2020

Partial Adjustment of Self-Reported Alcohol Consumption in the National Health and Nutrition Examination Survey: A Zero Inefficiency Stochastic Frontier Analysis

Isabella A. Yerby
University of Central Florida

PARTIAL ADJUSTMENT OF SELF-REPORTED ALCOHOL CONSUMPTION IN THE NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY: A ZERO INEFFICIENCY STOCHASTIC FRONTIER ANALYSIS

by

ISABELLA A. YERBY

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Economics
in the College of Business Administration
and in the Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term
2020

Thesis Chair: Richard Hofler, Ph.D.

ABSTRACT

Alcohol consumption is likely underreported in the National Health and Nutrition Examination Survey. The problem, common among most self-reported data, stems from social desirability bias. As a result of this bias, the true level of alcohol consumption, specifically heavy episodic drinking, or binge drinking, is undercounted. By applying zero inefficiency stochastic frontier analysis, we adjust for the inefficiency caused by survey participants underreporting their alcohol consumption. This paper serves to partially correct the estimates of heavy episodic drinking and to serve as an example that stochastic frontier analysis can be used outside of its standard application to correct underreported and self-reported data. The study concludes that among the sample of 2,901 NHANES participants, 27.51% underreported their alcohol consumption. It further concludes that among those who did originally admit to binge drinking, 69.51% underreported the true extent of their binge drinking episodes. However, of the people who did not report binge drinking, none of them underreported. While more research should be done to determine why the model stated that none of the nondrinkers underreported, this paper demonstrates a possible use of the zero inefficiency stochastic frontier model in regards to adjusting self-reported data.

ACKNOWLEDGEMENTS

I would like to thank my amazing committee for their constant support in this process.

To my Thesis Chair and mentor, Dr. Richard Hofler, for his never-ending support and assistance, not only in creating this work, but throughout my undergraduate career. It has been a blessing to work with you for these two years and I will carry with me the many lessons you have taught me for the rest of my life.

To Dr. Melanie Guldi, for her patience, guidance, and strategic insight on all matters. Thank you for being a role model of the academic I strive to be one day.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

The National Health and Nutrition Examination Survey (NHANES) is a widely cited study conducted by the Centers for Disease Control and Prevention (CDC). The information obtained from the study is used nationwide for tasks ranging from extensive medical studies to developing policy. The nature of NHANES is incredibly advantageous to researchers because it contains both self-reported and medical examination data. However, the literature suggests that NHANES questionnaire data likely contains inaccurate self-reported information. This has negative consequences for all applications of this valuable dataset. Alcohol consumption, for example, is likely underreported as a result of both the self-reported nature of NHANES and the social stigma attached to excessive alcohol consumption, such as binge drinking. Alcohol-related policy based on underreported data will likely be sub-optimal or ineffective due to a misunderstanding of the severity of alcohol consumption. Research on alcohol consumption that utilizes NHANES will suffer from the same problems, with results being inaccurate as a result of the data being a poor representation of the true level of alcohol consumption. This paper serves to demonstrate that stochastic frontier analysis can be used to partially correct for the inaccuracies of self-reported data. In order to do so, this paper utilizes zero inefficiency stochastic frontier analysis to estimate an adjusted level of alcohol consumption in the United States.

CHAPTER TWO: LITERATURE REVIEW

Alcohol Consumption Trends, Definitions, and Consequences

The level of alcohol consumption among adults in the United States has been slowly increasing in recent years. According to a meta-analysis of six national surveys from 2000 to 2016 conducted by Grucza et al. (2018), past-year alcohol consumption was estimated to be increasing 0.3% per year. This increase was statistically significant for women but not men, with Blacks having the largest increase among racial groups, and the 50 years old and older age group having the largest increase among age brackets (Grucza et al., 2018).

Heavy episodic drinking, more commonly known as binge drinking, has faced even larger increases than alcohol consumption as a whole. According to the same meta-analysis, there has been an annual increase of approximately 0.70% (Grucza et al., 2018). The trends are nearly identical to overall alcohol consumption trends for the groups affected. Once again, the increase has been highest for women, Blacks and Hispanics, and 50 years old and older (Grucza et al., 2018). There was also increased prevalence among those age 30 to 49 years old and those with postsecondary education (Grucza et al., 2018).

