Alcohol Income and Health: a complicated but desirable mix

By
Gillian Ormond

PhD Dissertation
Department of Economics
University College Cork
February 2014

Professor Connell Fanning, Head of School of Economics
Dr. Rosemary Murphy, Supervisor of Research

Dedication

This thesis is dedicated to the memory of my beloved dad, Seán Ormond, who passed away three weeks after I completed my Viva-Voce Examination.

<u>CONTENTS</u>

CHAPTER 5: ESTIMATION OF THE EFFECT OF ALCOHOL CONSUMPTION AS ORDERED DATA ON HOUSEHOLD INCOME IN IRELAND COMPARING LIMITED & FULL INFORMATION METHODS OF ESTIMATION

CHAPTER 6: AN INVESTIGATION INTO THE EFFECT OF ALCOHOL CONSUMPTION ON HEALTH STATUS AND HEALTH CARE UTILISATION

List of Tables

List of Figures

Declaration

I can confirm that this thesis is my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed: ______________________ Date: _________________

Gillian Ormond

ACKNOWLEDGEMENTS

I wish to acknowledge the support of everyone who helped me in the completion of this thesis. In particular I would like to thank the following:

- Dr. Rosemary Murphy for her excellent supervision of this thesis and for her constant support, advice, motivation, and friendship.

- My colleagues in the Health Service Executive, particularly Mr. Dermot Halpin for his support in finishing this thesis.

- All the staff in the Economics Department in UCC for their support and advice.

- My family and friends, in particular my brothers John and Pat Ormond along with their wives, for their support, encouragement and enthusiasm.

- Above all, my parents, Seán and Moira Ormond, for their constant love, support and encouragement in everything I do.

Abstract

The aim of this thesis is to examine if a difference exists in income for different categories of drinkers in Ireland using the 2007 Slán data set. The possible impact of alcohol consumption on health status and health care utilisation is also examined. Potential endogeneity and selection bias is accounted for throughout.

Endogeneity is where an independent variable included in the model is determined within the context of the model (Chenhall and Moers, 2007). An endogenous relationship between income and alcohol and between health and alcohol is accounted for by the use of separate income equations and separate health status equations for each category of drinker similar to what was done in previous studies into the effects of alcohol on earnings (Hamilton and Hamilton, 1997; Barrett, 2002).

Sample selection bias arises when a sector selection is non-random due to individuals choosing a particular sector because of their personal characteristics (Heckman, 1979; Zhang, 2004). In relation to alcohol consumption, selection bias may arise as people may select into a particular drinker group due to the fact that they know that by doing so it will not have a negative effect on their income or health (Hamilton and Hamilton, 1997; Di Pietro and Pedace, 2008; Barrett, 2002).

Selection bias of alcohol consumption is accounted for by using the Multinomial Logit OLS Two Step Estimate as proposed by Lee (1982), which is an extension of the Heckman Probit OLS Two Step Estimate. Alcohol status as an ordered variable is examined and possible methods of estimation accounting for this ordinality while also accounting for selection bias are looked at. Limited Information Methods and Full Information Methods of estimation of simultaneous equations are assessed and compared.

Findings show that in Ireland moderate drinkers have a higher income compared with abstainers or heavy drinkers. Some studies such as Barrett (2002) argue that this is as a consequence of alcohol improving ones health, which in turn can influence ones productivity which may ultimately be reflected in earnings, due to the fact that

previous studies have found that moderate levels of alcohol consumption are beneficial towards ones health status.

This study goes on to examine the relationship between health status and alcohol consumption and whether the correlation between income and the consumption of alcohol is similar in terms of sign and magnitude to the correlation between health status and the consumption of alcohol. Results indicate that moderate drinkers have a higher income than non or heavy drinkers, with the weekly household income of moderate drinkers being €660.10, non drinkers being €546.75 and heavy drinkers being €449.99. Moderate Drinkers also report having a better health status than non drinkers and a slightly better health status than heavy drinkers. More non-drinkers report poor health than either moderate or heavy drinkers.

As part of the analysis into the effect of alcohol consumption on income and on health status, the relationship between other socio economic variables such as gender, age, education among others, with income, health and alcohol status is examined.

CHAPTER 1
INTRODUCTION

1.1: Study Design

The aim of this thesis is to examine if a difference exists in income for different categories of drinkers in Ireland using the 2007 Slán data set. The impact of alcohol status on health status and health care utilisation is also examined. Potential endogeneity and selection bias is accounted for throughout.

Endogeneity is where an independent variable included in the model is potentially a choice variable and is determined within the context of the model (Chenhall and Moers, 2007). In relation to the study of alcohol on income and alcohol on health status, alcohol consumption is governed in part by unobserved factors which may also be important determinants of the dependent variables income and health status, implying the possibility that the drinking status variables may be correlated with the error term of the conditional demand equation (French and Zarkin, 1998; Hamilton and Hamilton, 1997; Di Pietro & Pedace, 2008; Barrett, 2002).

The most frequently used method of dealing with the problem of endogeneity is through using Instrumental Variable (IV) estimation however the main difficulty with the use of IV regressions is finding a sufficient number of suitable instruments (Knowles and Owen, 1997; Barrow and Sala-i-Martin, 1999; Cho, 1996; Milbourne *et al*, 2003). Given this difficulty in the estimation of alcohol on income the literature advocates the use of separate income equations for each category of drinker (Hamilton and Hamilton, 1997; Barrett, 2002).The same technique is applied in the study of health status and alcohol consumption.

Sample selection bias arises when a sector selection is non-random due to individuals choosing a particular sector because of their personal characteristics (Heckman, 1979; Zhang, 2004). In relation to categorising individuals based on their levels of alcohol consumption, selection bias may arise as people may select into a particular drinker group due to the fact that they know that by doing so it will not have a negative effect on their income or health (Hamilton and Hamilton, 1997; Di Pietro and Pedace, 2008; Barrett, 2002). One way of accounting for potential selection bias is to use the standard two step estimation proposed by Heckman (1979) whereby a persons propensity to drink is estimated through probit analysis. This in turn allows predicted values for the inverse Mills Ratio to be generated which is then included as an additional variable in the income equation. Lee (1982) extends the Heckman Probit OLS Two Step Estimate to a Multinomial Logit OLS Two Step Estimate, to allow for selection correction for polychotomous choices. Given that alcohol status is grouped into three categories, the estimation is carried out using a multinomial logit two step estimation which accounts for the potential selection bias of alcohol consumption. Alcohol status is then estimated as an ordered probit to account for the ordinal nature of alcohol consumption and different methods of estimation are analysed.

