Does Drinking Truly Cut Down Individual Income?

Ke Jiang [1, *, a, †], Xingpei Li [2, *, b, †], Zening Zhao [3, *, c, †]

[1] *Nanjing Audit University, School of Crowe Chinese Auditing.*
[2] *China Agricultural University, College of Economics and Management.*
[3] *Nanjing Audit University, School of Crowe Chinese Auditing.*
[*]*Corresponding author. Email: [a]182021002@stu.nau.edu.cn [b]2018311320406@cau.edu.cn, [c]181091102@stu.nau.edu.cn*
[†]*These authors contributed equally.*

ABSTRACT
This paper examines whether drinking frequency is related to residents' income. Based on National Longitudinal Survey of Youth (NLSY) data, we find an inverted U-shaped relationship between them. Then, residents' drinking frequency in 1989 is utilized as an instrumental variable to modify the endogeneity, verifying the result that moderate drinking will increase residents' income and the best drinking frequency is 15.03 times per month. We explain the conclusion from the mechanism of improving health status and increasing social intercourse opportunities. On this basis, a gender heterogeneity analysis is carried out on the impact of drinking frequency on income. According to the results, the income promotion effect and optimal drinking frequency of moderate drinking in men are significantly higher than those in women. Moreover, the income of women in childbearing period is mainly affected by marriage and children, while drinking behaviour is not closely related to it. These results shed light for the formation of residents' healthy drinking habits and provided substantial evidence for measuring the economic effect on alcohol consumption.

Keywords: *Alcohol consumption, Drinking frequency, Income, Gender difference.*

1. INTRODUCTION

1.1. Research Backgrounds And Gaps

Ever since ancient times, alcohol has been an important component of our table. For such a special drink, people have quite different opinions on it. Some believe that it is a good way to loosen up and promote friendship, while some others state that it is addictive and causes delay in business. Therefore, we want to investigate the impact of drinking behaviour on residents' income. Former researches have paid much attention to the harm of excessive drinking. Johansson *et al*. show that people addicted to alcohol have lower employment probabilities [1]. Sloan and Grossman also find that heavy drinking in early adulthood is negatively associated with earnings because of its poor working productivity [2]. Böckerman *et al.* considered that alcohol addicted person leads to increased absenteeism from work in the short run and the health problems in the long run [3]. These researches come to a conclusion that excessive drinking have a negative impact on labour market outcomes.

However, some researchers expound the benefits of moderate drinking according to medical theories. For example, Timothy *et al.* come to a conclusion that moderate drinkers have less possibility of CVD (Cardiovascular Disease) mortality than non-drinkers. These results are consistent when controlled most of the variables such as demographic factors, social factors, behavioural factors. [4]. Poli, *et al* conducted a dozens of experiments and cast light on moderate drinking, which may have a significant reduction to the risk of certain diseases [5]. Cao and Lei indicate that alcohol consumption needs to be moderate in biological interpretation and drinking at a frequency of more than 3 days are healthy [6]. It seems like moderate drinking is more healthy than non-drinking and heavy drinkers.

To correctly estimate the relationship between drinking behaviour and income, it is of great significance to select the appropriate variables and models. According to Francesco, when estimating the relationship between alcoholism and labour market outcomes, different models are adopted to run a regression have different results [7]. In addition,

previous scholars use samples from different countries to study residents' drinking behaviour. Yin finds that drinking has a significant positive impact on income in China [8]. Balsa and French, Jones and Richmond find a positive correlation between heavy drinking and labour force employment, which can be attributed to the greater risk preference and income effect in developing country by analysing data from Uruguay [9,10]. Moreover, Peters and Stringham specifically differentiate between non-social drinking and social drinking. Their research denotes that social drinking is positively related to high earnings [11]. In Australia, Barret believed that moderate alcohol users have higher wages than abstainers and heavy drinkers [12]. However, as a typical representative of developed immigration country, studies on drinking behaviour of American residents are relatively few.

1.2. Hypotheses

According to previous studies, we can preliminarily summarize several possibilities of the relationship between drinking consumption and income. The hypotheses are as follows:

H1: Residents' income has a negative correlation with alcohol consumption.

H2: Residents' income has a positive correlation with alcohol consumption.