Binge drinking is commonly defined as the consumption of five or more drinks in a row for men and four or more drinks in a row for women (Wechsler & Austin, 1998). The gender distinction between five and four drinks was popularized by Harvard School of Public Health College Alcohol Study as a result of recognizing physiological differences that cause women to absorb alcohol more rapidly than men (Wechsler & Austin, 1998). More specifically, binge drinking is defined by the National Institute for Alcohol Abuse and Alcoholism (NIAAA) as a session of drinking alcohol that brings the blood alcohol concentration to 0.08-gram percent or above (NIAAA, 2004). However, the four/five measurement is typically an adequate estimation of this standard (NIAAA, 2004).

These recent increases are incredibly concerning, as excessive alcohol intake is responsible for a large number of deaths and medical repercussions each year. For example, from 2006 to 2010, excessive alcohol consumption in the United States was responsible for an average 88,000 deaths annually (Stahre et al., 2014). In 2010, excessive alcohol consumption resulted in $249 billion in costs, while binge drinking specifically accounted for approximately $191.1 billion (Sacks et al., 2015). Excessive alcohol use is also linked to numerous negative medical consequences. More specifically, alcohol use is associated with health conditions such as HIV, pneumonia, epilepsy, heart disease (Rehm, 2011), liver diseases such as alcoholic liver cirrhosis (Rehm, 2010), and numerous other serious medical conditions. Alcohol consumption can lead to cancers of the liver, breast, esophagus, larynx, etc. (Rehm, 2011). There is, of course, an increase in alcohol use disorders and binging as a result of higher rates of alcohol consumption (Rehm, 2011). Frequent alcohol users are more likely to have poor mental health, such as depression (Collins, 2016). Binge drinking and excessive alcohol use can also lead to increases in risky behaviors such as unprotected sexual encounters (George et al., 2009), injuries, violence, and even fatalities.

Correlates of Alcohol Consumption and Binge Drinking

Numerous demographic, societal, and economic factors can increase the risk of excessive alcohol consumption and binge drinking. The following are commonly accepted correlates with increased levels of alcohol consumption.

Race and Ethnicity

Typically, Whites are more likely to drink than Blacks. This is true at most stages of life. For example, white college students are more likely to engage in binge drinking than black college students (Wade & Peralta, 2017). Despite this trend, older Blacks that do drink are more likely to be in a high-risk drinking group than Whites (Sacco et al., 2009). Overall, evidence suggests that being white increases the prevalence of alcohol consumption (Moore et al., 2005).

Age

Older individuals tend to drink more often, although younger people tend to drink more at one time (Roche et al., 2015b). As a result of this, despite the fact that older people have a tendency to drink daily or weekly, it is youth that are at risk for alcohol-related accidents and injuries associated with binge drinking (Roche et al., 2015a). However, binge drinking among older adults has been increasing in recent years which may alter these risk statistics. Overall, when referring to binge drinking and high levels of alcohol consumption, there is higher prevalence among those in early adulthood compared to those later in life, as levels of alcohol consumption decline with age (Kuntsche et al., 2017; Moore et al., 2005).

Gender

Typically, women drink less alcohol than men, thus binge drinking prevalence is higher among men than women (CDC, 2012). This difference between genders has been decreasing over time as a result of an increase in alcohol consumption among women, but currently, women continue to binge drink less than men (Kuntsche et al., 2017).

Education

Higher levels of education are associated with higher prevalence of binge drinking. However, among those who do binge drink, those with lower education tend to drink at the highest quantity and frequency (CDC, 2012; Kanny et al. 2018). This relationship may vary depending on age, with young adults with higher education participating in more binge drinking episodes, while middle-aged adults with higher education participate in fewer heavy drinking episodes (Lui et al., 2018).

Employment

Results on the effect of employment, or more specifically unemployment, are inconclusive (Backhans et al., 2012; Bryden et al., 2013). While some papers state that employed individuals drink more than those who are unemployed (Roche et al., 2015b), others indicate that the relationship is the opposite (Bolton & Rodriguez, 2009).

