0109)	0.0002 (0.0064)	0.0065*** (0.0019)
finc_pc	0.0012*** (0.0003)	0.0007 (0.0008)	0.0010 (0.0007)	0.0001 (0.0011)	0.0013*** (0.0003)
L.finc_pc	0.0006 (0.0004)	0.0011 (0.0008)	0.0012 (0.0010)	0.0010 (0.0008)	0.0006 (0.0004)
L.house_exp	0.0005 (0.0003)	-0.0008 (0.0011)	0.0001 (0.0016)	-0.0011 (0.0013)	0.0005* (0.0003)
L.mortgage	0.0003 (0.0002)	-0.0039* (0.0020)	-475.6702* (250.8702)	-0.0016 (0.0013)	0.0002 (0.0002)
L.trust	-0.0078** (0.0033)	-0.0125 (0.0100)	-0.0330* (0.0190)	-0.0063 (0.0104)	-0.0076** (0.0032)
L.gdppc	-0.0000*** (0.0000)	0.0000*** (0.0000)	-0.0000*** (0.0000)	0.0000*** (0.0000)	-0.0000*** (0.0000)
diff_gdppc	-0.0000*** (0.0000)	-0.0000 (0.0000)	-0.0001*** (0.0000)	0.0000 (0.0000)	0.0000*** (0.0000)
L.diff_gdppc	0.0000*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
12.provcd	-0.4480*** (0.1232)	-1.1657 (0.7479)	-3.1344*** (1.1635)	-0.3122 (0.7690)	-0.1410 (0.4185)
13.provcd	-1.2482*** (0.1301)	-2.5425** (1.0066)	-8.7069*** (1.8126)	-0.5986 (1.0414)	-0.1423 (0.3760)
14.provcd	-0.8256*** (0.1329)	-2.9337*** (1.1103)	-9.3716*** (2.3112)	-0.9487 (1.1268)	-0.0927 (0.4293)
15.provcd		-0.6652 (0.9829)	-4.6941*** (1.6028)	0.3733 (1.0073)	0.6973 (0.5197)
21.provcd	-0.7341*** (0.1176)	-3.2299*** (1.0855)	-7.2436*** (1.9406)	-1.8956* (1.1015)	-0.3378 (0.4425)
22.provcd	-0.8953***	-3.5041***		-1.6141	-0.2611

