

# Intergeneration Mobility: Facts and Factors

**Ken Chen**

*MA in Computational Social Science, University of Chicago*

## **Abstract:**

Intergenerational mobility has been an obnoxious issue that passes inequality from generation to generation. This has been a persistent problem for most of big economies. In this project, taking advantage of a Chinese income-expenditure dataset, we aimed to unpack the salient facts that shapes how the economic status of the parents and children diverges and how does that divergence move across generations. For this purpose, we proposed Principal Component Analysis to construct comprehensive measures of economic attainment, and based on this measure, we were able to derive the transition between two generations. Additionally, we established Random Forest Models and determined the most relevant factors to children's economic status, including education, employment status and of course, the parents' incomes. The paper discusses the severity of economic stickiness, implies some working factors that somehow offsets the effect of parents and enriches social fluidity.

## **Introduction:**

In addition to the growing social inequality problems that large-scale economies are struggling with, intergeneration mobility has become an issue that goes hand in hand. Relevant research has been prosperous for the US, and there is remaining space to look deep into China's issue.

## **Economic Inequality**

The entire world has witnessed great economic growth since the end of World War II. Of all the countries that survived the chaos, China has especially achieved remarkable economic success.

However, despite the fact that China's Open and Reform Policy has injected massive stimuli to its economy, inequality has stood out to be the next topic of concern. As documented in Gan's paper, China has reached an Income Gini of 0.61 by their measure (2013), compared to 0.45 of the United States (2007), and 0.38 of China's neighbor Japan (2011). Alongside, there is a rich set of studies that branch into different dimensions of inequality. Besides income inequality, Zhang and Eriksson (2010) have looked into opportunity inequality in China, and Li et al (2011) has studied the mechanism that governs the translation from income inequality to consumption inequality. If we think of inequality as an unwanted factor harmful to the social well-being, we have to come up with strategies to remove or mitigate its intensiveness.

### **Intergenerational Mobility and Education: a bird's-eye view**

Instead of direct taxation and wealth redistribution, scholars have also bought prominence to intergenerational mobility. Corak (2013) included in his paper a plot of intergenerational earning mobility against social income inequality. The relationship is positive, which alluded to us the importance of studying intergenerational mobility. As Corak put it in his paper, the central value held by the broad concept of 'American Dream' is not the "outcome equality", but the "opportunity equality". In other words, all people should have equal access to the channel where they can move from bottom to up when they invest their time and endeavor. Meanwhile, they should also share a similar risk of moving down. This paradigm also positively perceived by Chinese society. On top of all, education plays a critical role ensuring that people have more or less equal access to develop their intellectual capital and expertise for seeking well-paid occupations. From a theoretical perspective, Becker and Tomes (1979) have framed a theoretical construct to determine intergenerational mobility,

primarily in terms of personal earnings. In their world of intergenerational inheritance, children's equilibrium income is jointly determined by their "market and endowed luck, the own income and endowment of parents, and the two parameters, the degree of inheritability and the propensity to invest in children". In this paper, they also suggested employing intergenerational income elasticity as a measure of mobility. In their later empirical work (1986), they found the intergenerational income elasticity to be 0.2, which implied mild income stickiness over generations and good opportunities to climb the social hierarchy. Their approaches were further refined by replacing income observations at a certain time stamp with their lifetime earnings. This could be realized by introducing other exogenous variables such as education and the working industry as instruments to make that prediction of lifetime income (Haider, 2006). We may try to emulate the approach by bringing into the paper more computationally enhanced tools.

### **Intergenerational Mobility and Education: in the United States**

There exist a good set of literature of how education might correlate to intergenerational mobility in the US. Previous research has uncovered two importance facts: the overall increase in people's education level does not lead to significantly lower inequality level (Chusseau et al, 2012); But higher education do play a centric role linking students from poorer families to better upward opportunities. What is threatening this 'opportunity inequality' are the financial constraints that contain them from winning in this admission game (Haveman and Smeeding, 2006). Empirically, Raj Chetty et al (2017) has documented some important results regarding college education: parental income is a substantial determinant of children's access to college admission; conditional on what colleges children attend, those from rich families share similar earning outcomes with those from poor families. These statistics

coincide with what was found earlier in other literature. Furthermore, Chetty points out that the proportion of students coming from poor families dropped essentially at schools where there had been most bottom-to-top-quantile mobility. This trend is alarming if the US society wants to maintain a fluid social structure.