Income

Prevalence of binge drinking increases with household income. Similar to education levels, those with low income have the lowest levels of binge drinking, but of those who do binge drink, those with low incomes drink at higher quantities and frequencies (CDC, 2012; Kanny et al. 2018).

Marital Status

Individuals who have never been married or are divorced are more likely to consume excessive amounts of alcohol (Roche et al., 2015b).

CHAPTER THREE: RESEARCH DESIGN

<u>Data</u>

All data used in the following model have been collected from the 2015-2016 National Health and Nutrition Examination Survey (NHANES).

Overview of NHANES

NHANES is a program of studies collected by the National Center for Health Statistics, which is part of the Centers for Disease Control and Prevention. It is composed of two sections: the interview portion and the examination component. The interview section asks participants about their demographic, socioeconomic, dietary, and health status. The examination component includes medical and physiological measurements and laboratory tests. The sample is created to represent the U.S. population at all ages, and over-samples people 60 years of age and older, Blacks, and Hispanics. Everyone included is required to visit a physician for an examination, with tests becoming more extensive for older participants. The survey is designed to be simple and accessible, with the survey being conducted at the respondent's home, transportation provided to the medical examination if required, and compensation given for participation. [1]

Overview of Variables

The dependent variable of our model will be alcohol consumption, defined as the number of days the respondent had (4/5) or more drinks of any alcoholic beverage during the last 12 months. The model will include the following independent variables: race, age, gender, education, employment, income, and marital status. These are all known correlates with alcohol consumption, as supported above in the literature review. For the sake of completeness,

[1] All information regarding NHANES has been referenced from cdc.gov

employment status will be included in the estimated regression model despite the lack of consensus as to its true relation to alcohol consumption. Additionally, age will be restricted to observations that are 21 years old or older. The variables MARRIED and EMPLOYED are derived from the NHANES variable as defined below but will be converted into binary variables that are married or not and employed or not, respectively. Table 1 contains a description of these variables.

Table 1. NHANES Variables and Survey Questions

Variable Name	NHANES Variable Name	Questionnaire Survey Question/Variable Description
RACE	ridreth1	Reported race and Hispanic origin
AGE	ridageyr	Age in years at screening
GENDER	riagendr	Gender of the participant
EDUCATION	dmdeduc2	"What is the highest grade or level of school you have completed or the highest degree you have received?"
EMPLOYED	ocd150	"Which of the following [type of work] were you doing last week?"[2]
INCOME	indhhin2	Total annual household income
MARRIED	dmdmartl	Marital status
HED	alq141q	"In the past 12 months, on how many days did you have (4/5) or more drinks of any alcoholic beverage?"[3]

Flaws of Self-Reported Data

Self-reported data is subject to self-report bias and misreporting. This kind of bias can result from fear of social pressures, recall errors, or the sampling approach. In the case of alcohol consumption, there is likely to be underreporting as a result of a social desirability bias, the tendency to report socially desirable behaviors and not to report socially undesirable ones (Althubiati, 2016; Chung & Monroe, 2003). NHANES is not immune to this bias. As shown

[2] While this is the chosen estimator for employment status in NHANES, last week's employment may not be a good measure of the likelihood of being employed during the year. It may also not be a good measure of employment's relationship with alcohol consumption as defined by the HED variable, given that HED is recorded over a year and EMPLOYED is recorded over a week.

[3] Binge drinking is typically defined over an interval of two weeks to one month, not 12 months.

above, in order to collect data on alcohol use, a surveyor asks participants "In the past 12 months, on how many days did you have (4/5) or more drinks of any alcoholic beverage?" In order to obtain correct responses, participants must tell the truth. However, it is likely that multiple participants do not tell the truth when self-reporting and thus the overall level of alcohol consumption in NHANES is underreported. Underreporting of self-reported data can also be the result of participants misunderstanding survey questions, recall bias, or other various biases (Althubiati, 2016).

Methodology

Stochastic Frontier Analysis – Production Case

Stochastic Frontier Analysis (SFA) is an econometric method developed by Dennis Aigner, C.A. Knox Lovell, and Peter Schmidt (1977). The basis of SFA is the simple production function, a standard in economics that determines the maximum obtainable output given fixed inputs. The SFA model allows for the estimation of an unobserved frontier outcome (also called a latent outcome). This model allows for both the production case (where the observed outcome is less than the frontier outcome) and the cost case (where the observed outcome is greater than the frontier outcome).

