Upward Intergenerational Social Mobility

Leon Luo, Mark Fu, Tony Pappas (Team Social Mobilizers)

July 29, 2021

Project motivation and goals

The motivation:

- Intersection of data science and economics/social sciences

The goal:

- To present a model incorporating various socioeconomic factors predicting social mobility

What is intergenerational mobility?

Definition

Intergenerational mobility is the difference in income between child and parent (Chetty et al. 2020)

Data background

- Socioeconomic data
  - Education level, family income, etc.
- Via Opportunity Insights
  - Research and policy institute
- Collected from publicly available federal government records
- 2000+ data points

Data cleanup

- Two datasets from Opportunity Insights
- College tier: categorical to numerical conversion
  - Using the given 14 tier conversions accompanying the dataset
- Redefinitions (for ease of presentation)
  - Parent and child median incomes measured in thousands of dollars
  - Mobility rate measured as percentage
- Extraneous variables were deleted
- Merge by par_q1 and state
  - par_q1 is % of parents earning in 1st quintile. This variable was not used in our analysis and was merely used to merge datasets.

Predictor variables

We chose median child income, median parent income, and school tier as predictor variables. We chose these variables because we believe factors such as education and family background highly influence mobility rate.

Descriptive statistics

Name	Non-null count	Mean	Median	Standard dev
Mobility rate	2198	1.93	1.61	1.31
Median child income	2198	36.9	34.3	12.8
Median parent income	2198	77.7	74.3	28.5
School tier (1 - 14)	2198	7.11	6.00	2.26

Preliminary analysis

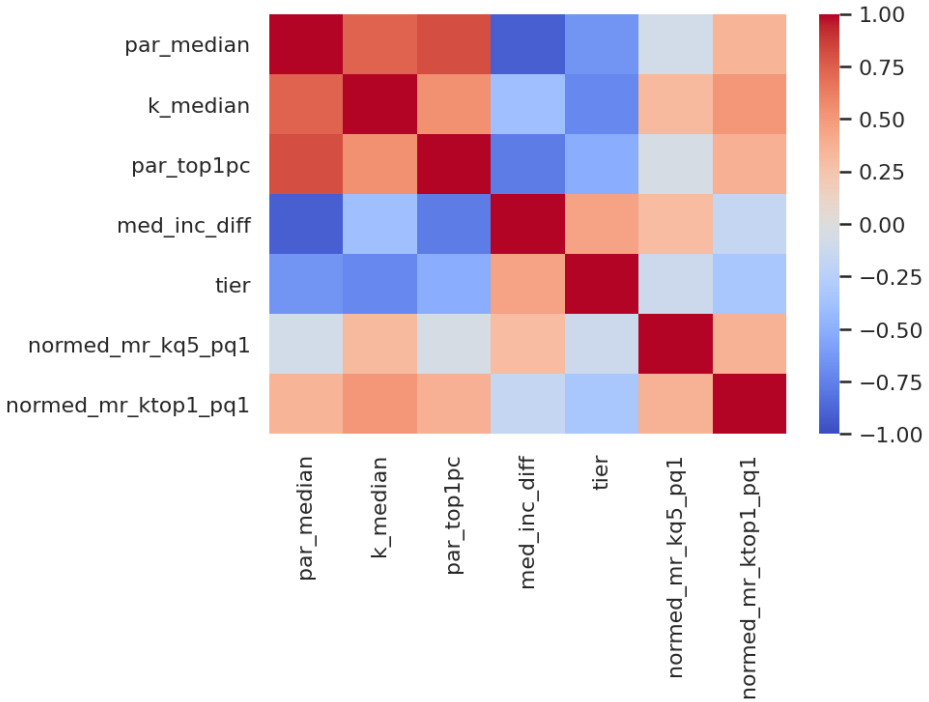

Figure 1: Correlation matrix. Note strong correlation between k_median and par_median, indicating collinearity.

