Intergeneration Mobility: Facts and Factors

By Ken Chen
Computational Social Science, University of Chicago

Background

Intimidating Facts? In addition to the growing social inequality problems that large-scale economies, like US and China, are struggling with, intergeneration mobility has become a an issue that goes hand in hand. Polarization from last generation is likely to persist into, or even enlarged during the next generation. Corak's (2016) work finds that the intergenerational income elasticity was 0.47 for the US, and 0.6 for China. **The first objective of this project is a descriptive study of the intergenerational mobility.**

Predictive Factors? There are a rich set of candidate factors that could offset this unwanted economic inheritance. Possibly, certain employment status, industries that people work in can somehow enrich social fluidity. Among them, education has been win most interest, which has been manifested to work positively for intergenerational mobility, but their access are profoundly intertwined with parents' economic status. **The second objective of the project is to determine the most influential factors relating to children's income.**

Data & Method

Data: My data is a survey dataset, China Household Survey The survey provides very detailed data on income and expenditure of the urban households in China. We used year 2002 to 2006. We merged the tables on the individual identifying index, and finally coined a dataset, where each record represents a child, while parent and family variables are merged at the tail accordingly.

Principal Component Analysis: Thanks to the dataset, we now have various sources of incomes, like labor income, income from self-organized business… Now we can construct a uniform metric that measures children and parents' intra-group economic status. This allows us to treat the multidimensional income metrics as univariate, and thus become a Y of the following model.

Random Forest Classifier: Based on what we found in the PCA step, we've divided up the children into five layers of economic classes. Then we came up with a random forest model, with additional power from other family and socioeconomic background information, to predict where children will end up being. Furthermore, we are interested in the most predictive factors.

*Our analysis focuses on children who are in the early phases of their professional careers, due to the cross-sectional structure of our dataset.

Principal Components
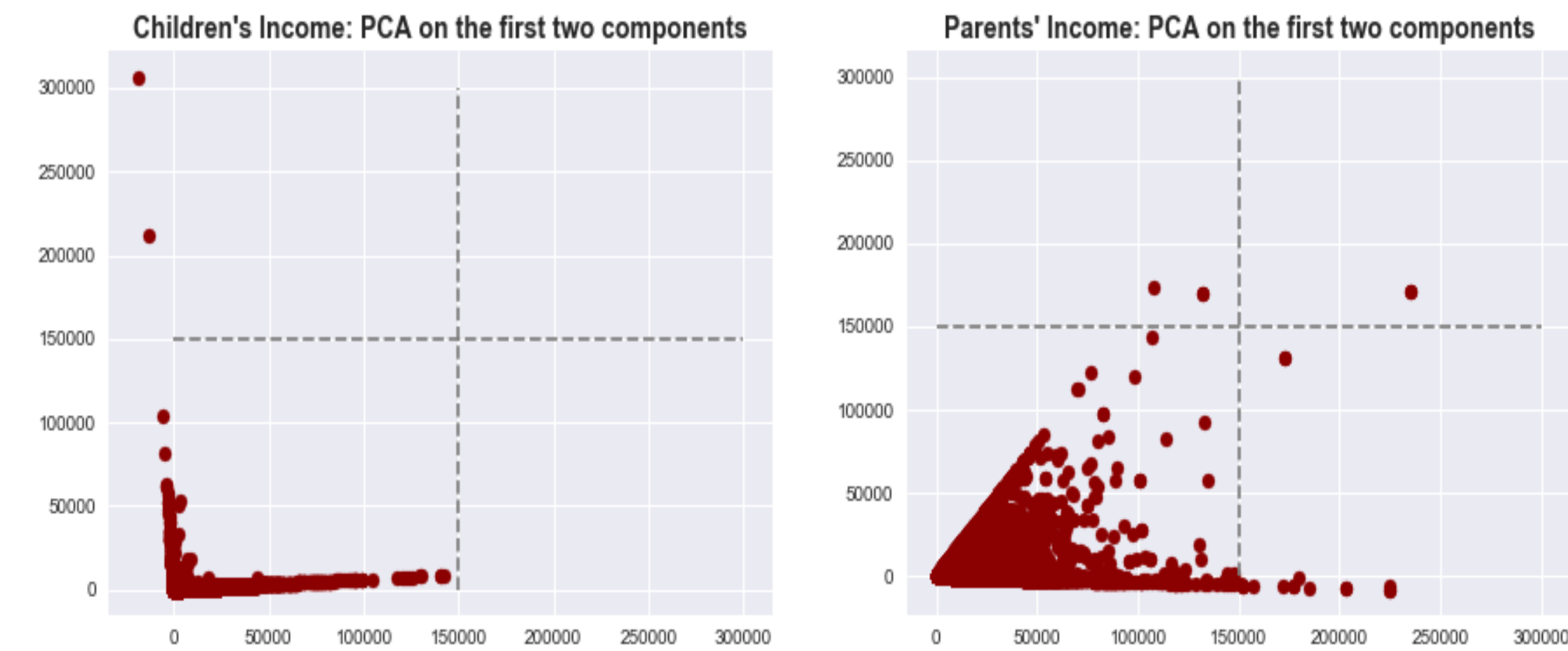
Theses are the scatterplots of where the children and parents lie on the first two components. The first component captures more information on labor income, while the second one captures private business for the children, while property income for the parents.