The existence of probable underreporting of alcohol consumption in the NHANES data (where the observed self-reported outcome can be less than the true, unobserved frontier outcome) calls for the stochastic frontier production case. An SFA model differs from a typical regression in the nature of its error term. The regression error term ε_i is broken into two parts: u_i and v_i, a non-negative error term and a two-sided, symmetric error term, respectively. The disturbance term u_i accounts for the fact that a firm's output is either on or below its frontier as a result of technical inefficiency (in the production case), while v_i is a random noise term. The stochastic frontier model for the production case can be written as

$$y_i = \boldsymbol{x_i'\beta} + v_\mathrm{i} - u_\mathrm{i}, \text{ for } i = 1, \ldots, \mathrm{n}, \tag{1}$$

where y_i is a scalar output, x_i is a k x 1 vector of covariates, β is a k x 1 vector of parameters, $v_i \sim i.i.d\ N(0, \sigma_v^2)$, and $u_i \sim i.i.d\ N^+(0, \sigma_u^2)$, which is a half-normal distribution, a normal mean-zero distribution truncated from below at zero.

In equation (1), y_i is the observed outcome and $\boldsymbol{x_i'\beta} + v_\mathrm{i}$ is the frontier or latent value. Adding u_i to y_i gives the unobserved latent value. This fact is the backbone of stochastic frontier analysis. Once these latent values have been estimated by estimating u_i, a large variety of analyses can be performed on the adjusted data.

A rather intuitive economic example would be a firm and its production function, which is determined by the uncontrollable happenstance impacting it, its production technology, and its own efficiency. This firm will either be producing on or below its production possibility frontier. That is, if it is producing below its production frontier, there is some inefficiency causing it to do so. In terms of stochastic frontier analysis, exogenous factors impacting the firm are contained within v. The term u_i is a measure of the distance between the observed level of output and the higher frontier level of output. In other words, u_i is essentially a measure of inefficiency, hence, if there is no inefficiency, $u_\mathrm{i} = 0$. By adding u_i to y_i, the observed output, it is possible to estimate the unknown frontier output level of the firm.

Zero Inefficiency Stochastic Frontier Analysis

For the purposes needed in this paper, the assumption that all firms are inefficient (i.e., all respondents underreport their alcohol consumption) is not adequate. It is possible that some firms will be efficient (i.e., some respondents truthfully report their alcohol consumption), thus $u_\mathrm{i} = 0$.

In order to model these two behaviors in one sample, it is necessary to utilize the zero inefficiency stochastic frontier model (ZISF) developed by Kumbhakar, Parmeter and Tsionas (2013). This allows for some firms (or observations) to be inefficient (some respondents underreport) and for others to be fully efficient (other respondents truthfully report). The derivations and equations that follow are summarized from Kumbhakar et al. (2013).

As previously explained, a standard stochastic frontier production model is written as

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_{\mathrm{i}} - u_{\mathrm{i}}, \text{ for } i = 1, \ldots, n,$$

where $v_i \sim i.i.d\ N(0, \sigma_v^2)$ and $u_i \sim i.i.d\ N^+(0, \sigma_u^2)$, which is a half-normal distribution. Unlike the original model, assume that some observations are fully efficient, so $u_{\mathrm{i}} = 0$. It is not possible to know which observations are efficient and which are not, thus it is necessary to develop a new model to determine this. From this need, the ZISF model can be written as

$$\begin{aligned} y_i &= \boldsymbol{x}_i'\boldsymbol{\beta} + v_{\mathrm{i}} \text{ with probability } p \text{ and} \\ y_i &= \boldsymbol{x}_i'\boldsymbol{\beta} + v_{\mathrm{i}} - u_{\mathrm{i}} \text{ with probability } (1 - p), \end{aligned} \tag{2}$$

where p is the probability of an observation being fully efficient ($u_{\mathrm{i}} = 0$) and $(1 - p)$ being the probability of an observation being inefficient ($u_{\mathrm{i}} > 0$).

