1-1-2016

# An Econometric Model of High School Graduation Rates in Oregon

Emily Mavrakis
*George Fox University*, emavrakis13@georgefox.edu

Elizabeth Miller
*George Fox University*, emiller13@georgefox.edu

An Econometric Model of High School Graduation Rates in Oregon

Emily Mavrakis, Elizabeth Miller
George Fox University

Abstract

This study aims to analyze potential factors causing low high school graduation rates in the state of Oregon. We developed an econometric model measuring six independent variables: violent crime rates; income per family, the percentage of the population that is Hispanic, the percentage of family households with only one parent, alcohol use in adults, and teen pregnancy. Our dependent variable was measured as the percentage of the population between 18 and 24 without a high school diploma. From our results we found two of our variables to be economically significant: family structure and alcohol. However, the coefficient for alcohol suggested that higher rates of binge drinking in adults lead to higher graduation rates, which contradicts expectations. This outcome is addressed within the paper. Overall, this study suggests that family structure has the largest and clearest impact on graduation rates in Oregon, out of the variables included in our model.

*Keywords*: graduation rates, high school, Oregon counties, family structure, education analysis

JEL Codes: J12, I21, I25

**Section One: Introduction**

      High school dropout rates are one of the largest problems for the state of Oregon. In fact, as of 2014, Oregon's four-year graduation rate was the fourth lowest in the country, at 72% (Hammond, 2015). A high school diploma is increasingly becoming the equivalent to a ticket into the successful economic world. On average in America, those without a high school diploma experience an unemployment rate of 8%, compared to an unemployment rate of 5.4% for those with only a high school diploma. In addition, individuals without a high school diploma have a median weekly income of  $493 compared to the median income of $678 for an individual with only a high school diploma (Bureau of Labor Statistics, 2015). Secondly, a high school diploma is a necessary stepping stone for college admittance, which has become very important for increasing one's earning potential in America, a high school diploma does not represent the same experience level as it once did. For example, in 1980, a person with a high school diploma could earn approximately $1,000,000 in their lifetime, in contrast to a person in today's economy, who could earn only $480,000 (Herold, 2012). Additionally, an increased number of citizens who have not graduated high school means a larger burden on the government and local community. It means taxpayers spending more on things social services as Medicare and Medicaid, and the government receiving less from the community. For example, "moving just one student from dropout status to graduation will yield over $200,000 in higher tax revenue" (Balfanz, 2012). Therefore, low graduation rates are detrimental to individuals and society as a whole.

      This paper explores the relationship between the low graduation rates in Oregon and possible contributing factors. We realize that there are similar studies that have been done to find the highest correlated factors to success or failure in school; these studies, however, conclude that the largest contributing factor is low attendance, with chronically absent students 7.4 times

more likely to drop out of high school (Attendance Works, 2013). In this paper, we would like to explore further, to see which county characteristics cause such problems.

There are many factors that influence one's choice to drop out of high school. The factors we have chosen to include are local violent crime rates; income per family, the percentage of the population that is Hispanic, the percentage of family households with only one parent, alcohol use in adults, and teen pregnancy. From our results, family structure has the largest and clearest impact on educational attainment. In this paper, we will provide an overview of how we have chosen to represent our variables, describe the methodology of how we have processed our data, and finally interpret our results and explain their significance.

**Section Two: Data Overview**

Our study aims to explore many factors that impact high school graduation rates in Oregon, which has one of the lowest high school graduation rates in the United States. Below are a discussion of our data, an explanation of our model, and a prediction of our results. Our study includes 35 observations, all the counties in Oregon excluding Wheeler County. Wheeler County was excluded since many of our datasets did not provide data for Wheeler County, because this is the smallest county in Oregon. Our model is cross sectional, and all data is from the year 2014.

