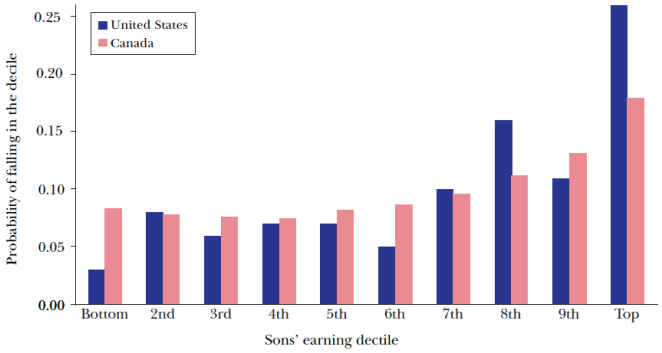
Research Proposal

State the research question:

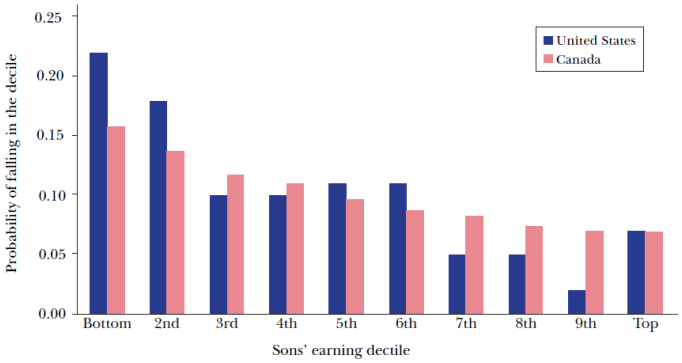
- Research Question? *How does education influence the economic intergenerational mobility?*
- Why is the question interesting? Though I will base my study upon a Chinese dataset, I believe this question has a broader range of interest, both for China and the USA. I believe the most essential component of the so called 'American Dream' is an individual's ability to succeed whatever the socioeconomic circumstance one was born in (Corak, 2013). The same is true for other developing countries such as China. However, recent data indicates that personal human capital investments and self-endeavor has become a less effective factor that can drive one up and different levels across the economic spectrum are more stagnantly separated. This fact is concerning all of us, and it is intriguing, as well as crucial to test the statement empirically and determine if there's any way to get over this social dilemma. For example, the two following graphs documented the intergenerational mobility in USA and Canada.

Figure 2
Earnings Deciles of Sons Born to Top Decile Fathers: United States and Canada



Source: Corak and Heisz (1999, table 6); Mazumder (2005, table 2.2).

Figure 3
Earnings Deciles of Sons Born to Bottom Decile Fathers: United States and Canada



Source: Corak and Heisz (1999, table 6); Mazumder (2005, table 2.2).

- What will be my contributions? My course research project will be comprised of a section that describe the magnitude of intergenerational economic mobility. As constrained by my data sample, I cannot perform an extensive search of the empirical evidence of the prominence of

such phenomenon based on my standalone. I will probably compare my findings with other external research. Additionally, I will try using some computationally enhanced methods to identify feature importance and feature impacts. Ideas I have in mind now include: use various income and wealth data to construct a more comprehensive measure of socioeconomic hierarchy; besides the widely applied Two-Stage-Least-Square method that studies the channel effects of education, I can replace the first stage prediction with a more advanced machine learning technique.

How to answer the research question:

- What data to use? For now I have two sets of data for candidacy. The first data produced by CHIP, short of China Household Income Project, which is composed of three rounds of surveys that collected data on "personal income and related economic factors in both rural and urban areas of the People's Republic of China" (Li, 2009). The most attractive attribute of the data is that it is a longitudinal one, which allows researchers to keep track of one family or individual over a longer horizon of time, and the lagged effect of parents' economic status is one of the variables of interest. But the records of background socioeconomic details are very constrained. The other dataset is the China Urban Household Survey data, which collected annual cross-sectional data from non-agricultural households in different cities and counties. It records a richer set of income and expenditure information and their subcategories, also more family-related matters.
- What theories to draw upon for interpretation of the data? 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 determined by parents' investments in children, as well we the endowment luck, in other words, the inheritability of wealth. The two aspects of latent variables lead to observations of education and parents' earning power. Admittedly, education and parents' income can not reflective the entirety of these concepts, but it doesn't hurt to implement our analysis through the lens, which has been widely employed across academia.

- What computational methods to develop to help our analysis? As I devised in my 'contributions' part, I will try using dimension reduction techniques to construct a more comprehensive measure of economic superiority, which should be able to aggregate more information in an effective way relative to conventional methods. Also, to unveil the channel effect through education, we plan to practice some more flexible models, such as single vector regression or CART, to testify more complicated and non-linear relationships among all the variables.

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.

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