

Regional Factor's Effects on High School Graduation Rates: Empirical Evidence Using U.S. State-Level Data

Richard Pandel, Jacob Lin

(Reh School of Business, Clarkson University, New York, USA)

Abstract: This paper focus on the effects of factors that affect family stability and their effects on high school graduation rates. We use state-level data through the years 2011-2018. We find that the unemployment rate has a high negative correlation with graduation rates. The other notable factor is the location of the state. Most interestingly high schools located in the west were found to have an average of 6.6% lower graduation rate than those in the Northeast. In the future, researchers may also look at household income, alcohol consumption rate, and crime rates to see their effects on the graduation rate.

Key words: high school dropout rate, economics performance, regional factor

JEL codes: I23, I25

1. Introduction

The reason for a student to drop out of high school is often related to home life. Some of these factors could be the crime rate, household income, and unemployment. Kearney M. S., and Levine P. B. (2014) look at the effects of low income on the dropout rate. They found that low-income youths are more likely to drop out. They found that less income inequality, less residential segregation, better primary schools, greater social capital, and greater family stability had the highest correlation. This paper looks to focus on the effects of family stability. Cavanagh Shannon and Fomby Paula (2019) argues that divorce is a major cause of instability for the child. Cavanagh Shannon and Fomby Paula (2019) also showed that low-income families have greater instability. With that in mind if the divorce rate in a state is high then you would expect to see a lower graduation rate. You would also expect that if the median income of a state was low that graduation rates would be lower.

This paper will attempt to explore the relationship of these variables with the graduation rate. The main variable in this paper is the median income. We will be exploring the effects of median income on graduation rates. This variable should be found to be significant in Kearney M. S., and Levine P. B. (2014) on its effects on the dropout rate. Cavanagh Shannon and Fomby Paula (2019) also found this to be a significant variable for creating instability for the family. This variable should have a strong correlation. The hypothesis for this paper is that the lower the median household income is then the lower the graduation rates should be as well. Figure one shows multiple samples of the relationship between the median household income and graduation rate. Looking at the data it seems that certain states follow the expected result. Where median income would increase so would the graduation rate.

Jacob Lin, Ph.D., Assistant Professor, Reh School of Business, Clarkson University; research areas: capital market research in accounting, empirical and theoretical corporate finance. E-mail: glin@clarkson.edu.

Although some states did not seem to show any correlation whatsoever. To find whether there is empirical evidence supporting the hypothesis. We took cross-sectional data in America by state. We also took historical data from the years 2011 to 2018. This is a short period to be looking at the states alone though. Therefore, we look at all the states combined when running the regression. To account for the different locations, we will also include a dummy variable for location. We used the Northeast areas, the base, the South, Mid-west, and the West where the other areas. Other variables that we looked at were the crime rate, unemployment, alcohol consumption rates, and divorce rates. The crime rate of an area could potentially have a relationship with the graduation rates. This could affect the environment that the students live in and could negatively affect their ability to finish high school. Alcohol consumption rates should also have a negative relationship with graduation rates. Balsa A. I., Giuliano L. M. and French M. T. (2011) found that alcohol consumption had a small effect on the student's GPA. This effect could also carry over to their ability to graduate high school as well. The unemployment rate is also an important variable. If the unemployment rate is high, then students may be more likely to drop out to get a job to help their families.

The findings of this study were surprising. In this study, we did a linear regression analysis. We used a 5% confidence interval to decide whether the variable was significant or not. We found that median household income, crime rate, and divorce rate were not significantly significant in this study. There could be multiple reasons why our findings are different from some other studies. One reason that may be the cause is that the state data was too large. If we looked more at the county level, then these factors may have had more significance in our model. However, we did find that alcohol consumption was statistically significant. The unemployment rate and the geographical areas were also found to be significant, excluding the Midwest area.

