# Inequality of Opportunity in China

**By**

**Haishan Yang**

**Submitted to**

**Central European University**

**Department of Economics and Business**

**In partial fulfilment of requirement for the degree of**

**Master of Arts in Economics**

**Supervisor: István Kónya**



**Budapest, Hungary**

**2019**

## Abstract

What is the degree of inequality in China? What is the major cause of this huge inequality? In this paper, by answering these questions, I firstly show the Gini index of income and education over the years. To do so, I divide the entire sample into 11 age groups and generate the Gini index separately. The results suggest that Gini index of education decreases dramatically from 0.506 for people born between 1940 and 1944 to 0.172 for people born between 1990 and 1994. Meanwhile, income inequality shows an alarming sign, remaining in a high level. By stratifying 31 Chinese provinces into three regions; East, Central and West. I find education inequality is severe in the central part of China due to limited educational resources. Then, I construct a sample scalar measure of inequality of opportunity which captures between-group inequality where groups are defined entirely by some factors that beyond individual controls such as the Hukou system of household registration, gender and family background. I find that the Hukou system plays a leading role in limiting both the earning potential and education achievement for people.



JEL classification: D31, D63, J62

Keywords: inequality; inequality of opportunity; China

## Acknowledgements

I would like to express my gratitude to Professor István Kónya, my research supervisor, for his patient guidance, enthusiastic encouragement, useful critiques of this research work and assistance in keeping my progress on schedule. I would also like to thank Professor Thomas Rooney for helping me with my writing. My grateful thanks are also extended to professor Jane Golley and professor Sherry Tao Kong for sharing the appendix of their paper. I want to express my thanks to the often unjustifiably unacknowledged STATA online community. Finally, I wish to acknowledge my classmates and Ph.D students at Central European University for exchanging research ideas, especially Dr. Franklin Maduko.

During my graduate study at Central European University, I encountered tremendous difficulties in my studies, especially in first semester due to my weak quantitative background. Many teaching assistants here helped me greatly, especially Luca Flóra Drucker and Balázs Vonnák. My classmate Benedek Tóth often saved me when I was struggling with problem sets. Professor Ádám Szeidl supported me in my Ph.D application and he is a source of inspiration for me to become a better researcher. Without them, I would not reach this final stage to write the thesis.

Finally, I wish to thank my family for their support and understanding throughout my studies.

## Contents

## Tables



## Figures

# I. Introduction

Equality is the core objective that many communist regimes try to achieve. In the political theory presented by the Chinese communist party and Constitution of the People's Republic of China, Chinese society is still in the early stage of socialism and is on the way toward communism. In this paper, I want to answer the question how far the Chinese society is from an ideal communist society where everyone is equal.

After the Chinese Communist Party took control on mainland China in 1949, the party launched a series of political movements to eliminate political opponents and so called "class enemies" such as businessmen. The party declared everyone should be equal in this social system, although one can argue that "All are equal but some are more equal than others".

In 1978, a dramatic political struggle unfolded following the death of Mao Zedong. Deng Xiaoping came to the power and initiated the Reform and Opening-Up policy. He publicly advocated "Our policy is to let some people and some areas get rich first to drive and help the backward areas." Since then, the Chinese economy has been growing phenomenally, as has the gap between rich and poor. (Zhao, S, 1993)

Income inequality in China has been widely studied. (Xie and Zhou, 2014) The Gini coefficient in China ranges from 0.45 to 0.55 by different studies. Unlike in the USA, where race and ethnicity play a substantial role in income disparity. China is a homogeneous in ethnicity in 2012: 91.51% of the population were Han Chinese. The main contributors for the income inequality are the rural-urban gap and difference in regional development. (Xie and Zhou, 2014)

From a longer historical horizon, the economy of southern China has been stronger than that in northern China since the Song Dynasty in the 10$^{th}$ century although historians recognize Chinese origin to Central China. However, the strong economic outperformance in the south in recent decades is considerable due to a 1979 policy. In that year, the Chinese government

started to designate several coastal cities as Special Economic Zones (SEZ). In these zones, a very different political and economic institution was allowed. The specific policies included: i) providing land with low price to foreign investors; ii) encouraging the growth of private company, and iii) deregulating financial control and labor movement. After the government found these policies worked very well in lifting the living standards, most of policies now are available nationwide. Yet, the south has the advantage of early opening-up and its geographic location give it advantages in export, maintaining much higher economic performance than the rest of China. (Wang J, 2013)

The divergence of economic development between the south east and the rest of China also has profound implication in income and education opportunity. In general, inequality in income is more severe in coastal cities due to occupational diversification. While inequality in education is South-eastern China is less than that in the rest of China, because there are more educational resources available from the rich cities.

This paper aims to do explore the development in inequality and inequality of opportunity in China. Using recent microdata, my original finding include Gini coefficient of education decreases over the year from 0.504 in the cohort born between 1940 and 1944 to 0.172 in the cohort born between 1990 and 1994. Moreover, Gini is 0.40 in Western China, 0.31 in Central China and 0.29 in the East. On contrast, income inequality remains high. Gini coefficient of family income in my entre sample is 0.543. Specifically: it is 0.509 in the West, 0.514 in the Central and 0.543 in the East. Given these preliminary results, I further study the inequality of opportunity. I find that the Hukou system of household registration plays a leading role in diverging the educational opportunity, while gender is detrimental in determining the income opportunity.

The remainder of the paper is structured as follows. Section 2 gives a short review of the inequality and inequality of opportunity literature. Section 3 relates a mathematical model

to gauge inequality of opportunity. Section 4 then introduces the data and empirical strategy to calculate the inequality of opportunity. Section 5 presents estimations results from the regressions. Section 6 discusses the trend of inequality of opportunity and section 7 concludes with summaries and policy recommendations.

## II. Literature Review

One important feature of many neoclassical growth models in macroeconomics predicts a convergence of economic performance between poor region/country and rich region/country. Although empirical findings from the country-level point to multiple conclusions, there is a consensus that homogeneity of institutions, technology promotes regional convergence. For example, Barro and Martin (1990) studied European regions, Japanese prefectures and the US. They find strong a tendency of convergence. However, this is not the case in China. Since the Reform and Opening-up in 1978, the regional inequality remains high and is growing bigger. Yang and Cai (in press) find that the ratio of urban-rural income and consumption are between 2 and 3.5 since the reform. The level is considerably higher than most of countries in the world. Other literature also documents that production per capita and consumption continue to diverge. In particularly, the initially rich coastal provinces were better in the beginning widen the advantages. (Kanbur and Zhang, 1999)

Tao Yang (2002) hypothesizes that fiscal policies, factor market distortions and regional development strategy are the main determinants for rising regional inequality in China. To support his argument, he gives an example of urban price subsidy. In many developing countries, social policies often target the poor: such as cash transfer to the poor or any kinds of indirect subsidy such as food stamps or free school meals for disadvantaged schoolchildren. Surprisingly, there is a social program in China to subsidize relatively wealthy urban citizens. In 1985, the government liberalized meat prices. To stabilize increasing meat prices for urban dwellers, the government subsidized 26.2 billion yuan to reduce them. It was just 7.9 billion

yuan in 1979 and reached a peak of 71.2 billion yuan in 1998. Urban citizen also benefit from credit transfer. From 1986 to 1990, the government subsidized 232.5 billion yuan to state-owned enterprise (SOE). From 1991 to 1995, the government subsidized 206.1 billion yuan to SOE. These two numbers account for roughly 19% and 9% of state revenue in the two periods. The preferential credit allocations to SOE assume a higher and growing urban wage even when the SOE was losing the money (Tao Yang, 2002).

The development of the Chinese economy as well as many economies in Soviet Union is basically a story of how to choose optimal allocation between efficiency and equity. There were extreme examples in these former Soviet countries to pursue equity as priority. On the other extreme, some people do see equity as a problem. In particularly, utilitarianists such as Jeremy Bentham (1789) believe that the goal of society is to achieve maximal output. It does not matter who enjoys the benefit. Under this philosophy, the social welfare function can be written as:

$$W = \sum_{i=1}^{n} U(Y_i) \quad (1)$$

where U represent individual utility and i represents individual. In other words, if transferring resources from person A to person B makes person A two units happier than it makes person B one units sadder, then it is a good deal. Under this philosophy, we are indifferent between welfare of Jeff Bezos and homeless people as long as the pie gets bigger. Nonetheless, the utility function is concave. One hundred euro more for Bezos is nothing but for people in Africa is a matter of life and death. Assuming everyone has same concave utility function, this model implies we better distribute income equally to everyone. In this sense, this often-viewed conservative social welfare can also tend to be a source for government intervention.

On the other hand, the more liberal parallel is called Rawlsian social welfare function proposed by John Rawls (1971). He argues that what we need to care about is to make the life

of worse-off members in our society as good as best-off members. This philosophy can be presented mathematically as:

$$W = \min(Y_1, Y_2, \ldots\ldots, Y_n) \quad (2)$$

In the late 1980s, philosopher G.A. Cohen started to formally separate the source of inequality into choice and responsibility, (Cohen 1989, p.993). Soon after, John Roemer (1993) put the idea into mathematical models. He defines those factors over which individuals have a measure of control as "effort", they can be the duration of study or the diligence one has. He further defines factors beyond individual control as "circumstances" (e.g. race, gender, birthplace). Given these definitions, an equal society is a place where people who exert the same effort receive the same outcome.

This approach is appealing because inequality of opportunity is more actionable information for policy makers. Intuitively, if Luca earns 1000 forints and Balázs earns 500 forints because Luca works harder or longer than Balázs, many people do not think it is a problem. However, If Balázs works as hard as Luca but earns just half of Luca's earning, many people would feel it is unfair. Thus, one could argue what society should address is the inequality of opportunity rather than inequality of outcome.

In fact, the difference between inequality because of individual effort or responsibility on the one hand, and those factors that beyond one's control on the other hand has become a burgeoning debate on many social sciences and increasingly in economic research. A sizable literature has been focusing on this topic. A wide variety of evaluation methodologies have been proposed to measure a range of outcome such as income, education attainment, health care and so on. Bourguignon, Ferreira, and Menéndez (2007) apply a linear model of income as a function of circumstances and efforts. Based on this model, they simulate counterfactual distribution of income where the effect of circumstances is suppressed. They decompose total income inequality in Brazil into components because of several observed circumstance variable

by comparing the real income distribution with different counterfactuals. Francisco and Jeremie (2008) study inequality of opportunity in Latin America using the similar methodology. They find that inequality of opportunity ranged from 0.13 in Colombia to 0.22 in Brazil, 0.34 in Guatemala. They also find that between one quarter and one half of observed consumption inequality is due to differences in opportunity at minimum. Hassine and Zeufack (2015) find that although there is a decline of consumption inequality over time in Sub-Saharan Africa, the increase in inequality of opportunity was more pronounced between households in Dar es Salaam because some households in the region under certain policies are able to take better advantage of the opportunity created by Tanzania's economic growth.