	(0.1338)	(1.2740)		(1.3023)	(0.5271)
23.provcd	-0.5046***	-3.6947***	-8.8219***	-1.9563*	-0.1753
	(0.1318)	(1.1477)	(2.1101)	(1.1746)	(0.4490)
31.provcd	0.0015	-0.9490**	-0.7417	-0.9431**	-0.5269*
	(0.0844)	(0.4556)	(0.7914)	(0.4810)	(0.3033)
32.provcd	0.4029***	-1.0398*	-2.6482***	-0.4009	-0.0243
	(0.0937)	(0.5400)	(0.9758)	(0.5532)	(0.2883)
33.provcd	-0.8680***	-0.9425	-5.4704***	0.2193	-0.0252
	(0.1113)	(0.6570)	(1.3852)	(0.6564)	(0.3413)
34.provcd	-0.9125***	-1.8123	-6.9272***	-0.1013	0.2306
	(0.1298)	(1.1593)	(2.0603)	(1.1510)	(0.5426)
35.provcd	-0.5080***	-2.0091**		-0.7878	-0.6316
	(0.1170)	(0.8530)		(0.8675)	(0.4718)
36.provcd	-1.1950***	-4.6453***	-9.3957***	-3.0272**	-1.7886***
	(0.1368)	(1.2028)	(2.0761)	(1.1830)	(0.5308)
37.provcd	-0.3663***	-2.3257**	-6.9665***	-0.7466	-0.0031
	(0.1108)	(0.9322)	(1.7532)	(0.9532)	(0.3938)
41.provcd	-1.0100***	-3.5315***	-9.2911***	-1.6809*	-0.7021*
	(0.1239)	(0.9897)	(1.8535)	(1.0024)	(0.3700)
42.provcd	-0.9875***	-1.4642	-5.8947***	-0.0605	0.0146
	(0.1275)	(0.8986)	(1.6309)	(0.9170)	(0.4451)
43.provcd	-0.3875***	-2.5397**	-7.9556***	-0.8248	0.0897
	(0.1246)	(1.0258)	(2.0833)	(1.0586)	(0.4251)
44.provcd	-0.8207***	-1.3844*	-5.7343***	-0.0445	-0.0369
	(0.1045)	(0.7205)	(1.3652)	(0.7269)	(0.3269)
45.provcd	-0.5156***	-3.0284**	-9.4511***	-1.0488	0.1208
	(0.1356)	(1.3155)	(2.3479)	(1.4106)	(0.5339)
46.provcd		0.4129		2.2300*	1.8077**
		(1.3189)		(1.3406)	(0.8047)
50.provcd	-1.0814***	-0.2396	-5.2640***	0.9871	0.5947
	(0.1489)	(1.0040)	(1.8079)	(1.0517)	(0.5680)
51.provcd	-0.9590***	-2.6549**	-9.4672***	-0.7540	-0.2973
	(0.1266)	(1.1342)	(2.4209)	(1.1555)	(0.5031)
52.provcd	-0.5543***	-2.4254**	-9.3909***	-0.3537	0.3256
	(0.1362)	(1.1531)	(2.2849)	(1.1764)	(0.5061)
53.provcd	-1.0859***	-3.9178***		-1.9569	-0.9399
	(0.1333)	(1.3574)		(1.3705)	(0.5873)
54.provcd		-1.6249		0.2860	0.1854
		(1.1546)		(1.1785)	(0.6072)
61.provcd	-0.6341***	-1.7734*	-7.2209***	-0.0468	0.3138
	(0.1212)	(0.9929)	(1.9283)	(1.0209)	(0.4418)
62.provcd	-0.7255***	-4.2822***	-9.6100***	-2.3599**	-0.4987
	(0.1365)	(1.1591)	(2.0723)	(1.1872)	(0.4135)
63.provcd		0.2230		2.2648	1.5951
		(1.5097)		(1.5118)	(0.9817)
64.provcd		-2.0604*		-0.3074	-0.2472

65.provcd	-0.7408** (0.3245)	(1.1596) -1.2934 (1.0216)	(1.1809) 0.4637 (1.0420)	(0.6854) 0.4316 (0.4570)
12L.provcd		2.7125*** (0.9102)	2.5111*** (0.9347)	-0.3781 (0.4344)
13L.provcd		5.2146*** (1.1641)	6.2965*** (2.0720)	4.4799*** (1.2309)
14L.provcd		5.6170*** (1.2518)	6.7040*** (2.5502)	4.8741*** (1.3155)
15L.provcd		4.7358*** (1.3812)		4.6058*** (1.3856)
21L.provcd		5.1361*** (1.1887)	4.9379** (2.2248)	4.8361*** (1.2227)
22L.provcd		5.9423*** (1.3767)	7.5746*** (2.2018)	5.2733*** (1.4373)
23L.provcd		6.2959*** (1.2793)	6.2374*** (2.3927)	5.7958*** (1.3346)
31L.provcd		0.7240 (0.5739)	1.1512 (0.9630)	0.4852 (0.5993)
32L.provcd		2.1835*** (0.6674)		2.0231*** (0.6807)
33L.provcd		2.6684*** (0.8118)	4.2282** (1.7717)	2.1666*** (0.8218)
34L.provcd		4.2886*** (1.3733)	5.1853** (2.5168)	3.5005** (1.3877)
35L.provcd		3.8846*** (0.9698)	5.2399*** (1.7334)	3.2665*** (1.0052)
36L.provcd		6.9930*** (1.2708)	7.2869*** (2.3476)	6.3163*** (1.2745)
37L.provcd		4.0779*** (1.0475)	5.1248** (2.0271)	3.4030*** (1.0922)
41L.provcd		6.0487*** (1.1653)	7.0628*** (2.2024)	5.3269*** (1.2155)
42L.provcd		3.6624*** (1.0811)	4.6567** (2.0848)	3.0338*** (1.1308)
43L.provcd		5.0142*** (1.1797)	6.1479** (2.3926)	4.3287*** (1.2381)
44L.provcd		3.2694*** (0.8595)	4.8771*** (1.7226)	2.5849*** (0.8832)
45L.provcd		5.5588*** (1.4514)	6.8209** (2.7172)	4.8034*** (1.5500)
51L.provcd		5.0660*** (1.2882)	7.0475*** (2.6753)	4.2917*** (1.3420)
52L.provcd		5.3956*** (1.3860)	6.6604** (2.6842)	4.6182*** (1.4489)
53L.provcd		6.3707***		5.6737***