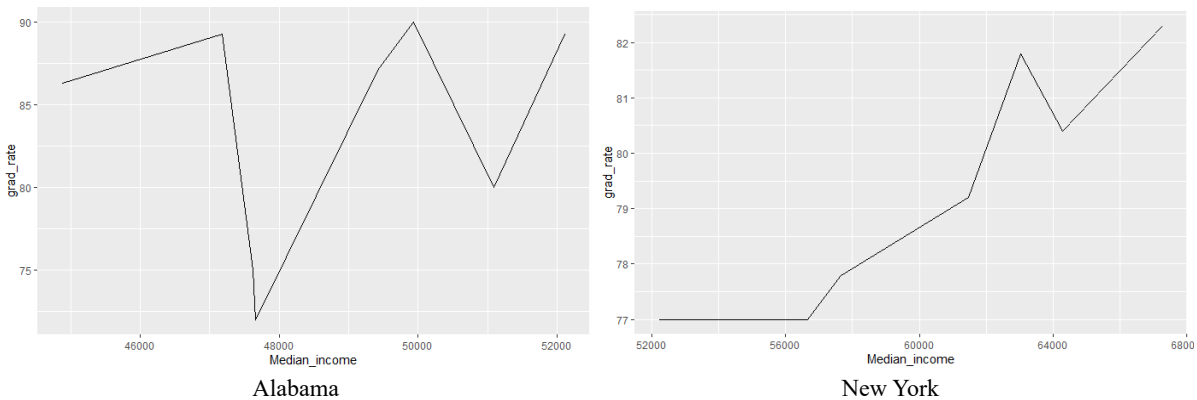


Figure 1 Median Household Income's Relationship With the Graduation Rate of Public Schools

2. Literature Review

There are multiple underlying reasons for students to drop out of high school. Kearney M. S., and Levine P. B. (2014) argued that income inequality and socioeconomic difference is major cause of a student deciding to drop out. Essential this paper found that students who had a low socioeconomic status were less likely to graduate. They also were less likely to go to college compared to students with similar test scores but a high socioeconomic status. Another major thing that leads to students wanting to go to college was knowing very few people who have been to college. This creates a situation where a student with low socioeconomic status was less likely to believe that graduating from school or going to college would benefit them. They then looked at states with a high-income inequality. They compared those states to other states with low-income equality. They looked at the chances of a student who lived with a family that made 30% lower than the average family and whether they also would have a

permanent income 30% lower. The study did find this to be the case. They were arguing that instead of income the bigger factor is income inequality. In their paper, they did prove this using empirical evidence.

Pong S. L. and Ju D. B. (2000) wrote a paper looking at the high school completion rates of families that lose income and mother-only families. They found that mother-only families did not necessarily have a lower graduation rate. However, they did find that families that went through a divorce and then became mother-only families did suffer lower graduation rates. They also found that if the family's income decreased that the graduation rate would be lower as well. They also found that there was a relationship between ethnicity and dropping out as well. They found that 8th graders that were Hispanic or Black and suffered from a family disruption were far more likely to drop out compared to White or Asian students who also suffered from family disruption.

Both studies seem to point towards the fact that income plays a role in the graduation rate. They both looked at different factories that were related to income and found that it played a significant role in the graduation rates.

3. Data and Descriptive Analysis

In this study, we used data from the CDC, FRED, the National Center for Educational Statistics, and the Federal Bureau of Investigation. Multiple data sets were needed to find all the data required. We then put the data onto a file and analyzed the data from there. Originally, we were going to use data back to 1986, unfortunately, we could not find the graduation rates of each state going back that far. The total amount of data points is 408. Looking at the graduation rate data, it should be noted that Idaho and Kentucky used a different way of measuring the data for 2011 and 2012. The numbers are still given in the data set, but this causes a small difference in the data. The median income data was collected through FRED and collected for each state and year. Although for the data analysis we used the difference between the natural log to give us a percentage change for each year in the state. We also tried to do it without taking the difference between the natural log to see if it gave better results. In either case, we found that the median income was not statistically significant. Alcohol consumption is measured per capita, this data was collected through OPENICSPR. The unemployment rate is also fairly straightforward data collected through FRED as well. The data for the crime was measured in nonviolent crime per 100z000 people. We could have used violent crime as a measure instead but felt that nonviolent crime would be nearly an equivalent measure. We also used a dummy variable to represent the different regions in the United States. We used four different regions with the Northeast region as the base region. We initially ran our analysis without the divorce rate but we found that a lot of our variables were not significant. We then decided to add the divorce rate to see if that would potentially make a significant change in the other variables. We had similar results with the divorce rate variable.