### **Intergenerational Mobility and Education: in China**

Due to China's contractive economic and diplomatic policies in the early years, studies on social mobility in China started off a bit late. Some early research, by Cheng and Dai (1995), revealed general patterns of intergenerational mobility in modern China. Their research was broken down into several economic groups. They've collected evidence advocating upward mobility of people from farm origins and manual working class. However, for the working class, mobility is relatively rigid. Wu and Treiman (2007) accounted for rural-urban mobility, which disadvantaged rural-origin men by exposing them to a greater possibility of downward movement. The mobility across different social classes might be uniquely due to China's Hukou system that directed the rural and urban population to different administrative regimes. For our interest in income mobility and education, there are a few papers in documentation. A recent study by Gong et al (2012) calculated the intergenerational income elasticity to be 0.63 (compared to 0.2 of United States in the 80<sup>th</sup>), indicating that parents' economic position has a very persistent effect on the children's. Golley and Kong (2013) especially placed their stress on education mobility over generations. Their study found that children of the rural and migrant populations are more likely to endure less education than their parents, while children in urban areas seem to at least maintain the same level of education as their parents. In this paper, we plan to measure the urban intergenerational economic mobility using

another data source, and especially investigates what role does education play in the setting.

## **Theory and Methods:**

### **Relevant theories**

One of the most well-known theories that shed light on intergenerational mobility was raised by Becker and Nigel (1979). They stated that children's fortunes are linked with their parents in substantial ways. Their economic well-being is jointly determined by their market, as well as the endowment luck. Two variables of interest are the degree of inheritability and the propensity to invest in children. If these parameters are all less than unity, then the income distribution between families will reach an equilibrium. Further researches have highlighted the influence of many other factors on mobility. Boudon and Raymond have synthesized a simulation model that shows education is positively linked to inequality of social opportunity. Therefore, our research will especially look into education and other related factors.

### **Data**

My data is a survey dataset, named China Household Survey Data, which is collected by the National Bureau of Statistics of China. Until 2006, the survey was initiated in 1986, and has been conducted for 5 rounds. It covered most of the country's provinces and cities. The samples were drawn through a multi-step selection procedure. The survey provides very detailed data on income and expenditure of the urban households in China. In fact, the dataset follows a rotating panel structure, with one-third of samples updated each year, and each new round will take an entirely new sample. One prominent issue with the dataset is that different rounds of surveys are collecting slightly different variables, and the variable labels are not consistent. For this research project, due to the constraint of time, we plan to

use only the fifth round of data (from 2002 to 2006).

The data can only be accessed once submitting access to the National Bureau of Statistics of China. I obtained the data in Apr. 2018. For each year, there are two tables. One of them is an individual based table, while the other is a household-based table. We merged the two kinds of tables, on the individual identifying index, so that people from the same household will share identical family variables. One major issue with the data is that children might leave the house to form a household entity when they are wealthy enough, and those who remain in the household might not be as economically successful as those who leave. Therefore, we filtered the merged table to contain only those who have just entered the workforce, by bounding the children's age between 16 and 30, and the parents' income variables were spread to become new columns of the table. We care most about the linkage between parents' income and children's income at the early phases of their careers.

### **PCA to coin measures of economic hierarchy**

Previous studies on intergenerational mobility have placed their focus on absolute terms of personal income, especially labor incomes. However, these metrics do not form a comprehensive outlook of one's economic well-being. Taken more dimensions of income variables into account, we might also be concerned about that what differentiated the economic well-being among the parents and the children might not be the same. Therefore, I'd like to urge the use of Principal Component Analysis to extract the principal component of the 'high-dimensional' income data, which can help to construct our measure of the economic hierarchy.