The density function of the convoluted error term is

$$f(\varepsilon_i) = p\ \left[\frac{1}{\sigma_v}\phi\left(\frac{\epsilon}{\sigma_v}\right)\right] + (1-p)\left[\frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right], \tag{3}$$

where $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, $\lambda = \frac{\sigma_u}{\sigma_v}$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and the cumulative density function of a standard normal random variable, respectively.

Using the density function of the error term as shown in equation (3) one forms the log likelihood function by the usual method. After obtaining the log likelihood function, parameters β, σ_u^2, σ_v^2, σ, and λ can be estimated.

Thus, with the necessary parameters estimated, the next goal is to estimate observation-specific inefficiency. The estimation of observation-specific inefficiency in ZISF utilizes a conditional mean function by Jondrow, Lovell, Materov, and Schmidt (JLMS) (1982). The modified JLMS estimator of u_i allows for p accounting for full efficiency. The modified version of the JLMS estimator for ZISF is given as,

$$E[u|\varepsilon] = (1-p)\frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2}\left[\sigma_0\frac{\phi\left(\frac{\varepsilon}{\sigma_0}\right)}{\Phi\left(\frac{-\varepsilon}{\sigma_0}\right)} - \varepsilon\right]. \quad (4)$$

When p = 0, (4) becomes the JLMS estimator of inefficiency. Next calculate the posterior estimate of the probability of an observation being fully efficient,

$$\breve{p}_i = \frac{\left(\frac{\hat{p}}{\hat{\sigma}_v}\right)\phi\left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}_v}\right)}{\left[\frac{\hat{p}}{\sigma_v}\phi\left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}_v}\right)\right]+(1-\hat{p})\left[\frac{2}{\hat{\sigma}}\phi\left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}}\right)\Phi\left(\frac{-\hat{\varepsilon}_i}{\hat{\sigma}_0}\right)\right]}. \quad (5)$$

With $\breve{p}_i$ and $\hat{u}_i$ estimated, it is now possible to develop the posterior estimate of each observation's inefficiency, defined by

$$\breve{u}_i = (1-\breve{p}_i)\hat{u}_i, \quad (6)$$

where $\breve{p}_i$ is the posterior estimate of the probability of being fully efficient as calculated in equation (5) and $\hat{u}_i$ is the JLMS estimator of inefficiency as calculated in equation (4), when $p = 0$. As explained with the base stochastic frontier model, once $\breve{u}_i$ is obtained, the latent outcome can be calculated by adding $\breve{u}_i$ to the observed y_i.

Applications to Alcohol Consumption

Unlike many components of NHANES, such as height and weight, the accuracy of self-reported alcohol consumption cannot be compared to medical examination data. Thus, the zero-inefficiency stochastic frontier model can be used to adjust for self-reporting bias, in that there is an underreporting of alcohol consumption by some, not all, respondents in NHANES. This is analogous to a production case. Instead of working with firms, each observation will be one individual participant in the survey. Individuals have a choice to tell the truth or to lie. Telling the truth is considered “efficiency,” whereas lying is considered “inefficiency.” The NHANES data involves some individuals who are honest about their alcohol consumption and others who are not. The standard SFA should not be used in this case because not every participant is going to underreport their alcohol consumption (i.e., not every observation is inefficient, which is what the standard SFA assumes.) Thus, the possibility of honesty, or efficiency, must be accounted for. By using zero inefficiency stochastic frontier analysis, it is possible to obtain estimates of $\breve{u}_i$ to partially adjust underreported observed values to be closer to their corresponding latent (truthful) values. When $\breve{p}_i = 1$ for a respondent, that person is telling the truth about their alcohol consumption and their “inefficiency” estimate in equation (6) equals zero. Conversely, when a person underreports their alcohol consumption, $\breve{p}_i < 1$, their “inefficiency” estimate in equation (6) is greater than zero. As there is no benchmark on which to test the accuracy of these estimations, all corrections will be a partial adjustment, not an absolute fix. The goal is to minimize the gap between the observed and the unknown latent outcome of alcohol consumption to allow for more accurate use of the NHANES data.