As part of the analysis of the effect of alcohol consumption on income, the relationship between other socio economic variables with both household income and alcohol status is examined. Similar studies have been carried out for other countries and in general findings have been that the financial welfare of moderate drinkers is greater than abstainers or heavy drinkers (French and Zarkin, 1995; Heien, 1996; Hamilton and Hamilton, 1997; Barrett, 2002). Some studies such as Barrett (2002) argue that the relationship between alcohol and earnings is similar to that of alcohol and health and that moderate drinkers have better health status than either abstainers or heavy drinkers.

Barrett (2002) states that based on the Grossman model (1972), the potential impact of health status on earnings represents a straightforward extension of the human capital framework of earnings determination and that alcohol consumption can influence ones health status, the consequences of which can influence ones productivity at work which may ultimately be reflected in an individuals earnings (Barrett, 2002). Ziebarth and Grabka (2009) similarly found a positive association between moderate alcohol consumption and wages. They argue that this could be due to the fact that given the findings in relation to the positive effect of alcohol on health, the correlation between alcohol and income is as a result of moderate drinkers being healthier and more productive. Secondly Ziebarth and Grabka (2009) state that moderate drinkers are more productive because they have greater life satisfaction which is then reflected in ones productivity.

Findings have tended to be that moderate levels of alcohol consumption are beneficial towards ones health status, compared with abstaining from or consuming heavy amounts of alcohol, which has a negative effect on health status. This gives rise to a U shaped curve or a partial U shaped curve referred to as a J shaped curve, showing a reduced relative risk of given diseases and in general better health for moderate consumers of alcohol compared with abstainers or heavy drinkers (Berger *et al*, 1999; Klatsky *et al* 2001; Bau *et al*, 2007). This study goes on to examine if the correlation between consumption of alcohol and income in Ireland is similar to the correlation between consumption of alcohol and health status in terms of sign and magnitude.

Drinkers are categorised based on the recommended weekly drinking levels of the Irish Health Promotion Unit (HSE, 2008). The recommended weekly limits of alcohol consumption for women is up to 14 standard drinks and for men up to 21 standard drinks per week and on any one occasion drink no more than 4 standard drinks for women and 6 for men. In the Slán 2007 dataset there is a substantial amount of information available on the amounts of alcohol people consume which allows respondents to be categorised into one of three categories of drinkers; non drinkers, moderate drinkers or heavy drinkers.

This research has not previously been conducted using Irish data. It contributes to the literature in the area of what variables influence ones alcohol consumption, the effects of drinking alcohol on household income, health status and health care utilisation, in particular focusing on the theoretical and econometric issues encountered in doing so. Explicit distinction is made between endogeneity and selection bias and the respective methods for dealing with these problems. Unlike the majority of previous literature (Hamilton and Hamilton, 1997; Barrett, 2002), the ordered nature of the dependent variable alcohol is accounted for in this study.

1.2: Purpose of the Study

The purpose of this thesis is to identify an endogenous free relationship between income and alcohol consumption; between health status and alcohol consumption, and between heath care utilisation and alcohol consumption. The effects of alcohol consumption in Ireland on household income, heath status and healthcare utilisation are analysed, while accounting for the endogenous relationship between these variables. Endogeneity not being accounted for would allow a correlation between the unobservables and the disturbance term to exist, resulting in biased and inconsistent estimators, the consequence being that less confidence would result in the estimates (Chenhall and Moers, 2007).

To proceed with the aim of this thesis firstly drinkers need to be categorised, based on the recommended weekly drinking levels of the Irish Health Promotion Unit (HSE, 2008).

Moderate Drinkers are those who had a drink in the last month, or in the week prior to the survey any woman who had up to 14 standard drinks and any man who had up to 21 standard drinks.

Heavy Drinkers are women who drank more than 14 drinks and men who had more than 21 drinks, in the week prior to the survey.

Non-drinkers are defined as those who never drank or who did not have a drink in the month prior to the survey. Dummy variables are established for the three categories of drinkers.

To begin with, the effect of alcohol consumption on household income is assessed using a multinomial logit two step model. As part of this analysis the relationship between both income and alcohol and a range of socio economic variables is looked at. Given that alcohol consumption could be viewed as ordered data, the analysis is also carried out by estimating alcohol consumption as an ordered probit model. Previous studies such as Hamilton and Hamilton (1997) and Barrett (2002) among others, have not accounted for this, and if ordinality is ignored then this may lead to a loss of efficiency and an increased risk of getting insignificant results (Harris *et al*, 2006).

The above estimation into the effect of individual alcohol consumption on household income in Ireland while accounting for the potential endogenous relationship between alcohol and household income, is then carried out using the Full Information Methods as well as the Limited Information Methods of Estimation, whereby both equations are estimated simultaneously. The method specifically used in this study is the Full Information Maximum Likelihood Method. Full information methods estimate all the equations in the model simultaneously, taking due account of all restrictions on such equations by the omission or absence of some variables and is seen to be a more efficient estimator (Greene, 1999, Gujarati, 2004; Pearce, 1986; Intriligator *et al,* 1996; Puhani, 2000).

The study then goes on to investigate the effect of alcohol consumption on an individuals health status and health care utilisation in Ireland while again accounting for the potential endogenous relationship between health status and alcohol consumption. Many studies have found a similar relationship in terms of sign and magnitude, between health status and alcohol consumption, as with income and alcohol consumption. An analysis is carried out to see if findings are similar to those in terms of the effect of alcohol on household income, in as far as the category of drinkers who are found to have the highest household income having the highest return on their health investment as proposed by Grossman (1972). Given that the measure of health status available for this study is self-assessed health, health care utilisation is used as an alternative measure. Numerous studies have been carried out into the relationship between health status and the frequency of use of health services and both are found to be highly correlated (Lim *et al*, 2005; Rotermann, 2006; Finkelstein, 2001).

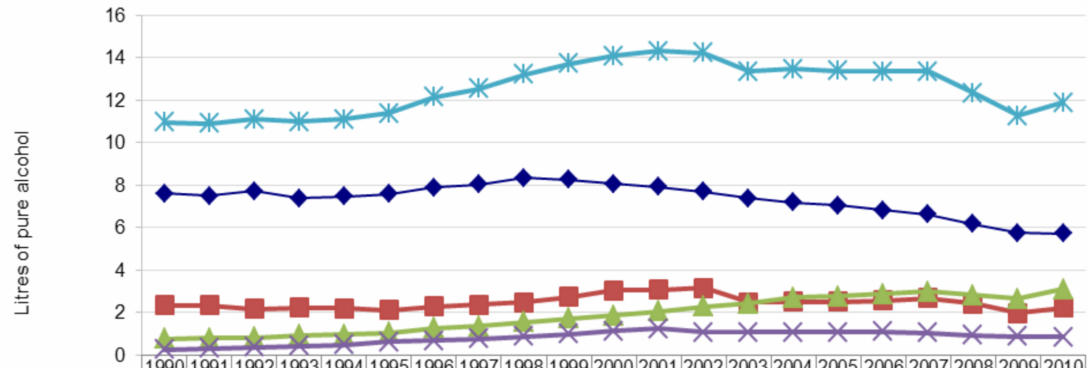
Results will provide an insight into the effects of alcohol consumption on income, health status and health care utilisation in Ireland. This information can then be used to further develop policies aimed at combating the problem of the misuse of alcohol consumption in Ireland.

