Research on the Characteristics and Factors Influencing Rural Household Registration and Migration—Based on the CGSS Data

Zixi Huang[1,*]

[1]University of Connecticut, 3 Discovery Drive Unit, 6079 Storrs, CT, 06269.

*Corresponding author's e-mail: 2692450085@qq.com

Abstract. The long-duration trend and tendency in China that people all rushed from rural areas to urban areas for better educational and income opportunities led to dramatic transformations in both areas, including but not limited to increase in general educational and income levels, improvement in social living standard, and better family well-beings. This paper will focus on this trend, and various factors influencing this trend, like education, income, and family status, and how do people change hukous, which are defined as the most significant factors.

Keywords: hukou, urban, rural, income, education, generation, three rural issues.

1   Introduction

Since the reform and opening-up policy enacted in China in 1978, Chinese industrialization and urbanization was processing in an unprecedented pace, and that corresponded with an upsurge occurred in people transforming from rural hukou to urban hukou, rushing toward better educational, medical, and residential resources. Hukou is a personally specific identifier in China to recognize the place of birth and residence and different Hukous are linked to various benefits and advantages. One's hukou belongs to the birthplace, but it can be transferred to other cities once the individual finds a lifetime occupation in that city or becomes a permanent resident there. As this trend kept going on, it aroused increasing social attention and interest to analyze various groups of social classes together with the factors, incentives, and family status within each kind of social classes and generations. Specifically, people who transferred from rural hukous to urban hukous reflect the Chinese development and social progress, in a way that the improvement in education and income for the general population could be projected by such people.

Based on an analysis conducted by Li Yu [3], at the group level, socioeconomic status is positively influenced by the macroscopic factor of national per capital GDP during the adult period of the same group, while at the individual level, it is positively influenced by the policy factors of spatial migration and hukou registration type,

together with some self-induced factors and education. However, as Yangyang Qi [6] emphasized, the discrepancy among family education and income level is constantly increasing, and therefore the gap between the rich and the poor. In this case, it is more significant to improve the intergenerational mobility.

This academic paper aims to study the significance of the hukou types in improving the education level and income, by researching differences among people with different kinds of hukou, including those who born in the urban areas, people who lived in the rural areas, and people who moved from rural areas to urban areas, and study the relative significance of different kinds of factors in driving such process. Among them, most research is conducted on the people with changed hukou. Also, an in-depth analysis would be conducted to some special groups, including those who changed hukou by various methods and durations, and those having inconsistent hukou, occupations, and residential cities, like those who lived in modern cities but with rural hukou and those who have extremely high incomes. The special groups often have uncommon characteristics and traits, and therefore they are studied separately.

2   Material

The main material used in this paper would be the CGSS yearly stat data, including 2015, 2017, 2018 and 2021. CGSS, named Chinese General Social Survey, was issued by Chinese government, beginning from 2003, to get a throughout understanding of the situation in different social strata. Below is a summary of the CGSS data base, and each dataset was clustered into three kinds of Hukous.

Table 1. The CGSS Dataset

Year	Number of Observants	Number of People born with Urban Hukou	Number of People with Rural Hukou	Number of People with Changed Hukou*	Number of Provinces **
2021	8055	1683	5597	625	19
2018	12755	3517	8004	1095	29
2017	12374	3515	7856	1003	31
2015	10739	2912	6931	896	29

*People with Changed Hukou: Those who changed their Hukou from Rural to Urban.