Far less has been done in terms of measuring the inequality in education. De Barros et al. (2009) use test in math and reading from PISA (Program for International Student Assessment) for high school students in Latin America as an outcome variable and parent's levels of education, parental occupation, gender and geographic location as circumstances. They conclude these five circumstances account for 28% of total inequality of score. Ferreira and Gignoux (2010) do a similar research for Turkey, and find the above circumstances account for 33% of overall inequality.

Little research focuses on the inequality of opportunity in China. Ying and Lv (2018) find that 23.2% of income inequality is driven by circumstances and the degree of inequality among female is higher than males. Zhang and Eriksson (2010) find that inequality of opportunity that account for Gini coefficient increased from 46% in 1989 and 63% in 2006. Notably, parental income and parent's profession determine about 2/3 of the inequality of opportunity.

Building on the existing literature, this paper is the first attempt, to my knowledge, to examine the inequality of opportunity both in year of education and income in China. Unlike many countries where ethnicity plays a role, Hukou system in China determines individual

opportunity largely. I also find that inequality of income and educational attainment does not go in same direction.

Before going to theoretical framework, a brief introduction of Hukou system is helpful to better understand the major barrier to achieve equality in China. In 1958, *People's Republic of China Hukou Registration* went into effect. The entire nation was divided into two groups: agricultural and urban. Under a command economy, to achieve industrialization, the Chinese central government gave residents with urban Hukou more benefits in terms of health care, education, food supply among many others. Moreover, it was very difficult, if not impossible for citizens to change their hukou status. Only 1.5% of the rural citizens could transfer their hukou status per year under official quota (Chan, 2014). In the following years, the policy became more restrictive. Rural citizens even could not move to the city freely, which partially caused the great famine in China. In 1978, after Reform and Opening-Up policy, the rural residents could move freely to the city and find a job. However, they couldn't enjoy many benefits as urban citizen have until now. Most of these migrant workers from rural areas have no insurance like most of urban citizens. Their children cannot attend local public school nearby where the work in the city.

## III. Theories of Inequality of Opportunity

In this paper, I set personal income and year of education as outcome variables. Like previous works on this topic, I assume income and year of education are determined by efforts and circumstances. The assumption can be written as:

$$y_i = f(c_i, e_i, u_i) \tag{3}$$

where y is outcome (i.e. income, education), c is circumstances and e is effort. I also assume effort is influenced by circumstances:

$$y_i = f(c_i, e_i(c_i, \varepsilon_i), u_i) \tag{4}$$

In the early work of Roemer (1998), he assumes people with favorable condition will work more efficiently. Thus, their effort tends to yield higher output. This means two people put same quantile of effort in their group, the degrees of effort are not equal.

By Roemer's definition (1998), the ideal society of full equality is:

$$F(y|c) = F(y) \tag{5}$$

Now there is a finite population of individuals, $i \in \{1, \ldots., N\}$. I divide the population into different types where individuals in the same type are in an identical circumstance with distributions $\{y_i^k\}$. There are k groups, given by $\Pi \in \{T_{1,\ldots..,} T_k\}$.

My objective is to measure the share of inequality in educational outcomes that can be attributed to inequality of opportunity, that is, deriving from unequal circumstances across the population, rather than unequal effort. To figure it out, between-type inequality provides us with a measure of inequality of opportunity, while within-type inequality could be interpreted as inequality of effort.

In previous works, two methods are mostly used to decompose total inequality into its 'between-type' and 'within-type' components. (Ferreira and Gignoux, 2011; Golley and Kong, 2016)

The first is an absolute scalar measure of inequality of opportunity,

$$IOA_D = I(\{\mu_i^k\}) \tag{6}$$

where $\{\mu_i^k\}$ is the distribution obtained from replacing each individual outcome, and $y_i^k$, with its type-specific mean, $\{\mu_i^k\}$. Intuitively, by doing so, I suppress within-group inequality by smoothing the outcome distribution, assigning to every individual the mean for their type. Then I can find out the inequality between types. There is also an associated relative measure,

$$IOR_D = \frac{I(\{\mu_i^k\})}{I\{y_i^k\}} \tag{7}$$

Which measures inequality of opportunity as a share of total inequality.

A second option is to suppress between-group inequality by replacing each individual outcome, $y_i^k$, with $y_i^k \frac{\mu}{\mu^k}$, where μ is the mean for the entire sample. This generate a standardised distribution, $\{v_i^k\}$, which leaves only within-group inequality $I(\{v_i^k\})$. In this case, indirect measure for the inequality of opportunity is:

$$IOA_R = I(\{v_i^k\}) \tag{8}$$

The relative measure is:

$$IOR_R = \frac{I(\{v_i^k\})}{I\{y_i^k\}} \tag{9}$$

Now there are three common used function for I{.}, generalized entropy index:

$$GE(\alpha)\begin{cases} \frac{1}{N\alpha(\alpha-1)}\sum_{i=1}^{N}\left[\left(\frac{y_i}{\overline{y}}\right)-1\right], & \alpha = 2 \\ \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{\overline{y}}, & \alpha = 1 \\ -\frac{1}{N}\sum_{i=1}^{N}\ln\frac{y_i}{\overline{y}}, & \alpha = 0 \end{cases} \tag{10}$$

In theory, GE(0) is the mean log deviation, GE(1) is the Theil index, and GE(2) is half the squared coefficient of variation. In this paper, I use GE(2) since some people have zero education, which prohibit me to use GE(0) and GE(1).

I am also interested in the partial contributions of each of the circumstance variables. To asses these, I use the to create a counterfactual distribution,

$$y_i^J = C_i^{j\neq J}\beta_I + \overline{C}^{j=J}\beta_i \tag{11}$$

That is, assigning the mean level of circumstance J to all individuals, except one circumstance at a time and then calculating both direct and indirect partial contribution to inequality of opportunity is:

$$IOA_p = I(\{y^J\}) \tag{12}$$

$$IOR_p = \frac{I\{y_i^k\} - I(\{y^J\})}{I\{y_i^k\}} \tag{13}$$

One major challenge to this theoretical design is the omitted variable bias. Although I do not have an "effort" variable in the regression, any effect of circumstances on outcome through theirs effects on efforts is captured by the regression coefficients of circumstances. Therefore, the regression likely captures the reduced-form influence of circumstances, both directly and indirectly through efforts. However, omitted circumstances can be a problem. These omitted variables can only bias estimates of inequality of opportunity downwards. For example, assume we do not have proxy of intergenerationally correlated ability, but father's education is correlated to the ability. Then the coefficient of parental education accounts for itself and the ability. The overall impact of circumstances on inequality of opportunity can still truthfully describe by the estimate, though it can still bias the total inequality of opportunity downwards.

Although the estimated total inequality of opportunity can only lower bound of the true value, the partial estimation is likely biased either way because many unobserved circumstances can determine outcome through the channel of regressors and they can correlate to them either positive or negative.

Another fundamental issue is how we deal with randomness or luck. Dworkin (1981) argues that there are two kinds of luck: brute luck and option luck. Brute luck is one that beyond individual control. It can be further divided into early and later brute luck. The former includes social background and genetics such as ability and talents. The later includes war or car accidents. As for option luck, it occurs when people intentionally take risk, which can be calculated or avoid. Lefranc et al (2009) proposes to include luck as third determinant of

advantage outcome, aside from circumstances and efforts. In this paper, I attribute luck to either circumstances or responsibility, but not a separate determinant.

## IV. Data and Empirical Strategy

The empirical analysis of this paper is based on the China Family Panel Studies (CFPS) survey, which is a national longitudinal general social survey project. The project collects data from three levels: individual, family and community. Until now, there were four rounds of survey conducted in 2010, 2012, 2014 and 2016. In the initial survey of 2010, the surveyor had a face-to-face interview with family members living in the sampled communities. Importantly and probably not commonly, the project team tracked and interviewed family members who were not present but still in same country. After collecting basic information of absent family members from present family members, the team was able to find out the family members elsewhere.

In this paper, I use the data from 2014 and 2016. The 2014 survey consists of 37147 individuals and the 2016 survey includes 36892 individuals. I merged the dataset of family information into an individual dataset. Notably, I did not append two datasets in 2014 and in 2016 into one dataset, because when I analysed the data, I found discrepancies between two datasets. It seems this large-scale survey is still developing. For example, in 2010 and 2012 survey, individual communist party membership and father's party membership are available, but it disappears in the later survey. Furthermore, number of siblings in the 2012 survey exists but it becomes family size in 2014. There is no variable of family income in the 2016 survey, but it exists in the 2014 one. The 2016 survey also contains the individual information of previous round of survey. To avoid errors at the expense of losing precision, I use 2016 survey to investigate inequality of education and 2014 survey to investigate inequality of income.

Table 1 provides descriptive statistics, column 1 shows illiteracy rate went down from 40.05% among people born between 1940 and 1944 to 4.65% for people born between 1990

and 1994. There was an increase in illiteracy rate among people born between 1945 and 1954. This is due to the Cultural Revolution when schools were shut down, and young people were encouraged to either go to countryside to learn from farmers or engage in Red Guards, a student-led paramilitary social movement. The sample also shows that for urban Hukou holder, the illiterate is almost eliminated, while for rural Hukou holder, there is still more than 5% of people who can't read. Born with a urban Hukou also have people nearly two times higher chance than rural people to have a college degree. Table 2 shows the income for urban citizen is higher than rural citizens in every age cohort. Male also earns more than female and. People in Eastern China earn more than people in Central China and Western China.

Given the available data, I select the following sets of circumstance variables: Hukou, gender, region and father's education. From Figure 3 and Figure 4, it turns out that all these factors mark huge difference between the two groups. The distribution of education outcome for people with urban Hukou first-order stochastic dominates the one with rural Hukou. The same dominance applies to all other three factors. Education attainment in East China is higher than in Central and Western China. It shows that as education level goes up, the disadvantage of female diminishes. A closer look at the data indicates that the discrepancy in early education stem from only the older generation. Many females born in the 1940s and the 1950s received no formal education. For younger generation, fortunately, the gap of education attainment starts to converge.

I categorize Hukou as a circumstance because it is reasonable to say that there is no inherent difference of innate abilities among rural and urban kids. Thus, the huge divergence of educational outcome can be fairly attributed to the Hukou system where rural children are unable to fulfil their true potential. The Hukou influences the education outcome through several channels: i) the quality of education in rural area is worse than that in the city. ii) Chinese public education is highly controlled under Minister of Education. Under the current

system, every major public university has fixed and disproportional quota for the local citizens. In rich municipality such Beijing and Shanghai, where urbanization rate is very high, and majority of the quota goes to the local urban citizen. In poor province such as Guangxi and Anhui, where higher education resource is rather limited, but most of the rural citizens must compete for spots in their province and for spots in other provinces higher difficulties. On the other hand, in china, people with rural Hukou who work in cities are often referred as "migrant works". Many of their rights are deprived.