1.3: Rationale

Alcohol consumption remains very high in Ireland and this places a huge cost on society. Currently what is primarily being used in Ireland to try and control excessive levels of alcohol consumption among individuals is a population based approach. The WHO (2007) describe a target based approach as an approach targeted at vulnerable populations as opposed to the population at large. The rationale of this thesis is to identify which is the most appropriate method to be used in combating excessive alcohol consumption; a target or population based approach.

Alcohol consumption in Ireland is relatively high and is placing a huge cost on society, on individuals and on businesses. The consumption of alcohol in Ireland, increased by 192% between 1960 and 2001, from an average of 4.9 litres per adult to 14.3 litres per adult. While there was a reduction in levels consumed between 2002 and 2009, the levels of alcohol consumed in 2010 is still 145% higher than the average amount of alcohol consumed per adult in 1960 (Ireland, 2012). Figure 1.3.1 depicts the levels of alcohol consumption per adult in Ireland between the years 1990 and 2010.

Figure 1.3.1 Levels of alcohol consumption in Ireland 1990-2010



	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Beer	7.6	7.5	7.7	7.4	7.5	7.6	7.9	8.0	8.3	8.3	8.0	7.9	7.7	7.4	7.2	7.0	6.8	6.6	6.2	5.8	5.7
Spirits	2.3	2.3	2.2	2.2	2.2	2.1	2.3	2.4	2.5	2.7	3.0	3.1	3.2	2.5	2.5	2.5	2.6	2.7	2.4	2.0	2.2
Wine	0.8	0.8	0.8	0.9	1.0	1.1	1.3	1.3	1.6	1.7	1.9	2.1	2.3	2.4	2.7	2.8	2.9	3.0	2.8	2.7	3.1
Cider	0.2	0.3	0.4	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.1	1.1	1.1	1.1	1.1	1.1	0.9	0.9	0.9
Total	11.0	10.9	11.1	11.0	11.1	11.4	12.1	12.5	13.2	13.7	14.1	14.3	14.2	13.3	13.5	13.4	13.4	13.4	12.4	11.3	11.9

(Source: Steering Group Report on a National Substance Misuse Strategy 2012)

According to an OECD report Ireland has the 10th highest consumption levels of alcohol per capita of 40 countries (Ireland, 2012). Ireland's per capita alcohol consumption was 11.3 litres per adult while the OECD average was 9.1 litres. Results are shown in Appendix A.

The cost of alcohol related problems in Ireland was estimated to be approximately €2.4 billion per year (Ireland, 2002) and this rose to €3.7bn in 2007, representing 1.9% of GNP that year (Ireland, 2002). Details of these costs are provided in Appendix A.

Over recent years there has been numerous policies developed and actions carried out which have helped control the levels of alcohol consumption in Ireland. In the past, most of the actions taken to address the high levels of alcohol consumption fall into one of three categories, price, availability and marketing which are the key factors on the supply of alcohol and impact on the volume and pattern of alcohol consumption (Department of Health, 2012). Ireland saw a decline in alcohol consumption seen for the first time in over 16 years in 2002 and 2003, which has been put down to an increase in excise duty on spirits (Joint committee on Arts, Sport, Tourism, Community, Rural and Gaelteacht Affairs, 2007). In 2001 and 2002 excise duty was introduced on cider and spirits respectively. This resulted in cider sales decreasing by 11.3% in 2002 while wine and spirits increased and beer remained relatively stable. In 2003, following the tax increase on spirits, sales of spirits decreased by 20% while wine sales increased by 8% and beer and cider showed only marginal changes. (Hope 2004). Drink driving has been a major policy area. Since the introduction of mandatory alcohol testing in 2006, there has been a 34% reduction in road accident deaths (Department of Health, 2012).

In February 2012, the Steering Group on National Substance Misuse Strategy (Ireland, 2012) made many recommendations in terms of policies and actions that can be developed to address the issue of misuse of alcohol. The majority of these recommendations are around the supply of alcohol such as increasing the price, introducing a social levy on drinks, increasing excise duty and further measures around drink driving and the phasing out of the sponsorship of sport and other large public events by the drinks industry (Ireland, 2012). Recommendations are also made around prevention and education programmes, treatment and rehabilitation programmes, and around research programmes to examine further the consequences of alcohol and the impact of alcohol policy measures.

The Alcohol Beverage Federation of Ireland (ABFI) are an umbrella organisation for the drinks industry manufacturers and suppliers in Ireland and work to promote and support an environment that encourages the sustainable and responsible development of the alcohol beverage manufacturing sector in Ireland. In a minority report by the ABFI (ABFI, 2012), they state that while they are an active and fully engaged member of the National Substance Misuse Strategy Group, they feel that the approach and recommendations of the National Substance Misuse Strategy Group to target problematic drinking, is primarily a population based approach which has reached its peak and they suggest using a target based approach as opposed to a population approach in order to achieve this.

The ABFI argue that alcohol is a legal licensed product and when consumed in a responsible manner can form part of a healthy, balanced lifestyle. While they agree with many of the recommendations of the steering group particularly in relation to prevention and education programmes and the enforcement of drink driving and age verification regulations, they have concerns in relation to the recommendations on pricing, the social responsibility levy, structural separation in retail outlets, restrictions on certain promotions, restrictions and bans on alcohol advertising, a ban on alcohol sponsorship and the introduction of new low risk weekly guidelines on alcohol consumption. These concerns are based primarily on the lack of available evidence to support many of these recommendations, the ineffectiveness of legislation to tackle misuse contrasted with the success of co-regulatory codes of practice that are already in place and are highly effective, and the devastating impact that many of the recommendations will have on jobs, businesses and livelihoods in the industry (ABFI, 2012).

The ABFI argues that the population based approach set out in the recommendations of the steering group, will not solve the misuse of alcohol consumption in Ireland, and while they accept that there is a problem of the misuse of alcohol in Irish society, there is not enough detailed reference to the positive effects of moderate consumption. Adams and White (2005) argue that where a J Shaped relationship between exposure and risk exists, population based approaches may not always be beneficial and may cause a negative effect to certain groups and this should not be overlooked. The World Health Organisation (WHO, 2007) suggests that there are advantages and disadvantages to both approaches, but that support for population approaches has declined in some countries.