**There are in total 34 provincial-level administrative regions in China.

2.1   Analysing the Data Distributions.

For each dataset, the individual statistics are analyzed, including, including education, income, family situation, gender, hukou, area, age, and the statistics for the same

individuals when they were 14. These are analyzed by drawing their corresponding one-dimension and two-dimension graphs, and also their correlations factors and graphs. In the first round of analysis, R studio and several packages are used to construct the graphs, and it was discovered that the data in 2015, 2017 and 2018 all have similar distributions and properties, including:

- The age distributions for three kinds of hukous are close to each other, with a medium around age 57.
- The education levels for people with urban hukous and transferred hukous are same, with a medium around 6(defined as high-school level), and the medium for those with rural hukous is much lower, around 4(defined as junior high school).
- They are classified into four kinds of areas, based on their overall education level.
- People with urban hukous have highest both individual incomes and family incomes, people with transferred hukous have medium levels of incomes, and people in rural hukous have the lowest portion of income. In general, the income for urban people and the changed-hukou people are quite similar.
- The unemployed people distributed as 2:1 for rural people versus urban people, and those who are working in agriculture mostly have rural hukous.
- There are 8 main approaches for people born with rural hukous to transfer to urban hukous, with the four most popular ones being changing occupations, family follow-up, attending higher schools, and governmental land acquisition.
- People with urban hukous generally have the best attributes, people with changed hukous have the medium attributes, and the rural people have the worst attributes.
- The education level is gradually increasing as time goes by. From 1949 to 2023, the time-frame is classified into six intervals, including 1949-1957, 1957-1966, 1966-1976, 1976-1992, 1992-2002, and 2002-current, and they are based on the Chinese policies and social situations, including foundation, before the cultural revolution, the cultural revolution, after the cultural revolution, reform and opening, and the new age. As time goes by, the average income and education status are both leveling up.

Fig. 1. Types of Changing Hukou vs. Education Level
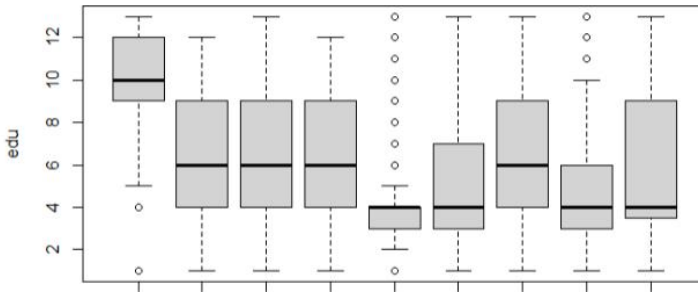
| School | Army | Work | Promotion | Land | Family | House | Reform |

This graph is an example showing the relationship between the education level and the methods of changing the hukous. As shown, people who changed hukous via working have the highest education level.

In contrast, the 2021 dataset has some different patterns, although there also exist some coincided ones. For the 2021 observants,

- The age, education, occupation, and area distributions for three kinds of hukous are similar to that for previous periods.
- the overall income situations are much lower for 2021 than for the previous periods.
- Parents' educational levels are much higher for people with urban hukous, with both a much higher rate and proportion than people with other hukous.
- People who have urban hukous and changed hukous have more similar properties and attributes, while only the people with rural hukous are still left behind.
- There were only three main kinds of changing hukous, including attending a higher school, join the army, and finding an advanced job.

One possible reason for the differences would be the recent COVID-19, which was destructive to the economy and social structure. It isolated people in different areas, limiting access for people in rural or urban areas to migrate, and therefore decreased the population mobility. Therefore, it is supposed that the people who have transferred hukous in 2021 have lived in modern cities for years and therefore they generally have similar properties to those who born with urban hukous. Therefore, the 2021 data was separated from the other datasets.

2.2   Analyzing the Special Trends for Separate Groups.

Some special trends and properties are discovered among people with uncommon characteristics, including those with changed hukous, who worked in urban cities with rural hukous, and people with income more than about 1 million, and it is also recommended to research these two groups separately to capture the patterns in the society.