Transition Matrix

These are the probabilities of how the poor stay poor, and how the rich retain themselves
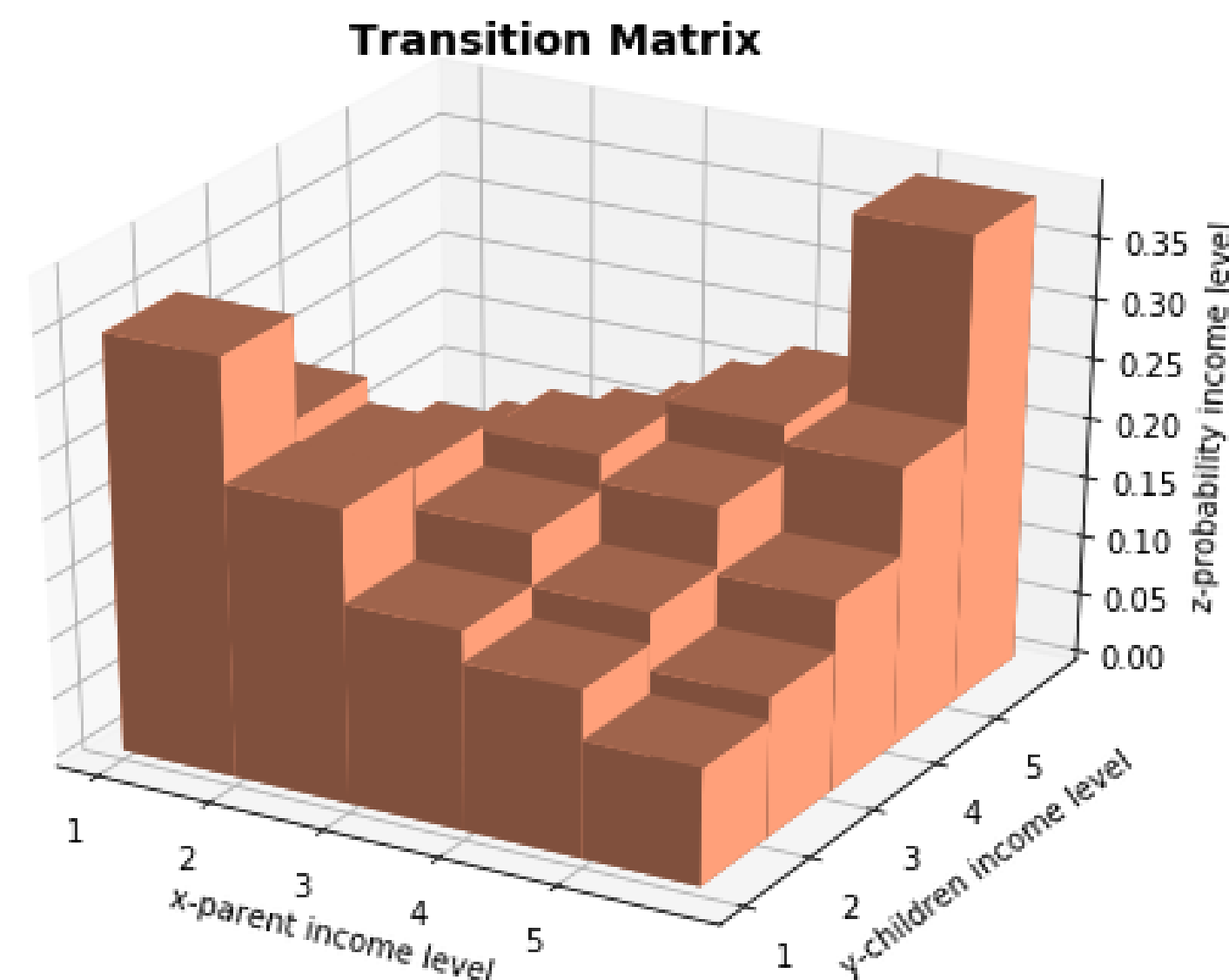
The x-axis represents parents' economic status, and the y-axis represents that of the children. The taller these columns, the more likely is the transition.