Previous studies for other countries have found that moderate drinkers earn more and have a better health status than either heavy drinkers or abstainers. If similar results are found in this study using Irish data, results of this thesis would aid the ABFI and government in formulating initiatives in order to bring heavy levels of alcohol consumption under control using the target based approach which would be agreeable to all stakeholders.

This study divides drinkers into 3 categories; non, moderate and heavy drinkers based on national recommendations. Different socio economic factors which have an effect on drinking have been identified, and this data can be used as an aid in identifying further target groups. Most importantly this thesis has found that alcohol does have an effect on income and health status, with both income and health status of moderate drinkers being higher compared to other categories of drinkers, ceteris paribus.

The ABFI highlight the fact that previous studies have shown that there are benefits to moderate levels of alcohol consumption. They argue that if population based approaches are adopted the benefits associated with moderate alcohol consumption are not being recognised. The implementation of population based polices could result in individuals who are moderate consumers of alcohol, reducing their levels of alcohol consumption and being worse off as opposed to better off as a result. Target based approaches would give more recognition to the benefits of moderate levels of alcohol consumption.

1.4: Data Description

The data to be used in this research will be taken from the 2007 Slán survey. Slán is a national health and lifestyle survey which was first undertaken in 1998. A further survey was carried out in 2002 and again in 2007, with the 2007 survey being the most comprehensive. There is no linkage between the surveys. This cross sectional survey is commissioned by the Department of Health and Children in Ireland. The survey and analyses were carried out by the National University of Ireland, Galway along with the Consortium consisting of the Royal College of Surgeons in Ireland, National University of Ireland, Cork and the Economic and Social Research Institute (ESRI). (Morgan *et al*, 2008). The main aims of these surveys are to

- Produce reliable data of a nationally representative cross-section of the Irish population in order to inform the Department of Health and Children's policy and programme planning.
- Maintain a survey protocol which will enable lifestyle factors to be measured and re-measured which will allow for trends and changes to be identified. This is a useful tool in the monitoring of the different policies that are being implemented and in the planning for any future policy changes.

In the Slán survey a cross section of the Irish adult population, aged 18 and over, are surveyed. The selection is a random sample which is proportionately distributed across counties, locality, gender and urban/rural locations. The 2007 survey was also proportionately distributed across age groups and social classes. All counties in the republic were represented.

In the 2002 questionnaire, there were eight sections which covered general health (including reported height and weight), physical activity, tobacco, alcohol use, illegal substance, accidents, household details, and dietary habits. The 2002 survey had a valid sample of 11,212 questionnaires. A national response of 5,992 (53.4%) was obtained. The gender distribution of the overall respondents was 41% (2,448) male, and 59% (3,526) female.

In the 2007 questionnaire there are nine sections which cover general health (including reported height and weight), mental health and well being, physical activity, diet and nutrition, smoking, alcohol status and other substances, injury, family-social networks & neighbours, body weight and measurement. The 2007 Slán survey was the largest survey to date. 10,364 people (62% response rate), aged 18+ were interviewed in their own homes, by researchers from the Economic Social and Research Institute (ESRI). A sub study on body size with 967 younger adults (aged 18-44 years) and a more detailed physical examination involving nurse assessment and blood urine sampling in 1,207 adults (aged 45 years and over) was also carried out. The sample was representative of the general population in Ireland when compared with the census 2006 figures and was further weighted to match the census for analysis. Most findings were analysed by gender, age and social class categories.

In relation to income , twenty-four categories are set out reflecting a households total net income per week, ranging from the lowest category of less that €86 per week to the highest of €1,535 or more per week and respondents are asked to select which income band is applicable to their household.

In the general health section of the Slán survey, respondents are asked to categorise their health as being excellent, very good, good, fair or poor. Respondents are also asked about the last time they consulted a General Practitioner.

1.5: Methods and Techniques

The relationship between alcohol use and income is examined for three categories of drinkers; non, moderate and heavy drinkers accounting for endogeneity and selection bias. This is initially carried out using the Lee Multinomial Logit OLS Two Step Estimate (Lee, 1982). This is a limited information method of estimation. It is assumed that individuals will choose an income and alcohol consumption combination that maximises utility, subject to given constraints. Income for an individual depends on a vector of human capital variables, sociodemographic characteristics along with ones drinking status similar to what was done in previous studies (Hamilton and Hamilton, 1997; Barrett, 2002). The Drinking Status equation uses the same variables as the income equation which accounts for the effect of income on drinking status, along with other exogenous variables that are hypothesised to be unique to the drinking decision. The Lee Multinomial Logit OLS Two Step Estimate derives the inverse mills ratio from the estimation of alcohol status, and this is then included in the income equation to account for potential selection bias. Income differentials between drinker types are then estimated which accounts for the potential endogeneity bias which may arise through unobserved factors which affect both alcohol use and income.

Given that alcohol consumption could be viewed as ordered data (Harris *et al*, 2006), the ordered nature of the dependent variable alcohol is accounted for by estimating alcohol status using an ordered probit model. Previous studies in this area have not accounted for the ordered nature of the data (Hamilton and Hamilton, 1997; Barrett, 2002). The income equations for the three categories of drinkers are estimated with an underlying ordered probit selection rule using both the Limited Information Method of estimation and the Full Information Method of estimation. The Full Information Maximum Likelihood method consists of finding parameter values that maximise the likelihood of the data by treating all equations and parameters jointly (Greene 2002).

In looking at the effect of alcohol consumption on health status, the health status equation is estimated for the three categories of drinkers using an ordered probit model accounting the potential selection bias of drinking. Similarly health care utilisation which is closely correlated to health status in that those with poorer health status tend to have higher levels of health care utilisation (Lim *et al*, 2005; Finkelstein, 200; Rotermann, 2006) is analysed in the same manner, whereby GP consultations is assessed through an ordered probit model accounting for the potential selection bias of drinking.

1.6: Organisation of Study

By examining the impact of alcohol consumption on income, health status and health care utilisation and by showing that there are advantages to moderate levels of alcohol consumption, a target based approach aimed at the misuse of alcohol consumption can be considered.

Chapter 2 reviews previous studies into the effect of alcohol consumption on income, health status and health care utilisation and reviews the literature in relation to the econometric issues that arise and techniques that can be adopted in such studies. Section 2.1 reviews how drinkers can be categorised and examines the factors that effect alcohol consumption and income. The issue of endogeneity and selection bias is reviewed along with the econometric techniques that can be adopted to account for such issues. Section 2.2 examines the alcohol status variable as ordered data and how estimation could be carried out while accounting for endogeneity and selection bias. Section 2.3 looks at the different methods of estimation for simultaneous equations. Limited Information Methods and Full Information Methods of estimation are assessed and compared. Section 2.4 looks at the concept of self assessed health status and health care utilisation as a measure of ones health. The variables that have been found to affect an individual’s health status and health care utilisation are examined along with the Grossman human capital model of the demand for health. Possible econometric techniques to look at the effect of a lifestyle variable such as alcohol on health, is reviewed. Conclusions are presented in section 2.5.