2.2.1 Properties for People with Changed Hukous.

There were eight main ways of changing hukous in years 2015-2018, including working, family follow-up, attending schools, land acquisition, purchasing houses, reformation, joining the army and becoming a cadre. Till 2021, only three of them remained as potential ways, including joining the army, working, and attending schools. After analyzing each kind of methods, it was found that for years 2015-2018, they can be classified as active ways (working, family follow-up, and attending schools), passive ways (land acquisition, real estate, and reformation), and joining the army. The active ways are followed more often than the passive ways (1854 active people vs. 731 passive people), which means that the opportunities of changing hukous are beneficial for those competent people to advance and seek for promotion. Population following the active ways generally have higher educational levels and incomes than those following the passive ways. Those who changed hukous by joining the army have random distributions, having people from all strata of the society with different levels of education and income. More specifically, those who followed the active ways have characteristics like the urban residents, while those who followed the passive ways have characteristics closer to the rural residents. Therefore, the way of changing hukous differentiates two strata within the changed-hukou population. Furthermore, the three main ways in 2021 all fall in the criteria of active ways, which is why the people with changed hukous in 2021 have characteristics similar to those with urban hukous. Therefore, it is recommended and decided to classify the ways of changing hukous into three categories, including active ways, passive ways, and joining the army, in the ground of future analysis.

2.2.2 Properties for People who Worked in Cities with Rural Hukous.

There are two main inconsistencies detected, including people who work in the cities with rural hukous, and who work in the rural areas with urban hukous. Since there are relatively few cases of people working in the rural areas with urban hukous, this situation is simply excluded. For people who work in the cities with rural hukous, they are under the similar trend as the people with changed hukous, but they are not so successful or lucky in that they fail to change their hukous even if they are working in those modern cities. For their characteristics, they typically have income and education level similar to the medium of the urban residents or changed-hukou residents, and their job situations and areas are quite relevant and similar to the counterpart of urban residents.
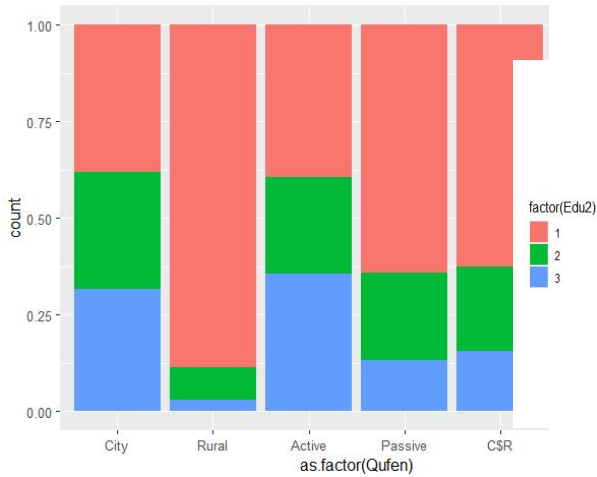
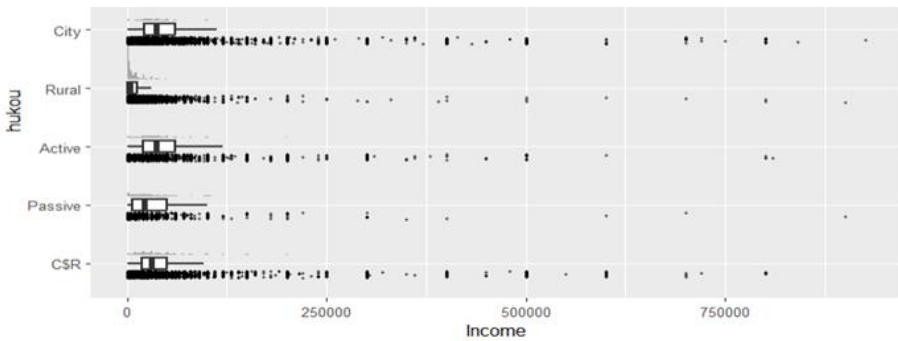**Fig. 2.** Five Types of people vs. Education level

Fig. 3. Five types of people vs. Income Level

These are two images showing the education and income level for each kind of people, including local city residents("City"), local rural residents("Rural"), people changed hukous in an active way("Active"), people changed hukous in a passive way("Passive"), and people who worked in cities with rural hukous.

