How is attaining a bachelor's degree in Georgia counties affected by median income and poverty rate?

Abstract

Using county-level data from Georgia in 2018, we attempt to understand the relationship between the percentage of the population with at least a bachelor's degree and median income, poverty, and population. In our final model, we found an increase in the total population is estimated to be associated with a positive change in the estimated median percentage of bachelor's degree attainment, holding all other predictors constant. The estimated effect of total population matches intuition because larger populations could correlate with urban areas, where higher education is a cultural or socioeconomic priority. Meanwhile, it is possible that changes in the poverty rate and median income could be associated with no effect on estimated median bachelor's degree attainment. This result was unexpected since having more wealth makes it easier to attain a bachelor's degree. Overall, our analysis suggests that bachelor's degree attainment is more complex than we can capture with this model.

Introduction

Beginning in the 1890s, Georgia passed a wide variety of "Jim Crow" laws that mandated racial segregation or separation in public spaces, effectively codifying the region's tradition of white supremacy. Despite the federal Civil Rights Act of 1964, public facilities in Georgia and throughout the region remained segregated in many areas well into the 1970s. Today, segregated residential patterns persist in Georgia and other Southern states (Hatfield, 2007). Georgia also has large amounts of income inequality, with the wealthiest 5 percent of Georgia households now having average incomes about 16 times as large as households in the bottom 20 percent and five times as large as those in the middle 20 percent (Tharpe, 2012). The history of racial segregation and inequality along with its large number of counties makes Georgia particularly interesting for studying sociological phenomena.

One way inequality manifests itself is in the quality of education students receive. In 2014, researchers found that low-income youth are more likely to drop out of school if they live in a place with a greater gap between the bottom and middle of the income distribution (Kearney and Levine, 2014). In 1990, researchers looked at income level, gender, ethnicity, and household composition as predictors of children's school-based competence. They found that income level and gender were the best predictors of child competence. However, as Black people are much more likely than white people to live in low-income households, this study highlights the inequities associated with Black and white children growing up with the unequal distribution of economic resources in America (Kupersmidt et al, 1990).

Our study investigates counties in Georgia counties in 2018 to understand the relationship between median income and bachelor's degrees, while controlling for poverty rate, and total population. We chose to look at 2018 because it is the year where the United States had the highest income inequality overall, but it was also before COVID so we do not have to consider the impacts of a global pandemic (US Census Bureau, 2023).

Data

We used data from the Vera Institute of Justice, which combines data from the U.S. Department of Justice Bureau of Justice Statistics, the National Corrections Reporting Program, and the National Center for Health Statistics to create a database about prison, incarceration rates, and demographics at the county and state levels. We also used data from the US Census, where respondents responded directly to surveys, and federal, state, and commercial entities provided data of interest. The level of analysis in this combined dataset is counties in the state of Georgia. Specifically, we explored *bachelor's degree* (percentage of a county that achieved at least a bachelor's degree, denoted *B*), *median income* in a county (in thousands of dollars, denoted *M*), *poverty rate* (percentage of a county living in poverty, denoted *P*), and *total population* (the total number of people in a county, denoted *T*).

Results

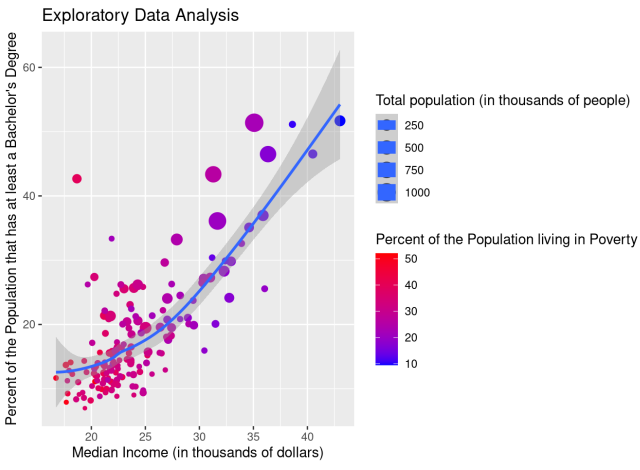

Figure 1: Percent of the population with at least a bachelor's degree vs. median income, with total population and poverty.

We created a scatterplot of the percent of the population that had at least a bachelor's degree against the median income in thousands and dollars, with a color gradient for the percent of people living in poverty and a size gradient for the total population. This plot shows the relationship between our chosen explanatory variables and how they relate to the attainment of bachelor's degrees. In general, our graph shows that as median income increases and poverty decreases, the percentage of the population in a county that has a bachelor's degree increases. Moreover, this graph makes it easy to spot large counties, which seem to in general have higher bachelor's degree attainment rates.