The Hukou system can affect income both directly and indirectly. The direct way is that i) companies prefer to hire local urban people because they suspect the mobility of local citizens are low and their records are more trackable. ii) urban citizen can be more patient to find a better job due to unemployment benefit. In many big cities, if you are an unemployed citizen with urban Hukou, you receive unemployment benefit. For example, in Shanghai, the unemployment benefit amount to 1815 Yuan for three years per month, roughly 235 euro, while migrant workers basically have no any unemployed benefit if they come to big cities and try to find a job (Shanghai Municipal Human Resources and Social Security Bureau, 2019). The indirect way includes absent health insurance for migrant workers as well as less social connection for them.

Secondly, intergenerational persistence plays a role in educational inequality in China. Golley and Kong (2012) find a strong correlation between parent's education and their children. More specifically, level of education between fathers with senior higher school or above and children in urban area is highly correlated. Most children with rural hukou are likely to complete high school despite the low education attainment of their parents. The reason for the persistence could be because educated parents pay more attention to their children's education. Moreover, better educated probably are willing to and able to exert more financial resources for the education of the kids. In sum, it is fair to say that children with well-educated parents

enjoy a more favourable condition in terms of their education opportunity among many others. (Golley and Kong, 2013)

Thirdly, gender discrimination is widespread. In China, the situation is more special due to the One Child policy. Nancy Qian (2006) shows survival rates for girls go up alone with growing female income, keeping male income constant and vice versa. Most relevantly, she finds a rising income for the mother increases the year of education for all children, while a rising income for the father has no effect on the education outcome for boys, but reduces the year of education for girls. Prior to 1980, due to limited resources, parents preferred allocating more education resource to male children, because once a daughter got married, she had no responsibility to take care of parents. After 1980, school education became accessible for everyone. The disadvantage diminished gradually. Nonetheless, for higher education, female students were still in a disadvantaged position. In Chinese junior high school, Chinese language, mathematics and English are compulsory courses. Apart from that, students must choose to go science group or liberal art group. For students in the science group, they will study physics, biology and chemistry. For students in the liberal art group, they will study history, political science and geography. Traditionally, parents and school teachers would encourage female students to go to the liberal arts group. In the early years after Reform and Opening-up policy in 1978, higher education resources were extremely precious and the opportunity to attend colleges is a privilege for few. the government allocated disproportional resources to advance the research on science and engineering in university. As a result, there were many more spots in university for students in science group. This trend started to reverse as Chinese government started to commercialize higher education. Many private liberal art schools emerged alongside with the founding of many schools of liberal art and social sciences within public universities.

Lastly, where you born also matters a lot. I categorize China into three parts: East, Central and West. In 1985, the Chinese government published "*Suggestions of the Central*

*Committee of the Communist Party of China on Formulating the Seventh Five-Year Plan for National Economic and Social Development*," in which the government put China into these three regions. This categorization is not purely based on geography, but more on economic policy and economic growth rate. In this paper, I use this definition as an important circumstance to determine the inequality of opportunity. From Figure 3 and Figure 4 we can see years of education and income in East first-order dominate that in the rest of China due to the fact that the number of higher educational institutions and major companies and factories disproportionally locates in the East.

Table 3 summarizes the major independent variables in the regression. Table 4 and Table 5 are regressions of education on various proxies for circumstances. Table 6 and Table 7 are regressions of income on same dependent variables plus years of education. After I run the regressions, I construct a distribution $\{\mu_i^k\}$ based on coefficients and residuals obtained from the regression. To obtain $IOA_D$, I plug in $\{\mu_i^k\}$ into formula of GE(2). For $IOA_D$, I firstly plug the original distribution of outcome into GE(2), creating a scalar vector. Then $IOA_D$ divides the vector and we have $IOA_D$.

For partial contribution, I run regressions without single variable, such as gender. Then I get GE(no gender). The partial contribution is:

$$IOR_p = \frac{I\{y_i^k\} - I(\{y^J\})}{I\{y_i^k\}} \tag{14}$$

where $\{y^J\}$ is GE(no gender) in the above example.

## V. Estimation Result

Table 4 to Table 7 show the results of the regressions. Table 4 and Table 5 show the regression of education on the circumstances I have described above. Table 6 and Table 7 show the regression of income on the circumstances plus years of education. Before I interpret the Table, I should make several crucial principles straightforward and clear.

Firstly, the independent variables in the regressions serve not only as the proxies of themselves, but broader circumstances that beyond one's control. This is because no one can figure out all circumstances into a regressions so omitted variables always exist. In this case, I interpret the coefficient of father's education as it captures effects of family backgrounds on earning and offspring's education, which may include intergenerational ability. They may include parent's social network that could help their child to find a better opportunity. They can also be family income that is used to get their children to a better school.

The coefficient of Hukou is also likely to absorb other effects. Obviously, even without the hukou system that deprives many rights of rural dwellers they would still have many barriers to education and earning. It can be that urban citizens are better nourished or have access to more public educational facilities such as library. The coefficient of Hukou could also absorb the social network that urban citizens build on for years but not through family networks.

However, the purpose of this paper is not to figure out the casual relationship between Hukou or father's education and offer's education or income. Rather, these dependents variables serve as proxies to absorb many uncontrolled factors, which will help us to gauge the inequality of opportunity. In other word, while there is a need to interpret with caution the regression coefficients as causal between a independent variable and the outcome, the omitted variable bias is not critical to the measurement of inequality of opportunity.

Secondly, it is important to recognize that temporal patterns in income cannot be seem as cohort trends. In other words, the variation in these regression coefficients across cohorts cannot be interpreted as evidence of changes over time, because I am not able to disentangle age, period and cohort effects on the coefficients as the outcome variables are not measured at the different points in time. Nevertheless, along with other evidence, I still can report some

general trends of what was happening. The results of equality of opportunity can also serve as reference for policy makers.

While income is changing over time, the level of education is fixed for most people after certain years. Table 4 column 1 shows controlling family background, gender and region, urban Hukou is associated with 2.6 more years of education compared to that of those with rural Hukou, and a strong positive correlation exists among all age cohorts with significant level at 1%. Male is associated with more years of education, although the correlation is weaker in the younger cohort. Compared with the rest of China, people in the East have a higher educational level. From the coefficient of family background, one can see that compared with those whose father is illiterate, father with some education helps a lot in determining offspring's education. Again, father's education here serves as a proxy for family background.

Table 6 and Table 7 indicate the regression of personal income on various circumstances and years of education. Ideally, family income is a better measurement for income inequality, because many housewives have no income but also contribute to the family equally. Unfortunately, there is not a very good measurement of family income in this dataset due to the traditional family structure. The ideal type of traditional Chinese family is so-called "four generations under one roof". Even until today, many Chinese couples live with their parent's after marriage, so when they answer questions about family income, many of them tend to include their parent's income. In this case, I use individual personal income as the outcome variable.

Another issue that arises from the dataset is that among 12205 samples with no missing data, 2932 individuals report that they have zero personal individual. Including this individual with zero income, I find the Gini coefficient of the entire sample is 0.756, which is very unprecedented. To better understand these individuals with zero income, I did a basic comparison between the group of individuals with zero income and the one with some income.

The former has a family yearly income of 63280 Yuan which is roughly equal to 8233 euro, while the latter has 76873 Yuan which is an equivalent of 10000 euro. There is no big difference between the percentage of people who own an urban hukou, which is 12% and 13% respectively. The gender ratio is balanced in both groups and the educational levels are also alike with 21% and 19% respectively.

Based above statistics, I am cautiously confident that not too much involuntary unemployment among those report their income is zero since these people are seemingly not so much worse off than the rest. With no further information available, I delete all the samples with zero income. It is worth noting that this strategy deserves more discuss in further researches.

It appears from Table 6 and Table 7 that gender and region are associated with higher earnings in the whole sample and every birth cohort. As mentioned in this chapter, females earn less because more males study in the STEM field in colleges. Coefficient of East region is significant at 1% in every cohort and full sample shows the career opportunity is much more promising in Eastern China due to public policy and geographic convenience regarding export. An urban Hukou is also associated with higher earnings, although this association is in a declining trend with the younger cohort. This will be discussed more intensively in the next chapter.

## VI. Inequality of Opportunity

Using the regression results from Table 4 to Table 7 and the method in chapter IV, I present measures of both the relative and absolute indexes of inequality of opportunity using the direct method. Table 8 panel A column 4 shows that relative inequality of opportunity grew from 0.265 among people in the cohort who were born between 1940 and 1944 to peak of 0.437 for those who were born between 1980 and 1984, while relative inequality of opportunity decreases from older cohort to younger cohort as it is seen in panel B column 4.

Table 8 column 1 and column 2 provide GE(2) and Gini coefficient for comparative purposes. These two indexes have a nearly identical trend. Specifically, the Gini coefficient of education declines from old cohorts to younger cohorts, while the Gini coefficient of income remains at a high level.

Table 8 column 2 and column 3 in panel A show that the Gini coefficient and $IOA_D$ go in the same direction. In other words, inequality in education decreased, the inequality between different groups of types also decreases. Take the cohort of people who were born between 1985 and 1989 for example, the Gini coefficient is 0.202 and absolute inequality of opportunity is 0.025. It means among 0.202 of Gini coefficient, at least 0.025 is explained by the circumstances I described above. The co-movement between Gini and $IOA_D$ indicate much of the decreases of inequality can be attributed to the decline in inequality of opportunity. On the other hand, how to explain a relatively high inequality of opportunity, while the Gini coefficient of education? Recall equation 9:

$$IOR_D = \frac{I(\{\mu_i^k\})}{I\{y_i^k\}}$$

Given $I\{y_i^k\}$, in my case, column 1 in Table 8 is in decreasing order and $I(\{\mu_i^k\})$, in column 3 is also in decline, the relative inequality of opportunity $IOR_D$ can only increase only if $I\{y_i^k\}$ decreases bigger than $I(\{\mu_i^k\})$ does. In other words, total inequality decreases more quickly than inequality of opportunity, which implies, in our theoretical framework, that younger cohorts exert more effort or exert effort more effectively.

It is more difficult to interpret panel B in Table 8 as income. While years of education is fixed at some age, personal income changes in different time periods. We see the Gini coefficient in column 2 panel B remains high in all age cohorts. There are two ways could affect the variation in panel B column 3, The first could be these circumstances such as Hukou, gender and regions have more influences in the older population. For example, male and female

might earn a similar amount of money in the early career, but the male has more opportunity to be promoted due to discrimination against female. The second could be also period effect. Maybe due to gradual liberalization of the Chinese economy, people face less barrier to achieve their full potential as seen from panel B column 3. Another possibility is that the determinants of earning are more complicated than the determinants of education. The circumstances I set do not cover the main determinants as the R-squared in Table 6 and Table 7 is relatively small.