CHAPTER FOUR: RESULTS

In order to find the model most compatible with our sample, we began by using Stata to create all 128 possible combinations of the independent variables that might belong in the best model. Then, these models were ranked by pseudo-R^2 and by AIC and BIC scores (i.e., the best model will have one of the highest pseudo-R^2 values and the lowest AIC and BIC scores). By these criteria, the best model contains the independent variables AGE, MARRIED, INCOME, and GENDER. Thus, the frontier model we estimated is:

$$HED_i = \beta_0 + \beta_1 AGE_i + \beta_2 MARRIED_i + \beta_3 INCOME_i + \beta_4 GENDER_i + v_i - u_i,$$

where the variables are as defined in Table 1 and the subscript *i* denotes an individual observation. The estimation results are in Table 2 below. The dependent variable, HED, is how many days the respondent had (4/5) or more drinks of any alcoholic beverage in the last 12 months. Overall, the model seemed reasonable as the signs on the coefficients are consistent with the literature: older age is associated with less binge drinking, being married decreases binge drinking, higher income leads to more binge drinking (although this variable is not significant at even the 0.10 level), and being female leads to less binge drinking.

Table 2. Results from Estimating Zero Inefficiency Stochastic Frontier Model

Variable	Estimates
AGE	-0.027* (0.0155)
MARRIED	-1.246** (0.543)
INCOME	0.012 (0.060)
GENDER	-2.203*** (0.510)
_Cons	7.262*** (2.165)

Standard errors in parentheses
***p<0.01, ** p<0.05, *p<0.1

The parameters of this model were estimated via maximum likelihood. These estimated parameters are used to determine the posterior probability of full efficiency, as defined in equation (5). The summary statistics for the estimated parameters and the posterior probability of full efficiency are listed below.

Table 3. Results for Selected Estimated Variables

variable	mean	sd	p5	p25	p50	p75	p95
ALSineff	0.988	0.000	0.988	0.988	0.988	0.988	0.988
postprob	0.704	0.437	0.000	0.000	0.993	0.994	0.994
u_i_v	0.293	0.432	0.006	0.006	0.007	0.988	0.988

The variable *ALSineff* is the estimate of inefficiency given by the standard SFA model as produced by Aigner, Lovell, and Schmidt (1977). The variable *postprob* is the posterior probability of full efficiency, $\breve{p}_\iota$, as defined by Kumbhakar, Parmeter, and Tsionas (2013) from the zero-inefficiency stochastic frontier model. The measure u_i_v is $\breve{u}_\iota$, the posterior estimate of each observation's inefficiency, as defined in equation (6).

If a participant did not report their true alcohol consumption, $\breve{p}_\iota < 1$ (*postprob* < 1). If they told the truth, then $\breve{p}_\iota = 1$. In order to indicate who most likely told the truth, we created a new binary variable titled *honest*. We assigned values of 0 and 1 to the new variable given the values of $\breve{p}_\iota$. That is, when $\breve{p}_\iota > 0.50$, then *honest* = 1, but when $\breve{p}_\iota \leq 0.50$, then *honest* = 0.[4] A value of 1 means the person told the truth, while a value of 0 means the person did not tell the truth. Variables following the same rules were created for those who claimed to have participated in a binge drinking episode as well as those who claimed to have never participated in a binge

[4] Additional cut off points of 0.70, 0.80, 0.90, 0.95 were tested as well. This did not change the result beyond approximately 10 observations being considered efficient or inefficient for *honestdrinker* and *honest*. There was no change in result for *honestnondrinker*.

drinking episode. These variables were named *honestdrinker* and *honestnondrinker,* respectively. The results are shown below.

Table 4. Results for *honest*

honest	Freq.	Percent	Cum.
0	798	27.51	27.51
1	2,103	72.49	100.00
Total	2,901	100.00	

Table 5. Results for *honestdrinker*

honestdrinker	Freq.	Percent	Cum.
0	798	69.51	69.51
1	350	30.49	100.00
Total	1,148	100.00	

Table 6. Results for *honestnondrinker*

Honestnondrinker	*Freq.*	*Percent*	*Cum.*
1	1,753	100.00	100.00
Total	1,753	100.00	

Of 2,901 observations, 798 received an *honest* value of 0, meaning 798 underreported their true level of alcohol consumption. Thus, in this sample, approximately 27.51% of participants in the NHANES survey underreported their alcohol consumption to some degree. Of the 1,148 who reported that they participated in at least one binge drinking episode per year, 69.51% underreported their total number of binge drinking episodes. However, as shown in Table 6, of the 1,753 participants that did not report binge drinking, none of them underreported. This is very likely to be false, especially given the results for those who reported drinking. This must be investigated further, to determine if it is a result unique to the sample, the NHANES data, or the zero-inefficiency stochastic frontier model.