Our final multiple linear regression model is:

$$log(\widehat{B}) = \widehat{\beta}_0 + \widehat{\beta}_1 log(M) + \widehat{\beta}_2 log(T) + \widehat{\beta}_3 P + \widehat{\beta}_4 log(M) * P$$

On the original scale, our model is as follows:

$$\widehat{B} = e^{\widehat{\beta}_0} \bullet M^{\widehat{\beta}_1} \bullet T^{\widehat{\beta}_2} \bullet e^{\widehat{\beta}_3 P} \bullet M^{\widehat{\beta}_4 P}$$

Below is a table with our estimates, standard errors, t-statistics, and p-values.

	Estimate	Standard error	t-statistic	p-value
Intercept	-1.60576	1.06750	-1.504	1.3457
Log Median Income	1.32003	0.31960	4.130	5.92e-05
Log Total Population	0.18067	0.02031	8.897	1.47e-15
Poverty	0.09115	0.03416	2.669	0.00843
Log Median Income * Poverty	-0.03298	0.01088	-3.032	0.00285

Since there is an interaction term in our model, the effect of median income on the estimated median change in the percentage of a county's bachelor's degree attainment rate depends on the poverty rate in that county. Similarly, the effect of the poverty rate on the estimated median change in the percentage of a county's bachelor's degree attainment rate depends on the median income in the county. Hence, in our discussion of the effect of either explanatory variable on our response, we will fix the corresponding explanatory variable at the median value in our data.

When the poverty rate is fixed at 31.67% and total population is held constant, we are 95% confident that a 10% increase in median income is associated with between a decrease of 3.18% and an increase of 8.86% in estimated median percentage of the county that has at least a bachelor's degree. Fixing median income at $22,732 and holding the total population constant, we are 95% confident that a 1% increase in poverty rate is associated with between a decrease of 5.6% and an increase of 3.5% in the estimated median percentage of bachelor's degree attainment in Georgia counties. Since both of our 95% confidence intervals for the effects of median income and poverty rate on estimated median bachelor's degree attainment include zero, increases in median income or poverty rate could result in a positive, negative, or no change in the estimated median percentage of the population with a bachelor's degree, holding the corresponding variables constant. This suggests that neither variable provides a simple solution to increasing bachelor's degree attainment in Georgia counties.

We are 95% confident that a 20% increase in total population is associated with an increase in the estimated median percentage of bachelor's degree attainment between 2.6% and 4.1%, holding median income and poverty rate constant. Though this effect is not very large, this positive relationship indicates that larger counties tend to have a bigger proportion of the population with a bachelor's degree. We speculate this observed effect could be the result of people moving to large counties after graduating from college, thus resulting in a higher bachelor's degree attainment rate.

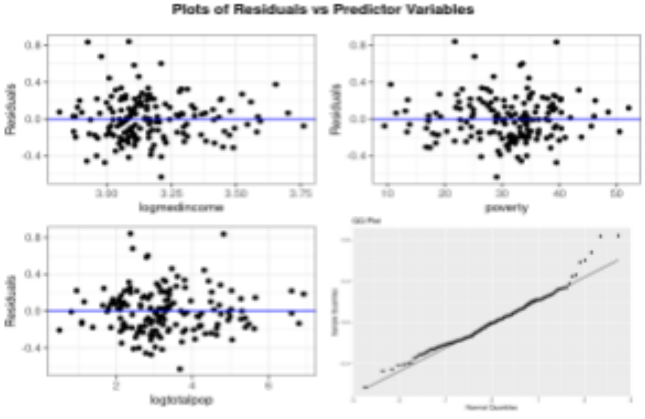

Figure 2: Residual and QQ plots for our final model to check for assumptions of homoscedasticity, normality, and linearity

As seen in Figure 2, Every residual plot is roughly cloud-shaped and centered at zero, with more points closer to 0 than far away from zero, indicating that our model satisfies the linearity and constant variance assumptions. The QQ plot indicates a heavy right tail, suggesting some right skew and potentially violating our normality assumption. However, since we are looking at median income, which is likely to have large positive outliers, this is somewhat expected.