Table 1 OLS Regression Results

Coefficients	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	80.7936646	19.4574274	4.152	4.03e-05 ***
Median_income	2.894e-05	2.784e-05	1.040	0.299155
Alcohol consumption per capita	-2.3700323	1.0304778	-2.300	0.021967 *
Unemployment rate	-1.5356318	0.1222739	-12.559	< 2e-16***
Nonviolent Crime rate	-0.0001089	0.0002200	-0.495	0.620812
Mid west	-0.6948948	0.7461370	-0.931	0.352249
South	-2.8032970	0.7658016	-3.661	0.000285***
West	-6.7653349	0.7223643	-9.366	< 2e-16***
Divorce_rate	0.0365353	0.0272940	1.339	0.181468

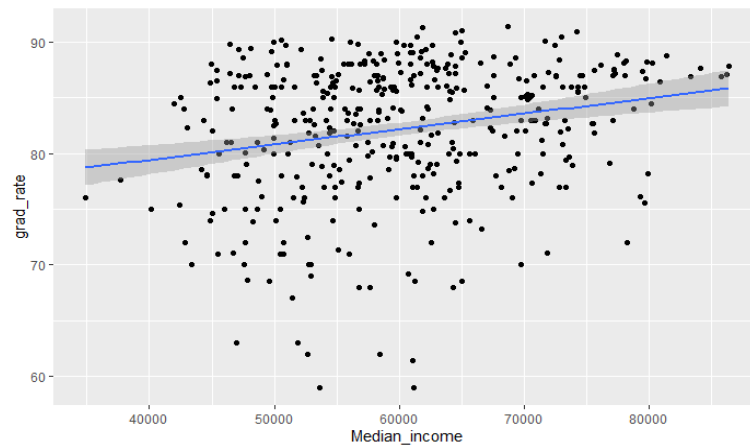


Figure 2 The Relationship Between Median Income and Graduation Rate

We can see that only alcohol consumption, the dummy variables of west and south, and unemployment were statistically significant in our analysis. The median income only had a t-statistic of 1.040. This is not within the 5 percent confidence interval. Looking at the summary statistics we can see that we can be very confident in the unemployment rate correlating with the graduation rates. It had a t-statistic of -12.559. This is within the 5% confidence interval. There also seems to be a strong negative relationship in the west for the graduation rate.

Not finding evidence that median household income played a role in the graduation rate was very surprising. We decided to bring up a chart to see what the relationship looked like and whether there was something wrong with our formula or if the correlation was weak. As you can see from looking at Figure 2, there appears to be a very weak positive relationship. The data is scattered everywhere, and it would be very difficult to be confident in using median income to predict the graduation rates with the data we have. The formula we used to model our data had an r-squared of 0.4847. This means that our data only accounts for 48.47 percent of the variation. This is not very high. It implies that there are other variables that we are missing.

4. Theory and Hypothesis

The theory we are focused on is that a lower family income will decrease the likelihood of graduating from high school. The idea is that a low family income will increase stress in the family and have other points of view. Some of these views may be that college and school are not worth it or they are simply unaware of the true benefits of schooling. Previous studies such as Kearney and Levine (2016) found that students with low socioeconomic status were less likely to go to college compared to similar students with high socioeconomic status. Their views differed as well. The person with the lower socioeconomic status generally felt that college would be a waste of time. A person with a higher socioeconomic status viewed it as a way where they would benefit. When comparing the two the study used similar GPAs to use that aspect as a control. This is like what our hypothesis was, just not as in-depth. We were using a simpler viewpoint of this by just using median income per household.

We hypothesized that the median household income would affect graduation rates. We used a 5% confidence level. We also looked at other factors that might affect the homelife of high school students. These factors included divorce rates, alcohol consumption rates, crime rates, and unemployment. We felt that these factors would increase the change and disruptions of the homelife. This would then distract the student from the school and may be more likely to drop out.