		(1.4712)		(1.5227)	(0.6175)
54L.provcd		5.9662***		5.4435***	-1.0602
		(1.4515)		(1.5310)	(0.8772)
61L.provcd		4.1784***	5.4191**	3.4436***	-1.1941**
		(1.1883)	(2.2623)	(1.2604)	(0.4775)
62L.provcd		6.9194***	6.7025***	6.2891***	-0.5046
		(1.3297)	(2.4522)	(1.3958)	(0.4586)
64L.provcd		6.0577***		5.5111***	-0.7166
		(1.4832)		(1.5302)	(0.8939)
50L.provcd					-1.9925***
					(0.5974)
65L.provcd					-0.8369
					(0.6511)
Constant	0.4968***	-5.5493***	1.8679	-7.6043***	0.8973***
	(0.1780)	(0.7274)	(1.5463)	(0.8093)	(0.1989)
Observations	56,650	56,244	51,298	56,244	56,691

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix VII. Estimates for Binomial Moving Choice Models with Interaction Term

VARIABLES	(1) Local Move	(2) Long Move	(3) Non-Hukou Location Move	(4) Hukou Location Move	(5) Binary Move
age	-0.0114*** (0.0027)	-0.0144* (0.0079)	-0.0232 (0.0158)	-0.0072 (0.0087)	-0.0137*** (0.0027)
age2	0.0064** (0.0027)	0.0031 (0.0081)	0.0021 (0.0182)	-0.0020 (0.0089)	0.0081*** (0.0027)
L.own	-0.2119*** (0.0214)	-0.1236* (0.0646)	-0.1972** (0.0986)	-0.0825 (0.0711)	-0.2145*** (0.0212)
L.urban	0.0770*** (0.0176)	0.0425 (0.0471)	-0.1055 (0.0862)	0.0516 (0.0498)	0.0762*** (0.0173)
L.agr_hukou	0.1841*** (0.0568)	-0.1358 (0.1532)	-0.4828 (0.2997)	-0.0063 (0.1880)	0.1782*** (0.0559)
L.eduy	0.0068*** (0.0019)	0.0013 (0.0059)	0.0140 (0.0113)	-0.0005 (0.0065)	0.0074*** (0.0019)
finc_pc	0.0013*** (0.0003)	0.0008 (0.0007)	0.0011* (0.0006)	0.0004 (0.0010)	0.0014*** (0.0004)
L.finc_pc	0.0009** (0.0004)	0.0010 (0.0008)	0.0011 (0.0011)	0.0009 (0.0008)	0.0009** (0.0004)
L.house_exp	0.0005 (0.0003)	-0.0008 (0.0011)	0.0001 (0.0014)	-0.0012 (0.0013)	0.0005* (0.0003)
L.mortgage	0.0003 (0.0002)	-0.0039** (0.0018)	-556.3398* (299.6961)	-0.0017 (0.0012)	0.0002 (0.0002)
L.trust	-0.0078** (0.0033)	-0.0129 (0.0101)	-0.0379** (0.0191)	-0.0065 (0.0105)	-0.0079** (0.0032)
L.gdppc	-0.0000*** (0.0000)	0.0000*** (0.0000)	-0.0000*** (0.0000)	0.0000*** (0.0000)	-0.0000*** (0.0000)
diff_gdppc	-0.0000*** (0.0000)	-0.0000 (0.0000)	-0.0001*** (0.0000)	0.0000 (0.0000)	0.0000*** (0.0000)
L.diff_gdppc	0.0000*** (0.0000)	0.0000*** (0.0000)	0.0001*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
12.provcd	-0.3889*** (0.1407)	-1.3187* (0.7469)	-3.4042*** (1.0951)	-0.4161 (0.7697)	-0.2383 (0.4229)
13.provcd	-1.3361*** (0.1573)	-2.7015*** (1.0148)	-8.9548*** (1.7389)	-0.6733 (1.0514)	-0.1474 (0.3761)
14.provcd	-1.0764*** (0.1557)	-3.0743*** (1.1088)	-9.5979*** (2.1762)	-0.9946 (1.1298)	-0.0908 (0.4306)
15.provcd		-0.7887 (0.9855)	-5.0207*** (1.5211)	0.3204 (1.0105)	0.6635 (0.5164)
21.provcd	-0.9388*** (0.1260)	-3.3599*** (1.0893)	-7.5972*** (1.8610)	-1.9419* (1.1020)	-0.3330 (0.4405)
22.provcd	-0.8257***	-3.6602***		-1.6736	-0.2204