The Slán National Health and Lifestyle survey is described in Chapter 3. All the variables used in the study are described as set out in the Slán survey and a detailed description of the dependent variables income, drinking status, health status and health care utilisation is provided. Standard Deviations and the mean values are set out for both the dependent and independent variables along with the minimum and maximum value in relation to each variable.

Chapter 4 presents a study on the effect of alcohol consumption on household income in Ireland. The econometric issues that arise, in particular the endogeneity and selection bias of alcohol consumption are assessed, and the techniques that can be adopted to deal with such issues. The Lee Multinomial Logit OLS Two Step Estimate is used which involves the estimation of the alcohol status equation in step one from which the inverse mills ratio is derived which is then included as an additional regressor in the income equation in step two. This estimation allows the relationship between household income and alcohol status with different personal and socio economic variables to be examined.

Chapter 5 considers the ordinal nature of alcohol consumption and the potential implications of not accounting for this. Methods of estimating the effect of alcohol on income treating alcohol as an ordinal variable while still accounting for endogeneity and selection bias are reviewed. Limited Information Methods of Estimation and Full Information Methods of Estimation are also reviewed with both methods used in the estimation of the effect of alcohol on income.

Chapter 6 presents an empirical study of the effects of alcohol consumption on health status and health care utilisation while accounting for the potential endogeneity and selection bias of alcohol. A variation of the Heckman two step model is used in the estimation of this effect, estimating alcohol consumption and health status and health care utilisation as an ordered probit. Differences in health status and health care utilisation for non, moderate and heavy drinkers is examined and the relationship between both alcohol status, health status and health care utilisation with a host of other personal and socio-economic variables such as age, gender, marital status, employment status and level of education, among others, is also assessed. The relationship between alcohol status and specific illnesses is also examined. Finally conclusions and recommendations are identified and discussed in Chapter 7.

Chapter 2

THEORETICAL ISSUES AND PREVIOUS RESEARCH

This thesis presents an empirical study of the effect of alcohol consumption on household income and healthcare in Ireland. Individuals are categorised into one of three categories of drinkers; non drinkers, moderate drinkers or heavy drinkers using the Slán 2007 Data Set. Section 2.1 focuses on the estimation of income and the factors that affect income which need to be accounted for in such an estimation. The issue of endogeneity and selection bias is examined and the possible methods of estimation that account for such issues reviewed. The categorisation of individuals into different drinking categories is assessed along with the factors that have an affect on the levels of alcohol an individual consumes. Section 2.2 looks at interpreting alcohol status as ordinal data and how estimation could be carried out while accounting for endogeneity and selection bias. Section 2.3 looks at the different methods of estimation for simultaneous equations. Limited Information Methods and Full Information Methods of estimation are assessed and compared. The concept of health and self assessed health as a measure of ones health status is discussed in Section 2.4. The Grossman human capital model of the demand for health is utilised to identify the factors that affect both health status and health care utilisation. Econometric techniques that could be utilised to look at the effect of a lifestyle variable such as alcohol, on health is reviewed. Conclusions are presented in section 2.5.

2.1: Alcohol and Income: An Introduction

This section reviews previous literature which focuses on measuring the effect of an individual's alcohol consumption on income. Endogeneity and selection bias are also discussed in detail given that these issues emerge in the literature, as problems that arise when trying to measure this effect. Alcohol consumption is estimated in step one and from this the inverse mills ratio is derived. The income regression is estimated in step two and the inverse mills ratio derived in step one is included as an additional regressor in the income equation. This two step method controls for selection bias. The income regressions are estimated by drinker type to control for possible endogeneity similar to what was done in previous studies (Hamilton and Hamilton, 1997; Barrett, 2002).

2.1.1: Income by Drinker type

In looking at the effect of alcohol consumption on an individual's welfare, different measures of welfare tend to be used, examples being wages, earnings and income among others[1]. Income in defined as a consumption opportunity gained by an entity within a specified timeframe, which is generally expressed in monetary funds (Barr, 2004). For households and individuals, it is the sum of all the wages, salaries, profits, interest payments, rents and other forms of earnings received, and for firms, income generally refers to net-profit (Barr, 2004).
Zietz and Zhao (2009) define household income as the sum of incomes of all household members from all sources and can vary significantly over time with changes in household composition and/or income opportunities of its members.

Hamilton and Hamilton (1997) and Barrett (2002) use earnings as a measure of welfare and set out the earnings equation for non, moderate and heavy drinkers shown in equation 2.1.1. Barrett (2002) defines earnings as the gross personal income annualised for the year prior to the survey date. Hamilton and Hamilton (1997) define earnings in their study as income before taxes, from wages during the year 1984.

$$\ln Y_{ij} = X\beta_j + u_{ij} \qquad \textbf{(2.1.1)}$$

Where: X_i vector of human capital variables & sociodemographic characteristics

β_j coefficients on the observable characteristics

u_{ij} $\sim N(0, \sigma j^2)$

[1] Different measures of individual's financial welfare are used in the literature, Barrett (2002) and Hamilton and Hamilton (1997) use earnings for example, while French and Zarkin (1995) use wages.

i indexes individuals where $i = 1,2,......N$

j indexes drinking status where $j = 1,2,3,$

Earnings for each individual will be hypothesised to depend on a vector X which consists of a vector of human capital and socio demographic variables. This is a very general specification, which allows for labour market returns for individual characteristics to differ by drinking status (Hamilton and Hamilton, 1997; Barrett, 2002).

By comparing the estimated β across drinker types it is possible to gauge whether financial welfare rewards for observed productivity related characteristics are greatest for moderate drinkers and lowest for heavy drinkers, as implied by the medical research on the health effects of alcohol consumption (Hamilton and Hamilton, 1997; Barrett, 2002). It is assumed that an individual will select the earnings drinking status combination that maximises expected utility (Hamilton and Hamilton, 1997; Barrett, 2002).

Vector X consists of many human capital and socio demographic characteristics which have been found in previous studies to affect earnings (Hamilton and Hamilton, 1997; Barrett, 2002). These are presented in Section 2.1.2

2.1.2: Factors affecting Income

There are many factors such as education, gender, age, marital status, among others which affect income at individual and household level (Barrett, 2002; Hamilton and Hamilton, 1997; French & Zarkin, 1995).

2.1.2.1: Education

Education can have a significant impact on ones financial welfare and generally studies show that an individual with a higher education tends to be financially better off than someone who does not have third level qualifications (Barrett, 2002; French & Zarkin, 1995; Heien, 1996). Van Ours (2004) in looking at the wage effects of tobacco and alcohol find that highly educated individuals earn about 35% more than individuals without an education.

