Jesse Hendren-Mills

July 2020

hendreje@mail.uc.edu

is Labor Market Value Affected by “bad” HABITS?

A Panel Data Research Approach

Table of Contents

[1. Introduction 1](#_Toc107264659)

[2. Literature Review 2](#_Toc107264660)

[3. Data 3](#_Toc107264661)

[4. Empirical Methodology 5](#_Toc107264662)

[5. Results 7](#_Toc107264663)

[6. Conclusion 9](#_Toc107264664)

[7. References 11](#_Toc107264665)

[**Figure 1: Dependent Variable Summary Statistics 4**](#_Toc107264638)

[**Figure 2: Multivariate Panel Regression Results 8**](#_Toc107264639)

[**Equation 1: Pooled Ordinary Least Squares Econometric Model 7**](#_Toc107264646)

[**Equation 2: First Differences Squares Econometric Model 7**](#_Toc107264647)

[**Equation 3: Fixed Effects Econometric Model 7**](#_Toc107264648)

[**Equation 4: Random Effects Econometric Model 7**](#_Toc107264649)

# 1. Introduction

In every state in the United States there are sin taxes. These particular taxes are implemented by the government to discourage citizens from buying the products that is deemed to be harmful to consume, such as alcohol or tobacco products. Through extensive medical research, these products have been linked to numerous cardiovascular health issues and often lead to life-long health issues. With this in mind, having life-long health issues tend to lead to higher medical expenditures but does the amount someone consumes of these products indicate their labor market value? For this project, I will be evaluating how smoking, alcohol consumption, education rates, African American state population percentage, female state population percentage, and an interaction term of African American and female population percentage to se the effect on household income. By using these six variables, I will try to answer the question if “bad” habits have an effect on labor market value. This will be achieved by using a panel data method that will consist of U.S. states as the individuals (i) from the years 2018 and 2020 (t).

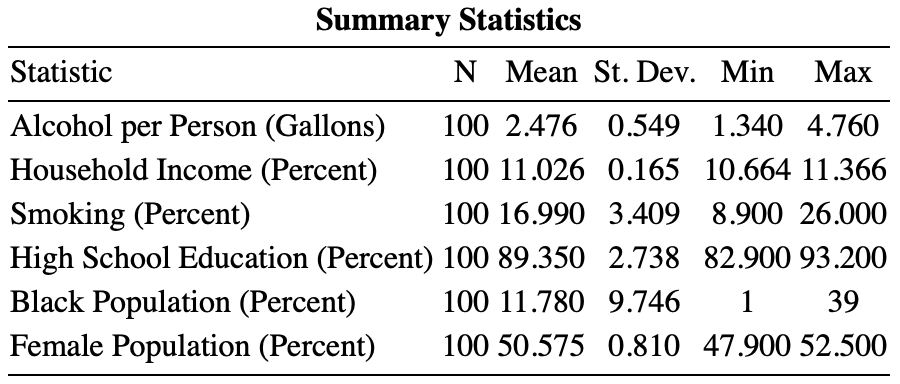
# 2. Literature Review

For this study, there were not any previous studies that had conducted a panel data study with these variables. However, using studies that have done a cross-sectional approach may be helpful in identifying deficiencies in my models. One issue was simultaneity, the percentage of individuals that smoke and drink.[[1]](#footnote-1) For this, I will attempt to get rid the problem by not including the percentage of individuals who drink but the amount in which individuals drink on average. This in turn doesn’t identify, if or if not, someone is drinking but the amount these groups of people consume on average. In previous studies, smoking and the volume of drinking has a major effect on an individual’s health.[[2]](#footnote-2) Some sources suggest that smoking has a negative effect, roughly 10 percent, on wages while alcohol consumption has a positive effect, roughly 24 percent, on wages.[[3]](#footnote-3) Utilizing a more dynamic model is essential due to the mixed results in one study.[[4]](#footnote-4) This study suggests that doing a simple multilinear regression is not enough to determine true causality. That a first difference method could provide a more wholesome result and give a better understanding to the question. To further understand the true effect of these two variables on household income, it would be beneficial to take out the unobserved and time effects these variables experience to get a true effect. One of the issues that was stated in previous literature reviews that causes concern is the possible endogeneity of these variables. Alcohol consumption and smoking could be considered habits that are influenced by social functions or social norms. This empirical study recognizes this and will attempt to isolate as much endogeneity that could be present.

# 3. Data

The data used for this project is a collection of data from multiple sources and compares data from all states of the U.S. in the two given years. As stated earlier, this analysis considers the effects smoking, alcohol consumption, and education above or equal to a high school diploma have on household income. Income is measured as a percentage of United States dollars by taking a log and was found from two different sources[[5]](#footnote-5)[[6]](#footnote-6) for each year. The smoking rate variable is showing how much of the population in a given state smokes in a given year.[[7]](#footnote-7)[[8]](#footnote-8) Alcohol consumption is measured in gallons consumed and is an average for its given state.[[9]](#footnote-9)[[10]](#footnote-10) Lastly, educational attainment is a percentage of a given state that has received a high school diploma or higher for each year.[[11]](#footnote-11)[[12]](#footnote-12) Additionally, adding demographic variables can provide an insight to whether these particular groups of people, African Americans and females, are effected. Below are the summary statistics of the given variables.