Concerning the independence assumption, our model may also have some problems with spatial independence. Counties that are next to each other are likely similar in their culture and beliefs, which could dictate attitudes toward higher education and policies that influence poverty levels, population sizes, and income levels. However, adjusting for spatial data is beyond the scope of this project. According to the adjusted R^2 of our model, our multiple linear regression model accounts for 69.05% of the variance in median proportion with at least a bachelor's degree at the county level in 2018 Georgia.

Plotting Cook's Distance, Leverage, and Studentized Residuals had some mixed results. The Cook's Distance had no outliers that exceeded the critical threshold of one. The Leverage, which measures the effect of an outlier in the x-direction, had thirteen values above the critical threshold of 0.063. Of the thirteen, the counties with the five highest leverage were Forsyth County, Fayette County, Randolph County, Clay County, and Oconee County. Overall, we found no reason to remove any of these outliers due to their Leverage, though the absolute effect of the model estimates was smaller for all our explanatory variables except for the log of the total population. Only two counties were flagged in our assessment of the Studentized Residuals – Chattahoochee County and Clarke County had Studentized Residual values greater than our threshold of 3. Removing both these outliers had marginal to no impact on the overall regression model, and since we found no theoretical reason to remove them from our analysis, we opted to keep them in the model.

When fitting our model, we tried to balance model validity, having a model that was able to be interpreted, and optimizing the different measures of AIC, R^2, and statistical significance through ANOVA. Looking at our model without the interaction term between median income and poverty, the variance inflation factor of all the variables is less than 4, so there is little concern regarding collinearity.

Discussion

We have a few intuitive conjectures for why higher median incomes, larger populations, and lower poverty rates in counties are associated with a higher proportion of the county getting at least their bachelor's degree three associations. For example, wealthier people are likely to have more resources to invest in their own and their children's education, and larger counties could include a larger city or a university, easing the barriers to achieving a bachelor's degree. Lower poverty rates could mean that the people in a county are past the subsistence threshold required to survive, and can spend more time and energy towards secondary pursuits such as education.

A major limitation of our analysis is that this is not a causal analysis, but rather a study of associations. For example, earning more money may promote higher rates of attaining a bachelor's degree and people who receive more education may be more likely to earn higher incomes. Also, our observed median poverty rate of 31.67% we set our variable P to is very high. Similarly, the observed median median income of $22,732 we set our variable M to seems extremely low, though Georgia may have a very low cost of living overall, so less income per person is required. Both of these phenomena could suggest systemic inequity in Georgia affects the economic and educational prosperity of its population.

Our final model doesn't investigate how the biases and limitations in our analysis further the exclusivity of higher education. Partly, this is because our model has limited explanatory power and doesn't include all of the contributing factors to bachelor's degree attainment. Legislation based on our analysis would further the exclusivity of higher education and the disparities between accessibility to these opportunities between different communities within a county. Community outreach and equitable resources for students aspiring to go to college from all socioeconomic backgrounds, for example, are important to see rises in bachelor's degree attainment. Research could be conducted into the racial inequality of higher educational attainment if we include additional explanatory variables that stratify the population by race. Considering the spatial independence issues of our analysis could help to determine if there are associations between states that impact the attainment of bachelor's degrees.

References

Hatfield, E. (2007, June 1). *Segregation.* New Georgia Encyclopedia. https://georgiaencyclopedia.org/articles/history-archaeology/segregation/.

Kearney, M. & Levine, P. (2014, June). *Income Inequality, Social Mobility, and the Decision to Drop out of High School.* National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w20195/w20195.pdf.

Kupersmidt, J. B., Patterson, C. J., & Vaden, N. A. (1990). Income Level, Gender, Ethnicity, and Household Composition as Predictors of Children's School-Based Competence. *Child Development*, 61(2), 485–494. https://doi.org/10.2307/1131109.

Tharpe, W. (2012, December 2019). Growing Inequality Threatens Georgia's Economy. The Georgia Budget & Policy Institute. https://gbpi.org/growing-inequality-threatens-georgias-economy/#:~:text=After%20decades%20f%20expanding%20inequality,20%20percent%20(see%20Figure).

US Census Bureau. (2023, September 13). *Household income distribution according to the Gini Index of income inequality in the United States from 1990 to 2022 via Statista.* Retrieved November 17, 2023, from https://www.statista.com/statistics/219643/gini-coefficient-for-us-individuals-families-and-households/

US Census Bureau Data. (2021). https://api.census.gov/data/2021/acs/acs5/variables.html

Vera Institute of Justice. (2023). *About: Data Sources and Processing.* https://trends.vera.org/methodology/.