Finally, Table 9 presents the partial contribution to inequality of opportunity. Based on equation 14 in chapter IV, I calculate variations of inequality of opportunity attribute to three major circumstances: Hukou, gender, and East region. Because of endogeneity, the output can be biased in either way. Assuming the bias in every cohort is in the same fashion, I do can find some causes for the trends of the partial contributions. The most notable feature in Table 9 is that Hukou contributes a lot to inequality of education, but not so much in income.

There are several reasons for the discrepancies. Firstly, rural Hukou directly disadvantages children, but effects less on income after controlling the education. For example, a migrant couple work in the city, and they have to a child. The child will still have a rural Hukou and it is extremely difficult for the child to get in a public school in the city. Many of these children eventually go back to their parent's village and are taken care of by their grandparents and enroll in village schools with much lower quality. Secondly, for Hukou's partial contribution to inequality of opportunity in income, it appears people with urban Hukou who were born between 1950 to 1964 have considerable advantage. Note that the age of retirement is around 55 years ago in China, so these even the youngest people in this cohort has already retired. Many of them with urban Hukou grew up in a China where State-owned enterprise (SOE) was dominant. Due to connections in the city and discrimination against rural Hukou, majorities of jobs in SOEs are taken by the urban Hukou holders. In China, the social security system is far away from comprehensive even today. Most the workers in private

companies in this cohort had no insurance, no pension at all. That is why there is a 10% to 32% of the contribution of Hukou to inequality of opportunity in income in the older cohorts, not insignificant in the younger cohorts.

Column 3 in Table 9 shows an as the advantage of males is decreasing over time. This is probably because as educational resources are more available to more people and the financial burden of education becomes less of concern. In contrast, gender still plays a significant role in determining personal income as shown in column 6. As discussed before, disproportional males study STEM fields could contribute partial to this high inequality of opportunity in income for female.

## VII. Conclusions, Limitation and Policy Recommendations

This paper begins with a theoretical review of different perspectives on equality. Starting from Jeremy Bentham to John Rawls. Then I applied Roemer's approach on equality to divide the determinants of outcome into "efforts" and "circumstances."

By doing some basic statistics comparison and calculating the Gini coefficients based on China Panel Survey (CFPS) for 2014 and 2016, this paper has provided a mixed picture of the inequality in China. As the expansion of educational resources, especially the higher education, inequality in educational attainment decreases dramatically. On the other hand, a very conservative estimation based on personal income displays that income inequality remains at a high level. I further studied inequality of opportunity. I focus on four determinants for the inequalities. They are Hukou, gender, region and father's education. A clear conclusion is that equality is far from being achieved by Roemer's definition. It means:

$$F(y|c) \neq F(y) \tag{15}$$

Figure 3 and Figure 4 clearly confirm the above inequality. Moreover, in terms of education, Figure 5 and Figure 6 provides another perspective. Figure 4 indicates that among those disadvantaged groups such as people with rural Hukou or are female. The within-group

inequality is more severe, which implies that there is a lack of a mechanism to help most worse-off individuals to catch up in these disadvantaged groups. On the other hand, within-group inequality in income has no apparent differences.

The regression outputs from Table 4 to Table 7 indicate that all these circumstances such as urban Hukou, male and Eastern region are associated with longer years of education and higher earning. However, the omission of unobserved variables – weather 'efforts' or circumstances' --- meant that some of the coefficients are likely to be biased and cannot be treated as causal.

Next, I find that absolute inequality of opportunity decreases from the older cohort to the younger cohort. For education, the decline of inequality of opportunity might attribute to the expansion of education resources. For income, age effect and period effect may play roles in same direction. The former suggests older people's income is subjected to the influences of circumstances more due to the occupational differences and types of companies where they work for. The period effect suggests a shift from a planned economy to a more market-oriented economy with private companies employ most employees nowadays, the younger cohorts are less restricted by many social and institutional barriers. Lastly, I find Hukou is the major course for the inequality of opportunity in education, while gender is the major cause of inequality of opportunity in income. Noticeably, the effect of Hukou maybe includes rural-urban inequality which is common to many developing countries while the effect of gender cannot be entirely explained by gender discrimination in the job market but also stem from the fact the males disproportional obtain more degrees in STEM fields.

In sum, the novelty of this paper is to employ theories and techniques in a growing field of inequality of opportunity and provide a relatively comprehensive picture of inequality in both in income and educational attainment in China and the causes of them.

However, there are three major limitations in the field of measuring inequality of opportunity. They are robustness, accuracy, and dimensionality (Ferreira and Peragine, 2015). Specifically, in this paper, I could only measure a lower bound of inequality of opportunity, but how far the lower bound from the true value of it. One study by Niehues and Peichl (2014) uses panel data and fixed effect to try to get lower and upper bounds. Due to data limitation, I could not do the same exercise in this paper. Secondly, the literature of inequality of opportunity is very much in its infancy. Measuring Inequality of opportunity is still the a stage such as measuring outcome inequality before Lorenz Curve. There is no unifying principle in the field of inequality of opportunity., Different economists use different circumstances to capture the inequality of opportunity. Thus, I couldn't really do a comparison between my results and other results using a similar methodology in China or elsewhere.

In addition to the limitation in the theoretical framework, there are also several constraints in the case of China. Firstly, the Gini coefficients of net family asset and net family income are 0.771 and 0.567 respectively, compared to 0.43381 for that of personal income. This suggests, the degree of wealth inequality can be much more severe than income inequality. However, due to a typical Chinese family structure and vagueness of survey questions, I could not do a very good analysis of wealth inequality. Second, unlikely the public university system in Western Europe where governments fund public school more evenly, the quality of higher educational institutions varies widely and the Chinese government classifies public universities into several tiers. Those universities in top tier receive disproportional fund from the government. Consequently, the return of education in these few top universities is much higher. I use years of education as the outcome variable to measure inequality, but the quality of education does not go into the calculation.

Despite these limitations, this paper does provide some actionable information for policymaking.

First and foremost, a drastic modification, if not the abolishment of the Hukou system is needed to address rural-urban inequality. The government should focus on compensating migrant workers from various aspects such as health care, and children's education.

Secondly, the rising inequality of opportunity in income between genders in China deserves further study and attention from the government.

Lastly, many regional inequalities documented in this paper were likely caused by public policies from the Chinese central government. Putting one region as a "special economic zone" might disadvantage other areas. From the perspective of equality, the government should figure out a way to compensate for those disadvantaged areas. For example, Figure 2 shows that in recent year, some provinces in Central China have had high Gini coefficients in education, especially Sichuan province. Figure 3 indicates income inequality in the youngest cohort is most severe in Hubei and Jilin. The reasons behind the high Gini coefficients in these regions can vary from area to area. From the central government's perspective, to invest more in educational resources in Central China and give career support to certain regions are possible policies.

## References:


**Lefranc, Arnaud, Pistolesi, Nicolas and Trannoy, Alain** (2006). "Inequality of opportunities vs. inequality of outcomes: Are Western societies all alike?" Working Papers 54, ECINEQ, *Society for the Study of Economic Inequality*.

**Bentham, J., and Lafleur, L. J.** (1948). *An introduction to the principles of morals and legislation***.** New York: Hafner Pub. Co.

**Cohen, G.** (1989). *On the Currency of Egalitarian Justice.* Ethics, 99(4), 906-944

**Dworkin, Ronald** (1981). "What is equality? Part 2: Equality of resources." *Philosophy and Public Affairs* 10 (4):283 - 345.

**Bourguignon, F., Ferreira, F., and Menendez, M.** (2007)**. "**Inequality of opportunity in Brazil" *Review of Income and Wealth*, 53(4), 585–618.

**Ferreira, F. and Peragine, Vito**. (2015). "Equality of Opportunity: Theory and Evidence," Working papers 27, *Società Italiana di Economia Pubblica*.

**Ferreira, F., and Gignoux, J.** (2011). "The measurement of inequality of opportunity: Theory and an application to Latin America." *Review of Income and Wealth*, 57(4), 622–657.

**Ferreira, F. and Gignoux, J.** (2010). "Inequality of opportunity for education: the case of Turkey," Welfare and social policy analytical work program; working paper no. 4. Washington DC: World Bank.

**Golley, Jane and Kong, Sherry Tao**, (2018). "Inequality of opportunity in China's educational outcomes," *China Economic Review*, Elsevier, vol. 51(C), pages 116-128.

**Golley, Jane and Kong, Sherry Tao,** (2013). "Inequality in Intergenerational Mobility of Education in China," *China and World Economy*, vol. 21(2), pages 15-37, March.

**Hassine Belghith, Nadia and Zeufack, Albert,** (2015) "Inequality of Outcomes and Inequality of Opportunity in Tanzania." World Bank Policy Research Working Paper No. 7260.

**Niehues, Judith and Peichl**, **Andreas** (2014). "Upper bounds of inequality of opportunity: theory and evidence for Germany and the US," *Social Choice and Welfare, The Society for Social Choice and Welfare*, vol. 43(1), pages 73-99, June.

**Kam Wing Chan** (2014) "China's urbanization 2020: a new blueprint and direction" *Eurasian Geography and Economics*, 55:1, 1-9

**Qian, Nancy** (2008). "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance," *The Quarterly Journal of Economics*, Oxford University Press, vol. 123(3), pages 1251-1285.

**Barro, Robert J. and Sala-i-Martin, Xavier** (1978): "Unanticipated Money, Output, and the Price Level in the United States," *Journal of Political Economy*, Vol. 86, no. 4 (1978): 549- 580

**Rawls, John,** (1971). *A theory of justice.* Cambridge, Mass.: Belknap Press of Harvard University Press

**Roemer, John E.** (1993). "A pragmatic theory of responsibility for the egalitarian planner." *Philosophy and Public Affairs* 22 (2):146-166

**De Barros, R. P., Ferreira, F.,Molinas, J. and Saavedra, J.** (2009). "Measuring inequality of opportunity in Latin America and the Caribbean." Washington DC: Palgrave and Mac-Millan and The World Bank.

**Yang, D. T., and Cai, F.** (in press). The political economy of China's rural – urban divide

**Kanbur, R., and Zhang, X.** (1999). "Which regional inequality? The evolution of rural – urban and inland – coastal inequality in China from 1983 to 1995." *Journal of Comparative Economics*, 27(4), 686 – 701.

