American Alcohol Consumption from 2013-2023

Jolie Liu

Simmons University

ABSTRACT

Alcohol consumption has been studied through examining individual demographics or larger alcohol market information. This study brings together those two perspectives and further investigates the interaction between economic theories such as demand theory and Keynesian Economics, as well as contradictions in current literature over trends and predictors of alcohol consumption in individuals. Data from the Bureau of Labor Statistics (BLS) included information about the alcohol consumption habits and demographics of individuals, while data from the St. Louis Federal Reserve (FRED) included information about national economic indicators such as the Economic Policy Uncertainty (EPU) index and the consumer price index (CPI) of alcohol. Logistic regression was used to predict the log odds of purchasing alcohol for a particular consumer unit. Both CPI and EPU were not statistically significant in the regression, indicating that the price of alcohol and uncertainty in the economy are irrelevant as one of the deciding factors whether a consumer unit will purchase alcohol.

INTRODUCTION AND BACKGROUND

Alcohol remains a unique consumer good. Although alcohol is consumed in social and celebratory events, it is neither a seasonal nor necessary good, whose consumption can indicate the prevalence of alcoholism and provide crucial cultural insight regarding the consumption behavior of Americans. Remaining a wide cultural staple and a large market, the alcohol industry rakes in over 122.6 billion dollars annually in America and employs over 2 million people by direct retail sales of alcohol alone (American Beverage Licensees 2018). Supporting the power of this lucrative industry are the many individuals and households that take part in its consumption. As such, one essential economic theory that provides an understanding of

consumer choices outlines the relationship between the likelihood of individuals to purchase a good and a good's price as inversely proportional: individuals are less likely to purchase a good if the price is high and vice versa (law of demand). Another critical theory comes from Keynesian Economics, a branch concerned with patterns of consumption at a national level and poses the idea that the relative state of the economy is an important factor in consumption. Individuals that are pessimistic about the general state of the economy are more motivated to hold on to their current savings and consume less (Keynes). By studying the alcohol consumer choice, it will unlock new information on the cultural consumption of Americans, may indicate the the prevalence of alcoholism, may shed light on the reaction of Americans during economic downturns, and investigates the consumption behavior of a good that impacts all Americans in both large and small ways, whether it is through being offered at dinner, through sobriety, or through the massive impact alcohol has on the American economy. In particular, this paper will examine the price of alcohol and public perception of the state of the economy as relevant indicators contributing to the probability of whether a household will purchase alcohol or not and will focus on combining individual demographics as well as nationwide economic indexes as factors of consumption to produce a more comprehensive understanding of individual consumption behavior, and to better understand the nature of that relationship when it comes specifically to American alcohol. In this paper, I argue that price and economic uncertainty are not important factors to determining the probability of a person purchasing alcohol.

LITERATURE

Previous studies have been done on the manner that individuals consume alcohol. One study in particular informs the bulk of my model: this study was done by taking consumption and

demographic information from the Bureau of Labor Statistics (BLS), particularly from their Census Expenditure Survey (CE) data, with a two step model. Paulin's study (Paulin 2003) primarily shows factors that influence the probability of an individual purchasing alcohol in general in the first step, and then delves into the probability of individuals purchasing specific types of alcohol, and whether they were purchasing away or from home in the second step (Paulin 2003). I will be focusing on Paulin's first step model. Most importantly, he discovered variables that were statistically significant for predicting the probability of alcohol consumption are age, income, sex, and marital status. Patterns were discovered such that people tend to drink more as their age increases up to around age 55, men drink more than women, people with higher incomes tend to drink more and spend more money on alcohol, and single people also drink more than married people. Interestingly, Paulin's study included race as a variable, but this was not found statistically significant (Paulin 2003), contrary to other literature that indicates trends on race and alcohol consumption (Harris et al. 2021) that lead to the inclusion of race in my model. Blacks score higher in frequency and severity in alcohol consumption than Whites (NIH 2021), and higher frequency of high risk drinking has been correlated to Hispanic and Native American populations (Chartier & Caetano 2010). Economic demand theory posits that the price of a good impacts consumption, informing my model and leading to the inclusion of price as a variable, while Keynesian economics posits that the state of the economy has a large impact on consumer behavior, notably that people tend to spend less when economic conditions are more uncertain. Some information contradicts this Keynesian thought, providing evidence that a fair amount of people use alcohol as a coping mechanism for economic hardship, and these people are more likely to develop alcohol addictions, contributing to long-term higher consumption (Mackinnon 2014). However, both sources support the concept that alcohol

consumption is impacted by the population's perception of the state of the economy, even if they are uncertain of the relationships' particular nature. As such, I've included an uncertainty index for the general American economy.