**Figure 1: Dependent Variable Summary Statistics**



The data that was collected for this analysis, individually, would be considered to be cross-sectional data. By combining two cross sectional data sets over two separate time periods, I was able to make a panel data set. By utilizing a panel data method on these data points, I can control for time and individual effects of these variables over time. Additionally, we can see if the variables are endogenous with each other and test for heteroskedasticity, serial correlation, and inconsistency.

# 4. Empirical Methodology

The panel data methods that will be used during this project are pooling ordinary least squares (POLS), first-difference (FD), fixed-effects (FE), and random-effects (RE). By using panel data, we can estimate consistent individual characteristics by upholding all of the six ordinary least squares assumptions.[[13]](#footnote-13) These assumptions are crucial to uphold due to statistical properties and to avoid bias or false positive estimations.

First, I will test if exists in the model. By testing if exists we can determine whether to utilize POLS or the other models available. If does exist, homoskedasticity is violated and therefore eliminating the POLS model. With first difference, fixed-effects, and random effects models remaining, utilizing a Hausman Test will help determine which of the remaining models are viable. In a Hausman Test, is being tested for correlation with the other independent variables. If is correlated with the other independent variables, another OLS assumption is violated, zero conditional mean assumption. With this assumption broken we can no longer consider random effects to be the best model for this cause. Finally, the last two models available would be first-difference and fixed-effects models. For this, we want to test the covariance of each model’s idiosyncratic error, , over t periods and see if those are equal to zero. These errors will be regressed on their own errors in previous time periods to provide a metric called rho, ρ. If ρ is equal to zero, saying that there is no covariance between the idiosyncratic error of any time period, then fixed-effects is the best estimated. If ρ is not zero than first difference is the best estimator.

**Equation 1: Pooled Ordinary Least Squares Econometric Model**

**Equation 2: First Differences Squares Econometric Model**

**Equation 3: Fixed Effects Econometric Model**

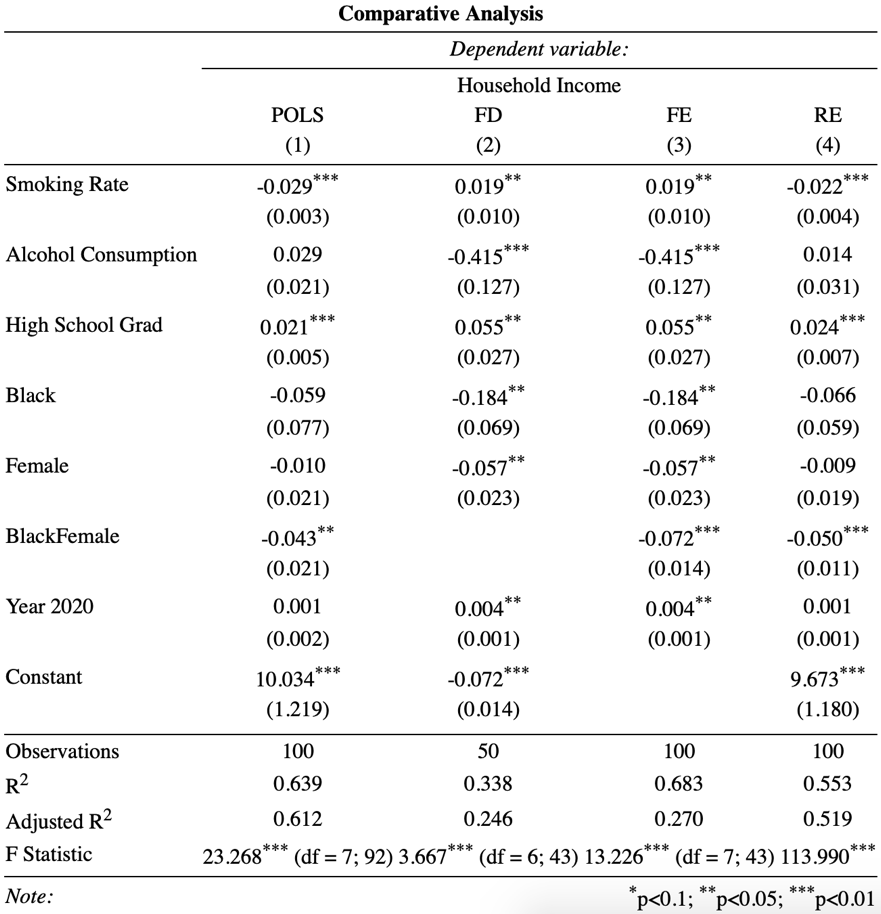
**Equation 4: Random Effects Econometric Model**

# 5. Results

After implementing the empirical methodology listed out above, the results below do provide a wide variety of implications. We can see that high school education is statistically significant in every model but find that the standard deviation varies in individual models. Additionally, alcohol consumption is not statistically significant in any of the models provided. This could be due to alcohol consumption not having an impact on labor market value because of cultural undertones both good and bad. During the empirical process, was found to exist which violated the property of homoskedasticity. Further evaluation showed that in-fact was correlated with the other independent variables and violated the zero conditional mean assumption. With both of these tests conducted, POLS and random-effects were no longer considered to be viable statistical models to predict the actual effects of the independent variables on the dependent variable. Lastly, when test for the presence of idiosyncratic error over multiple time periods, rho was found to be statistically significant at the 99 percent confidence interval. This meaning that rho is likely to not equal zero and therefore concluding that first difference is the best estimator for this study.