H3: Residents' income has a peak when people have moderate alcohol consumption.

We will focus on the inner relations between alcohol consumption and labour market outcomes, then try to give out an appropriate explanation. In order to estimate the relationship more accurately, this paper decides to innovate the previous research from the following aspects: firstly, the samples. We decided to focus on the American youth group to enrich the research in this field. Moreover, teenagers are more likely damaged their health than adults by drinking too much. Thus, casting an eye on younger people's drinking behaviour has important implications. Secondly, the variables, especially the key explanatory variables. We do not choose dummy variables or categorical variables to measure residents' drinking behaviour like previous researches. Actually, the continuous variable indicator of drinking frequency is applied as the core explanatory variable, and more details will be illustrated in the data description part. Thirdly, as for the models, we not only use Ordinary Least Square (OLS) and fixed effect model to do the regression analysis, but also analyse the possible reasons leading to the conclusion. Furthermore, we discuss the endogeneity problem and gender heterogeneity problem.

2. DATA AND MODELS

2.1. Data description

Our data set comes from the National Longitudinal Survey of Youth (NLSY) and includes information on labour market outcomes, alcohol consumption, and assorted demographics for individuals in each of 2 years -1989 and 1994 which includes series of sub-indicators. The data are restricted to young adults who are between the ages of 24 and 32 in 1989 (and 29-37 in 1994). Every individual has a particular identifier (range 1-6105).

We measure outcomes of individuals in labour market by the indicator of monthly wage and salary incomes. We use the logarithmic form of average monthly income **(*Income*)** in calendar year (in dollars) as our explained variable. Compared with wages, it more comprehensively measures the real income of residents.

As for the selection of the key explanatory variable, we find that most of the research used dummy or categorical variables to measure alcohol consumption. The choice of such variables is ambiguous for its range of standard. In order to accurately measure individual's alcohol consumption, we decided to choose a continuous variable. Specifically, ***Days*** means the number of days in the last month the individual has had at least 1 drink, reflecting the drinking frequency. On a further note, it is believed that the income effect is mostly related to the days a man drink rather than the amount he consumes. What we need is a comparatively long-term indicator to measure and run a regression of the long-term dependent variable.

As illustrated above, firstly, we consider the individual variances and social demographical, e.g., sex, race, people's health condition and their marital and education status. Some of the variables will be the standard we choose to conduct heterogeneity analysis. After comprehensive discussion, we use ***sex, race, higrad, health*** and ***marst*** to control demographic variables, as listed in Table 1. Furthermore, we assume that individuals' original family factor can make a big difference as well, especially to his/her current income and living condition. In this case, we specifically include ***wdad14***, ***dadwork, dadhgc*** to see whether this individual lived with his/her dad, and his/her dad's education level and inauguration.

The statistic description is also shown in Table 1. In 1994, the average monthly salary of the surveyed group is about $1736, and the number of monthly drinking varies from 0 to 31, with an average of 5.36. The gender ratio of surveyed group is relatively average, more than 92% of respondents are in good health, and the average length of education is about 13 years.

Table 1 the Definition and Description of Variables

Name	Definition	Mean	Std
Dependent var			
Income	ln(income/12+1)	1761.135	1696.194
Key Explanatory var			
Days	Days last month the individual has had >1 drink	5.361	7.222
days2	The square term of days	28.740	52.157
Control by individual			
Race	1=hispanic;2=black;3=others	2.351	0.768
Sex	1=male 2=female	1.511	0.450
higrad	Years of education	13.117	2.413
health	1=the individual has a health problem, else 0	0.079	0.268
numkid	The number of children	1.286	1.268
marst	1=never married; 2=married with a spouse present;3=other	1.936	0.677
Control by original family			
wdad14	1= lived with their father when 14, else 0	0.717	0.451
dadwork	1=the individual's father worked when 14,else 0	9.535	5.285
dadhgc	Years of education one's father had	0.757	0.429

2.2. Models

As discussed above, we build a cross-sectional model to better describe the relationships between labour markets outcomes and alcohol consumption. Because the data set only has two years data, cross-sectional model for each year are more sustainable.