**Dependent Variable:** <u>Education (Edu):</u>  We are using education as our dependent variable in our regression model. Specifically, we are using the percentage of the population between 18-24 that were less than a high school graduate for 2014, published by the US Census Bureau.  There is a large range across the 35 counties measured in our study (the maximum of this dataset is 39% in Morrow County, and the minimum is 3.9% in Benton County). Both the median and mean of this dataset are 19.3%, showing that the data is evenly divided around the mean. The standard deviation is 7.15, so 95.4% of the counties fall between 5.02% and 33.6%.

| Education (2014) | |
| --- | --- |
| Percentage of the Population with Lower than a High School Education | |
| Mean | 19.32 |
| Standard Error | 1.21 |
| Median | 19.30 |
| Standard Deviation | 7.15 |
| Minimum | 3.9: Benton County |
| Maximum | 39.3: Morrow County |

*Measured for 35 counties in OR, not including Wheeler County*

**Independent Variables:**

Violent Crime rates (VCR): Our first independent variable is violent crime rates. We are using violent crime rates (the number of reported violent crime offenses per 100,000 people), published by County and Health Rankings and Roadmaps. The mean of the data is 167.8 and the median is 137. The dataset has a very large standard deviation of 109.28, so 95.4% of our data fall between 0 and 386.36. Multnomah county has the highest violent crime rate of 497. There are also two counties with a violent crime rate higher than 300, but the majority are under 300. Sherman County is the smallest county included in our study with a population of 1,790, which means that there may have been a few violent crimes in Sherman County but they are given a score of zero when this is standardized to be out of a population of 100,000. This may be a problem for our study, as the data for this county on this measure is not completely accurate.

| Violent Crime Rates (2014) | |
| --- | --- |
| Number of violent crimes per 100,000 people | |
| Mean | 167.80 |
| Standard Error | 18.47 |
| Median | 137.00 |
| Standard Deviation | 109.28 |
| Minimum | 0: Sherman County |
| Maximum | 497: Multnomah County |

*Measured for 35 counties in OR, not including Wheeler County*

Median Income (Inc): Another independent variable included in our model is the median household income for each county for the year 2014 reported in 2014 inflation-adjusted dollars from the US Census Bureau. This dataset has a mean of $45,364.31 and a median of $43,685,

meaning more of the counties fall below the mean than above. With a standard deviation of

$7,597.88, 95.4% of the data falls between $30,168.55 and $60,560.07. Washington county has

the highest median household income of $65,272, and is one of only two counties (the other

being Clackamas) with a median household income over $60,000. Malheur County has the

lowest household median income of $34,380, and several of the counties have median household

incomes under $40,000.

| Income Summary Statistics (2014) Median Household Income in 2014 Inflation-Adjusted dollars | |
|---|---|
| Mean | $45,364.31 |
| Standard Error | $1,284.28 |
| Median | $43,685.00 |
| Standard Deviation | $7,597.88 |
| Minimum | $34,380: Malheur County |
| Maximum | $65,272: Washington County |

*Measured for 35 counties in OR, not including Wheeler County*

Ethnicity (Eth): We are also including a measure of ethnicity in our model, specifically

the percent of the population in each county that is Hispanic, published by the U.S. Census

Bureau for 2014. We decided only to include Hispanic in our study instead of including different

races as the Hispanic population makes up the largest minority in the state of Oregon, and

therefor would be most likely to be statistically significant. The mean across the 35 counties is

11.33%, while the median is 7.91%. This suggests that more counties in Oregon have a Hispanic

population below 11.33%, and that some of counties above the mean have a very high

percentage. This is in fact true, as only five of the counties had a percentage over 20, and three of

these counties had a percentage over 30 including Morrow County, with the maximum of

33.18%. Wallowa County has the minimum of 2.5%. The standard deviation is 8.44 percentage

points, so 95.4% of the counties fall between 0 and 28.21.

| Ethnicity Summary Statistics (2014) Percentage of the Population that is Hispanic | |
|---|---|
| Mean | 11.33% |
| Standard Error | 1.43% |
| Median | 7.91% |
| Standard Deviation | 8.44% |
| Minimum | 2.50%: Wallowa County |
| Maximum | 33.18%: Morrow County |

*Measured for 35 counties in OR, not including Wheeler County*

Family Structure (FamStr): Another variable we are including is family structure, in the

form of percentage of households with children that are single-parent households, as published

by the US Census Bureau in 2014. The mean of the data set is 33.03% and the median is 33.26%.