	(0.1466)	(1.2723)		(1.3072)	(0.5328)
23.provcd	-0.5263***	-3.8487***	-9.1603***	-2.0249*	-0.1343
	(0.1381)	(1.1532)	(2.0136)	(1.1817)	(0.4502)
31.provcd	-0.1524*	-1.0521**	-0.9139	-1.0227**	-0.5928**
	(0.0903)	(0.4641)	(0.7831)	(0.4892)	(0.2997)
32.provcd	0.7043***	-1.1465**	-2.7553***	-0.4770	-0.0913
	(0.1214)	(0.5482)	(0.9397)	(0.5610)	(0.2870)
33.provcd	-0.9000***	-1.0128	-5.6167***	0.1973	-0.0303
	(0.1674)	(0.6640)	(1.3193)	(0.6645)	(0.3436)
34.provcd	-1.2982***	-1.9249*	-7.1745***	-0.1086	0.2093
	(0.2220)	(1.1625)	(1.9619)	(1.1615)	(0.5429)
35.provcd	0.0156	-2.0801**		-0.8211	-0.6775
	(0.1773)	(0.8683)		(0.8811)	(0.4848)
36.provcd	-1.8100***	-4.7706***	-9.5981***	-3.0709**	-1.8434***
	(0.2409)	(1.2138)	(2.0337)	(1.1956)	(0.5396)
37.provcd	-0.8986***	-2.4203***	-7.3867***	-0.7759	-0.0542
	(0.1423)	(0.9333)	(1.6774)	(0.9558)	(0.3934)
41.provcd	-1.0376***	-3.6249***	-9.5015***	-1.6975*	-0.7173*
	(0.1332)	(0.9965)	(1.7476)	(1.0117)	(0.3706)
42.provcd	-0.8939***	-1.6045*	-6.1905***	-0.1200	0.0177
	(0.1480)	(0.9048)	(1.5454)	(0.9177)	(0.4554)
43.provcd	-0.6920***	-2.5661**	-8.2244***	-0.7813	0.0923
	(0.1369)	(1.0604)	(2.0209)	(1.0948)	(0.4275)
44.provcd	-0.7208***	-1.4815**	-6.0020***	-0.0816	-0.0798
	(0.1142)	(0.7320)	(1.2998)	(0.7387)	(0.3277)
45.provcd	-0.8044***	-3.1622**	-9.7773***	-1.0953	0.1496
	(0.2315)	(1.3150)	(2.2414)	(1.4075)	(0.5399)
46.provcd		0.3064		2.2435	1.8390**
		(1.3578)		(1.3917)	(0.8194)
50.provcd	-1.2077***	-0.4270	-5.5824***	0.8292	0.3335
	(0.2060)	(1.0194)	(1.6931)	(1.0635)	(0.5900)
51.provcd	-0.3882***	-2.7400**	-9.7340***	-0.7634	-0.2380
	(0.1455)	(1.1406)	(2.3606)	(1.1618)	(0.5084)
52.provcd	-0.6987***	-2.4906**	-9.7191***	-0.3147	0.3538
	(0.1847)	(1.1633)	(2.1905)	(1.1874)	(0.5056)
53.provcd	-1.4852***	-4.0069***		-1.9571	-0.9728
	(0.2029)	(1.3389)		(1.3507)	(0.5938)
54.provcd		-1.7494		0.2326	0.1545
		(1.1638)		(1.1902)	(0.6106)
61.provcd	-1.0193***	-1.9315*	-7.5004***	-0.1284	0.3250
	(0.1486)	(0.9976)	(1.8319)	(1.0266)	(0.4421)
62.provcd	-0.9704***	-4.4436***	-9.9265***	-2.4187**	-0.5035
	(0.1484)	(1.1676)	(1.9519)	(1.1969)	(0.4153)
63.provcd		0.1486		2.2789	1.5625
		(1.5116)		(1.5127)	(0.9972)
64.provcd		-2.1840*		-0.3373	-0.2483