For education level, level 1 is the basic level (from non-educated to junior high school), level 2 is the medium level (from high school to technical school), and level 3 is the high level (university and above).

It is clear here that the city residents and people who changed hukous in an active way share the same traits of education, and the people who worked in cities with rural hukous are similar to those who changed hukous in a passive way. The rural people are generally in the lowest education level.

The income distribution is also similar to the education distribution, except for the incomes for the people who worked in cities with rural hukous are closer to the city residents.

Properties for people who worked in cities with more than 950,000 incomes. Among the observants, a group of outliers is also found to mislead the data trend analysis, which are people with extremely high incomes. The threshold 950,000 is set by using the outlier analysis formula and some subjective observations. The general population have incomes below 950,000 while a few people have such high incomes that it is recommended to be analysed separately. In the analysis specially conducted to the high-income people, it is found that people with high incomes fall in all ranges of education level and family situation, which means that some of them are highly educated and intellectual while some of them are uneducated and live in rural areas. Their family status and birth places also vary randomly. Therefore, the extremely high income is not deemed useful in researching the social situations and general trends, and people with high incomes are temporarily segregated.

3   Methodology

3.1   Correlation Analysis

A series of potentially valuable factors is selected, and a correlation analysis is conducted, aiming to identify those factors whom are more correlated and contributed to the individual development such as income and education level.
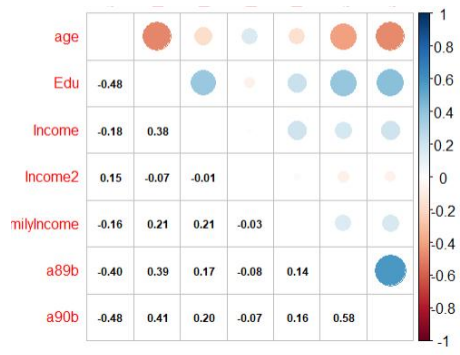

Fig. 4. Correlation Analysis

This is an example for the correlation plot for the population in year 2015-18, and similar graphs and analysis are also conducted specifically for people with different kinds of hukous and in different areas.

*A89b is father's education level.

*A90b is mother's education level.

In the correlation analysis, it is found that the observant's education and income are highly correlated, and Family's income is also correlated to parents' education

levels, but in a medium degree. Father's education level and mother's education level are very highly correlated, and it corresponds to a traditional social phenomenon and tendency to match the family and personal status. The correlation between the elder generation and the younger generation is also analyzed, and it is found that the only significant factor is the education level, which means that the education level is likely and supposed to be passed by to the next generation, and the next generation typically have education levels one degree higher than the parents (For instance, people are likely to attend graduate schools if their parents attended undergrad school.). The income level would not make huge contribution to the next generation, and people with high/low family incomes could have low/high personal education and income. It contradicts to a common assumption that parents with higher income levels could educate their children better, as indicated by the social data that about half of the rich families have their children uneducated. For some other factors, including areas, time eras, birth places, hukous, and occupations, none of them significantly contribute to the education and income level. Therefore, further research will still focus on observant's education and income level, parents' education level, and the hukous.

3.2   Linear Model and Statistical Analysis.

For each dataset, a set of linear models is built, treating the income and education level as the dependent variables, and using all explanatory variables, including age, income and education, family's education and income, area, hukou, gender, era, how did they change hukou (only for those who changed hukou), and the same statistics for them in their age 14 (for instance, their hukou in age 14). Several linear models are built via R studio, and below is a list of noticeable coefficients in the linear model. Below is a part of the model result, and some factors, like education, contain too many levels so they are not listed.

