Quantifying the Effects of Alcohol Consumption on Labor Force Participation

Jeremy Chan

ECN 140 Term Project

998284653

Using National Longitudinal Survey of Youth data, the author explores the relationship between labor force participation and frequency of excessive alcohol consumption. The analysis includes controls for individual demographic variables, geographic demographic variables, and family demographic variables. The author found that the most comprehensive model created from available variables points to an increase in probability of labor force participation when drinking in excess one to five times a month compared to zero times, controlling for other variables. Examination of higher frequencies as well as causality testing was inconclusive.

**Introduction**

With the United States’ economic recovery from the Great Recession in full swing, it is often of interest to researchers to look at labor force participation and unemployment rates as signs of economic recovery. However, there exist factors external to the economy which affect the rate at which people are employed. Factors such as alcoholism, family size, or geographic area can greatly affect an individual’s decision to participate in the labor force. Moreover, there may be inferences that can be drawn about a person’s employment status based on some of their demographic features. It is therefore of interest to construct a model which allows for the examination of employment rate individuals based on various demographic effects.

Specifically, this paper aims to examine the relationship between excessive alcohol consumption and employment rates of individuals using data from a longitudinal study conducted by the National Longitudinal Survey of Youth (NLSY). In the context of this paper, being in the labor force is defined as the rate of people who are employed or seeking employment as a proportion of the total number of observations. To accomplish this, a model will be constructed which models the effects of various frequencies of excessive alcohol consumption and the effect each level has on the probability of an individual being in the labor force. This model will be contrasted with simpler models to better capture the effects of excessive alcohol consumption. Following the construction of the model, the model’s fit will be evaluated through a variety of basic methods to assess the accuracy of this model moving forward. Finally, future considerations on improvements to the model will be discussed.

**Data**

The data used in this paper comes from a longitudinal study conducted by the NLSY in both 1989 and 1994. Specifically, this paper focuses on data from 1994 in order to present a more modern model of the relationship between alcohol consumption and employment.

The primary labor variable of interest is labor force participation. Originally coded as a four-category variable (employed, unemployed, out of labor force, in active armed forces), the variable was condensed into a single, binary output variable. There were no observations in the “In Active Armed Forces” group, so the “Employed” (*emp*) dummy variable was coded as a 1 if the original employment status was employed or unemployed and a 0 if the original status was out of the labor force. This codification allows us to take into account natural unemployment and capture the probability that an individual is in the labor force (i.e., either employed or unemployed and seeking work) given their alcohol consumption tendencies. It should be noted that the assumption was made that the “Unemployed” group encompassed people who were indeed seeking work and not out of the labor force. **Table 1**, which displays the descriptive statistics for each group, can be found in the appendix.

The primary alcohol consumption variable was a six-category variable entitled “drnk6m” which indicated the number of times in the past month that an individual had 6 or more drinks (noted as excessive alcohol consumption) in one sitting. This variable was split into 7 dummy variables, each named group x where x took the values of 0 to 6. The table below explains the groups:

|  |  |
| --- | --- |
| Group | Excessive Alcohol Consumption (6+ drinks) Frequency |
| 0 | No occurrences in the past month |
| 1 | Once in the past month |
| 2 | Two or three times in the past month |
| 3 | Four or five times in the past month |
| 4 | Six or seven times in the past month |
| 5 | Eight or nine times in the past month |
| 6 | 10 or more times in the past month |

The descriptive statistics for each group can be found in **Table 2** in the appendix. Additionally, a bar graph displaying the proportion of each group which is employed can be found in **Figure 1** in the appendix.

Individual characteristics captured in the regressions included the following characteristics: age, gender, race, health, the number of years of education completed, family size, intelligence test percentile, and marital. Descriptive statistics and an explanation of measurement methods of these variables can be found in **Table 3** in the appendix.

Geographic characteristics captured in the regressions included the following characteristics: region and local unemployment rate. Descriptive statistics and an explanation of measurement methods of these variables can be found in **Table 4** in the appendix.

Family characteristics captured in the regressions included the following characteristics: Father’s level of education, mother’s level of education, father’s working status, and mother’s working status. Descriptive statistics and an explanation of measurement methods of these variables can be found in **Table 5** in the appendix.

**Methods**

To estimate the effects of the frequency of excessive alcohol consumption on labor force participation, a simple regression was created to see the raw effects. The equation of the regression is presented below (with group numbers corresponding to those presented in the data section):

This regression captures the raw effects of frequency of excess alcohol consumption on the probability of being in the labor force. In this regression, it is expected that, in general, the betas will take on negative signs, as labor force participation should decline in correlation with an increasing frequency of excess alcohol consumption. However, this regression does not account for a variety of other factors which may affect the probability of being in the labor force outside of the alcohol consumption frequency group.