My paper aims to contribute to the existing knowledge on alcohol consumption in four main ways. First of all, I would also like to contribute new discoveries to conflicting areas in the literature, such as the contradictions such as whether race is an important factor that predicts alcohol consumption, whether alcohol consumption tends to decrease or increase with income, and whether people react to downturns of the economy by consuming more or less alcohol. Secondly, by primarily basing my model on Paulin's 2003 study and using the same data source (BLS), I will add to the range of the regression period. While Paulin's study only utilized data in the year 2000, my regression and data will cover a far larger range from 2013-2023 with the objective of obtaining statistics that are more relevant to general American alcohol consumer trends, rather than those relating to a specific year, and allows the inclusion of the changing status of price and economic indexes over time. Thirdly, I hope to contribute by combining national economic indexes with individual demographic and statistics. Existing studies focus on individual statistics, explore questions such as: what characteristics of individuals impact if they purchase alcohol? Furthermore, economic theories and research have been conducted on the general nature of consumption, and general trends in the Alcohol market. The aim of this study is to create a novel model incorporating the interaction between these two different types of existing and studied theories on consumption; to analyze consumption from both the perspective of an individual's consumption patterns and national macro patterns of consumers.

METHODS

The equation below is the model that my regression will be based upon:

$$Buy_i=\beta_0 + \beta_1 Age_i+\beta_2 Sex_i+\beta_3 Income_i+\beta_4 Married_i +\beta_5 Price_i +\beta_6 Taste_i+\beta_7 EPU_i+\beta_8 Black_i+\beta_9 Asian_i +\beta_{10} PI_i+\beta_{11} NA_i+\beta_{12} Multi_i+\beta_{13} COVID_i+\varepsilon_i$$

In the model above, I will be using logistic regression as referenced in Paulin's study (Paulin 2003) which follows an exponential pattern. The dependent variable, Buy_i, is the log odds that an individual consumer unit will purchase alcohol. Age_i is the year in age of an individual and appears as a decimal in the regression, and Sex_i is going to be an indicator variable for biological sex of the individual (0 for male, 1 for female). $Income_i$ is the amount of salary of the consumer unit before taxes or other family grouping deductions. $Married_i$ is the indicator variable for marital status (0 for single, 1 for married). $Price_i$ accounts for the general price of alcohol by using the consumer price index(CPI). CPI takes on a value by evaluating the average urban prices of alcohol in major cities comparatively to the average prices from 1982-1984 as the baseline. $Taste_i$ is meant to account for the preference of the individual for alcohol and is represented by an indicator variable accounting for the amount in dollars that a household spends on alcohol (0 for below and 1 for above $30). EPU_i stands for the Economic Policy Uncertainty Index that measures how uncertain people perceive the economy to be in the United States, in a particular year, based on newspaper coverage of policy, disagreement among economic forecasters, and the number of federal tax codes meant to expire, since a higher number of codes expiring means a higher uncertainty of future policy. EPU ranges from around 0 to 1000. For the variable representing race, there will be 5 dummy variables to represent 6 categories of race: White, which is the omitted variable, Black as $Black_i$, Asian as $Asian_i$, Pacific Islander as PI_i, Native American as NA, and Multiracial as $Multi_i$. Although the literature includes that Hispanic

populations tend to have higher rates of alcohol dependency (Chartier & Caetano 2010), this was not a group that was indicated in the BLS CE data. (The BLS also includes a 7th category for "other" in the race category, however no entries exist with this value.) The last variable included is $COVID_i$ which is a dummy variable that represents COVID's timeline (0 for before COVID, 1 for after COVID), using March 2020 as the breakpoint. The subscript i represents the ith individual as crosssectional data, each data entry capturing the responses of different consumer units at various points in time. The data is also in quarterly form, and CPI and EPU, which were measured by month, were aggregated into quarterly data.

The data that I used was collected from multiple databases and sources. First of all data was found through the BLS from their annual Consumer Expenditure Surveys, particularly using the section that is from 2013 to 2023 that details the demographic (race, sex, income, age) and expenditure information (including alcohol spending) for many individual households by quarter: PUMD Data Files : U.S. Bureau of Labor Statistics (bls.gov). Lastly, the information on the CPI and EPU, respectively, is from the St. Louis Fed's website: Consumer Price Index for All Urban Consumers: Alcoholic Beverages in U.S. City Average (CUSR0000SAF116) | FRED | St. Louis Fed (stlouisfed.org) and EPU - Economic Data Series | FRED | St. Louis Fed.