**Wang, J.** (2013). "The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities." *Journal of Development Economics* Volume 101, March 2013, Pages 133-147

**State Council of the People's Republic of China** (1985). "*Suggestions of the Central Committee of the Communist Party of China on Formulating the Seventh Five-Year Plan for National Economic and Social Development"*

**Ying and Lv** (2018). "What Are the Reasons for Unequal Opportunity in Urban Residents' Income Distribution." *Statistics Research* Vol. 35, No. 9

**Yu Xie and Xiang Zhou** (2014). "Income inequality in China," *Proceedings of the National Academy of Sciences*, 111 (19) 6928-6933

**Zhang, Yingqiang and Eriksson, Tor** (2010). "Inequality of opportunity and income inequality in nine Chinese provinces, 1989-2006," *China Economic Review*, Elsevier, vol. 21(4), pages 607-616, December.

**Zhao, S**. (1993). "Deng Xiaoping's Southern Tour: Elite Politics in Post-Tiananmen China." *Asian Survey*, 33(8), 739-756.

Table 1: Descriptive Statistics: Education

| Birth cohort | Highest degree obtained (% of child cohort) Illiterate | primary High | Junior school | High or above | College | Average years of schooling | number of observations |
|---|---|---|---|---|---|---|---|
| Panel A: Whole sample | | | | | | | |
| 40-44 | 40.05 | 28.28 | 14.32 | 6.27 | 3.08 | 4.81 | 845 |
| 45-49 | 49.21 | 28.87 | 15.57 | 5.64 | 1.50 | 4.70 | 1400 |
| 50-54 | 50.20 | 21.73 | 19.04 | 6.73 | 2.29 | 4.84 | 2006 |
| 55-59 | 36.65 | 14.21 | 27.20 | 19.92 | 2.03 | 6.42 | 1978 |
| 60-64 | 23.45 | 18.67 | 33.14 | 20.37 | 4.37 | 7.73 | 2405 |
| 65-69 | 27.77 | 28.28 | 30.59 | 8.99 | 4.36 | 6.86 | 2981 |
| 70-74 | 24.29 | 27.24 | 30.91 | 10.36 | 7.20 | 7.37 | 2944 |
| 75-79 | 16.14 | 23.79 | 34.21 | 14.40 | 11.45 | 8.58 | 2236 |
| 80-84 | 10.07 | 21.65 | 36.33 | 14.81 | 17.13 | 9.70 | 2323 |
| 85-89 | 6.49 | 18.11 | 33.42 | 16.63 | 25.35 | 10.65 | 2959 |
| 90-94 | 4.65 | 13.47 | 27.28 | 23.49 | 31.11 | 12.02 | 2665 |
| Panel B: sample with urban household registration | | | | | | | |
| 40-44 | 26.29 | 27.89 | 19.92 | 15.94 | 9.96 | 7.59 | 251 |
| 45-49 | 27.06 | 25.97 | 28.14 | 14.50 | 4.33 | 7.15 | 462 |
| 50-54 | 24.48 | 19.93 | 34.27 | 13.46 | 7.87 | 7.66 | 572 |
| 55-59 | 15.87 | 12.46 | 31.57 | 34.30 | 5.80 | 8.88 | 586 |
| 60-64 | 5.30 | 9.35 | 31.31 | 39.72 | 14.33 | 10.64 | 642 |
| 65-69 | 6.27 | 13.01 | 36.99 | 24.76 | 18.97 | 10.32 | 638 |
| 70-74 | 5.14 | 9.86 | 33.92 | 25.81 | 25.27 | 10.92 | 748 |
| 75-79 | 2.26 | 7.92 | 27.46 | 29.40 | 32.96 | 11.91 | 619 |
| 80-84 | 1.58 | 5.06 | 21.04 | 23.89 | 48.42 | 12.93 | 632 |
| 85-89 | 1.38 | 5.84 | 12.90 | 20.28 | 59.60 | 13.67 | 651 |
| 90-94 | 1.36 | 3.68 | 7.95 | 24.42 | 62.60 | 14.36 | 516 |
| Panel C: sample with rural household registration | | | | | | | |
| 40-44 | 57.24 | 28.45 | 11.95 | 2.19 | 0.17 | 3.63 | 594 |
| 45-49 | 60.13 | 29.10 | 9.38 | 1.28 | 0.11 | 3.48 | 938 |
| 50-54 | 60.46 | 22.45 | 12.97 | 4.04 | 0.07 | 3.71 | 1434 |
| 55-59 | 45.40 | 14.94 | 25.36 | 13.86 | 0.43 | 5.38 | 1392 |
| 60-64 | 30.06 | 22.06 | 33.81 | 13.33 | 0.74 | 6.67 | 1763 |
| 65-69 | 33.63 | 32.44 | 28.85 | 4.69 | 0.38 | 5.91 | 2343 |
| 70-74 | 30.72 | 33.08 | 29.90 | 5.17 | 1.13 | 6.16 | 2204 |
| 75-79 | 21.46 | 29.87 | 36.88 | 8.66 | 3.22 | 7.24 | 1617 |
| 80-84 | 13.25 | 27.85 | 42.05 | 11.41 | 5.44 | 8.34 | 1691 |
| 85-89 | 7.93 | 21.58 | 39.21 | 15.60 | 15.68 | 9.69 | 2308 |
| 90-94 | 5.44 | 15.82 | 31.92 | 23.27 | 23.55 | 11.33 | 2149 |

Table 2: Descriptive Statistics: Income

| birth cohort | mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation | mean | Standard deviation |
|---|---|---|---|---|---|---|---|---|
| | Whole sample | | Urban | | Rural | | Male | |
| Whole | 28773 | 268 | 34229 | 521 | 25460 | 282 | 32417 | 364 |
| 40-44 | 10406 | 1145 | 10362 | 2137 | 10419 | 1365 | 11451 | 1342 |
| 45-49 | 14962 | 1129 | 15471 | 1738 | 14763 | 1422 | 16207 | 1348 |
| 50-54 | 20280 | 1048 | 27428 | 2493 | 16635 | 860 | 22716 | 1277 |
| 55-59 | 25322 | 894 | 30756 | 1505 | 20233 | 915 | 27057 | 1004 |
| 60-64 | 26765 | 692 | 31239 | 1126 | 22817 | 800 | 30534 | 894 |
| 65-69 | 27240 | 746 | 32421 | 1513 | 24218 | 765 | 31308 | 1033 |
| 70-74 | 29445 | 811 | 32880 | 1216 | 26923 | 1077 | 34284 | 1217 |
| 75-79 | 32207 | 852 | 37807 | 1649 | 28328 | 846 | 37670 | 1243 |
| 80-84 | 36499 | 974 | 43111 | 1804 | 31331 | 962 | 40891 | 1330 |
| 85-89 | 33148 | 668 | 38529 | 1374 | 30221 | 695 | 38146 | 863 |
| 90-94 | 23564 | 607 | 24043 | 1600 | 23427 | 633 | 27007 | 924 |
| | Female | | West | | Central | | East | |
| whole | 23160 | 364 | 23633 | 464 | 26068 | 428 | 32463 | 424 |
| 40-44 | 6750 | 1721 | 8160 | 1376 | 32207.9 | 4.81 | 18030 | 1742 |
| 45-49 | 10078 | 1387 | 9776 | 1512 | 12382 | 1731 | 20862 | 1233 |
| 50-54 | 11547 | 979 | 21460 | 4125 | 18678 | 1731 | 28271 | 1342 |
| 55-59 | 18757 | 1854 | 20919 | 1810 | 21090 | 1096 | 30513 | 1137 |
| 60-64 | 19813 | 957 | 20888 | 1131 | 24412 | 991 | 30045 | 1254 |
| 65-69 | 21259 | 984 | 22355 | 1153 | 26091 | 1127 | 30045 | 1254 |
| 70-74 | 22616 | 851 | 24725 | 1334 | 25681 | 1281 | 36692 | 1386 |
| 75-79 | 24878 | 992 | 25316 | 1469 | 29946 | 1331 | 36692 | 1386 |
| 80-84 | 30537 | 1372 | 26866 | 1137 | 30760 | 1425 | 42649 | 1553 |
| 85-89 | 26774 | 997 | 28090 | 1325 | 30416 | 1259 | 36892 | 950 |
| 90-94 | 19431 | 695 | 20454 | 1152 | 22559 | 1045 | 25556 | 933 |

Table 3: Definition of Variables

| Variables | Definition | Mean or percentage | Standard deviation | Minimum value | Maximum value |
|---|---|---|---|---|---|
| Year of education | Numerical | 8.00 | 4.748 | 0 | 22 |
| Father with primary school | Categorical Primary school=1 | 24% | 0.43 | 0 | 1 |
| Father with junior high school | Categorical Junior high=1 | 27% | 0.43 | 0 | 1 |
| Father with senior high school of above | Categorical Senior high school=1 | 15% | 0.35 | 0 | 1 |
| Hukou | Categorical Urban hukou=1 | 22% | 0.42 | 0 | 1 |
| gender | categorical Male=1 | Male 53.43% | 0.51 | 0 | 1 |
| Family size | Numerical | 4.54 | 2.11 | 1 | 19 |
| Province | Categorical, include 31 provinces dummy | Gansu (12.46%) Henan (12.44%) …… | None | 0 | 1 |
| Birth cohort | Categorical, include 11 dummies | Birth between 1990-1994 (9.38%) …… | None | 0 | 1 |
| Relative income level | Categorical, include 5 dummies, | Medium level (37.85%) Relatively low (25.10%) ….. | none | 0 | 1 |