These two numbers being so close tells us that the data is fairly evenly distributed, which is

backed-up by the fact that the minimum of 15.11% in Lake County and the maximum of 48.34%

in Jefferson County are similar distances from the median. The standard deviation is 6.46%, so

95.4% of the counties' fall between 20.11% and 45.95%.

| Family Structure Summary Statistics (2014) Percentage of Households that are Single-Parent Households | |
|---|---|
| Mean | 33.03% |
| Standard Error | 1.09% |
| Median | 33.26% |
| Standard Deviation | 6.46% |
| Minimum | 15.11%: Lake County |
| Maximum | 48.34%: Jefferson County |

*Measured for 35 counties in OR, not including Wheeler County*

Alcohol Use (Alcohol): We are also including excessive alcohol use as a variable,

measured by the percentage of adults reporting binge or heavy drinking as reported by County

Health Rankings of Oregon. We could not find specific data on drug use or smoking habits per

county but we believe they are closely correlated with alcohol binging, so have decided to use

the alcohol variable as a general substance use variable. The mean is 19.6% and the median is

19%, showing, as in the previous variable, that the selection is fairly evenly distributed. This is

reiterated by the small standard deviation of 1.56%, showing that 95.4% of the data is located

between 16.48% and 22.72%. The minimum is 17% in Crook County and the maximum is 24%

in Multnomah County.

| Alcohol Use Summary Statistics (2014) Percentage of Adults who report heavy drinking | |
|---|---|
| Mean | 19.6% |
| Standard Error | 0.26% |
| Median | 19% |
| Standard Deviation | 1.56% |
| Minimum | 17%: Crook County |
| Maximum | 24%: Multnomah County |

*Measured for 35 counties in OR, not including Wheeler County*

Teen Pregnancy (TeenPreg): The final variable we are including is teen pregnancy rates,

the number of births per 1000 female population ages 15-19, as reported by County Health

Rankings of Oregon. The mean is 31.4 and the median is 30, suggesting that slightly more

counties have a teen pregnancy rate of less than 31.4. However, there are two counties, Gillam

and Sherman, who qualify as having teen pregnancy rates of 0, possibly because their pregnancy-

population ratio is not significant enough to be translated into a population of 1,000. If we

disregard Gillam and Sherman Counties, the minimum teen pregnancy rate is 9 in Benton

County and the maximum rate is 65 in Malheur county. The standard deviation is 14.00, showing

that 95.4% of counties have a teen pregnancy rate between 3.4 and 59.4, and 68% of counties

have a rate between 17.4 and 45.4.

| Teen Pregnancy Summary Statistics (2014) Number of Births per 1000 female population ages 15-19 | |
|---|---|
| Mean | 31.4 |
| Standard Error | 2.37 |
| Median | 30 |
| Standard Deviation | 14.00 |
| Minimum | 0: Gillam & Sherman Counties |
| Maximum | 65: Malheur County |

*Measured for 35 counties in OR, not including Wheeler County*

**A Priori Hypotheses**

We predict that an increase in VCR, Eth, FamStr, Alcohol, and TeenPreg to cause an increase in Edu. All of these factors significantly impact the lives of children and the environment they grow up in. These factors will lead to a greater percentage of the population having less than a high school education. We predict that Inc will have the opposite effect. We believe the coefficents will be largest for Inc and FamStr as these two factors play a large role in the opportunities and support a child is given. We think that Alcohol and/or TeenPreg may not be statistically significant, as the rates for these are small across the counties in Oregon, compared to the other variables.