The data went through processing to be usable in the model. I have randomly sampled 0.5% of the aggregated data to improve efficiency running the models due to computational limitations. Altogether, my model uses 8,495 data entries. In addition, there were missing data points, most notably, for census questions that individuals left blank. Any individuals with blank values for any of the independent variables in the model have been removed accordingly.

Empirical Results

Model Variable	Coeffecient	Standard error	Pvalue	Significance
intercept	-2.3	2	0.125	
Age	0.0054	0.0395	0.4455	
Income	0.0000004	0.0000008	0.309	
Price (CPI)	-0.00136	0.0081	0.434	
EPU	-0.0003	0.00106	0.39	
Sex	-0.16	0.011	0.075	**
Married	0.0074	0.11	0.25	
Black(race)	-0.37	0.19	0.004	*
Native American(race)	-0.13	0.032		
Asian(race)	-0.76	0.29		
Pacific Islander(race)	-0.96	1.02		
Multiracial	-0.34	0.42		
Taste	0.3	0.04522		
COVID	-0.089	0.27	0.369	

Figure 1 shows all the coefficients from the logistic regression model, their standard error, as well as their pvalue and significance from a one tailed test. One asterisk "*" show the coefficient is statistically significant at the 5% significance level. Two asterisks "**" show the coefficient is statistically significant at the 10% significance level.

After running the logistic regression, the two variables CPI and EPU were found not statistically significant, with p-values 0.43 and 0.39 respectively, concluding that both the price of alcohol and uncertainty in the economy are not important predictors of whether an individual purchases alcohol or not. This conclusion sparks intrigue as it contradicts the majority of the well documented economic theories regarding consumption. Perhaps this speaks to the unique status of alcohol as a good such that is not subject to general consumption trends or theories. These results may also be due to income's role in alcohol expenditue. Although not statistically significant in this model, income was statistically significant for Paulin's regression (Paulin

2003). It may be that because people with a higher income consume more alcohol than those with lower incomes, economic uncertainty and rising prices impact high income populations less and therefore have little to no effect on their spending. This possible explanation may also illuminate the lack of effect of COVID-19 on alcohol expenditure as well, since economic conditions disproportionately impacted people of color and low income populations, rather than those of higher income. Lastly, this could also be due to dependency in alcohol consumption. It is important to note that my model does not distinguish between alcoholism and the regular purchasing of alcohol: perhaps a lack of effect from price and economic uncertainty partially reflects society's alcohol dependency issue, as alcohol dependency causes alcohol to be essential to the lives of individuals who, in turn, purchase alcohol regardless of the cost.

The two variables stood out as statistically significant based on this model: race and sex. Sex is statistically significant on the 10% significance level and the positive sign of the coefficient reflects earlier findings (Paulin 2003), namely, that women tend to drink less than men. The interpretation of the coefficient is if the individual is female, then they are 14% less likely to consume alcohol (comparatively to being male). Race is made up of many indicator variables that were tested together using the Likelihood Ratio Test (found at the bottom of the appendix), and was found to be statistically significant at the 1% level of significance (0.004 pvalue), which stands contrary to Paulin, whose race value largely returned as statistically insignificant. Furthermore, the interpretations of the various indicator variable coefficients for race are unexpected: if an individual is Black, they are 30% less likely to purchase alcohol compared to a white individual, if an individual is Native American, they are 13% less likely to purchase alcohol compared to a white individual, if an individual is Asian, they are 53% less likely to purchase alcohol compared to a white individual, if an individual is Pacific Islander,

they are 62% less likely to purchase alcohol compared to a white individual. These interpretations counter existing literature that indicate that ethnic minorities tend to have higher rates of alcohol issues and dependency (Chartier & Caetano 2010). However, my model does not disgtinguish between alcohlism and regular consumption. As such, it is possible that, while ethnic minorities tend to have higher rates of alcohol *dependency,* overall, white individuals still tend to consume alcohol more for regular consumption purposes.

Another unexpected result is the coefficient for $Married_i$, which while statistically insignificant, is in the wrong direction, which also challenges existing literature that married couples are less likely to consume alcohol than single individuals (Paulin 2003). This means it is "notable" that $Married_i$ in this model is positive, meaning that if an individual is married, they are 0.74% more likely to consume alcohol. In addition, a couple of coefficients including $Married_i$ that were measured as statistically significant in Paulin's study do not seem statistically significant in this regression such as variables for income and age. Perhaps much has changed within the last two decades, culturally, socially and economically that has impacted how individuals perceive alcohol at different incomes, ages or marrital statuses.