Table 4: Regression Table: Education (part 1)

| Birth cohort: | Whole | 40-45 | 46-49 | 50-54 | 55-59 | 60-64 |
|---|---|---|---|---|---|---|
| Variable: | Dependent variable: year of education | | | | | |
| Hukou | 2.616*** | 3.258*** | 3.226*** | 3.419*** | 3.060*** | 3.415*** |
| | (0.0593) | (0.316) | (0.224) | (0.209) | (0.208) | (0.164) |
| Gender | 1.356*** | 2.600*** | 1.910*** | 2.616*** | 2.824*** | 2.139*** |
| | (0.0507) | (0.257) | (0.191) | (0.168) | (0.188) | (0.152) |
| East | 0.827*** | 1.329*** | 1.221*** | 1.085*** | 1.700*** | 0.776*** |
| | (0.0638) | (0.321) | (0.241) | (0.218) | (0.242) | (0.200) |
| Father with primary school | 2.688*** | 2.203*** | 1.392*** | 1.686*** | 1.704*** | 1.407*** |
| | (0.0620) | (0.320) | (0.242) | (0.209) | (0.208) | (0.171) |
| Father with junior high school | 4.324*** | 2.314** | 2.094*** | 1.280*** | 1.447*** | 1.493*** |
| | (0.0690) | (0.905) | (0.519) | (0.384) | (0.370) | (0.283) |
| Father with senior high school above | 4.663*** | 2.473*** | 2.576*** | 2.424*** | 1.082** | 2.011*** |
| | (0.0846) | (0.797) | (0.499) | (0.404) | (0.464) | (0.313) |
| Central | 0.551*** | 0.610* | 0.613** | 0.890*** | 1.501*** | 0.841*** |
| | (0.0677) | (0.333) | (0.263) | (0.230) | (0.271) | (0.206) |
| Constant | 4.097*** | 1.114*** | 1.400*** | 1.218*** | 2.190*** | 4.529*** |
| | (0.0635) | (0.274) | (0.210) | (0.190) | (0.225) | (0.192) |
| Observations | 25,271 | 833 | 1,381 | 1,985 | 1,945 | 2,309 |
| R-squared | 0.288 | 0.313 | 0.278 | 0.285 | 0.252 | 0.274 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Regression Table: Education (part 2)

| Birth cohort: | 65-79 | 70-74 | 75-79 | 80-84 | 85-89 | 90-94 |
|---|---|---|---|---|---|---|
| Variable: | Dependent variable: year of education | | | | | |
| Hukou | 3.690*** | 3.882*** | 3.597*** | 3.930*** | 3.226*** | 2.212*** |
| | (0.158) | (0.166) | (0.178) | (0.172) | (0.168) | (0.168) |
| Gender | 1.269*** | 1.706*** | 1.250*** | 0.657*** | 0.423*** | -0.342** |
| | (0.131) | (0.133) | (0.156) | (0.152) | (0.145) | (0.153) |
| East region | 1.010*** | 1.776*** | 2.186*** | 1.949*** | 1.552*** | 1.020*** |
| | (0.160) | (0.167) | (0.204) | (0.202) | (0.186) | (0.191) |
| Father with primary school | 1.713*** | 1.504*** | 1.330*** | 1.301*** | 1.314*** | 1.750*** |
| | (0.150) | (0.160) | (0.195) | (0.223) | (0.232) | (0.269) |
| Father with junior high school | 1.974*** | 2.032*** | 2.033*** | 1.780*** | 2.268*** | 2.576*** |
| | (0.218) | (0.193) | (0.226) | (0.223) | (0.223) | (0.248) |
| Father with senior high school above | 2.310*** | 2.634*** | 2.698*** | 2.340*** | 3.127*** | 3.208*** |
| | (0.226) | (0.258) | (0.268) | (0.249) | (0.244) | (0.272) |
| Central region | 1.152*** | 0.892*** | 1.364*** | 1.442*** | 0.714*** | 0.640*** |
| | (0.173) | (0.175) | (0.219) | (0.212) | (0.195) | (0.202) |
| Constant | 3.760*** | 3.535*** | 4.443*** | 5.632*** | 7.051*** | 9.128*** |
| | (0.153) | (0.160) | (0.204) | (0.244) | (0.244) | (0.279) |
| Observations | 2,799 | 2,610 | 1,872 | 1,851 | 2,203 | 2,045 |
| R-squared | 0.290 | 0.360 | 0.378 | 0.364 | 0.291 | 0.212 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Regression Table: Income (part 1)

| Birth cohort: | Whole | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 |
|---|---|---|---|---|---|---|
| Variable: | Dependent variable: year of education | | | | | |
| Hukou | 0.161*** | -0.184 | 0.0396 | 0.400*** | 0.461*** | 0.213*** |
| | (0.0207) | (0.383) | (0.192) | (0.108) | (0.0702) | (0.0719) |
| Gender | 0.447*** | 0.806* | 0.515* | 0.678*** | 0.710*** | 0.613*** |
| | (0.0199) | (0.420) | (0.265) | (0.170) | (0.108) | (0.0702) |
| East region | 0.283*** | 0.343 | 0.725*** | 0.0426 | 0.315*** | 0.455*** |
| | (0.0260) | (0.297) | (0.257) | (0.160) | (0.104) | (0.0862) |
| Father with primary school | 0.0461** | -0.393 | -0.0100 | -0.152 | 0.123 | -0.0145 |
| | (0.0234) | (0.258) | (0.178) | (0.140) | (0.0830) | (0.0781) |
| Father with junior high school | 0.0149 | -0.666*** | 0.185 | -0.256 | 0.0129 | 0.0761 |
| | (0.0273) | (0.195) | (0.152) | (0.260) | (0.112) | (0.104) |
| Father with senior high school above | 0.0385 | | 0.0529 | 0.336 | 0.189 | -0.0781 |
| | (0.0318) | | (0.351) | (0.544) | (0.177) | (0.0989) |
| Year of education | 0.0493*** | 0.0298 | 0.0543** | 0.0285* | 0.0401*** | 0.0644*** |
| | (0.00288) | (0.0447) | (0.0241) | (0.0145) | (0.00997) | (0.0127) |
| Central Region | 0.0681** | -0.159 | 0.209 | -0.0798 | -0.122 | 0.172* |
| | (0.0278) | (0.310) | (0.307) | (0.160) | (0.120) | (0.0893) |
| Constant | 8.925*** | 8.192*** | 8.064*** | 8.703*** | 8.496*** | 8.489*** |
| | (0.0361) | (0.452) | (0.359) | (0.177) | (0.163) | (0.152) |
| Observations | 9,273 | 36 | 128 | 374 | 579 | 895 |
| R-squared | 0.130 | 0.223 | 0.175 | 0.139 | 0.240 | 0.200 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Regression Table: Income (part 2)

| Birth cohort: | 65-69 | 70-74 | 75-79 | 80-84 | 85-89 | 90-94 |
|---|---|---|---|---|---|---|
| Variable: | Panel A: national wide | | | | | |
| Hukou | 0.109* | 0.130** | 0.151** | 0.0558 | 0.101** | -0.0910 |
| | (0.0592) | (0.0593) | (0.0635) | (0.0568) | (0.0496) | (0.0862) |
| Gender | 0.470*** | 0.452*** | 0.514*** | 0.423*** | 0.530*** | 0.332*** |
| | (0.0537) | (0.0501) | (0.0545) | (0.0480) | (0.0470) | (0.0663) |
| East region | 0.320*** | 0.395*** | 0.301*** | 0.336*** | 0.234*** | 0.255*** |
| | (0.0645) | (0.0645) | (0.0738) | (0.0615) | (0.0668) | (0.0905) |
| Father with primary school | 0.00524 | -0.0556 | 0.0816 | 0.0118 | -0.0355 | 0.0764 |
| | (0.0666) | (0.0543) | (0.0631) | (0.0609) | (0.0665) | (0.0812) |
| Father with junior high school | 0.0282 | 0.0113 | -0.0886 | 0.0113 | -0.0359 | -0.0543 |
| | (0.0755) | (0.0711) | (0.0808) | (0.0582) | (0.0571) | (0.0900) |
| Father with senior high school above | 0.0654 | -0.0159 | 0.149* | 0.00516 | -0.0288 | -0.0403 |
| | (0.0921) | (0.0926) | (0.0855) | (0.0740) | (0.0599) | (0.108) |
| Year of education | 0.0438*** | 0.0476*** | 0.0301*** | 0.0634*** | 0.0551*** | -0.000298 |
| | (0.00725) | (0.00813) | (0.00789) | (0.00787) | (0.00813) | (0.0104) |
| Central Region | 0.173** | 0.0484 | 0.0890 | 0.0812 | 0.0981 | 0.137 |
| | (0.0689) | (0.0721) | (0.0761) | (0.0695) | (0.0659) | (0.0932) |
| Constant | 8.991*** | 9.036*** | 9.223*** | 9.026*** | 8.991*** | 9.384*** |
| | (0.0810) | (0.0933) | (0.0924) | (0.105) | (0.0968) | (0.127) |
| Observations | 1,146 | 1,265 | 1,012 | 1,098 | 1,524 | 934 |
| R-squared | 0.139 | 0.166 | 0.149 | 0.189 | 0.132 | 0.039 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Total Inequality of Opportunity

| Birth cohort | Total inequality | | Inequality of opportunity | |
|---|---|---|---|---|
| | GE(2) | Gini | $IOA_D$ | $IOR_D$ |
| | Panel A: year of education | | | |
| Whole | 0.176 | 0.329 | 0.050 | 0.284 |
| 40-44 | 0.426 | 0.506 | 0.113 | 0.265 |
| 45-49 | 0.392 | 0.489 | 0.105 | 0.268 |
| 50-54 | 0.422 | 0.506 | 0.114 | 0.270 |
| 55-59 | 0.274 | 0.410 | 0.080 | 0.292 |
| 60-64 | 0.152 | 0.298 | 0.041 | 0.270 |
| 65-69 | 0.178 | 0.325 | 0.055 | 0.310 |
| 70-74 | 0.165 | 0.314 | 0.060 | 0.364 |
| 75-79 | 0.119 | 0.265 | 0.047 | 0.395 |
| 80-84 | 0.087 | 0.225 | 0.038 | 0.437 |
| 85-89 | 0.069 | 0.202 | 0.025 | 0.362 |
| 90-94 | 0.051 | 0.172 | 0.017 | 0.333 |
| | Panel B: personal individual | | | |
| Whole | 0.402 | 0.403 | 0.059 | 0.147 |
| 40-44 | 0.214 | 0.336 | 0.122 | 0.573 |
| 45-49 | 0.362 | 0.421 | 0.108 | 0.298 |
| 50-54 | 0.501 | 0.434 | 0.116 | 0.232 |
| 55-59 | 0.361 | 0.390 | 0.139 | 0.385 |
| 60-64 | 0.299 | 0.387 | 0.107 | 0.359 |
| 65-69 | 0.433 | 0.395 | 0.053 | 0.122 |
| 70-74 | 0.480 | 0.395 | 0.065 | 0.135 |
| 75-79 | 0.355 | 0.380 | 0.054 | 0.152 |
| 80-84 | 0.392 | 0.388 | 0.065 | 0.166 |
| 85-89 | 0.310 | 0.380 | 0.062 | 0.199 |
| 90-94 | 0.310 | 0.401 | 0.018 | 0.059 |