**Section 3: Methodology**

Theoretically, living in an area with high violent crime rates would lower the safety of children, causing negative psychological consequences and making it harder for children to attend and excel in school. Higher violent crime rates may suggest more high school students are involved in crime, for example gangs may be more prevalent in these areas, causing students to drop out of school. Living in a single parent home and living with parents who consume high amounts of alcohol can also lower graduation rates as these children may face more negative psychological consequences, may carry extra burdens at home, and may have less parental support, making attending and succeeding in school more difficult. Ethnicity may play a role as well, minorities are often disadvantaged in numerous ways, including educational success. Teen pregnancy rates were included since finishing high school is extremely hard to do for a teen who is pregnant and/or raising a child. Living in a family with higher income should allow the child access to better resources to help them succeed in school, and there is a higher chance their parents will have higher educational attainment and therefore encourage this in their children.

To model this relationship we used a non-linear OLS regression model. This model,

Model (1),was based on the following equation:

$Ln\_Edu = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(VCR) + \boldsymbol{\beta}_2(VCR^2) + \boldsymbol{\beta}_3(Inc) + \boldsymbol{\beta}_4(Eth) + \boldsymbol{\beta}_5(FamStr) + \boldsymbol{\beta}_6(Alcohol) + \boldsymbol{\beta}_7(TeenPreg)$

Ln_Edu is the natural log of the percent of the population between 18-24 with lower than

a high school education, VCR is violent crime rates, $VCR^2$ is violent crime rates squared, Inc is

median household income, Eth is the percentage of the population that is Hispanic, FamStr is the

percentage of households with children that are single parent homes, Alcohol is the percentage of

adults reporting binge or heavy drinking, and TeenPreg is the number of births per 1000 female

population ages 15-19.

After trying many different models with the datasets, we decided this model was the best

for several reasons.  Four variables in our model, Inc, $VCR^2$, FamStr, and Alcohol, are

statistically significant at the 99% level, and one variable, VCR, is statistically significant at the

95% level. In the basic linear OLS model only FamStr and Alcohol are statistically significant.

In addition, the adjusted R-squared for our model is higher than the adjusted R-squared in the

linear model: 0.636 compared to 0.489, showing that our model is capturing much more of the

variation in Y than the non-linear model.

Additionally, we compared this model to many other nonlinear models. When we

changed just the dependent variable to be a natural logarithm, income became statistically

significant at the 90% level. In these models, replacing any of the independent variables with the

natural log of these variables does not change the statistical significance of any of the variables.

However, when we added the square of VCR to this model, the output improved greatly. We

compared our model with the model:

$Ln\_Edu = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(VCR) + \boldsymbol{\beta}_a(Inc) + \boldsymbol{\beta}_3(Eth) + \boldsymbol{\beta}_4(FamStr) + \boldsymbol{\beta}_5(Alcohol) + \boldsymbol{\beta}_6(TeenPreg)$

The model we chose has a higher R-squared and more statistically significant variables. Before adding $VCR^2$, the log linear model had an adjusted R-squared of 0.572 with two statistically significant variables at the 99% level, FamStr and Alcohol, and one at the 90% level: Income. After adding $VCR^2$, the adjusted R-squared is .636, and many more of the variables are statistically significant as listed above.

## Section 4: Results and Interpretations

The table below shows the OLS output of our model in the second column, (1). The third

column, (2) shows results from the same model, excluding "Alcohol."

| OLS Regression of High School Graduation Rates in Oregon | | |
|---|---|---|
| Dependent Variable: Ln_edu, 35 observations | | |
| **Regressor** | **(1)** | **(2)** |
| Inc | 1.68014e-05*** (5.84086e-06) | -9.80787e-06 (7.23926e-06) |
| TeenPreg | 0.00524128 (0.00578591) | 0.00735486 (0.00836905) |
| VCR | −0.00304897** (0.00116611) | −0.0022883 (0.00161332) |
| sq_VCR | 7.50401e-06*** 2.59141e-06 | 3.17536e-06 (3.15348e-06) |
| Eth | −0.009105 (0.009207) | −0.0102585 (0.0117772) |
| FamStr | 0.0191604*** (0.00616908) | 0.0232886** (0.0111291) |
| Alcohol | −0.225662*** (0.0584225) | |
| Constant | 6.76652*** (1.61686) | 3.49467** (1.4031) |
| R-squared | 0.711210 | 0.417268 |
| Adjusted R-squared | 0.636338 | 0.292397 |
| These regressions were estimated from 2014 for every county in Oregon besides Wheeler County. The data was gathered from the U.S. Census Bureau and County Health Rankings and Roadmaps. Statistical significance levels are represented at the *** 1%, **5%, and *10% level. | | |