Preliminary analysis

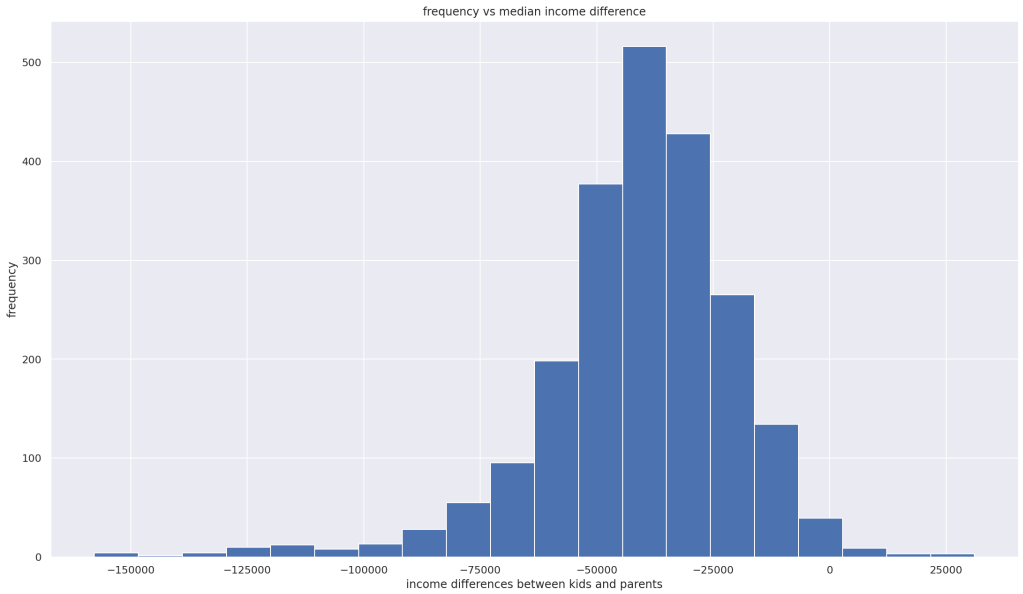

Figure 2: Histogram shows majority of children make less than parents. This is accounted for because parent income is defined by household, while child income is defined by individual.

The predictive model

We chose to use multiple linear regression because we wanted to assess the strength and importance of relationship between mobility and each of the predictor variables.

Two of our predictor variables, k_median and par_median, are collinear. We thus present two regressions to separate the collinear predictors.

The predictive model

OLS Regression Results

Dep. Variable:	normed_mr_kq5_pq1	R-squared:	0.118
Model:	OLS	Adj. R-squared:	0.117
Method:	Least Squares	F-statistic:	147.1
Date:	Thu, 29 Jul 2021	Prob (F-statistic):	1.07e-60
Time:	02:18:17	Log-Likelihood:	-3573.9
No. Observations:	2198	AIC:	7154.
Df Residuals:	2195	BIC:	7171.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.6438	0.212	-3.043	0.002	-1.059	-0.229
k_median	0.0468	0.003	15.983	0.000	0.041	0.053
tier	0.1183	0.017	7.102	0.000	0.086	0.151

Omnibus:	1212.278	Durbin-Watson:	1.760
Prob(Omnibus):	0.000	Jarque-Bera (JB):	14425.849
Skew:	2.356	Prob(JB):	0.00
Kurtosis:	14.632	Cond. No.	320.

Figure 3: Regression including predictors k_median and tier. Note that despite a lower $R^2 = 0.118$, the low pvalues all indicate significant results.

The predictive model

OLS Regression Results

Dep. Variable:	normed_mr_kq5_pq1	R-squared:	0.063
Model:	OLS	Adj. R-squared:	0.062
Method:	Least Squares	F-statistic:	73.19
Date:	Thu, 29 Jul 2021	Prob (F-statistic):	1.70e-31
Time:	02:18:17	Log-Likelihood:	-3641.3
No. Observations:	2198	AIC:	7289.
Df Residuals:	2195	BIC:	7306.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.2382	0.193	21.919	0.000	3.859	4.617
par_median	-0.0132	0.001	-10.480	0.000	-0.016	-0.011
tier	-0.1811	0.016	-11.429	0.000	-0.212	-0.150

Omnibus:	1307.920	Durbin-Watson:	2.011
Prob(Omnibus):	0.000	Jarque-Bera (JB):	20783.742
Skew:	2.500	Prob(JB):	0.00
Kurtosis:	17.211	Cond. No.	595.

Figure 4: Regression including predictors par_median and tier. Note that despite a lower $R^2 = 0.063$, the low pvalues all indicate significant results.

Conclusions

- We conclude that child income is positively correlated with mobility. This makes intuitive sense; if the child earns more, they will be more likely to earn more than their parents.
- Parent income is negatively correlated with mobility. This result is also intuitively correct; if the parents earn more, it will be harder for the child to earn more than them.
- Finally, school tier is both positively and negatively correlated with mobility, depending on the regression model. This can be explained by examining **const**. In the first model, with tier showing positive correlation, const = -0.64, while in the second model, with tier showing negative correlation, const = 4.2. Thus, the two regressions fit different intercepts but still move in the same direction.

Future Research Questions

- How would social mobility be affected by location, race, gender?
- What should parents do to set the next generation up for positive social mobility?
- What does a wealth gap look like in terms of social mobility?
- How can social mobility solve a wealth gap?
- How does policy affects social mobility?