The impact of education on income can also vary by the drinker type. Hamilton & Hamilton (1997) find that the effect of education on earnings can differ between different categories of drinkers in that being a college graduate is more beneficial for non and moderate drinkers but not for heavy drinkers. Barrett (2002) on the other hand shows that having a university degree has a significant effect on earnings but this applies across all drinker types.

Grossman (1972) also highlights education as a factor that influences ones health in that educated people have better knowledge around the importance of good health and know what are the ways to achieve and maintain good health. Good Health in turn means that people are able to work more, are absent from work less and are more productive which results in higher income (Grossman, 1972).

2.1.2.2: Age

A person’s age can have a significant impact on ones income (Nunes, 2008). Findings tend to show that very often income of young people, starting off in their working career, tends to be less, however as they get older and gain more experience, this can then have a positive impact on their income (Nunes, 2008). The cycle then starts to change again as people near retirement age in so far as they have often peaked in terms of their income which tends to be earning close to the maximum possible, hence very often people in this age category tend to experience a drop or a levelling off in earnings (Barrett, 2002).

Contrary to these findings Van Ours (2004) finds that age has a positive effect on wages for both males and females and that for every year they grow older, male’s wages increase with 1.2% and females increase 0.6% annually. Other studies show that not only does age affect ones income but in terms of analysing different categories of drinkers, this affect can vary, depending on the levels of alcohol they consume (Hamilton & Hamilton, 1997). Hamilton & Hamilton (1997) show that for non and moderate drinkers earnings increase when workers are in their 30's or 40's and then drops off slightly when workers are in their 50's. They find for heavy drinkers earnings increase only very slightly when they are in their 30's and 40's and than falls after that.

Barrett (2002) finds that age earnings profile is steepest for heavy drinkers peaking at the age of 30-34 years and declining steeply over older age intervals and that moderate drinker's peak at the age of 45-49 years and there is a slower decline over later years.

2.1.2.3: Race

Previous studies show that income for white workers is higher than those of other races (Yang, 2007; Hogan and Perrucci, 2007). The average income of black workers is 11.06% lower than that of white workers, while Hispanic workers' average income is lower than that of white workers by a sizable gap of 41.25%, in a study by Yang (2007). Deaton and Lubotsky (2003) also find that black people have a lower income than white people.

Mullahy and Sindelar (1996) show that race is an important determinant of employment and unemployment, and that in relation to males, whites have higher employment propensities than non-whites. For females on the other hand race is not a significant determinant in employment propensities, but non white females do have greater unemployment propensities (Mullahy and Sindelar, 1996). Contrary to these findings Berger and Leigh (1988) find that race differences in terms of wages are insignificant.

2.1.2.4: Gender

Financial welfare differences exist between males and females in that men are better off than women, even when taking account of people with the same characteristics (Mullahy and Sindelar, 1996; Heien, 1996; Zhang, 2008; Miyoshi, 2008). Mullahy and Sindelar (1996) find that men are more likely to be employed and less likely to be unemployed. Zhang (2008) carried out a study of gender based employment and income differences in Urban China, and discovers that men do enjoy advantages in human capital and political connections; however findings do show that neither of these factors explains the observed gender gaps in employment and income. Instead results point to the importance of family/work conflict. Zhang (2008) shows that, controlling for age and other background characteristics, female disadvantages in income and employment exist only for wives and mothers. Working wives and mothers spend much more time on housework than do working husbands and fathers. There was little evidence of a gender gap to be explained among the unmarried and non-parents. Miyoshi (2008) finds that there are significant wage differences between males and females in Japan that cannot be explained by differences in observable characteristics therefore female workers will not receive the same wage even if they have the same characteristics as males. Reasons cited for this gender wage gap is due to the fact the full time work experience and seniority which do affect wages is shorter for females than for men. Similarly Napari (2009) finds significant disparities in wage development between genders during the first ten years in the labour market. After ten years the size of the gender wage gap more than doubles. Having controlled for characteristics such as education, region, firm size, a significant part of the gender wage gap remained unexplained. These finding were similar to those of Loprest (1992).

2.1.2.5: Marital Status

Several studies have looked into the effect of marital status on income and findings show that marriage for men is positive in terms of contributing to their income however it is not very significant for women who are married (Berger and Leigh, 1988; Mullahy and Sindelar, 1996). Mullahy and Sindelar (1996) also find that being married has a little impact on unemployment. In a study of the wage advantages to married men compared with single men, Schoeni (1995) discover that there was an advantage to married men in all 14 countries studied.
This gain in terms of wages to married men compared with single men was estimated by Ahituv and Lerman (2007) to be 18%, and the gain in being married versus being divorced is 19%. Entering remarriage led to a 13% advantage over those remaining divorced. Ahituv and Lerman (2007) also find that an increase in wages can have an influence on marital status. A 10% increase in wages, led to a

6% increased likelihood of entering marriage, 6% increase in the chances of staying married and less than 2% chance of getting remarried. Loh (1996) estimated the marriage earnings premiums to men, using census data 1940-1980, had been consistently significant at percentages ranging from 11% in 1959 to 23% in 1969.

Madalozzo (2008) carried out a study into income differentials between married women and those who remain single or cohabite. The author based the study on women only, due to the fact that much research has been carried out into the effect of marital status on income for males; however this has not been the case for females. Results show that there is a statistically significant gap between married and cohabiting women in the range of 49% to 53%, favouring the cohabiting women. When comparing married women with single women, single women's income is 25.6% higher than if they were married (Madalozzo, 2008).

Table 2.1.1 summarises the findings from different studies into the effects of marriage on income for both men and women. In general studies show that for men marriage has a positive effect on income. For women some studies show the effects to be insignificant while Modalozzo (2008) find marriage to have a negative effect on income.

Table 2.1.1. Findings from previous studies into the Effects of Marriage on Income for Men and Women

Effect of Marriage on Income for Men and Women		
	Men	**Women**
Berger & Leigh (1988)	Positive	Not significant
Mullahy & Sindelar (1996)	Positive	Not significant
Schoeni (1995)	Positive	-
Ahituv & Lerman (2007)	+18%	
Loh (1996)	+11%-23%	-
Modalozzo (2008)	-	-25.6% (negative effect)

(Source: Authors own)

Mullahy and Sindelar (1996) find that family size is not a significant determinant of employment status outcome for males, however for females family size is significant in so far as it has a negative impact on people's employment possibilities. In looking at gender based differences in employment and income, Zhang (2008) find that while gender gaps in employment and income exist, this is not the case among those unmarried and among those without children. Zhang (2008) states that it is family formation that is the reason behind gender based differences in income and employment. Gaps are concentrated among married women and women with children, even when the same level of human capital and political capital exists as with men and even when women work as many hours in paid labour as men.