Our income metric consists of several subcategories:

$$A = (Labor_{inc} \ SOB_{inc} \ Property_{inc} \ Labor_{inc})$$

To utilize the PCA component, we determine to use the first principle component as the measure of

economic status. In other words, we are solving the following optimization problem:

$$\arg \max_x \text{Var}(Ax)$$

Then, by projecting each individual's income vector onto  $x$ , we can obtain our measure. This is going to be done separately for the parents and children. Let us name the measure as 'economic score'.

The first two principal components have explained over 90 percent of the variance. And these are the scatterplots of where the children and parents lie on the first two components.

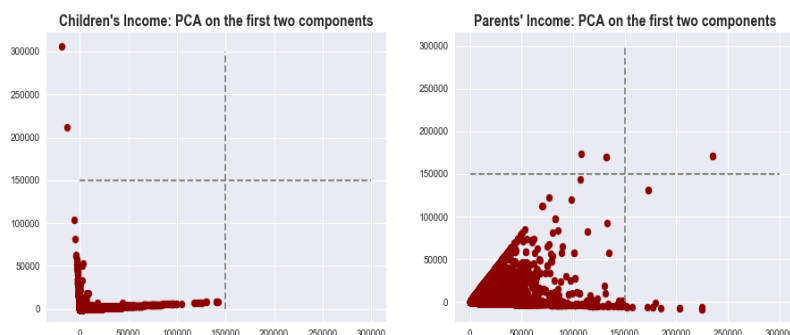


Figure 1

The first component captures more information on labor income, while the second one captures private business for the children, while property income for the parents. And also, we've depicted how their intra-group scores are distributed:

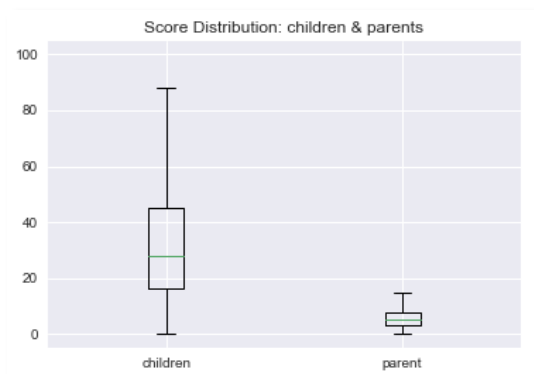


Figure 2

We can see that the children's generation is distributed more sparsely while there is apparent clustering inside of the parents' group.

Following the method from John et al (2004), then we divide up both generations into five layers, which were represented by level 1 to level 5. And here we present the transition matrix, where the x-axis stands for the parents' economic level, y-axis stands for children, and the z-axis represents the empirical probability of transition from parents' level to the children's level accordingly. Columns on the diagonal stand higher than the surroundings. This is especially the case for the very rich and the very poor. Children mostly inherit their parents' economic position, and the possibility of moving up or down is relatively weak.

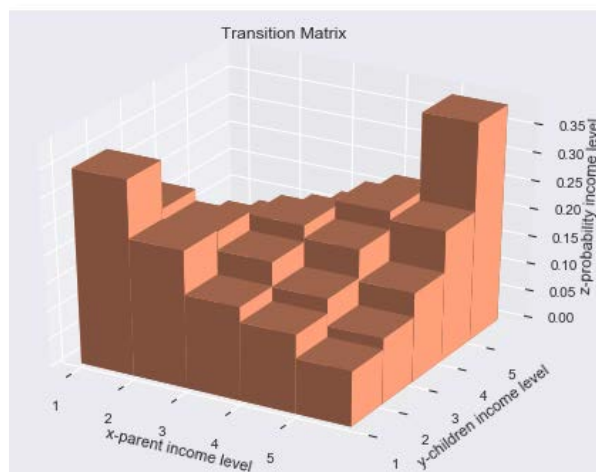


Figure 3

## Predictive models

This project is also concerned about identifying what factors are more contributive to intergeneration mobility most and how much of that is realized via the channel of education investment. Thus, we proposed two models to see detect such effects.