65.provcd	-1.1753*** (0.3594)	(1.1634) -1.4231 (1.0295)		(1.1880) 0.4020 (1.0518)	(0.6920) 0.3882 (0.4592)
11bL.provcd		0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
12L.provcd		3.5327*** (1.2554)		3.0873** (1.2323)	-0.1341 (0.4443)
13L.provcd		6.2133*** (1.4438)	5.7380*** (2.0326)	5.1988*** (1.4858)	-1.2559*** (0.4207)
14L.provcd		7.0429*** (1.5126)	6.6388*** (2.5141)	5.9782*** (1.5514)	-1.0105** (0.4685)
15L.provcd		6.0429*** (1.6408)		5.5156*** (1.6233)	-0.9709 (0.8350)
21L.provcd		6.2380*** (1.4477)	4.5761** (2.2037)	5.6413*** (1.4443)	-0.6577 (0.4652)
22L.provcd		7.0896*** (1.6277)	8.0952*** (2.1827)	6.1790*** (1.6635)	-0.7249 (0.5595)
23L.provcd		7.4631*** (1.5428)	5.9789** (2.3970)	6.6644*** (1.5718)	-0.4919 (0.4826)
31L.provcd		1.7942* (1.0318)	0.7022 (1.2011)	1.2777 (1.0000)	0.4801 (0.3132)
32L.provcd		3.2781*** (1.0763)		2.9009*** (1.0372)	0.7522** (0.3149)
33L.provcd		3.4014*** (1.2266)	4.0935** (1.7367)	2.6527** (1.1982)	-0.8620** (0.3897)
34L.provcd		5.3149*** (1.6209)	5.2000** (2.4351)	4.4186*** (1.6276)	-1.6709*** (0.6034)
35L.provcd		4.9794*** (1.3557)	5.1434*** (1.7353)	4.1620*** (1.3500)	0.5853 (0.5244)
36L.provcd		8.1933*** (1.5480)	7.2606*** (2.3887)	6.9656*** (1.5477)	-0.0355 (0.5646)
37L.provcd		4.9091*** (1.3900)	5.3115*** (1.9885)	4.0474*** (1.3911)	-0.9302** (0.4265)
41L.provcd		6.7707*** (1.4529)	7.0793*** (2.1309)	5.8394*** (1.4777)	-0.4887 (0.4144)
42L.provcd		4.8212*** (1.3843)	4.5525** (2.0894)	3.6735*** (1.4130)	-1.0792** (0.4938)
43L.provcd		5.9206*** (1.5023)	5.5752** (2.4175)	4.9949*** (1.5368)	-0.9479** (0.4653)
44L.provcd		4.0207*** (1.2229)	4.4272** (1.7813)	2.9099** (1.2142)	-0.6975** (0.3557)
45L.provcd		6.7395*** (1.6935)	6.8440*** (2.6408)	5.5638*** (1.7676)	-1.1321* (0.6063)
51L.provcd		6.2611*** (1.5410)	6.7492*** (2.5683)	5.1996*** (1.5733)	-0.3270 (0.5438)
52L.provcd		6.7736***	6.7096**	5.7282***	-1.2325**