The correlation coefficient is a standard method used for noting pairwise collinearity issues. However, since the majority of this data is categorical, correlation coefficients may not indicate much about categorical data and VIF has been run below to the left:

Variables	VIF
age	1.1
income	1.1
CPI	3.3
EPU	1.6
Sex	1.1
Married	1.1
Black	1.1
Native American	2
Asian	1.1
PI	1
Multi	1
Taste	2
COVID	4.2

Figure 2 shows all the VIF coefficients for all independent variables in the regression. All have been rounded to the nearest tenths place

Most of the VIF coefficients are relatively low. Age, income, EPU, Sex, Married, and all the race variables except Native American all have coefficients approximately 1 which cannot rule out multicollinearity, although it illustrates that the multicollinearity is low or moderate. The

highest multicolinnearity is found in CPI (3.3) and COVID (4.2). However, these results are expected since CPI is a national index for prices and COVID-19 is an event that has expected influence over the economy. As such both these variables will remain in the model.

Confusion Matrix

Prediction	Reference: No buy	Reference: Buy
No buy	1182	0
Buy	81	428

Figure 3 shows a confusion matrix, detailing how many predictions the model correctly identified

To evaluate the accuracy of this logistic regression model, a confusion matrix was run on 1,691 entries of data that were not used to train the model. The accuracy of this regression is 95%, which is incredibly high. It is also important to note that this accuracy score may be inflated by one of my predictor variables "taste," which accounts for how much a person has spent on alcohol. Clearly, if an individual has spent over $30 on their alcohol purchase, they must have had to have purchased alcohol. This accuracy also gives evidence that race and sex are relevant predictors of alcohol consumption.

CONCLUSION

Altogether, CPI and EPU are discovered to be irrelevant to the model, showing up as statistically insignificant in the regression. This indicates that CPI, and therefore the price of alcohol, as well as EPU as a measure of uncertainty in the state of the economy are not relevant predictors of whether or not a consumer unit will purchase alcohol. These are unexpected results, countering essential and extensively studied theories on economic consumption, namely, demand theory and Keynesian Economics.

Directions for future studies include expanding on the limitation of this model. First of all, some coefficients that were relevant in Paulin's study in 2003 were discovered to be statistically insignificant in this regression such as marital status and income, and age. This has a possibility of suggesting an omitted variable bias, however this could also be attributed to the different time frame the novel model is based on. More research needs to be done on isolating the variables' impact and exploring the possibilities and reasons behind this shift, especially in marital status which takes on an opposite sign than expected.

Other future studies include exploring alternative measures of economic propseperity and public opinion. This model utilized EPU, however many other variables were researched beforehand that may also signify economic prosperity, including GDP growth and unemployment rate. Lastly, more studying should be done on the distinction between alcoholism and regular purchasing of alcohol in order to fully utilize this information to better serve the public, reduce alcoholism and promote healthy alcohol customs.

Appendix

Note:variable names have changed are named differently in Rstudio. The name equivalents will be listed: Value = Age, FSALARY = Income, USEPUINDXD = EPU, dummysex = Sex, dummymar = Married, MEMBRACE1 = White, MEMBRACE2 = Black, MEMBRACE3 =Native American, MEMBRACE4 = Asian, MEMBRACE5 = Pacific Islander, MEMBRACE6 (implicit) = Multiracial/other

Logistic Regression (double sided pvalue test) (figure 1)

```
> LOGISTIC = glm(data = train, dummyALC~VALUE+FSALARYM+cpi+USEPUINDXD+dummysex+dummymar+MEMBRAC
E2+MEMBRACE3+MEMBRACE4+MEMBRACE5+MEMBRACE6+ dummybuy+dater, family = "binomial")
> summary(LOGISTIC)

Call:
glm(formula = dummyALC ~ VALUE + FSALARYM + cpi + USEPUINDXD +
    dummysex + dummymar + MEMBRACE2 + MEMBRACE3 + MEMBRACE4 +
    MEMBRACE5 + MEMBRACE6 + dummybuy + dater, family = "binomial",
    data = train)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.295e+00  1.996e+00  -1.150  0.25024
VALUE        5.440e-03  3.950e-02   0.138  0.89046
FSALARYM     4.017e-07  8.029e-07   0.500  0.61684
cpi         -1.359e-03  8.151e-03  -0.167  0.86761
USEPUINDXD  -3.025e-04  1.060e-03  -0.285  0.77537
dummysex    -1.593e-01  1.093e-01  -1.457  0.14519
dummymar     7.397e-02  1.131e-01   0.654  0.51296
MEMBRACE2   -3.719e-01  1.920e-01  -1.937  0.05275 .
MEMBRACE3   -1.331e+01  3.197e+02  -0.042  0.96680
MEMBRACE4   -7.632e-01  2.895e-01  -2.636  0.00838 **
MEMBRACE5   -9.617e-01  1.016e+00  -0.947  0.34380
MEMBRACE6   -3.358e-01  4.226e-01  -0.794  0.42692
dummybuy     3.085e+01  4.522e+02   0.068  0.94562
dater       -8.934e-02  2.667e-01  -0.335  0.73768
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10348.8  on 8494  degrees of freedom
Residual deviance:  2793.4  on 8481  degrees of freedom
AIC: 2821.4
```