We defined a random forest classifier, including variables of education and parents' income, to see how well these factors can help predict how economically accomplished the children can be. We turned the question into a classification problem, by dividing the parents and children into five layers based on their score on our measure.



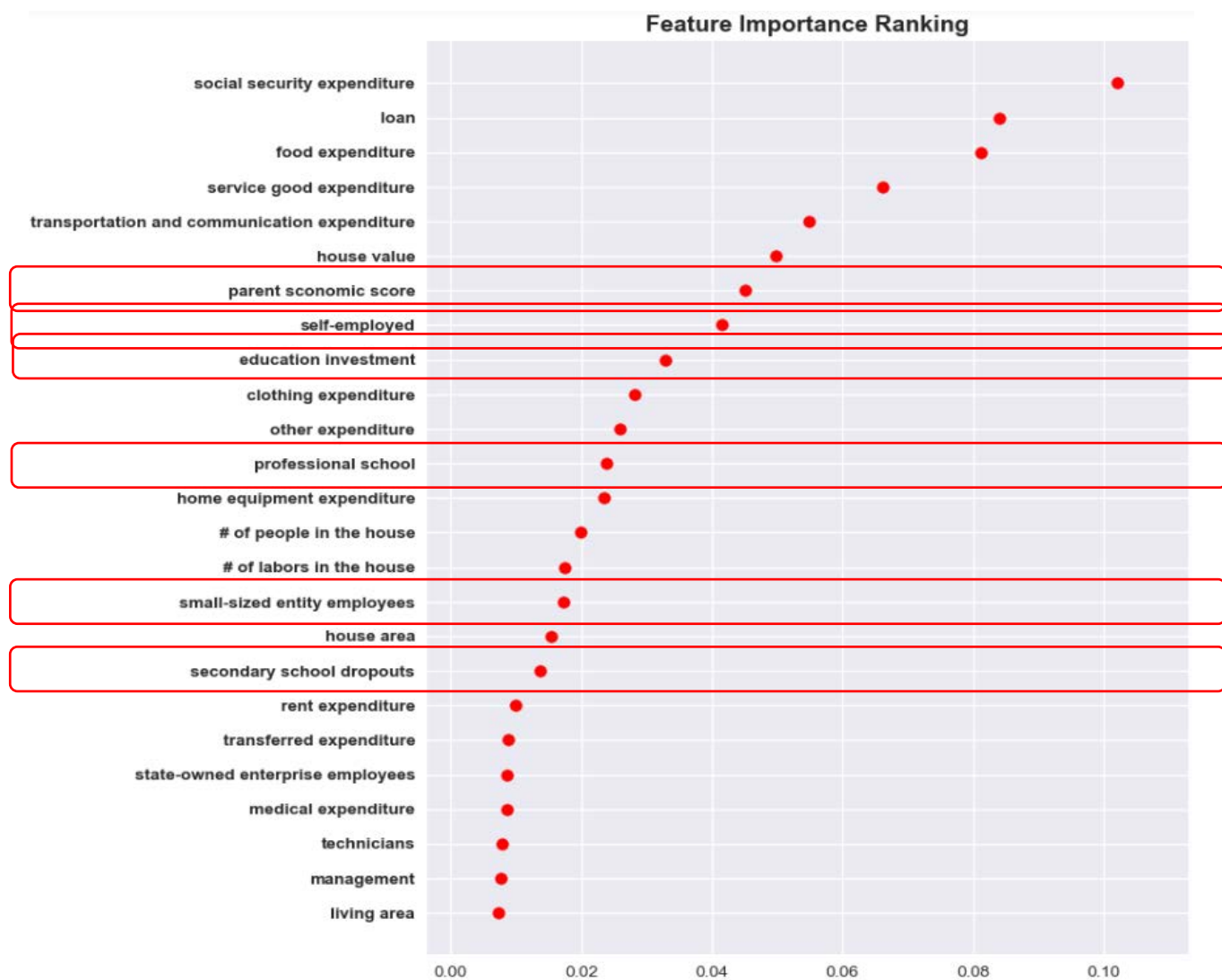


Figure 4

Out of the 348 socioeconomic and family background variables, we've selected out the 25 that are most predictive of children's economic well-being. Apparently, parents' economic attainment serves to be a major determinant of children's earnings. The influence is even beyond our expectation. Another two types of variables, namely the employment status and education turn out to be the most salient socioeconomic factors that help predict children's earnings. Purposefully, we've plotted the PDP/ICE plot for the two types of education that pops up in the above feature plot:

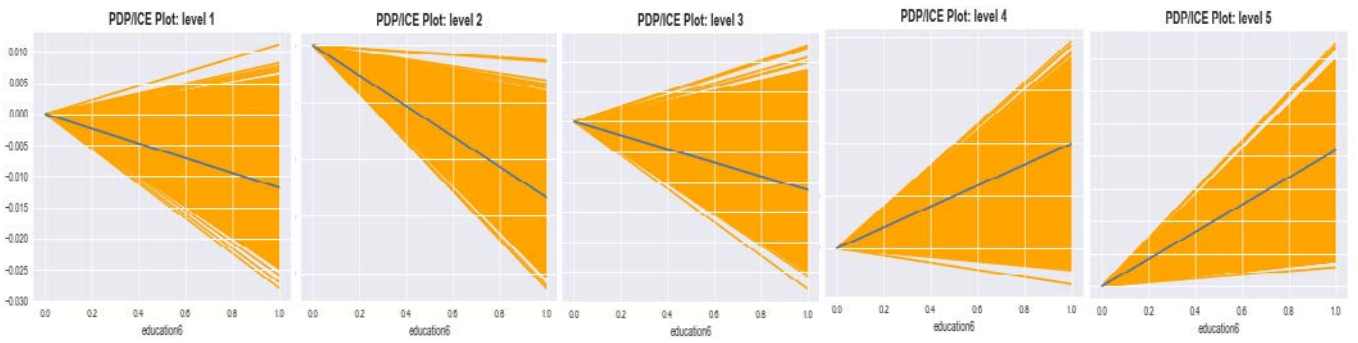


Figure 5: professional education

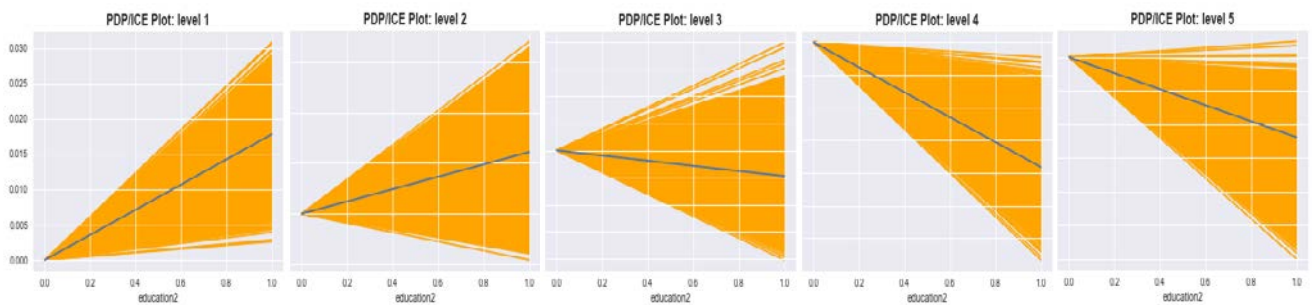


Figure 6: secondary school dropouts

These plots implies that with higher level of education, children are more likely to end up being on the higher level of economic scores, and people with inadequate education are more likely to end up being poor. This partially verifies our assumption that education can somehow enrich social mobility and working against parents' rigidity influence in a positive way.