	(1.6432)	(2.6151)	(1.6854)	(0.5609)
53L.provcd	7.9565***		6.9852***	-0.6487
	(1.7077)		(1.7424)	(0.6369)
54L.provcd	7.1489***		6.2078***	-1.2065
	(1.6902)		(1.7447)	(0.8852)
61L.provcd	4.7474***	5.5302**	3.8370**	-1.4967***
	(1.4666)	(2.1884)	(1.4917)	(0.4846)
62L.provcd	8.0758***	6.5688***	7.0363***	-0.6285
	(1.5892)	(2.3799)	(1.6351)	(0.4645)
64L.provcd	7.1967***		6.2274***	-0.8596
	(1.7092)		(1.7370)	(0.8985)
11bL.provcd#	1.4427*		1.1018	0.1812
1L.agr_hukou				
	(0.8625)		(0.8019)	(0.2368)
12L.provcd#1	0.5886		0.4333	-0.4911***
L.agr_hukou				
	(0.5161)		(0.5152)	(0.1860)
13L.provcd#1	0.1832	0.4980	0.0452	-0.1646
L.agr_hukou				
	(0.2372)	(0.4225)	(0.2721)	(0.1065)
14L.provcd#1	-0.4692*	-0.5153	-0.5496**	-0.0572
L.agr_hukou				
	(0.2436)	(0.4597)	(0.2762)	(0.1003)
21L.provcd#1	-0.0075	0.3847	-0.1213	0.0100
L.agr_hukou				
	(0.2369)	(0.4699)	(0.2765)	(0.0710)
22L.provcd#1	-0.0901		-0.3780	-0.5070***
L.agr_hukou				
	(0.3272)		(0.3859)	(0.1207)
23L.provcd#1	-0.4892		-0.5651	-0.4159***
L.agr_hukou				
	(0.3362)		(0.3571)	(0.0878)
31L.provcd#1	0.0442	0.1202	0.0499	0.1696**
L.agr_hukou				
	(0.4539)	(0.8777)	(0.5243)	(0.0802)
32L.provcd#1	0.0460		-0.1102	-0.6832***
L.agr_hukou				
	(0.3822)		(0.3870)	(0.1099)
33L.provcd#1	0.4264		0.2879	-0.1683
L.agr_hukou				
	(0.4245)		(0.4386)	(0.1594)
34L.provcd#1	0.1230		-0.2474	0.1285
L.agr_hukou				
	(0.4077)		(0.4334)	(0.1932)
35L.provcd#1	0.0040		-0.1873	-0.9489***
L.agr_hukou				

		(0.4610)		(0.4884)	(0.1800)
36L.provcd#1		-0.1137	-0.6244	0.1090	0.3184*
L.agr_hukou					
		(0.2407)	(0.5582)	(0.4616)	(0.1775)
37L.provcd#1		0.3330		0.1079	0.3435***
L.agr_hukou					
		(0.4117)		(0.4083)	(0.1075)
41L.provcd#1		0.4670**		0.2501	-0.2293***
L.agr_hukou					
		(0.1985)		(0.2285)	(0.0716)
42L.provcd#1		-0.0831		0.1855	-0.5158***
L.agr_hukou					
		(0.3316)		(0.4268)	(0.1423)
43L.provcd#1		0.1999	0.6331	0.0085	0.1747*
L.agr_hukou					
		(0.2422)	(0.4911)	(0.2790)	(0.0920)
44L.provcd#1		0.4635**	0.4125	0.5190	-0.4172***
L.agr_hukou					
		(0.2362)	(0.3697)	(0.3218)	(0.0797)
51L.provcd#1		-0.1280	0.2086	-0.2503	-1.0403***
L.agr_hukou					
		(0.2795)	(0.5019)	(0.3189)	(0.1001)
52oL.provcd#		0.0000	0.0000	0.0000	0.0000
0bL.agr_huko					
u					
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
52L.provcd#1		-0.3722		-0.5562*	-0.1852
L.agr_hukou					
		(0.2763)		(0.3004)	(0.1398)
53oL.provcd#		0.0000	0.0000	0.0000	0.0000
0bL.agr_huko					
u					
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
53L.provcd#1		-0.6600**		-0.8224***	0.0712
L.agr_hukou					
		(0.2933)		(0.3131)	(0.1556)
11b.provcd#1	-0.5726***				
L.agr_hukou					
		(0.2127)			
12.provcd#1L.	-0.5972***				
agr_hukou					
13.provcd#1L.	-0.1815				
agr_hukou					
		(0.1105)			
14.provcd#1L.	0.0184				
agr_hukou					