To account for various individual characteristics, the following regression attempts to account for individual demographic factors as described in the data section:

where Xi is a vector of observed individual characteristics. This approach allows for keeping individual demographics constant and observing the effects of alcohol consumption frequency on probability of labor force participation. Again, it is expected that all off the betas related to excessive alcohol consumption frequency are negative.

The first two regressions do not account for geographic differences, as labor force participation may differ when living in different sections of the country. To that end, a third regression was created, taking the form:

where Gi represents a vector of geographic demographic factors as described in the data section. As in the first two regressions, it is expected the betas related to frequency of alcohol consumption to all be negative and increase as the frequency of excess consumption increases. This regression fails to capture one more aspect of an individual’s characteristics which may affect their labor force participation, so an additional regression was created.

The final, most comprehensive regression created built upon the regression presented in (3) by adding some basic family demographics and takes the form:

where Fi represents a vector of family demographics which were added as a control as described in the data section. In keeping with the first three regressions, the betas are expected to be negative in sign and increase in magnitude as the group number increases. Essentially, the group with the highest frequency of excessive alcohol consumption should have the lowest probability of being in the labor force. By controlling for the other external factors, this regression equation should be able to answer if the probability of being in the labor force is lowered by an increased frequency of excessive alcohol consumptions.

**Results**

The fourth regression was selected as the final model because of the comprehensive modeling it accomplished; it accounted for all of the additional demographic factors, had the highest R2, and the accuracy of the model was relatively high when predicting observed values. The confusion matrices for each model are shown in **Tables 6-9** in the appendix. While this is the most comprehensive model given the available data, there is the possibility for omitted variable bias, as the R2 value is only around .18. Additionally, there may be further interactions between demographic variables and the excessive alcohol consumption frequency groups which were not captured in this regression. For instance, there might be a bias for increased excessive alcohol consumption associated with a parents’ frequency of excessive alcohol consumption.

In the fourth and most comprehensive regression, we see that groups 1, 2, and 3 all have an increased probability of labor force participation above group 0 while groups 4, 5, and 6 all have decreased labor participation, holding all other demographic factors constant. From groups 0 through 2, the increase in probability of labor force participation over group 0 dropped. In group 3, however, increased the probability of labor force participation over group 0 by a similar amount to the increase when going from group 0 to group 1. Groups 4 through 6 sees the same exponential curve as they all decrease the probability in labor force participation, but the effect starts large and becomes smaller, holding all else equal.

The results of all four regression models are presented in **Table 10** in the appendix. In general, through all four models, the magnitude of the effect on labor force participation probability shrank as the frequency of excessive alcohol consumption increased. The exception to this is that across all regressions, group 3 (excessive alcohol consumption four to five times a month) saw a relatively large increase in probability of labor force participation. Additionally, once demographic controls were added to the regressions, decreased labor force participation probability was observed in groups three through 6. Interestingly, in regression 1, all frequencies of excessive drinking above zero saw an increase in probability of labor force participation. Theoretically, this may be due to conflating so many demographic variables together. Furthermore, the omission of controlling for the “health” variable seemed to have a large effect, as that variable seemed to have the largest impact on labor force participation across all regressions. It is expected that upon including “health” into all regressions, the signs would stabilize as consistent across all regressions.

**Conclusions**

Overall, the results of the model from regression 4 did not match the expectations of negative betas at every level and an increasingly larger effect as the frequency of excess alcohol consumption increased. While probability of labor force participation did decrease in groups 4 through 6, probability of labor force participation actually increased in groups 1 through 3 in comparison to group 0, holding all else constant. These results ultimately point to the need for more analysis as to why the regressions did not fit the expected form and how to improve them to better capture the true causal reasons for probability of labor force participation. At first inspection, it appears that frequency of excessive alcohol consumption is not a primary determining factor of labor force participation.

In summary, the data does not conclusively point to a single, generalizable conclusion about the relationship between increased frequency of excessive alcohol consumption and the probability of labor force participation. Instead, the only statistically valid conclusions are that individuals who consumed six or more drinks in a single sitting between one and five times a month saw an increased probability of labor force participation. Further research is needed to conclude that this is a causal relationship and that the drinking affects the labor force participation — rather than the other way around — and to verify that the effects of alcohol consumption are indeed statistically significant when included in context with other demographic variables. Specifically, the health variable seemed to have a much larger impact and may warrant additional examination in the context of these regressions.