## Conclusion:

In China, children's income distribution is marked with a higher degree of variance, and the intergenerational fluidity is quite tight. Especially for the very poor and very rich, there is very weak possibility of moving upward or downward. In our random forest model, parents' economic status turns out to be very predictive of children's economic performance. Professional education is most likely to increase the opportunity of getting into the highest or second highest economic layer. The self-employed are more likely to stay at the bottom for the children in their early careers.

## **Discussion:**

One of the weaknesses of the entire computation design is the interpretability of the predictive model. The permutation feature importance plot decides the importance of features without showing directions. Although the PDP/ICE plots helps to picture the correlation in a fancy fashion, but the plots themselves do not fit well into a conditional context interpretation that can extract causality, which is should be the centric of social science studies if we aim to produce actionable policy recommendations. An explainable social science research design should contain components of both. Further extension of the project could be causal estimation employing the selected features from the tree-based model outputs. If we divide up the population by parents' income, the effects may as well be different. Robustness of the causal estimation can be computationally strengthened by tree-based analogous strategies like Athey and Imbens (2015) took to estimate heterogenous causal effects.

## **Reference:**

Becker G. &Tomes N.,1979, "An equilibrium theory of the distribution of income and intergenerational mobility". *The Journal of Political Economy*, 87( 6):pp. 1153 – 1189.

Corak M.,2016, "Inequality from Generation to Generation: The United States in Comparison". Iza Discussion Papers.

1. Becker, G. &Tomes, N.,1979, "An equilibrium theory of the distribution of income and intergenerational mobility". *The Journal of Political Economy*, 87( 6): pp.1153 – 1189.
2. Becker, G & Tomes, N, 1986. "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, University of Chicago Press, vol. 4(3), pp. 1-39, July.
3. Boudon, R. (1974) *Education, Opportunity, and Social Inequality*. New York: Wiley

4. Cheng, Yuan & Dai, Jianzhong, 1995, "Intergenerational Mobility in Modern China." *European Sociological Review*, Volume 11, Issue 1, May 1995, pp.17–35
5. Chetty Raj, Friedman John N., Saez Emmanuel, Turner Nicholas, Yagan Danny, "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility," Stanford University Working Paper, 2017.
6. Chetty, Raj, et al. "The Fading American Dream: Trends in Absolute Income Mobility Since the 1940s." National Bureau of Economic Research, 2016, rev. March 2017. <http://www.nber.org/papers/w22910>
7. Corak M.,2016, "Inequality from Generation to Generation: The United States in Comparison". Iza Discussion Papers.
8. Golley, J. and Kong, S. T. (2013), Inequality in Intergenerational Mobility of Education in China. *China & World Economy*, 21: 15-37. doi:10.1111/j.1749-124X.2013.12013.x
9. GONG, H., LEIGH, A. and MENG, X. (2012), INTERGENERATIONAL INCOME MOBILITY IN URBAN CHINA. *Review of Income and Wealth*, 58: 481-503. doi:10.1111/j.1475-4991.2012.00495.x
10. Gan, Li, "Income Inequality and Consumption in China." Unpublished, 2013
11. Jin, Y., Li, H., & Wu, B. (2011). Income inequality, consumption, and social-status seeking. *Journal of Comparative Economics*, 39(2), pp.191-204.
12. John P. Formby, W.James Smith, Buhong Zheng, 2004. Mobility measurement, transition matrices and statistical inference, *Journal of Econometrics*, Volume 120, Issue 1, 2004, Pages 181-205
13. Shi, Li. Chinese Household Income Project, 2002. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-08-14. <https://doi.org/10.3886/ICPSR21741.v1>
14. Susan Athey and Guido Imbens. 2015a. Machine learning methods for estimating heterogeneous causal effects. ArXiv e-print number 1504.01132.
15. Susan Athey and Guido Imbens. 2015b. A measure of robustness to misspecification. *American Economic Review*. 105, 5 (May 2015), 476-80. DOI=10.1257/aer.p20151020.
16. Zhang, Y., & Eriksson, T. (2010). Inequality of opportunity and income inequality in nine Chinese provinces,1989-2006. *China Economic Review*, 21(4), pp.607-616.
17. Distribution of Family Income. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html>