5. Empirical Methodology

After seeing the results of our data, we need to omit certain variables that we planned to use. We had to omit the median household income, divorce rate, and nonviolent crime rate. These were, unfortunately, the variables we were most interested in. After looking at our analysis we could not be confident in putting them into our linear regression model. The empirical model is

$$\text{Graduation rate} = e - \alpha AC - \alpha_2 U - \alpha_3 MD - \alpha_4 S - \alpha_5 W \quad (1)$$

Where AC is the alcohol consumption rate per capita, U is the unemployment rate, MD is the midwest, S is South, and W is west. The model should show a negative relationship with all the variables. graduation rate. The R-squared value of this model is 0.4812. This means that our model accounts for 48.12 percent of the variation in the graduation rate. Leaving out the other variables that we originally tested simplified our model. It also excluded variables that were found not to be significant. There should be no consequences in omitting those variables with the data we have. Our model should not be suffering from heteroskedasticity, the variables are not highly correlated with one another.

6. Results and Discussion

The estimates of our model are

$$\text{Graduation rate} = 97.02 - 2.42AC - 1.56U - 1.06MD - 3.07S - 6.79W \quad (2)$$

The signs of our model and values seem to be plausible. They are all negative and have relatively small effects on the graduation rate, per unit wise. Looking at the coefficient we can see that a high school from the Northeast, holding other factors constant, can expect a graduation rate of 97.02 percent. The first variable is the alcohol consumption rate. It shows that you can expect the graduation rate to drop by 2.42 for everyone's increase in the alcohol consumption rate, holding other factors constant. This is about what we expected. We expected that alcohol consumption should decrease the graduation rates. The unemployment rate also has a negative relationship. We can see that for every percentage point increase in unemployment that we can expect a 1.56 decrease in the graduation rate, holding other factors constant. Now looking at the dummy variables we can see that if the high school is located in the Midwest that you can expect a graduation rate of 95.96, holding other factors constant. It should be noted that this variable was not found to be significant, but we could not omit one of the variables and leave in the other two which were found to be significant. Now if you are located in the South, you can expect the graduation rate to be about 93.99, holding other factors constant. The west area has the lowest graduation rate, according to our model you can expect to have a graduation rate of 90.23, holding other factors constant.

As previously stated, the R-squared is about 48 percent. This tells us that our model accounts for 48 percent of the variation of graduation rates. This was lower than what we were expecting. With that said we found a lot of variables that we initially thought would be statistically significant were not.

Our hypothesis was found to not have statistically significant evidence. We have to reject our hypothesis that an increase in median income would also increase the graduation rates. We found that the median income only had a t-statistic of 1.04 which was too low for our 5 percent confidence test. We however did find that alcohol consumption per capita and unemployment did have a negative relationship with graduation rates. The divorce rate and crime rate were found to be insignificant in our analysis.

Overall, our results were not similar to the literature that we looked at. We were expecting that median income to be a significant variable, but it was not. We also found that divorce rates were not significant either. The literature we read claimed that these were significant factors in the dropout and graduation rates. There are some differences in our data. We looked at it from a much broader and simplified way. They used much more complex variables to make the data fit better. For example, Kearney and Levine (2016) used income inequality rather than income. We wanted to see if having a lower income would affect it the same way income inequality did.

7. Conclusion

This paper looked at the effects that median income had on the graduation rate. We also looked at other variables that could affect the family life of high school students. We used State-level data from the years 2011 to 2018. Our results were very surprising. We did not find a lot of significant variables that we tested. We theorize that we should have used smaller-scale data sets. Statewide data may have averaged out our data so that the effects of the variables did not show as we were expecting. County-level data may be a better way of doing this analysis. The other literature also looked at this from a smaller scale, once again this differs from our paper. However, we did find that Alcohol consumption per capita and unemployment were negatively correlated with the graduation rate. If we were to do our analysis again, we would need to find data on a smaller level and also find other variables that could affect graduation rates.

References

- Balsa A. I., Giuliano L. M. and French M. T. (2011). "The effects of alcohol use on academic achievement in high school", *Economics of Education Review*, Vol. 30, No. 1, pp. 1-15.
- Cavanagh Shannon and Fomby Paula. (2019). "Family instability in the lives of American children", *Annual Review of Sociology*, Vol. 45, doi: 10.1146/annual-soc-073018-022633.
- Kearney M. S. and Levine P. B. (2014). "Income inequality, social mobility, and the decision to drop out of high school (No. w20195)", National Bureau of Economic Research.
- Pong S. L. and Ju D. B. (2000). "The effects of change in family structure and income on dropping out of middle and high school", *Journal of Family Issues*, Vol. 21, No. 2, pp. 147-169.