Table 2. Model Coefficients and Significance

Model of Income	Selected model Coefficient	Model of Education	Selected Model Coef.
Family Income	2.960e-06***	Family Income	2.960e-06***
Age	-5.741e-02***	Age	-5.741e-02***
Gender	-5.019e-01**	Gender	-5.019e-01***
Era1	-4.784e-01***	Era1	-4.784e-01***
Era2	-7.718e-02*	Era2	-7.718e-02(no*)

It is better to classify the education level into three categories, since the difference inside each education level is so minuscule that grouping them together captures the overall trend while not losing information.

- People with the best family income level polarized education levels and there are two possible explanations for this phenomenon. The most probable reason is that the people with highest education levels are

devoted into their own interests, like art, scientific researches and sports, or just never need to work.

- A three-dimensional relationship occurs among era, hukou, education, where people born in current eras or in urban areas have higher education level and therefore income level.
- Area, personal education, and era contribute the most to the income level.
- Parents' education, personal income, era and hukou contribute the most to the education level.

3.3   Causal Analysis

In the correlation analysis, it was found many variables have high correlations, and in the causal analysis it would be determined which variables lead or mainly influence other variables.
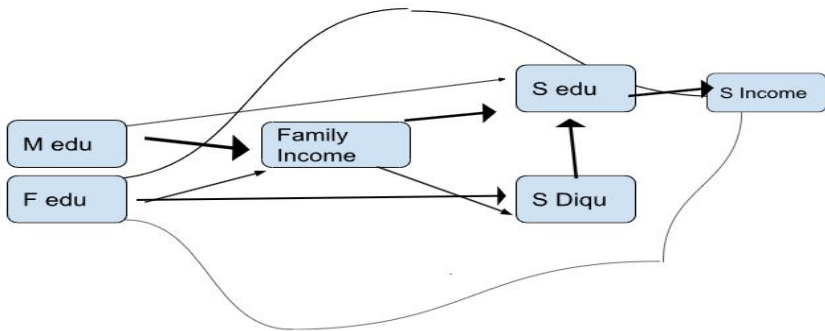

Fig. 5. Assumption of Causal Analysis

This is the assumption of the causal relationship. It is assumed that parents' education influences the family income, and then children' born areas and education. Lastly, children' area, hukou, and education sort their income levels.

Different directions of causal analysis are conducted, and different linear models are built, examining whether each variable is capable of explaining other variables, by checking the proportion of explanation, R square, standardized coefficient, standard error, T value and P value. The model and link with the lowest standard error for the coefficient and lowest T or P values is deemed the most trustworthy link.

For the 2021 normal income group, the transformed link has the following changes: mother's attribute is the most significant variable, while father's situation and family status also make slight contributions. The people with changed hukous differentiate in that father contributes less while mother takes even higher proportions. For the local urban citizens, family income no longer contributes to any of the children' characteristics, which implies that the urban education and job opportunity is approaching equal and justified.

For year 2015-2018, the trends are similar for the urban people, people with changed hukous in active ways, and the only exception is the people who changed hukous in passive ways. People with changed hukous in passive ways or with extremely high income cannot be sufficiently explained by any factor.

In general, the population distribution has the following causal relationship:

- Mother's education plays the crucial role, and influences family status and children' attributes, which is possibly caused by the previous policy that the hukou for a new-born mainly relied on mother's hukou. Only since recently can people' hukous follow their fathers' hukou.
- Father's education plays a minor role in the family.
- The family's area is randomly distributed, but the area could influence the children' characteristics including both the education and income level.
- The education influences the income level in a positive way.

4   Discussion

4.1   Personal Development.

As shown in both the analysis and the model, the same characteristics for the individuals at age 14 are not significant, which means the born place or the family situations does not necessarily determine or fix one's future achievement, and today our society gives certain equality and chance to everyone. Another possibility is that characteristics at younger age are controlled and reflected by family's situation, which is then guaranteed by parents' education level, and therefore the performance of the next generation have been pre-determined by their parentages. In this case, the analysis and the model both assigned the first weight to the parents' situations, because the parents' attributes directly and indirectly link to any other factor. Between parents' attributes, mother's education often plays a more crucial role than father's situation, which is potentially due to the social tradition that mother is mainly in charge of taking care of the children.