	(0.1055)
21.provcd#1L.	0.0038
agr_hukou	(0.0718)
22.provcd#1L.	-0.5284***
agr_hukou	(0.1228)
23.provcd#1L.	-0.4037***
agr_hukou	(0.0890)
31.provcd#1L.	0.1505*
agr_hukou	(0.0807)
32.provcd#1L.	-0.6707***
agr_hukou	(0.1090)
33.provcd#1L.	-0.2385
agr_hukou	(0.1632)
34.provcd#1L.	0.1487
agr_hukou	(0.2007)
35.provcd#1L.	-0.9502***
agr_hukou	(0.1800)
36.provcd#1L.	0.4324*
agr_hukou	(0.2223)
37.provcd#1L.	0.3386***
agr_hukou	(0.1079)
41.provcd#1L.	-0.2694***
agr_hukou	(0.0729)
42.provcd#1L.	-0.5416***
agr_hukou	(0.1466)
43.provcd#1L.	0.1747*
agr_hukou	(0.0928)
44.provcd#1L.	-0.4755***
agr_hukou	(0.0814)
45.provcd#1L.	0.0298
agr_hukou	(0.2024)

50.provcd#1L. agr_hukou	-0.1175 (0.2239)				
51.provcd#1L. agr_hukou	-1.0461*** (0.1013)				
52.provcd#1L. agr_hukou	-0.1228 (0.1444)				
53.provcd#1L. agr_hukou	0.1597 (0.1706)				
61.provcd#1L. agr_hukou	0.2357** (0.1186)				
50L.provcd					-1.7422*** (0.6317)
65L.provcd					-1.1110 (0.6882)
45L.provcd#1 L.agr_hukou					0.0344 (0.2026)
50L.provcd#1 L.agr_hukou					-0.1045 (0.2251)
Constant	0.6212*** (0.1826)	-6.4085*** (1.0818)	2.7154* (1.6491)	-8.2585*** (1.1098)	0.9308*** (0.2031)
Observations	56,649	56,177	46,621	56,177	56,688

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix VIII. Heckman Selection Estimates for Moving Intention

VARIABLES		(1)	
rationed_ move	age	0.0027 (0.0075)	
	age2	-0.0001 (0.0085)	
	L.female	-0.0727** (0.0349)	
	L.own	-0.0074 (0.0473)	
	L.urban	-0.0110 (0.0418)	
	L.agr_hukou	0.0874* (0.0488)	
	L.eduy	0.0018 (0.0050)	
	L.finc_pc	-0.0000 (0.0013)	
	L.house_exp	-0.0004 (0.0008)	
	L.mortage	0.0088 (0.0061)	
	L.trust	0.0096 (0.0081)	
	L.gdppc	0.0000 (0.0000)	
	diff_gdppc	0.0000 (0.0000)	
	L.diff_gdppc	-0.0000 (0.0000)	
	Constant	-0.0125 (0.2797)	
	select	house_cost	-0.0001 (0.0001)
		finc_pc	0.0017*** (0.0003)
		agr_hukou	0.1720*** (0.0401)
		own	-0.4086*** (0.0351)
		Constant	-2.4861*** (0.0410)
	athrho	0.5607*** (0.1753)	
	lnsigma	-0.9283*** (0.0864)	
	Observations	98,177	

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Appendix IX. Hypothesis Tests Conducted in Section 5

Table IX.1. Joint Equality in Probit Model

	Chi2	P-Value
age	4.43	0.2188
age2	4.51	0.2111
L.own	4.42	0.2194
L.urban	8.83	0.0316
L.agr_hukou	6.34	0.0963
L.eduy	1.70	0.6367
finc_pc	1.11	0.7746
L.finc_pc	0.05	0.9974
L.house_exp	1.79	0.6173
L.mortage	12.32	0.0064
L.trust	2.49	0.4768
L.gdppc	57.33	<0.0001
diff_gdppc	20.33	0.0001
L.diff_gdppc	45.55	<0.0001