Conclusions

- In China, the fluidity is very tight. Especially for the very poor and very rich, there is very weak possibility of moving upward or downward.

- In out 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.

Feature Importance

- **Here we plot the 25 most important features among the 348 variable**

- The parents' economic status indeed plays a crucial role influencing children's income. The influence was beyond our expectation.

- Education so also quite decisive of children's economic wellbeing. 'education6' is professional education and is positively linked to economic status; 'education2' represents secondary school dropouts, and is negatively related

 * Hyperparameter settings are determined through five-fold cross-validation. We decided to go for 500 decision trees, and the maximal depth of each tree is set 5.

PDP/ICE Plot
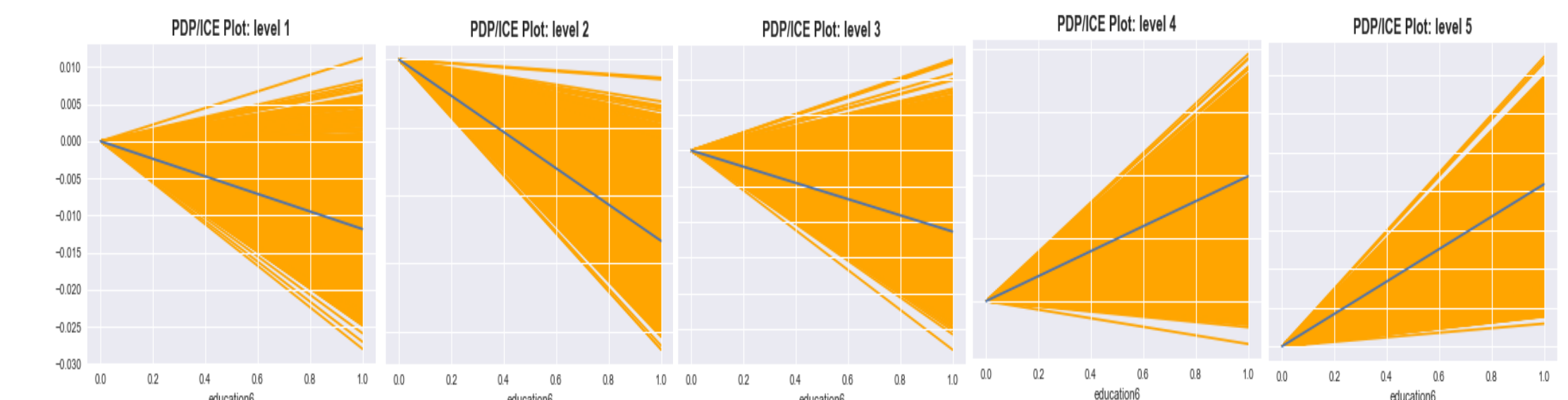
Professional Education → ← Secondary School Dropout