CHAPTER FIVE: CONCLUSIONS

As demonstrated, zero inefficiency stochastic frontier analysis can be used to adjust for underreporting of self-reported data. As shown by alcohol consumption in NHANES, it is likely that a large portion of participants underreport their alcohol consumption, particularly concerning binge drinking. For the entire sample of 2,901 observations, 798 individuals, roughly 27.51%, underreported their binge drinking. Furthermore, in Table 5, of the 1,148 people who reported binge drinking, 798 underreported their number of binge drinking episodes. This is a 69.51% underreport. Table 5 demonstrates the possible severity of underreporting in self-reported alcohol data.

It is incredibly unlikely that all 1,753 people who claimed they never had a binge drinking episode, never binged. Thus, given the current results, it is possible that the estimate of dishonesty in Table 4 should likely be higher than it is. It is also possible that the 69.51% underreporting could be an inaccurate estimate, albeit still a good example of the use of the model. Perhaps more accurate results for the drinkers could be created by eliminating nondrinkers from the data and estimating the underreporting of only those who claimed to have already binge drank at least once. More research should be done to determine why the zero-inefficiency stochastic frontier model did not yield sensible results for nondrinkers.

In terms of public health policy, these results may imply that current policies based off NHANES are inadequate because the efforts are based on inaccurate estimates. However, the results for nondrinkers show a difference in frequency of binge drinking, not necessarily a difference in prevalence. These individuals are already considered binge drinkers, thus the more interesting observations would have been those who claimed to have never binged. Adjusted values for those nondrinkers could yield different levels of prevalence of binge drinking, which could have a significantly larger policy impact than the current results. This is further motivation for additional research because sensible results indicating the true level of underreporting of the prevalence of alcohol consumption could reveal a need for more intensive intervention policy addressing binge drinking.

This study is not without limitations. There is the possibility of overreporting in addition to the observed underreporting. However, in the case of alcohol consumption, this overreporting is likely to be significantly lower than the underreporting as a result of social desirability bias. Additionally, due to the nature of NHANES, it is possible that the rest of the data used to estimate the model in the study also fall victim to the flaws of self-reported data. However, the main goal of the study is to demonstrate an additional use of SFA beyond its standard application. While more research should be made in regard to this approach and the results for nondrinkers, there is evidence that SFA, and zero inefficiency stochastic frontier analysis in particular, can be used for a wide variety of uses beyond the standard applications.

REFERENCES

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and Estimation Of Stochastic Frontier Production Function Models. *Journal of Econometrics*, *6*(1), 21-37.

Althubaiti, A. (2016). Information bias in health research: definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, 9, 211.

Backhans, M. C., Lundin, A., & Hemmingsson, T. (2012). Binge drinking—a predictor for or a consequence of unemployment?. *Alcoholism: Clinical and Experimental Research,* 36(11), 1983-1990.

Bolton, K. L., & Rodriguez, E. (2009). Smoking, drinking and body weight after re-employment: does unemployment experience and compensation make a difference?. *BMC Public Health*, 9(1), 77.

Bryden, A., Roberts, B., Petticrew, M., & McKee, M. (2013). A systematic review of the influence of community level social factors on alcohol use. *Health & Place*, 21, 70-85.

Centers for Disease Control and Prevention (CDC. (2012). Vital signs: binge drinking prevalence, frequency, and intensity among adults-United States, 2010. *MMWR. Morbidity and mortality weekly report, 61(1),* 14.

Chung, J., & Monroe, G. S. (2003). Exploring social desirability bias. *Journal of Business Ethics*, 44(4), 291-302.

