Income Mobility Through a College Education in the United States

Margaret Perry, Rutendo Madziwo, Zainab Rizvi

Abstract

This analysis focused on students from families that were in the bottom 20% of the income distribution when the student entered college and moved to the top 20% of the income distribution within ten years of graduation. We looked at how this mobility is affected by the diversity of the campus, the tier of the college, the choice of majors, endowment, and the mean parental income for that college. We chose to analyze this data through multiple regression and checked for conditions for regression and inference which were satisfied. We further analyzed the relationship between tier and income mobility using Analysis of Variance (ANOVA) regression. We found that, on average in the population, the percentage of students in STEM majors, international students and higher parental median income have a positive coefficient in the regression equation whereas the percentage of black students has a negative coefficient. We also found that tiers are more statistically significantly different from each other in the top range (1-8) than the bottom range (8-12). For corresponding figures, please see the attached appendix.

Background

As of August 2013, 66.7% of college students in the U.S.A were graduating with college debt (Chris 2013) and even though students from low-income families are still given the opportunity to study, they report finding it even more difficult to repay their debt (Sandy 2003). We were hence interested in finding the income mobility achieved by students going to different colleges. This research is important to us because it can help students, particularly low-income students make decisions on colleges that they attend motivated by whether their college education will help them earn enough money to achieve income mobility. We are also interested in finding out which qualities are embodied by colleges that are most likely to produce graduates that achieve income mobility. This analysis might help the education system in introducing or investing more in these factors at other colleges in order to increase overall economic mobility in the country. We hypothesizedthat parents' median, the tier of the college, endowment, diversity of the college, and percentage of STEM students would affect the rates of income mobility.

Descriptive Statistics

We lost around 1400 observations due to missingness. The predictors that had the most missing values were our response variable and the percentage of STEM students. Since these statistics are acquired from various sources, we did not consider it appropriate to fill in the missing data. After removing the incomplete observations, we were left with 718 colleges: 94 from the Northeast, 157 from the West, 158 from the Midwest and 309 from the South. While this is a much smaller dataset than all of the colleges in the United States, we believe that this analysis is still helpful in describing the general patterns in seeing which factors affect income mobility.

The median parental income per college was recorded for all the colleges in our dataset. There were 12 tiers in our dataset which split the colleges into tiers and these tiers were scaled on a 1- 12 range where 1 = Ivy Plus and 12 = Less than two year schools of any type (Please see the appendix for a list of all the tiers). The average endowment of assets per college had 45 missing observations and endowment was highly skewed. To account for the college's diversity, we looked at two main factors - the percentage of black students at the college, and the percentage of international students at the college. Our response variable, the percentage of children who reach the top 20% of the income distribution post-graduation among children with parents in

the bottom 20% of the income distribution, had a mean of 17.30% (SD = 10.82%). Our final model included a combination of explanatory variables that define characteristics related to institutional characteristics such as endowment and tier as well characteristics about the student body such as the percentage of international students and the median income of the parents.

Analysis Strategy

To determine the best model for the regression we used best subset selection and stepwise regression methods. The result of best subsets can be seen in Fig. A. Both methods showed that parents' median, the tier of the college, endowment, diversity of the college, and percentage of STEM students would affect the college's income mobility after students graduate. In addition to these variables we added an interaction term between tier and median parental income, because we are interested in seeing how these two variables affect the regression as they are expected to oppositely affect the response variable. We used a correlation matrix and the variance inflation factor to test for multicollinearity and found that multicollinearity did not exist between any of these variables. As shown in Fig. B, we also plotted a residuals versus fitted model and qqplot to check for equal variance of errors and normality within the model. For the model as a whole all of these assumptions were met. However this did bring to light three outliers but none of the points were shown to be leverage points based on Cook's distance. We also ran the our regression model after removing these outliers and did not find a statistically significant difference in the r-squared value and the p-values of individual predictors so decided to keep all the observations for the final model. The independence condition is satisfied because colleges are autonomous institutions - even if they are part of consortiums and leagues, the major decision making lies with their adminsitration.