As stated in the previous section, Inc, VCR$^2$, FamStr, and Alcohol are statistically significant at the 99% level, and VCR is statistically significant at the 95% level in the first model. However, only two of our variables are economically significant: Family Structure and Alcohol. The other coefficients are very close to zero. An increase in the family structure by 1 percentage point leads to a 1.91% increase in the total population that does not have a high school diploma. Since having a high school diploma is so essential for the outcomes of an individual, we believe this 2% increase in the population with a high school diploma is a big enough difference to be considered economically significant.

According to our model, an increase in the amount of adults reporting heavy drinking by one percentage point leads to a decrease of 22.57% in the total population that does not have a high school diploma. This coefficient is very large and therefore economically significant. However, this finding contradicts the expectation on how parent's alcohol consumption impacts their children's educational attainment. One would expect higher rates of alcohol consumption in adults to increase the amount of adults reporting heavy drinking, but according to our results, this is not the case.

There are several reasons that could be causing this result from our data. First, our model could be subject to omitted variable bias. There could be another variable correlated with heavy alcohol consumption that affects high school graduation rates that we have not included in our model. There could also be a problem in the data itself. The data is self-reported, and since heavy alcohol consumption is viewed very negatively by society, many individuals may not have honestly reported their drinking habits.

Because of this problem in the variable "Alcohol" we also decided to include the same model with the Alcohol variable removed. By removing this variable the output changes

significantly. In this model there is only one variable that is statistically significant: family structure. However, now family structure is only significant at the 95% level. In addition, the adjusted R-squared drops from .636 to .292. This shows that the variable Alcohol was explaining a lot of the variance in Y in the first model.
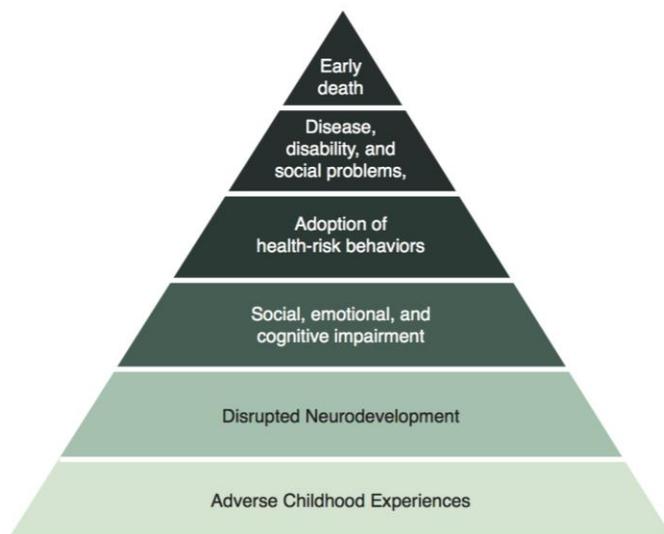
While we tried to include a model with many statistically and economically significant variables, finding variables that are not significant tells and important story as well. For example, violent crime rates and income are statistically significant, but in real world the impact these variable have on graduation rates is extremely small. Therefore, our model rules out violent crime and income as problems causing the low high school graduation rates Oregon has been experiencing. Teen pregnancy rates and ethnicity also do not contribute to causing low high school graduation rates, as these variables are not even statistically significant.