Collins, S. E. (2016). Associations between socioeconomic factors and alcohol outcomes. *Alcohol Research: Current Reviews,* 38(1), 83.

George, W. H., Davis, K. C., Norris, J., Heiman, J. R., Stoner, S. A., Schacht, R. L., ... & Kajumulo, K. F. (2009). Indirect effects of acute alcohol intoxication on sexual risktaking: The roles of subjective and physiological sexual arousal. *Archives of Sexual Behavior*, 38(4), 498-513.

Grucza, R. A., Sher, K. J., Kerr, W. C., Krauss, M. J., Lui, C. K., McDowell, Y. E., ... & Bierut, L. J. (2018). Trends in adult alcohol use and binge drinking in the early 21st-century United States: a meta-analysis of 6 National Survey Series. *Alcoholism: Clinical and Experimental Research*, 42(10), 1939-1950.

Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2-3), 233-238.

Kanny, D., Naimi, T. S., Liu, Y., Lu, H., & Brewer, R. D. (2018). Annual total binge drinks consumed by US aduselflts, 2015. *American Journal of Preventive Medicine*, 54(4), 486-496.

Kumbhakar, S. C., Parmeter, C. F., & Tsionas, E. G. (2013). A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172(1), 66-76.

Kuntsche, E., Kuntsche, S., Thrul, J., & Gmel, G. (2017). Binge drinking: Health impact, prevalence, correlates and interventions. *Psychology & Health*, 32(8), 976-1017.

Lui, C. K., Kerr, W. C., Mulia, N., & Ye, Y. (2018). Educational differences in alcohol consumption and heavy drinking: An age-period-cohort perspective. *Drug and Alcohol Dependence*, 186, 36-43.

Moore, A. A., Gould, R., Reuben, D. B., Greendale, G. A., Carter, M. K., Zhou, K., & Karlamangla, A. (2005). Longitudinal patterns and predictors of alcohol consumption in the United States. *American Journal of Public Health*, 95(3), 458-464.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). (2004). NIAAA Council

Approves Definition of Binge Drinking. *NIAAA Newsletter,* No. 3, Winter 2004

Rehm, J. (2011). The risks associated with alcohol use and alcoholism. *Alcohol Research & Health*, 34(2), 135.

Rehm, J., Taylor, B., Mohapatra, S., Irving, H., Baliunas, D., Patra, J., & Roerecke, M. (2010). Alcohol as a risk factor for liver cirrhosis: a systematic review and meta-analysis. *Drug and Alcohol Review*, 29(4), 437-445.

Roche, A., Kostadinov, V., Fischer, J., & Nicholas, R. (2015a). Evidence review: The social determinants of inequities in alcohol consumption and alcohol-related health outcomes. *Australian's National Research Centre on AOD Workforce Development and Flinders University*.

Roche, A., Kostadinov, V., Fischer, J., Nicholas, R., O'Rourke, K., Pidd, K., & Trifonoff, A. (2015b). Addressing inequities in alcohol consumption and related harms. *Health Promotion International*, 30(suppl_2), ii20-ii35.

Sacco, P., Bucholz, K. K., & Spitznagel, E. L. (2009). Alcohol use among older adults in the National Epidemiologic Survey on Alcohol and Related Conditions: A latent class analysis. *Journal of Studies on Alcohol and Drugs*, 70(6), 829-838.

Sacks, J. J., Gonzales, K. R., Bouchery, E. E., Tomedi, L. E., & Brewer, R. D. (2015). 2010 national and state costs of excessive alcohol consumption. *American journal of preventive medicine*, 49(5), e73-e79.

Stahre, M., Roeber, J., Kanny, D., Brewer, R. D., & Zhang, X. (2014). Peer reviewed: Contribution of excessive alcohol consumption to deaths and years of potential life lost in the United States. *Preventing chronic disease*, 11.

Wade, J., & Peralta, R. L. (2017). Perceived racial discrimination, heavy episodic drinking, and alcohol abstinence among African American and White college students. *Journal of Ethnicity in Substance Abuse*, 16(2), 165-180.

Wechsler, H., & Austin, S. B. (1998). Binge drinking: the five/four measure. *Journal of Studies on Alcohol*, 59(1), 122-124.