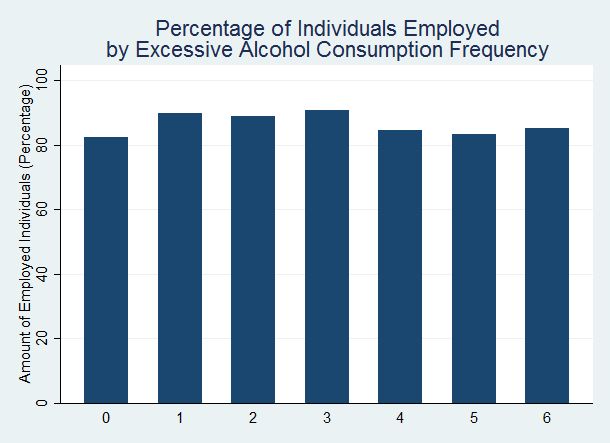
**Appendix**

**Table 1**. Descriptive statistics for labor force participation. Mean and SD are recorded as proportions and should be scaled by 100 times to obtain a percentage.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **SD** | **Min.** | **Max.** |
| In Labor Force | .844049 | .362836 | 0 | 1 |

**Table 2**. Descriptive statistics for frequency of excessive alcohol consumption. Variables recorded as frequency of having six or more drinks in one sitting in the past month. Mean and SD are recorded as proportions and should be scaled by 100 times to obtain a percentage.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **SD** | **Min.** | **Max.** |
| Group 0 (0 times) | .6670156 | .4713163 | 0 | 1 |
| Group 1 (1 time) | .0936005 | .2912939 | 0 | 1 |
| Group 2 (2 or 3 times) | .117823 | .3224225 | 0 | 1 |
| Group 3 (4 or 5 times) | .0506878 | .2193758 | 0 | 1 |
| Group 4 (6 or 7 times) | .0212321 | .1441678 | 0 | 1 |
| Group 5 (8 or 9 times) | .0098684 | .0988559 | 0 | 1 |
| Group 6 (10 or more times) | .0397727 | .1954394 | 0 | 1 |



**Figure 1.** This figure shows the employment percentages for each group of excessive alcohol consumption frequency. A value of 0 on the x-axis meant that an individual had not had 6 or more drinks in one sitting in the past month, a value of 1 indicated once, a value of 2 indicated 2 or 3 times, a value of 3 indicated 4 or 5 times, a value of 4 indicated 6 or 7 times, a value of 5 indicated 8 or 9 times, and a value of 6 indicated 10 or more times. According to the graph, employment was fairly high and stable across all groups.

**Table 3**. Descriptive statistics for individual characteristic variables used in the labor-alcohol model. Male, Black, Hispanic, Other Race, Health, Never Married, and Married are all represented as proportions and should be multiplied by 100 to obtain a percentage. Age is coded in years. Gender is coded as a dummy variable where a male individual is recorded as 1. Race is coded as three dummy variables where a value of 1 for *race1* indicates a Hispanic individual, a value of 1 for *race2* indicates a Black individual, and a value of 1 for *race3* indicates all other races. Health is coded as a dummy variable where a value of 1 indicates that the individual has a health issue limiting the type or amount of work they can do. The number of years of education completed is coded as an integer from 0 to 20, inclusive. Family size is coded as an integer. Intelligence test percentile is coded as a value between 1 and 99, inclusive. Marital status is coded as three dummy variables where a value of 1 for *marst1* indicates an individual has never been married, a value of 1 for *marst2* indicates an individual is married with a spouse present, and a value of 1 for *marst3* encompasses all other marital statuses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **SD** | **Min.** | **Max.** |
| Age | 32.89653 | 2.220882 | 29 | 37 |
| Male | .4982057 | .5000342 | 0 | 1 |
| Black (*race1*) | .1848086 | .3881712 | 0 | 1 |
| Hispanic (*race2*) | .2873804 | .4525743 | 0 | 1 |
| Other Race (*race3*) | .527811 | .4992633 | 0 | 1 |
| Health | .0783493 | .2687405 | 0 | 1 |
| Family Size | 3.190789 | 1.604193 | 1 | 13 |
| Never Married | .2654007 | .4415794 | 0 | 1 |
| Married (with Spouse) | .5337919 | .4988941 | 0 | 1 |
| Intelligence Test Percentile | 40.63382 | 28.80908 | 1 | 99 |

**Table 4**. Descriptive statistics for geographic characteristic variables used in the labor-alcohol model. All variables besides Local Unemployment are recorded as proportions and should be multiplied by 100 to obtain a percentage. Region is coded as a dummy variable where a value of 1 in *reg1* indicates an individual living in the Northeast, a value of 1 in *reg2* indicates an individual living in the North Central, a value of 1 in *reg3* indicates an individual living in the South, and a value of 1 in *reg4* indicates an individual living in the West. Local unemployment rate is coded as a midpoint of a range to avoid individual identification.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **SD** | **Min.** | **Max.** |
| Northeast | .1719498 | .377365 | 0 | 1 |
| North Central | .2444677 | .4298033 | 0 | 1 |
| South | .3779904 | .4849215 | 0 | 1 |
| West | .2055921 | .4041638 | 0 | 1 |
| Local Unemployment | 7.327303 | 2.763158 | 1.5 | 16.5 |