$$Income_i = \alpha_1 days_i + \beta_1 X_i + \beta_2 Y_i + u_i \quad (1)$$

where ***i*** represents different individuals. X_i means the controllable variables which reflect differences within individuals, Y_i means the controllable variables which reflect differences within original families. β_1 and β_2 serve as each coefficient of these two group of controllable variables. u_i is the error term. $Income_i$ means monthly wage of each individual. α_1 is the coefficient of $days_i$.

$$Income_i = \alpha_1 days_i + \alpha_2 days2_i + \beta_1 X_i + \beta_2 Y_i + u_i \quad (2)$$

To prove whether the **H3** is true or not, we create a quadratic term of days ($days2_i$) and adding it at the end of the regression in every process (Actually, it strengthens the credibility of our models.), α_2 is the coefficient of $days2_i$.

To better analyse the time impact on drinking and how it differs between two years. A fixed-effect model is also constructed to further strength our regression results. Moreover, the fixed-effect model will help us eliminate the influence of unobservable factors.

$$Income_{it} = \alpha_1 days_{it} + \alpha_2 days2_{it} + \beta_1 X_{it} + \beta_2 Y_{it} + u_{it} \quad (3)$$

Above is Equation (3), only adding ***t*** to differ from different years of each individuals.

3. RESULTS AND DISCUSSION

3.1. Regression Results

What we have in dataset is 1989 and 1994, two-year data. We conduct cross-sectional regression to testify our hypothesis. Initially, we estimate the impact of drinking days on wage in year 94:

Using ***income*** as the dependent variable, Column 1 gives a linear estimation between ***income***, ***days***. We control other demographic and family variables, e.g., the individual's ***sex***, ***race***, ***higrad*** and ***numkid***. Apart from individual's sex and race, the education background can contribute to individual's income a lot. The number of kids is controlled because we believe that the more kids an individual has, the more time he/she has to allcoate to caring instead of working, which would definitely lead to less income. Excluding these factors out of the model will overestimate the impact of drinking days on income. Meanwhile, we also control variables from one's original family. Considering that an individual's family of origin would greatly affect his/her mind, his value, and decision in future career. We adopt ***wdad14*** and ***dadhgc*** to control these factors since in most family the dad will always play a role of children's life model and influence children greatly. With all these variables controlled, the ***days*** and ***income*** show insignificant and positive relationship. The point estimation indicates that those people who drink 1 day more per month will increase their monthly wage by 0.007 percent. The estimation is biased since the positive relationship between alcohol consumption frequency and income is against people's consensus. Other coefficients are worth noting as well. The condition of an individual's disease is negatively and significantly related to his/her wage,

and more years of education a person has received, the higher wage will be. In the meantime, a married person is likely to earn more than unmarried people, and a Black person tends to have less month wage than Hispanic or White. The regression also indicates that averagely people who have one more kid to raise will receive 0.3 per cent wage lower than other people. The control variables of original family proved that an individual with a father's accompany before age 14 is likely to enjoy a higher salary than those who don't. Besides, the education years his/her father received seems to have a positive though insignificant effect on an individual's wage. All the estimation of control variables seem to be reasonable, but the estimation of key variables is not accepted. Therefore, we can reject the **H1** that the quality of work has a negative relationship with alcohol consumption, and **H2** which predicts a positive relationship between alcohol consumption and quality of work. In summary, the correlation between work quality and alcohol consumption is not linear.

In the column 2 we modify our model to non-linear structure. After adding ***day2***(the square of days) into our model, the estimation of ***days*** and ***days2*** became relatively significant, with R-square enhanced as well. Meanwhile, the coefficients of other control variables had little change. Thus, we can reach the conclusion that the relationship between alcohol consumption frequency and monthly wage is non-linear. Light to moderate drinking frequency can increase an individual's monthly wage, while if drinking too many days each month, the individual's monthly wage would decline undoubtedly. When an individual drinks about 11.8 days per month, then his/her monthly wage would reach the peak. Hence, Column 2 significantly supports the **H3**, that the quality of work will increase to a peak with moderate alcohol consumption.

Column 3 further test our **H3** with panel data. We use 2 years panel data (89 and 94) to test the non-linear relationship between alcohol consumption frequency and wage. After the Hausman test, we adopt fixed effect model to regress this non-linear model. The results shows that if the individual drinks 10.75 days per month, then his/her wage would reach the peak. Because the fixed-effect model cannot identify the people whose drinking pattern stay unchanged. So, the result is only made for reference purpose only.