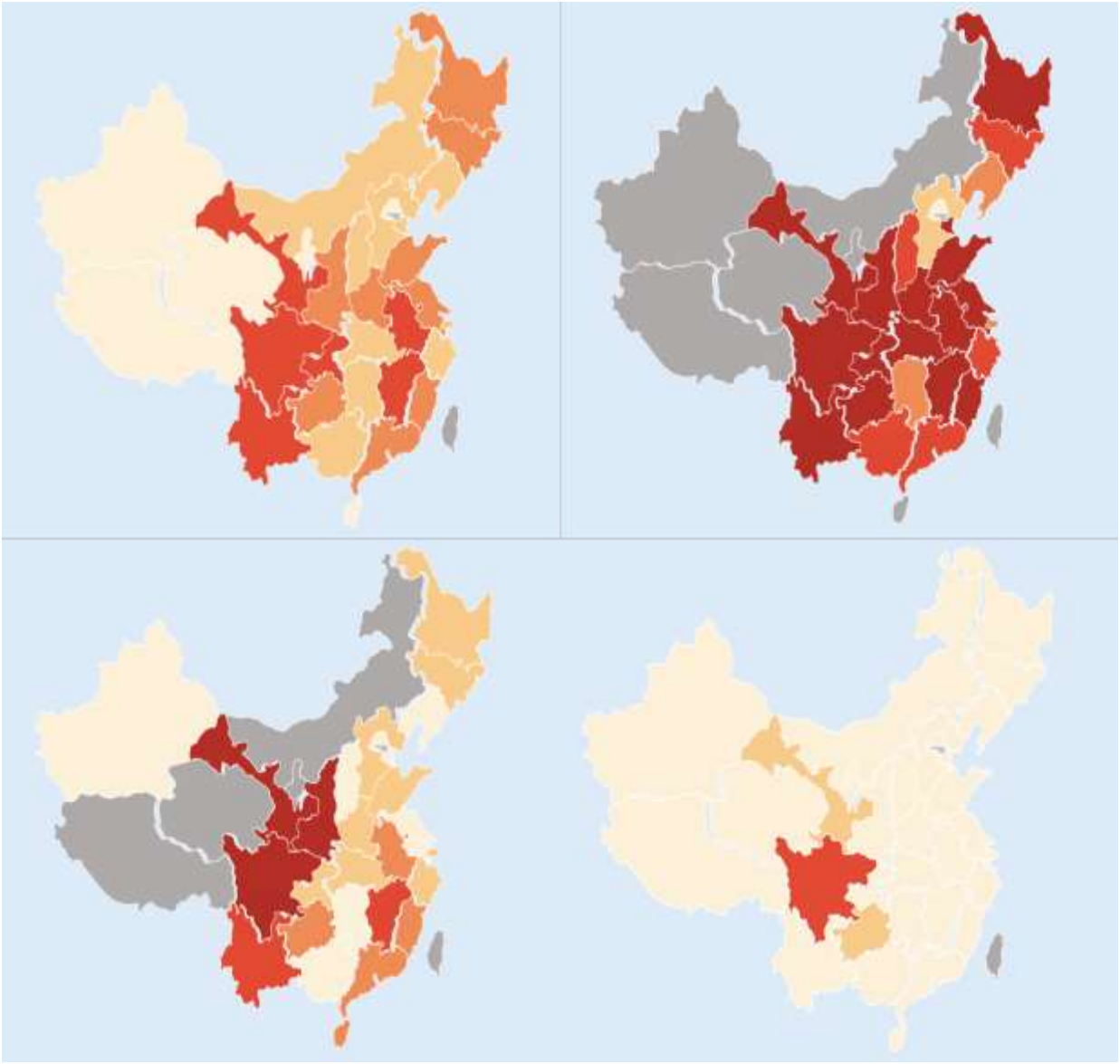
Note: the first column shows birth cohorts. Starting from group of people who were born between 1940 and 1944 to the younger cohorts.

Table 9: Partial Contribution to Inequality of Opportunity

| | Education | | | Income | | |
|---|---|---|---|---|---|---|
| Birth cohort | Hukou | Gender | East | Hukou | Gender | East |
| Whole | 18.52 | 16.57 | 1.81 | 3.63 | 29.40 | 9.26 |
| 40-44 | 30.88 | 32.99 | 7.91 | 6.35 | 85.31 | 12.01 |
| 45-49 | 43.6 | 25.03 | 9.95 | -3.03 | 15.90 | 42.00 |
| 50-54 | 34.89 | 35.52 | 7.53 | 10.24 | 21.40 | -0.79 |
| 55-59 | 31.28 | 42.30 | 11.65 | 32.09 | 35.62 | 9.42 |
| 60-64 | 39.52 | 26.18 | 3.21 | 10.43 | 30.38 | 22.39 |
| 65-69 | 42.86 | 15.41 | 3.96 | 3.01 | 28.13 | 19.91 |
| 70-74 | 35.28 | 19.06 | 4.67 | 0.10 | 22.00 | 10.31 |
| 75-79 | 35.28 | 12.73 | 6.70 | 1.26 | 40.31 | 3.67 |
| 80-84 | 49.82 | 10.03 | 8.10 | -1.39 | 24.31 | 10.00 |
| 85-89 | 39.31 | 6.91 | 8.34 | -4.89 | 42.00 | 2.62 |
| 90-94 | 22.12 | 6.95 | 2.48 | 0.07 | 62.84 | 9.60 |

Note: the first column shows birth cohorts. Starting from group of people who were born between 1940 and 1944 to the younger cohorts.

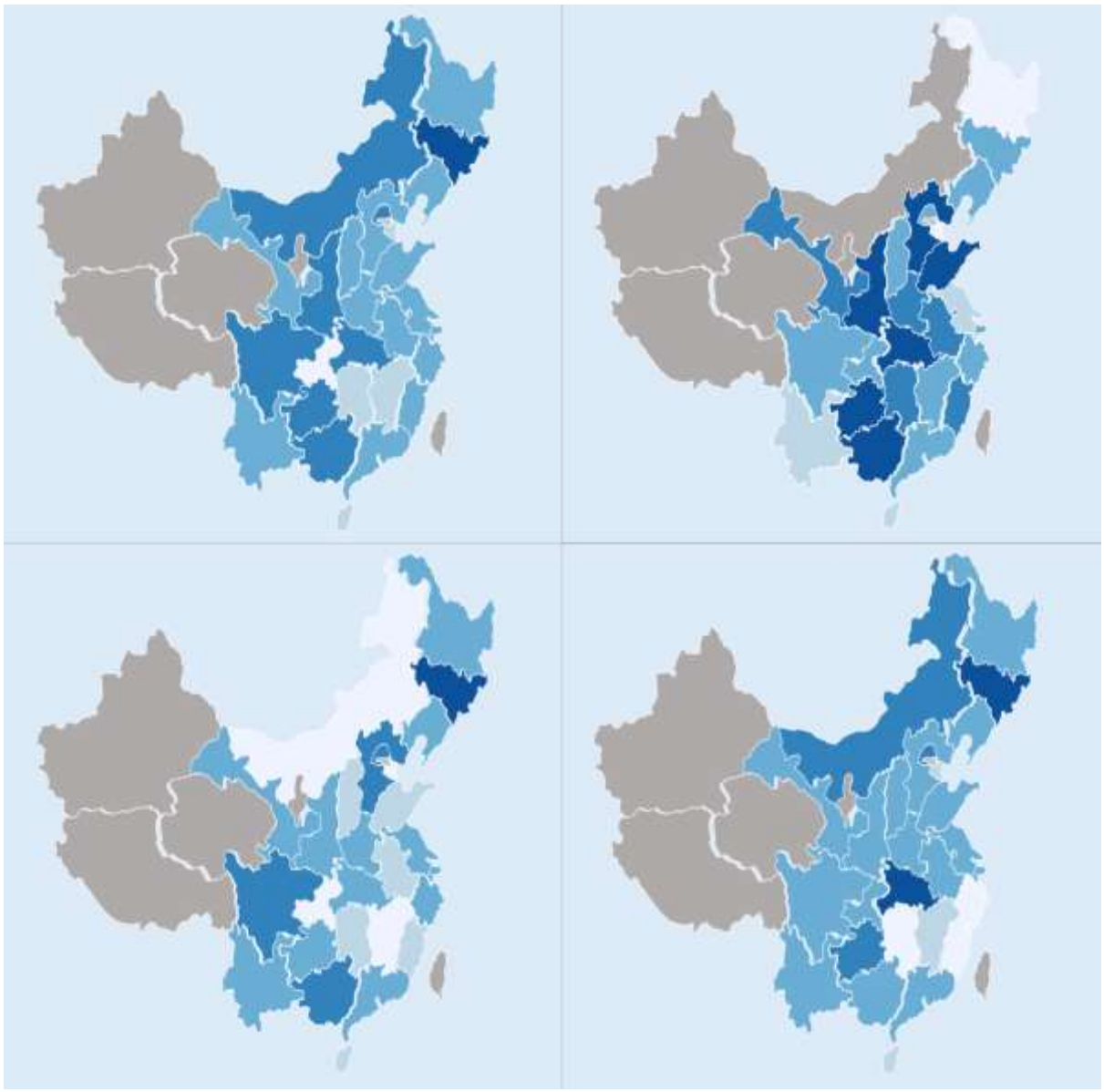
Figure 1: Development of the Gini coefficient in Education



Note: the map documents the change of within-province Gini coefficient (in percentage) in education over the time, those migrants from other provinces are also included. The lower left corner is the full sample map. The upper left corner is the people in the cohort who born between 1950 and 1960. The upper right corner is the people in the cohort who born between 1961 and 1980. The lower right corner the people in the cohort who born between 1981 and 1994.

| | Min. | Max. |
|---|---|---|
| | 0 | 23 |
| | 23 | 28 |
| | 29 | 35 |
| | 35 | 46 |
| | 46 | 68 |

Figure 2: Development of the Gini Coefficient in Education



Note: the map documents the change of within province Gini coefficient (in percentage) in education over the time, those migrants from other provinces are also included. The lower left corner is the full sample map. The upper left corner is the people in the cohort who born between 1950 and 1960. The upper right corner is the people in the cohort who born between 1961 and 1980. The lower right corner the people in the cohort who born between 1981 and 1994.

| | Min. | Max. |
|---|---|---|
| | 0 | 32 |
| | 32 | 35 |
| | 35 | 40 |
| | 40 | 45 |
| | 45 | 51 |