**Table 5**. Descriptive statistics for family characteristic variables used in the labor-alcohol model. Dad working at 14 and Mom working at 14 are recorded as proportions and should be multiplied by 100 to obtain a percentage. Father’s level of education is coded as an interaction term which takes the value of 0 if the individual did not live with their father at age 14 or the number of years of education an individual’s father has if the individual lived with their father. Mother’s level of education is coded as an interaction term which takes the value of 0 if the individual did not live with their mother at age 14 or the number of years of education an individual’s mother has if the individual lived with their mother. Father’s working status is an interaction term taking the value of 0 if the individual did not live with their father at age 14 or 1 if the individual lived with their father and their father worked at at age 14. Mother’s working status is an interaction term taking the value of 0 if the individual did not live with their mother at age 14 or 1 if the individual lived with their mother and their mother worked at at age 14.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **SD** | **Min.** | **Max.** |
| Dad’s Education | 7.529157 | 6.200652 | 0 | 20 |
| Mom’s Education | 9.76256 | 4.548345 | 0 | 20 |
| Dad working at 14 | .6695574 | .4704077 | 0 | 1 |
| Mom working at 14 | .5134569 | .4998562 | 0 | 1 |

**Table 6**. Confusion matrix for regression (1). Note: All observations were predicted to be employed according to this model. Total incorrect predictions: 1,043

|  |  |  |  |
| --- | --- | --- | --- |
| Actual Status -> | Unemployed | Employed | Total |
| Prediction: Employed | 1,043 | 5,645 | 6,688 |
| Total | 1,043 | 5,645 | 6,688 |

**Table 7**. Confusion matrix for regression (2). Total incorrect predictions: 958

|  |  |  |  |
| --- | --- | --- | --- |
| Actual Status -> | Unemployed | Employed | Total |
| Prediction: Unemployed | 194 | 109 | 303 |
| Prediction: Employed | 849 | 5,536 | 6,385 |
| Total | 1,043 | 5,645 | 6,688 |

**Table 8**. Confusion matrix for regression (3). Total incorrect predictions: 963

|  |  |  |  |
| --- | --- | --- | --- |
| Actual Status -> | Unemployed | Employed | Total |
| Prediction: Unemployed | 197 | 117 | 314 |
| Prediction: Employed | 846 | 5,528 | 6,374 |
| Total | 1,043 | 5,645 | 6,688 |

**Table 9**. Confusion matrix for regression (4). Total incorrect predictions: 961

|  |  |  |  |
| --- | --- | --- | --- |
| Actual Status -> | Unemployed | Employed | Total |
| Prediction: Unemployed | 199 | 117 | 314 |
| Prediction: Employed | 844 | 5,528 | 6,374 |
| Total | 1,043 | 5,645 | 6,688 |

**Table 10**. Results from the regressions presented in the methods section. Values for each variable are marked by regression number in the column. Robust standard errors, corrected for heteroscedasticity, are presented in parentheses below each variable. Other demographic features are not listed in the regression, but noted in the table. Significant effects are marked as follows: \* significant at 10% level, \*\* significant at 5% level, \*\*\* significant at 1% level.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Regression 1** | **Regression 2** | **Regression 3** | **Regression 4** |
| Group 0 (0 times) | .9273741 \*\*\* (.0220148) | .3275628 (.3502502) | .4876514 (.3542824) | .4630074 (.3580044) |
| Group 1 (1 time) | .3415368 \*\*\* (.0713776) | .2922777 \*\*\* (.0809158) | .2984774 \*\*\* (.0813007) | .2980868 \*\*\* (.0815498) |
| Group 2 (2 or 3 times) | .2969972 \*\*\* (.0631803) | .1935202 \*\*\* (.0690338) | .1947237 \*\*\* (.0689406) | .1951796 \*\*\* (.069092) |
| Group 3 (4 or 5 times) | .4045325 \*\*\* (.0977884) | .4443506 \*\*\* (.1151363) | .4466745 \*\*\* (.115585) | .4488687 \*\*\* (.1161596) |
| Group 4 (6 or 7 times) | .0881435 (.1293636) | -.0559846 (.1305149) | -.0646864 (.1308887) | -.0623975 (.1313956) |
| Group 5 (8 or 9 times) | .0400475 (.1849322) | -.0715619 (.1818845) | -.0647624 (.1835336) | -.0553729 (.182585) |
| Group 6 (10 or more times) | .1236815 (.096988) | -.0103424 (.1089025) | -.0202141 (.1086962) | -.0247628 (.1100822) |
| Individual Demographics | - | X | X | X |
| Geographic Demographics | - | - | X | X |
| Family Demographics | - | - | - | X |