In conclusion, we reject **H1, H2** and accept **H3**. We believe that moderate alcohol consumption is beneficial to improve the quality of work, and averagely drinking 11.8 days per month can receive the greatest enhancement to one's work quality.

Table 2 Cross-sectional & Panel regression coefficients

Var		OLS		Fixed effect
		(1)	(2)	
days		0.007	0.071***	0.043***
days2			-0.003***	-0.002***
sex		-1.199***	-1.151***	-
higrad		0.249***	0.244***	0.129**
numkid		-0.299***	-0.298***	-0.065*
marst:	2	0.926***	0.936***	0.344***
	3	0.564***	0.561***	0.430***
race:	2	-0.524***	-0.509 ***	-
	3	-0.137	-0.140	-
wdad14		0.269***	0.265***	-
dadhgc		0.015**	0.013*	-
cons		4.333***	4.208***	4.078***
R^2		0.163	0.1664	0.0949
N		6105	6105	12210

(P-value,0.1=*,0.05=**,0.01=***)

3.2. Instrumental Variables and Endogenous Test

According to the above regression results, we derive a preliminary conclusion: there is an inverted U-shaped relationship between drinking frequency and residents' income, which rises first and then falls. Nevertheless, we can't ignore the endogenous problem between them. The possible sources of endogenous drinking are as follows: First, it is caused by omitted variables. Yin believes that people with poor judgment and self-control are more likely to have the habit of drinking, and this ability is often unobservable [8]. In addition, the survey data of NLSY did not display the types of drinking of residents. Generally, we think that the frequency of drinking beer is higher than that of rare wines (e.g., wine), and those who drink rare wines are more likely to have higher income. Second, there may be a reciprocal causation error between drinking frequency and residents' income. Besides, based on the effect of drinking frequency on income that we observe, drinking consumption as a normal commodity may also change with the increase of income. In addition, high-income groups may have more social intercourses, which increases their drinking frequency.

In this case, we choose the drinking frequency of residents in 1989 as a instrumental variable for two-stage estimation. We believe that residents' drinking habits do not change easily. Due to the time interval (5 years), the drinking frequency in 1989 will not be significantly related to other variables and random error terms in the current period. In other words, the drinking frequency of residents in 1989 will significantly affect the drinking frequency in 1994, but not directly affect the income of residents in 1994, i.e., it is suitable as a instrumental variable. Since the original equation contains the square term of drinking frequency, we also

introduce the square term of drinking frequency in 1989 as the corresponding instrumental variable.

Subsequently, we perform descriptive statistics and tests on instrumental variables, as shown in Table 3. In 1989, the monthly average number of drinking consumption of residents was 5.19, slightly lower than the 5.36 in 1994, and the standard error is 6.98. Because the model has heteroscedasticity, Durbin-Wu-Hausman endogenous test is carried out, and the test value is 6.41, which rejects the hypothesis that the drinking frequency in 1989 is an exogenous variable. If the minimum eigenvalue statistic of instrumental variable test is 202.34, which is much higher than the critical value of 7.03 under 10% error, it shows that the model does not have weak instrumental variable problem. Additionally, the number of instrumental variables is equal to the number of endogenous variables, i.e., there is no over-identification problem. The above test results show that it is necessary and appropriate to use the above instrumental variables to estimate the relationship between drinking frequency and income. According to the 2SLS estimation results, the monomial and quadratic coefficients of drinking frequency are significant below the 1% level, indicating that there is still a significant inverted U-shaped relationship between drinking frequency and residents' income. The best drinking frequency is 15.03 times per month, which is higher than OLS estimation of 13.89 times per month.