Results

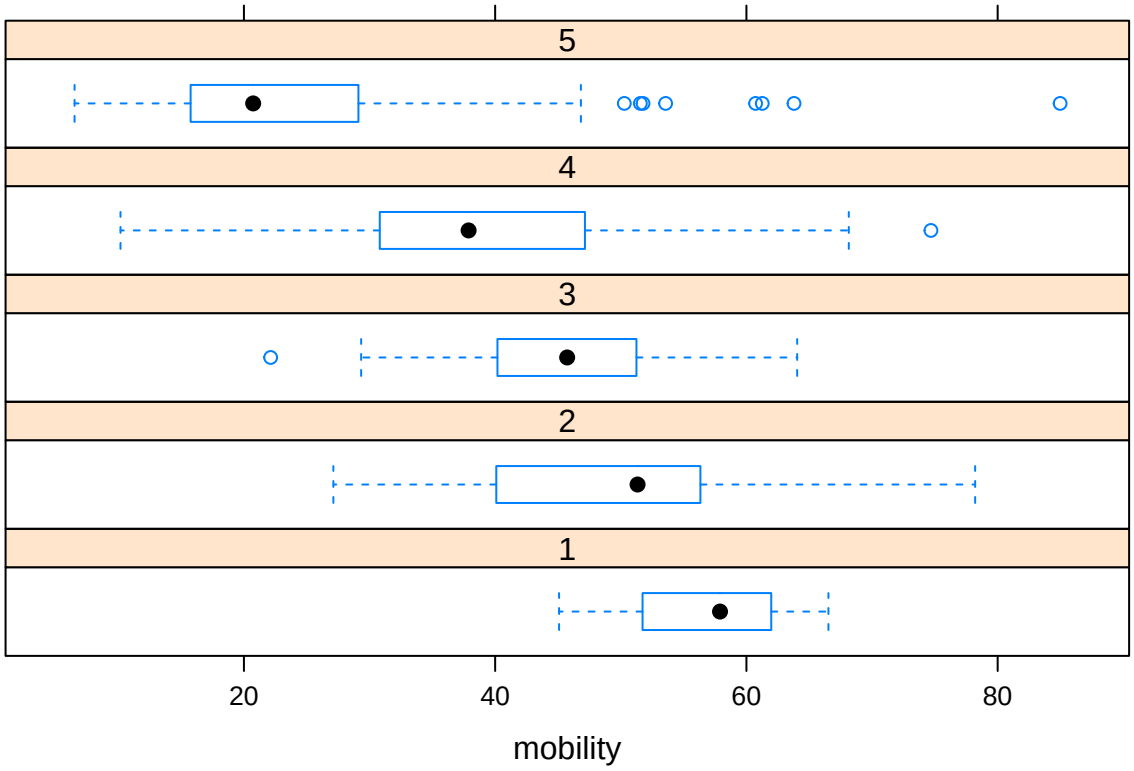
As shown in Fig. C, the adjusted r-squared value is 0.6829 which means that 68.29% of the variability in the response can be explained by this model. All the explanatory variables are statistically significant and have p-values below our alpha level of 0.05.

On average in the population:

- Holding all else constant, an additional percentage point increase in STEM students results in a 0.2167% increase in income mobility
- Holding all else constant, an additional percentage point increase in foreign students increases the percentage of income mobility by 5.08%
- Holding all else constant, an additional percentage point increase in black students decreases income mobility by 3.361%
- Holding all else constant, a one unit change in tier increases income mobility by 1.494%
- Holding all else constant, a dollar increase in the median parental income results in a 0.0000405% increase in mobility

where income mobility is defined as the percentage of students from the bottom 20% of the income strata who make it to the top 20% of the income strata within 10 years of graduation.

We looked at the tier vs. mobility plot in Fig. D and considered the change from one tier to the next as continuous and ordered. Since we were interested in looking at linear trends rather than difference of means at each level of tier, we decided to assume that the relationship between tier and our response is linear. We were surprised to find that a one unit increase in tier has a positive relationship with income mobility.



We were also interested in seeing how income mobility varied between tiers. To investigate this further, we decided to run an Analysis of Variance Regression (ANOVA) and perform a Fisher's Least Significant Difference (LSD) test to see the variation from tier to tier and the output is displayed in Fig. E. We found that tiers are more statistically significantly different from each other in the top range (1-8) than the bottom range (8-12) but there are long tails as can be seen in the box plots above.

Discussion

Our ANOVA results showed us that even though that Ivy-Plus schools (tier 1) are significantly different from the rest of the tiers and have mobility rates of 60% on average, there are some notable exceptions and we looked into them further. One such example is the State University of New York at Stony Brook (S that has similar rates of mobility but also admits a lot more low income students - specifically, 16.4% of students at Stony Brook are low income compared to 3.8% at the Ivies. Our model does not take into account low income access which we feel is an important factor in evaluating colleges.

We concluded that low-income students are not mismatched at selective colleges as can be seen by the significantly higher mobility rates. Moreover, we also concluded that the share of international students does not bring down rates of mobility but rather gives it a boost.

Finally, previous literature has focused on the "birth lottery" being the main decider for income mobility. We found that Mobility rates range from 0.9% at the 10th percentile to 6.5% at the 90th percentile across colleges. To put these numbers in perspective, the average bottom-to-top-quintile mobility rate in U.S. is currently 1.7% indicating that a top-tier college education still seems the best way to overcome the "birth lottery".

In future work, we would also like to directly include the income distribution of the student body in our analysis rather than the generic median income predictor we used here. We would also like to work with more recent data and include factors like average financial aid awarded to each student.

References

Chris, Denhart. 2013. "How The $1.2 Trillion College Debt Crisis Is Crippling Students, Parents And The Economy." *Forbes*, August. Forbes, 261–63. https://www.forbes.com/sites/specialfeatures/2013/08/07/how-the-college-debt-is-crippling-students-parents-and-the-economy/.

Sandy, Baum. 2003. "How The $1.2 Trillion College Debt Crisis Is Crippling Students, Parents And The Economy." *Journal of Student Financial Aid*, December. Journal of Student Financial Aid. https://publications.nasfaa.org/cgi/viewcontent.cgi?article=1068&context=jsfa.