2.1.2.6: Number of people in household

Household income is the sum of incomes of all household members (Zietz and Zhao, 2009), hence the number of people working in the household will affect household income. In Ireland, those over 16 years of age can be legally employed in regular full time jobs (Protection of Young Persons Employment Act, 1996).

2.1.2.7: Occupation

Barrett (2002) finds there are significant differences in earnings across industries for each drinker type. The general pattern of occupational earnings differentials are consistent across the three drinker types; administrative and professional occupations receive the highest earnings, and clerical and service occupations pay higher earnings than blue collar jobs (Barrett, 2002). The magnitude of the occupational earnings differences is substantially lower among non-drinkers while the returns to white collar occupations are greatest for moderate drinkers. After conditioning on other observable factors, non drinkers receive a substantial public sector earnings premium while heavy drinkers receive a large public sector earnings penalty (Barrett, 2002). In an analysis of female earnings, Ressler and Waters (2000) discover that those employed in management, sales and labour earn significantly more than craft and service occupations.

2.1.2.8: Region

Substantial differences in earnings by state in Australia exist (Barrett, 2002) and similarly this is the case in Canada (Hamilton and Hamilton, 1997). Barrett (2002) finds that differences are less pronounced among abstainers however for moderate and heavy drinkers, workers in Queensland and South Australia earn less than their counterparts in New South Wales. Ressler and Waters (2000) find that female earners in urban areas earn more than those in rural areas.

2.1.2.9: Health

"The greatest health is wealth" (Virgil, 70BC-19BC).

Grossman (1972) argues that if one can improve their health status they are then in a position to work more and this then results in ones income increasing. He adds to this theory by saying that an increased wage rate increases ones returns from healthy days so workers will therefore tend to increase their optimal capital stock of health.

Grossman (1972) looks at education as a factor allowing a person to improve the efficiency with which one can produce investments in health and that education raises the marginal product of the direct inputs. A given investment in health can be generated at less cost for an educated person and therefore they experience a higher rate of return to a given stock of health. Educated people tend to be healthier. Grossman (1972) argued that better educated people tend to be economically more efficient producers of health; they have the know how needed to stay healthy, they have better knowledge in terms of knowing how to use medical and other market inputs and their own time in order to produce health and therefore increase income. In a more recent article by Grossman (2008) he states that an increase in schooling raises the efficiency of the production process and that more educated people have more information about the true nature of the production function such as not smoking and what constitutes an appropriate diet. He argues that an educated person responds to new knowledge more rapidly. Grossman (2000) refers to studies into the spread of HIV/AIDS epidemic and the fact that after more than a decade of prevention campaigns about the dangers of the HIV/AIDS epidemic in Uganda, there has been a significant change in the HIV/education gradient.

Grossman (1972) argued that from the demand side educated people tend to recognise the benefits of improved health and they have a greater demand for health relative to other goods and a greater appreciation for the benefits brought about by being healthy. There is both a demand effect and a supply effect from education.

Mullahy and Sindelar (1996) in line with Grossman's theory show that health status affects ones participation in the labour force. Similarly Nunes (2008) find that health status is higher among men, lower age groups, individuals who smoke less, individuals who smoke more and individuals in higher education levels.

2.1.2.10: Alcohol and Income

Much research has been carried out into effects of alcohol on ones financial welfare and whether or not individuals with similar characteristics, are financially better or worse off which can be attributed to the level of alcohol they consume (Zarkin *et al*, 1998; Hamilton and Hamilton, 1997; Barrett, 2002). Many studies have categorised drinkers into categories of non-drinkers, moderate and heavy drinkers (Hamilton and Hamilton, 1997; Barrett, 2002), while some studies have broken these categories down further (Zarkin *et al,* 1998). In general findings appear to be that there is a positive association between moderate alcohol consumption and an individual's financial welfare, compared with non and heavy consumption of alcohol and an individuals financial welfare (French and Zarkin, 1995; Zarkin *et al*, 1998; Hamilton and Hamilton, 1997; Barrett, 2002).

French and Zarkin (1995) in their study of the effect of alcohol consumption on wages find that moderate alcohol users have higher wages than abstainers and heavy drinkers and that wages peaked at approximately 1.5 to 2.5 drinks per day on average, showing an inverse U-Shaped relationship between alcohol consumption and wages.

Heien (1996) in his study, using the National Household Survey on Alcohol Abuse, shows that moderate drinkers earn more than abstainers, however while the effect of moderate alcohol consumption on earnings is statistically significant, it is not as significant as other human capital variables such as education or age. Interestingly they do find that ex-drinkers earn less than lifetime abstainers.

These earlier studies such as those carried out by Heien (1996) and French and Zarkin (1995) however were limited in so far as drinking status is treated as exogenous; therefore the estimated impact of alcohol consumption on income may reflect the reverse effect of income on alcohol consumption (Zarkin *et al*, 1998; Di Pietro & Pedace, 2008; Hamilton and Hamilton, 1997; Barrett, 2002). Ziebarth and Grabka (2009) argue that if endogeneity and selection bias is not accounted for it may mean simply that people with certain characteristics self select themselves into different drinking habits. More recent studies have addressed the issue of endogeneity of drinking status when looking at its effect on earnings (Zarkin *et al*, 1998; Hamilton and Hamilton, 1997; Barrett, 2002). Findings of such studies proved to be similar to those of French and Zarkin (1995) and Heien (1996). Hamilton and Hamilton (1997) find that moderate alcohol consumption leads to increased earnings relative to abstention, however heavy drinking leads to reduced earnings relative to moderate drinking. There is a striking difference between heavy drinkers and other workers in terms of the shape of their age earnings profile (Hamilton and Hamilton, 1997). Heavy drinkers possessed flatter age earnings profiles and receive lower returns to higher education than other drinker types.

In his analysis of data from the Australian National Health Survey, Barrett (2002) also finds that moderate drinking leads to a significant earnings premium, but that drinking heavily leads to an earnings penalty. Other studies with similar findings are Lye and Hirschberg (2004) who show that earnings premium to drinkers was maximised at an average daily consumption level at approximately four drinks. Kenkel and Ribar (1994) estimate that heavy drinking led to 12% earnings penalty and alcohol abuse to a 30% earnings penalty for young men.

With some studies while there proved to be a positive correlation between alcohol use and financial welfare, the drop associated with heavy consumption of alcohol compared with moderate consumption, is not found (Zarkin *et al*, 1998; Bastida, 2006). Zarkin *et al* (1998) in their study looked at men and women separately and categorised drinkers into eight different categories, showing that male alcohol users have higher wages, approximately 7%, than non-drinkers, and this apparent wage premium is approximately the same over a wide range of alcohol consumption. Zarkin *et al* (1998) also show that the estimated alcohol use premium for women is approximately half as large as men.