Table 3 test of instrumental variables

Var	2sls	OLS
days	0.173***	0.071***
days2	-0.006***	-0.003***
DW-test	6.4069***	
Weak instrument test	202.342	
10% bias test	7.03	
N	6105	6105

(P-value,0.1=*,0.05=**,0.01=***)

3.3. Influencing Mechanism Analysis

Based on the unbiased estimation of the model, we have reached the conclusion that moderate drinking will increase residents' income. Next, this paper will analyse the specific mechanism of how drinking frequency affects residents' income. We believe that moderate drinking may affect income by improving health status and increasing social intercourse opportunities. In order to verify the above conjecture, this paper introduces the health indicators of residents into the regression equation, and compares the regression results with the original equation. The results are shown in Table 4. After introducing health indicators, the influencing coefficient of drinking frequency on income decreases from 0.071 to 0.054. Taking into account the interference of the square term of drinking frequency in the equation, this paper performs a separate regression on drinking frequency and health status and finds that the two are significant at the 1% confidence level, indicating that the income effect of moderate drinking is partly caused by improving health status. Increasing social and entertainment opportunities is also a possible influencing mechanism. Yin used the classification variable of drinking frequency to measure the social function of drinking, and reached the conclusion that moderate drinking can increase income through social communication channels [8]. In this paper, the frequency of drinking as the core explanatory variable shows great significance, which also supports the above conclusion.

Table 4 Influencing Mechanism Analysis

Var	OLS		2sls	
	(1)	(2)	(1)	(2)
days	0.071***	0.053***	0.173***	0.148***
days2	-0.003***	-	-	-
		0.002***	0.006***	0.005***
health	-	-	-	-2.159***
		2.222***		
Control	Y	Y	Y	Y
R^2	0.166	0.209	0.158	0.202
N	6105	6105	6105	6105

(P-value,0.1=*,0.05=**,0.01=***)

4. GENDER HETEROGENEITY ANALYSIS

According to the affecting mechanism of drinking frequency on residents' income, this paper further considers the gender difference of this effect. We generally believe that men are the main consumers of alcohol, while women often need to take on greater reproductive and family responsibilities to support their spouses to engage in paid work, and they are less involved in business relations, wine parties and other activities, i.e., they drink less frequently than men. Based on the differences of gender roles in labour participation, we speculate that there may be gender heterogeneity in the influence of drinking frequency on income.

The results of the heterogeneity analysis are summarized in Table 5. In 1989, the impact of drinking frequency on the income of both men and women showed a significant inverted U-shaped relationship. The best drinking frequency of men (13.5 times per month) was higher than that of women (11.3 times per month). In 1994, the difference of the best drinking frequency between men and women was more obvious (15.75 times per month for men and 8.25 times per month for women). Nevertheless, there was no statistical significance between women's drinking frequency and income in 1994.

Table 5 Gender Heterogeneity analysis

Var	Year 89		Year 94	
	male	female	Male	female
days	0.060***	0.077***	0.082***	0.049
days2	-0.002***	-0.004***	-0.003***	-0.003**
higrad	0.134***	-0.261***	0.194***	-0.271***
marst 2	0.638***	-0.370***	1.022***	-0.576***
3	0.105	0.471***	0.458***	0.616***
race 2	-0.437***	-0.022	-0.501***	-0.444***
3	0.163	0.070	0.178	-0.346**
numkid	0.058	-0.779***	-0.032	-0.599***
wdad14	0.260***	0.109	-0.256**	0.224*
dadhgc	0.007	-0.009	0.008	-0.012
cons	4.115***	2.217***	3.162***	2.378***
N	6105	6105	0.082***	6105

(P-value,0.1=*,0.05=,0.01=***)**

For the same interviewed group, why didn't the significance between women's drinking frequency and income in 1989 remain to 1994? We find that the core explanatory variable of drinking frequency is correlated with the control variables of "***numkid***" and "***marst***". After excluding these two variables, the relationship between drinking frequency and income becomes significant. On this basis, the explanation given in this paper is: the surveyed group were all aged between 24 and 32, just at the childbearing age. From 1989 to 1994, the average number of children raised by this group increased from 1.19 to 1.56, which means that 37% of women gave birth to children within five years. Women's drinking behaviour is affected by reproduction and no longer has a significant correlation with income.

In order to verify the above inference, this paper takes the changes of ***numkid*** as the standard, and divides the female samples in 1994 into two categories for heterogeneity analysis, as given in Table 6. The results show that the frequency of drinking still has a significant impact on the income of the non-reproductive group in this period, while there is no significant correlation between the drinking frequency and income among the women group who gave birth to children in this period. This supports our inference, that is, the drinking frequency of non-reproductive women has a significant impact on income (inverted U-shape), while the income of reproductive women is mainly related to marriage, children and other indicators, and the drinking behaviour is weak in explaining it. For other control variables, the impact of educational years and native families on income does not show obvious gender differences.