4.2   Family's Income—a Hidden but Significant Factor.

In most previous analysis or researches, the parents' income was deemed not significant, so do the previous correlation analysis and models in section 3. But, in our further hierarchical analysis, it is found that family's income level does contribute to the next generation's well-being, and its significance may be diluted by some outliers. For instance, as Yeboah [7] found, the people under poverty, especially the younger generation, tend to migrate more frequently, which implies that family income and situation is a determining factor in population circulation. Most of the highly educated families have high income levels, therefore income contributes directly to the link between parents' and observants' education levels. But in the macro social situation, there are many new monies who squander and do not pursue any further improvement in their well-beings, therefore their next generation tends to have low income and

education levels. This drives down the significance of the family income level, which should be higher.

4.3   Other Potential Correlations.

For those with high incomes, people with rural hukous tend to have higher income levels and low education level, while people with urban hukous often have lower income and higher educational level. This trend is similar for both generations. This reveals a potential social phenomenon that people in the rural area often receive a huge number of subsidies if the government has a need to take over their undeveloped lands, and they are usually uneducated so they squandered with their children uneducated, and could not change hukous to the urban areas.

4.4   Limitations, Innovation and Future Expectation.

In the research, there are a few limitations, including the limited resources for years during COVID, and some missing variables and information in year 2015. The causal analysis could have been done in a higher level, using more criteria like the Pearson Correlation, but it is only conducted in a medium level due to the time limit and limited pages here. But, since it was found that many previous researches did not conduct lots of data manipulation, this research did a lot in segregating the data based on their characteristics, a lot of correlation analysis for various combinations of variable sets, and many distribution patterns, to thoroughly analyse the dataset. Therefore, it would be beneficial to utilize the dataset to conduct a better casual analysis together with different models.

5   Conclusion

Therefore, in our analysis about the rural migration and hukou registration, it is determined that inheritance relationship plays the first role in passing down all other factors, and then education and income are the most significant ones directly contributing to the migration process, causing population stratification. Education generally plays a more influential role in the heritance relationship and population migration than income. Other factors, like area and occupation, are indirectly controlled by education and income. Rural people migrating in a positive way generally have better attributes than those who migrate in a passive way, and education level can contribute to the income level. The COVID had restricted the population circulation and fixed the distribution of rural and urban citizens to some degree.

References

1. Dongting Xu. (2021 June 8). Research on the Intergenerational Transmission of Chinese Residents' Educational Credentials Education from the Perspective of Educational Equity. A desertion submitted to Nanjing University of Finance and Economics.
2. Fei Li. (2015 June 6). China Academic Journal Electronic Publishing House. Research on Migrant Workers' Rural-urban migration action: Base on Survey Data from Zhongshan, Guangdong Province.
3. Li Yu, Lili Cao, and Meiyu. Yu. (2021 June 10). China Academic Journal Electronic Publishing House. Intergenerational mobility and influencing mechanism of social and economic status of agricultural population in megacities—hierarchical linear analysis based on CGSS data.
4. Puqin Wang. (2020 July 2). China Academic Journal Electronic Publishing House. Hukou Registration and Time at Immigration: The Status Attainments of Urban Migrants in Shanghai (1977 – 2014).
5. Weichao Huang. (2021, Sept 23). China Academic Journal Electronic Publishing House. Social cognition, social status and well-being of migrant workers—empirical analysis based on CGSS 2017.
6. YangYang Qi. (2018 Sept 2). China Academic Journal Electronic Publishing House. Intergenerational mobility of China's socioeconomic status: Based on CGSS data.
7. Yeboah, T. (2018). Young People and Migration in Ghana. https://doi.org/10.17863/CAM.22663.
8. Yu Xiang. (2019, June 10). Research on the influence of urbanization on farmers' income—analysis based on multidimensional perspective. Southwestern University of Finance and Economics.