Figure 3: Cumulative Distribution of Year of Education by Background

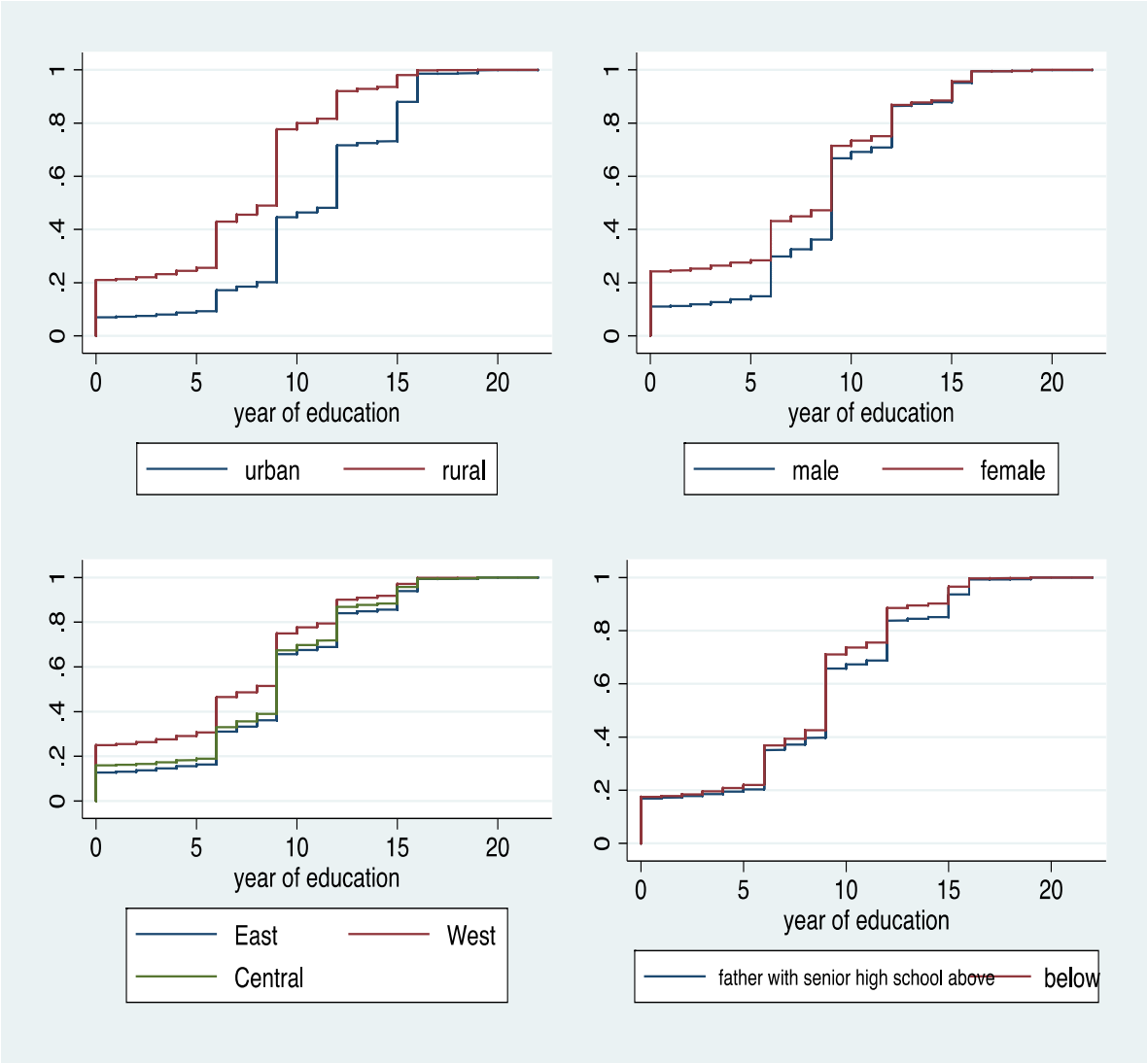


Figure 4: Cumulative Distribution of personal income by Backgrounds

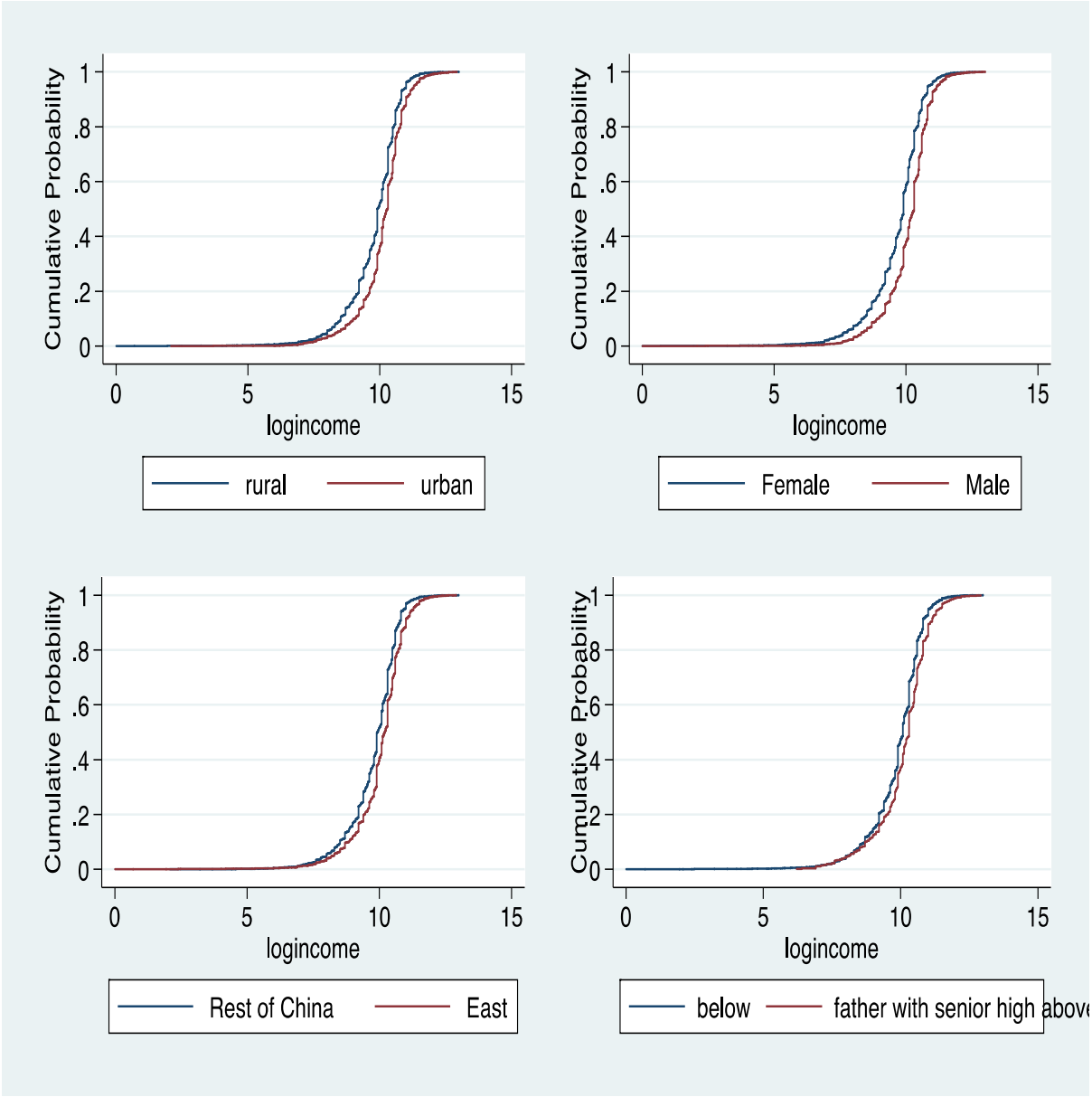


Figure 5: Lorenz Curve for Education by Backgrounds

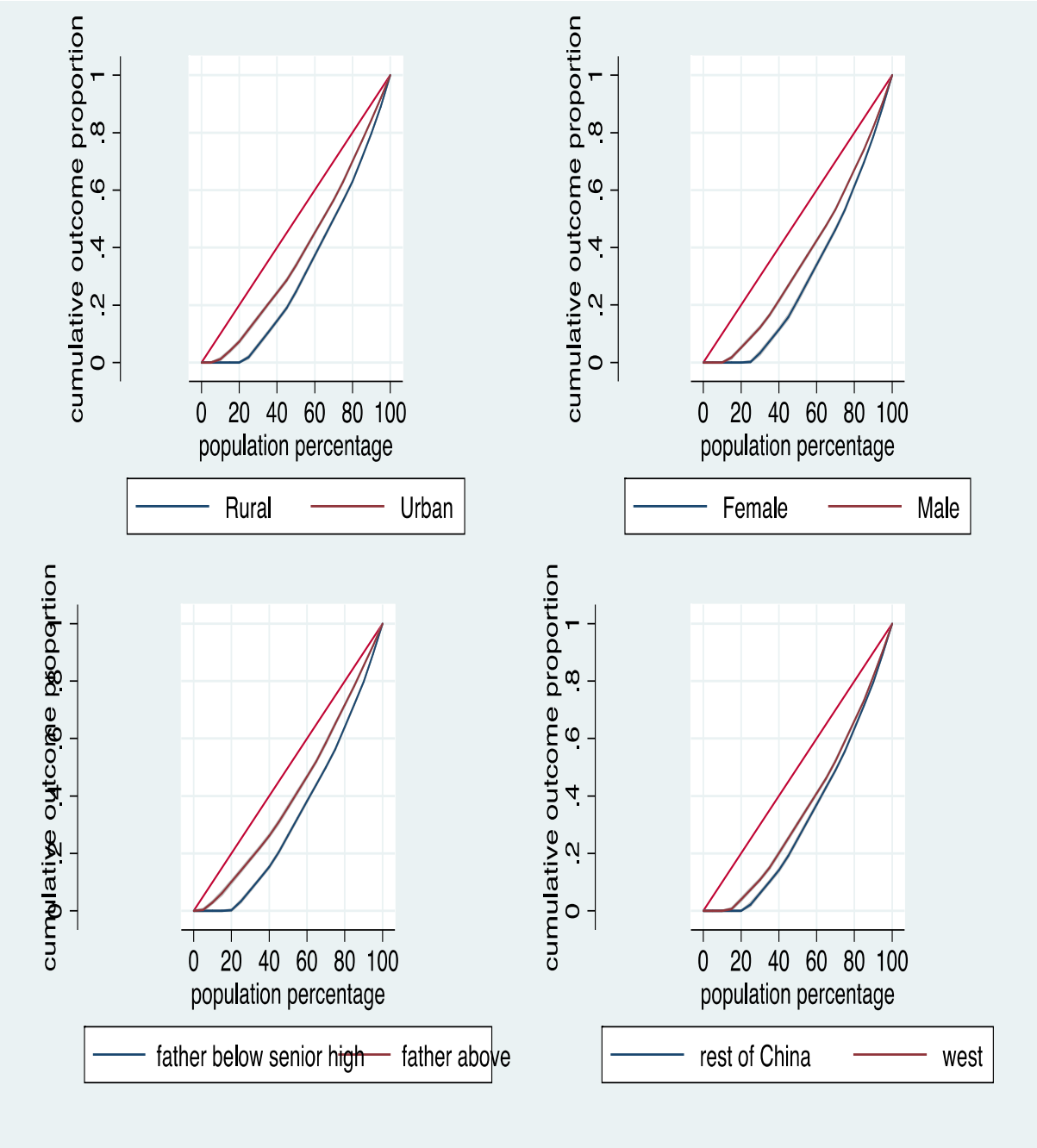


Figure 6: Lorenz Curve for Income by backgrounds