To further evaluate the fixed-effects model we have chosen, a coefficient test was conducted to test for robustness in serial correlation. The results show that while the year and education estimators decreased from a 99 percent confidence interval to a 95 percent confidence interval, these are considered to be the most accurate estimates for this model.

**Figure 2: Multivariate Panel Regression Results**



# 6. Conclusion

The results from previous studies suggest that the consumption of alcohol and smoking do have a significant impact on someone’s labor market value. Other studies utilize dynamic models to determine these variables true effects over multiple time periods. In this paper, the results were surprisingly similar. By utilizing a panel data model, smoking and alcohol consumption were considered to be statistically significant at the 99 percent confidence interval by settling on the first difference estimator. This measurement is considered to fulfill traditional OLS assumptions and has robust standard errors. Additionally, being a female and, to a less extent, being an African American woman did yield statistically significant results at the 95 percent confidence interval. There is also the possibility of smoking and alcohol consumption to be endogenous of each other which would be hard to isolate their true effects within any model and cause misleading results.

# 7. References

**Auld, M. Christopher. 2005. "Smoking, Drinking, and Income." *The Journal of Human Resources* (University of Wisconsin Press) 40 (No. 2): 505-518.**

**Coppola, Gianluigi, Niall O'Higgins, and Claudio Pinto. 2015. *Smoking, Drinking, Never Thinking of Tomorrow: Income and Risky Choices amonst Young Adults in the UK.* Discussion Paper No. 9249, Bonn: Institute for the Study of Labor.**

**Federal Reserve Economic Data. 2019. *2018 Educational Attainment.* Accessed April 19, 2020. https://fred.stlouisfed.org/tags/series?t=educational%20attainment%3Bsecondary%3Bstate&ob=pv&od=desc.**

**Grucza, Richard A. Ph.D., M.P.E, and Laura J., M.D. Bierut. 2008. "Cigarette Smoking and the Risk for Alcohol Use Disorders Among Adolescent Drinkers." *National Institutes of Health* 2046-2054.**

**Kaiser Family Foundation. 2018. *Adults Who Report Smoking by Gender.* Accessed May 7, 2020. https://www.kff.org/other/state-indicator/smoking-adults-by-gender/?currentTimeframe=0&selectedDistributions=all-adults&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.**

**Statista. 2018. *Median household income in the United States by state 2018.* Accessed May 1, 2020. https://www.statista.com/statistics/233170/median-household-income-in-the-united-states-by-state/.**

**—. 2019. *Per capita alcohol consumption of all beverages in the U.S. by state 2018.* Accessed May 20, 2020. https://www.statista.com/statistics/442848/per-capita-alcohol-consumption-of-all-beverages-in-the-us-by-state/.**

**Wooldrdge, Jeffrey M. 2018. *Introductory Econometrics: A Modern Approach, Seventh Edition.* Boston: Cenage Learning, Inc.**

**World Population Review. 2020. *Alcohol Consumption By State 2020.* February 1. Accessed March 15, 2020. https://worldpopulationreview.com/states/alcohol-consumption-by-state/.**

**—. 2020. *Educational Attainment By State 2020.* Accessed May 7, 2020. https://worldpopulationreview.com/states/educational-attainment-by-state/.**

**—. 2020. *Median Household Income By State 2020.* January 15. Accessed March 20, 2020. https://worldpopulationreview.com/states/median-household-income-by-state/.**

**—. 2020. *Smoking Rates by State 2020.* March 3. Accessed April 18, 2020. https://worldpopulationreview.com/states/smoking-rates-by-state/.**

1. (Auld 2005) [↑](#footnote-ref-1)
2. (Grucza and Bierut 2008) [↑](#footnote-ref-2)
3. (Auld 2005) [↑](#footnote-ref-3)
4. (Coppola, O'Higgins and Pinto 2015) [↑](#footnote-ref-4)
5. (World Population Review 2020) [↑](#footnote-ref-5)
6. (Statista 2018) [↑](#footnote-ref-6)
7. (Kaiser Family Foundation 2018) [↑](#footnote-ref-7)
8. (World Population Review 2020) [↑](#footnote-ref-8)
9. (Statista 2019) [↑](#footnote-ref-9)
10. (World Population Review 2020) [↑](#footnote-ref-10)
11. (World Population Review 2020) [↑](#footnote-ref-11)
12. (Federal Reserve Economic Data 2019) [↑](#footnote-ref-12)
13. Linear in Parameter, Random Sampling, No Perfect Collinearity, Zero Conditional Mean Assumption, Homoskedasticity, and Normality [↑](#footnote-ref-13)