**Conclusion**

Our study aimed to find the relationship between a variety of factors and the low high school graduation rates in Oregon. We used a log linear regression model and found that an increase in family household with only one parent present had a negative impact on high school graduation rates. Our model also suggested that alcohol use in adults would increase high school graduation rates, which was very contrary to intuition, and may show a flaw in our model, or be representing changes in another variable we did not include. However, this also creates a need for future research. We hope to look further into alcohol use in Oregon and find several other datasets measuring alcohol and substance abuse in teens and adults, if available, to include in our model. This may help solve a possible problem with omitted variable bias and allow us to check if the problem lies in our chosen data set.

Overall, by finding that family structure has the most clear and significant effect on graduation rates, we conclude that factors related to the child's home environment growing up are incredibly important for their future life outcomes. From our study, the factors that influence the child's likelihood to graduate college are things out of the child's control, but factors their parents have some control over.

To further explore the importance of stability in a childhood home, we looked at a similar study done by the Centers for Disease Control titled *Adverse Childhood Experiences*, or ACEs. In this study, the researchers interviewed children, teens, and adults to discover correlations between traumatic or unstable childhoods with a lack of general wellness as adults. A series of ten questions were asked, including questions such as, "Were your parents ever separated or divorced?" and "Did a parent or other adult in the household often or very often swear at you, insult you, put you down, or humiliate you? or Act in a way that made you afraid that you might be physically hurt?" (ACES Too High News, 2012). Many of their results were astonishing: people who answered "yes" to four or more of the questions have a 16% higher chance of becoming adult alcoholics and a 95% chance of being on antidepressants by the age of 50, were 20% more likely to have attempted suicide, and were 40% more likely to become pregnant as a teenager.



This study created summary model, featured here, which explains the mechanism by which Adverse

Childhood experiences influence health and well-being throughout a person's lifespan (CDC-Kaiser Permanente, 2016). The study also stated that a large amount of these negative outcomes can be avoided or lessened with the presence of at least one stable adult in a child's life. Overall, our study on high school graduation rates in Oregon shows that the stable support from home and the environment a child grows up in are very important in determining a child's future outcomes. In addition, our study reveals the many factors that are not causing the low graduation rates in Oregon, paving the way for future research on other factors not included in this study.

References

ACES Too High News. (2012) ACES got your ACE score? Retrieved from

    https://acestoohigh.com/got-your-ace-score/

Attendance Works. (2012, September 13). Utah Study: Chronically absent students 7.4 times

    more likely to drop out. Retrieved from http://www.attendanceworks.org/utah-study-

    chronically-absent-students-7-4-times-more-likely-to-drop-out/

Bureau of Labor Statistics. (2015) Earnings and unemployment rates by educational attainment.

    Retrieved from http://www.bls.gov/emp/ep_chart_001.htm

CDC-Kaiser Permanente. (2016). Adverse childhood experiences (ACEs). Retrieved from

    https://www.cdc.gov/violenceprevention/acestudy/

County Health Rankings and Roadmaps. (2014). Excessive drinking. Retrieved from

    http://www.countyhealthrankings.org/app/oregon/2014/measure/factors/49/data

County Health Rankings and Roadmaps. (2014). Teen births Retrieved from

    http://www.countyhealthrankings.org/app/oregon/2014/measure/factors/14/map

County Health Rankings and Roadmaps. (2014). Violent crime. Retrieved from

    http://www.countyhealthrankings.org/app/oregon/2014/measure/factors/43/data?sort=sc-

    0

Hammond, B. (2015, January 29). Oregon fails to improve dismal graduation rate. *The*

    *Oregonian.* Retrieved from

    http://www.oregonlive.com/education/index.ssf/2015/01/oregon_fails_to_improve_disma

    l.html

Herold, Benjamin. "A 'collapsed' Job Market for Dropouts." The Notebook [Philadelphia] Feb.

2012. The Notebook. http://thenotebook.org/february-2012/124481/collapsed-job-

market-dropouts

United States Census Bureau. (2014). ACS demographic and housing estimates [DP05].

Available at

http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t

United States Census Bureau. (2014). Educational attainment [S1501]. Available at

http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t

United States Census Bureau. (2014). Households and families [S1101]. Available at

http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t

United States Census Bureau. (2014). Income in the past 12 months [S1901]. Available at

http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t