VIF Test (figure 2)

```
> vif(LOGISTIC)
     VALUE   FSALARYM        cpi USEPUINDXD   dummysex   dummymar  MEMBRACE2  MEMBRACE3
  1.053315   1.144574   3.335050   1.633056   1.033715   1.103079   1.027036   1.995713
 MEMBRACE4  MEMBRACE5  MEMBRACE6   dummybuy      dater
  1.016671   1.002393   1.004499   1.995713   4.281472
```

Confusion Matrix (figure 3)

```
> confusionMatrix(test$dummyALC, predicted)
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1182    0
         1   81  428

               Accuracy : 0.9521
                 95% CI : (0.9408, 0.9618)
    No Information Rate : 0.7469
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8808

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9359
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 0.8409
             Prevalence : 0.7469
         Detection Rate : 0.6990
   Detection Prevalence : 0.6990
      Balanced Accuracy : 0.9679

       'Positive' Class : 0
```

Likelihood Ratio Test

```
Likelihood ratio test

Model 1: dummyALC ~ VALUE + FSALARYM + cpi + USEPUINDXD + dummysex + dummymar +
    MEMBRACE2 + MEMBRACE3 + MEMBRACE4 + MEMBRACE5 + MEMBRACE6 +
    dummybuy + dater
Model 2: dummyALC ~ VALUE + FSALARYM + cpi + USEPUINDXD + dummysex + dummymar +
    dummybuy + dater
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  14 -1396.7
2   9 -1405.2 -5 17.062   0.004384 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

References

American Beverage Licensees. (2018.)America's Beer, Wine & Spirits Retailers Create 2.03 Million Jobs & $122.63 Billion in Direct Economic Impact. *American Beer, Wine and Spirits Retailers.*[*America's Beer, Wine & Spirits Retailers Create 2.03 Million Jobs & $122.63 Billion in Direct Economic Impact – American Beverage Licensees*]

Chartier, K., & Caetano, R. (2010). Ethnicity and health disparities in alcohol research. *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism*, *33*(1-2), 152–160.

Economic Policy Uncertainty Index. (2012). Policyuncertainty.com. https://www.policyuncertainty.com/methodology.html

Harris, J. C., Mereish, E. H., Faulkner, M. L., Assari, S., Choi, K., Leggio, L., & Farokhnia, M. (2022). Racial Differences in the Association Between Alcohol Drinking and Cigarette Smoking: Preliminary Findings From an Alcohol Research Program. *Alcohol and alcoholism (Oxford, Oxfordshire)*, *57*(3), 330–339. https://doi.org/10.1093/alcalc/agab038

Mackinnon, S. P., Kehayes, I. L. L., Clark, R., Sherry, S. B., & Stewart, S. H. (2014, January 1) *Testing the four-factor model of personality vulnerability to alcohol misuse: A three-wave, one-year longitudinal study*. Penn State. https://pure.psu.edu/en/publications/testing-the-four-factor-model-of-personality-vulnerability-to-alc#:~:text=The%204-factor%20model%20of%20personality%20vulnerability%20identifies%204,with%20distinct%20mechanisms%20and%20motivations%20for%20alcohol%20misuse.

Maverick, J. B. (n.d.). *Which economic factors most affect the demand for consumer goods?*. Investopedia. https://www.investopedia.com/ask/answers/042815/which-economic-factors-most-affect-demand-consumer-goods.asp#:~:text=The%20demand%20for%20some%20consumer%20goods%20increases%20or,are%20high%2C%20the%20demand%20for%20consumer%20goods%20increases.

Paulin, G. (2003). *Consumer Expenditure on Alcohol 2000*. Consumer Expenditure Survey Anthology, 2003. https://www.bls.gov/cex/anthology/csxanth7.pdf