Table 6 Changes in numkid

Var	94 women OIS	
	unchanged	changed
Days	0.072***	0.010
days2	-0.003***	-0.001
Higrad	0.370***	0.298***
race 2	-0.348*	-1.026***
3	-0.018	-0.500*
wdad14	0.112	0.422*
Dadhgc	0.049***	-0.008
Cons	0.471***	1.345***
N	1916	1206

(P-value,0.1=*,0.05=**,0.01=***)

5. CONCLUSIONS

In summary, we use NLSY data to investigate the impact of drinking frequency on residents' income, and concludes that there is an inverted U-shaped relationship between drinking frequency and residents' monthly income, i.e., moderate drinking will increase residents' income. In this paper, the drinking frequency of residents in 1989 is used as the instrumental variable of drinking frequency of residents in 1994, and the best drinking frequency is estimated by 2SLS: 15.03 times per month. From the perspective of influencing mechanism, moderate drinking can improve the income level by improving health status and increasing social intercourse opportunities.

Considering the differences of gender roles in labour participation, this paper analyses the gender heterogeneity. According to the results, there is a significant inverted U-shaped relationship between the drinking frequency and income of men and non-reproductive women, and the best drinking frequency of men is higher than that of women. The income level of women during the childbearing period is closely related to indicators, e.g., marriage and children, and the drinking behaviour is not the main factor affecting their income. **These results offer a guideline for the**

development of alcohol industry and the formation of residents' healthy drinking habits.

REFERENCES

[1] Edvard Johansson, et al. "The association of alcohol dependency with employment probability: evidence from the population survey ‘Health 2000 in Finland’." Health Economics 16.7(2007): doi:10.1002/hec.1201.

[2] Sloan Frank A and Grossman Daniel S. "Alcohol consumption in early adulthood and schooling completed and labor market outcomes at midlife by race and gender" American journal of public health 101.11(2011): doi:10.2105/AJPH.2010.194159.

[3] Petri Böckerman, et al."Alcohol Consumption and Long - Term Labor Market Outcomes." Health Economics 26.3(2017): doi:10.1002/hec.3290.

[4] Timothy S. Naimi, et al. "Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. adults." American Journal of Preventive Medicine 28.4(2005): doi:10.1016/j.amepre. 2005.01.011.

[5] Alessio Lomuscio and Edoardo Pirovano. "A Counter Abstraction Technique for the Verification of Probabilistic Swarm Systems".Autonomous Agents and MultiAgent Systems.Ed. , 2019,

[6] CAO Ruihong and LEI Zhenhe. " On Drinking and Health." LIQUOR-MAKING SCIENCE & TECHNOLOGY .02(2019):135-142. doi:10.13746/j.njkj.2018153. (In Chinese)

[7] Francesco Renna."Alcohol Abuse, Alcoholism, and Labor Market Outcomes: Looking for the Missing Link." Industrial & Labor Relations Review 62.1(2008): doi:10.1177/001979390806200105.

[8] Yin Zhichao and Gan Li." Smoking, Drinking , and Earnings in China." ECONOMIC RESEARCH JOURNALS 45.10(2010):90-100+160. doi:CNKI:SUN:JJYJ.0.2010-10-008. (In Chinese)

[9] Ana I. Balsa and Michael T. French."Alcohol use and the labor market in Uruguay." Health Economics 19.7(2010): doi:10.1002/hec.1520.

[10] Alison Snow Jones and David W. Richmond. "Causal effects of alcoholism on earnings: estimates from the NLSY." Health Economics 15.8(2006): doi:10.1002/hec.1109.

[11] Bethany L. Peters and Edward Stringham. "No booze? You may lose: Why drinkers earn more money than nondrinkers." Journal of Labor Research 27.3(2006): doi:10.1007/s12122-006-1031-y.

[12] Garry F. Barrett. "The Effect of Alcohol Consumption on Earnings[I wish to]." Economic Record 78.240(2002): doi:10.1111/1475-4932.00041.