A study of an older cohort of people, aged 37 years and over, by Bastida (2006), show similar findings in so far as there is a positive association between alcohol consumption and earnings, however she did find that this appears to be the case for all levels of alcohol consumption from moderate to heavy. In this study, however the author did not take account of the problem of endogeneity.

While Berger and Leigh (1988) again find that drinking is associated with higher wages, because they only categorised people as drinkers and non-drinkers, they are unable to indicate whether moderate drinkers have the highest wage premiums. Similarly Van Ours (2004), when looking at the effect of alcohol and tobacco on wages, looked at drinkers and non-drinkers and find that for females drinking alcohol did not have an impact on wages. For males on the other hand there is a positive effect and the wages of male consumers are about 10% more than non-drinkers.

Mullahy and Sindelar (1996) take a slightly different approach in so far as they look at the effects of alcohol consumption on employment and unemployment and discover that for both men and women problem drinking results in reduced employment and increased unemployment. These findings also correspond to results from previous research carried out by Mullahy and Sindelar (1991, 1993) however this previous research did not treat drinking status as endogenous.

Balsa and French (2010) in their study of the labour market consequences of heavy or abusive drinking in Latin America specifically Uruguay find a positive association between heavy drinking and absenteeism, particularly for female employees. They find a positive relationship between heavy drinking and labour force participation or employment. This result was mostly driven by men and weakened when considering more severe measures of abusive drinking. As possible explanations for such findings, Balsa and French (2010) suggest that employment leads to greater alcohol use through an income effect, that the Uruguayan labour market rewards heavy drinking, or that labour market characteristics typical of less developed countries, such as elevated safety risks or job instability, lead to problem drinking.

There are many variables that impact on financial welfare and for this reason all these variables would need to be accounted for in the formulation of an income equation and estimated for when looking at the effect of alcohol on income (Barrett, 2002; Hamilton and Hamilton, 1997; French & Zarkin, 1995). In looking at the relationship between income and alcohol, endogeneity can arise and needs to be accounted for as was done in the studies by Barrett (2002) and Hamilton and Hamilton (1997) into the effect of alcohol on earnings.

2.1.3: The Endogeneity Issue

> *"Endogeneity leads to biased and inconsistent estimators within equations used to test theoretical propositions, which makes inferences problematic and consequently reduces the confidence we have in drawing from research".* (Chenhall and Moers, 2007)

Endogeneity refers to the fact that an independent variable included in the model is potentially a choice variable, and variables can be jointly determined which as a result leads to correlation between the unobservables and the disturbance term (Chenhall and Moers, 2007). Endogeneity does not occur when an independent variable is not determined by other parameters and variables in the model

(Chenhall and Moers, 2007). Endogeneity exists when the model includes an endogenous explanatory variable that is determined within the context of the model (Chenhall and Moers, 2007). It is likely to be apparent when studies place a choice variable on the right hand side of the equation that is specified to test, whether the choice variable is associated with the specified outcome (Chenhall and Moers, 2007). According to Milbourne *et al* (2003), the possible endogeneity of the right hand side variables has two implications one econometric and the other interpretation;

- Firstly, the parameter estimates will be biased and inconsistent and the model will fit too well (Milbourne *et al*, 2003). Therefore the magnitude of the parameter estimates will be unreliable, as will the measures of goodness to fit.

- Secondly, it makes interpretation of the parameter estimates difficult (Milbourne *et al*, 2003). It would be impossible to say whether drinking status affects income or whether income has an effect on drinking status.

If bias is unlikely to be a major problem, then normally Ordinary Least Squares would be used, however using an Ordinary Least Squares (OLS) regression of a demand equation with suspected endogeneity of the regressors would lead to estimates that are biased and inconsistent (Chenhall and Moers, 2007; Milbourne *et al*, 2003; Mullahy 1999, Greene, 2002)

The most common way to deal with the problem of endogeneity is through using Instrumental Variables (IV) estimation (Knowles and Owen, 1997; Barrow and Sala-i-Martin, 1999; Cho, 1996). An instrument is a proxy for the endogenous explanatory variable X that is highly correlated with X but is uncorrelated with the error term of the demand equation (Gujarati, 1995). Brookhart *et al* (2010) in looking at studies of medical intervention, define the Instrumental Variable (IV) Approach as identifying a variable that is assumed to be related to the treatment or endogenous independent variable, but is neither directly related to the dependent variable, or indirectly related via pathways through unmeasured variables. Brookhart *et al* (2010) state that an instrumental variable is an observed variable that generates variation in the exposure akin to randomised assignment. They state that the requirements of an instrumental variable are that the IV should affect treatment; should be a factor that is as good as randomly assigned; should be related to the outcome only through its association with treatment, thus an instrumental variable should have no direct or indirect effect on the outcome.

Ziggy and Shields (2001) in looking at the impact of alcohol consumption on occupational attainment in England state that an endogenous relationship exists between alcohol consumption and occupational attainment, which results in the error terms being correlated with one of the explanatory variables. They account for this existence of unobserved heterogeneity using instrumental variables, whereby they look for a variable that is correlated with alcohol consumption but not with occupational attainment. They state that the instrumental variable estimation (IV) accounts for endogeneity and allows a more accurate assessment of the true impact of alcohol consumption on occupational attainment. Ziggy and Shields (2001) do state that a practical difficulty with IV estimation is finding an instrument or set of instruments which are significant determinants of the endogenous variables but not a significant determinant occupational attainment. In looking for a variable that is correlated with the drinking variable but not with the error term, they use three different instruments; number of dependent children, long term non acute illnesses and smoking. They use the different variables due to the fact that different instruments can provide different results (Ziggy and Shields, 2001).

Similarly Milbourne *et al* (2003) argue that the main difficulty with the use of IV regressions is finding a sufficient number of suitable instruments. Ziebarth and Grabka (2009) argue that the instrumental variable approach is very limited and that while the consequences of weak correlation between the instrument and endogenous variable are well understood and while distinct tests are available, the exogeneity assumption of the instrument is not directly testable, rendering the rest of the analysis mostly a matter of belief. The most frequently used way of controlling for endogeneity in the estimation of alcohol consumption on income is to estimate separate wage equations for each category of drinker, treating alcohol consumption endogenously (Hamilton and Hamilton, 1997; Barrett, 2002).

Leigh and Schembri (2004) state that when looking at choice and outcome variables, some of the variation in the treatment variable can be affected by reverse causality, unobserved variables or measurement error. The Instrumental Variable approach is a solution to this problem. Leigh and Schembri (2004) state that a valid instrument should satisfy two requirements; first that it is logically related to and statistically correlated with a choice variable; secondly, there is no reason why the instruments should be directly related to the outcome other than the instruments effect on the choice.

Leigh and Schembri (2004) when looking at the factors that affect the health production function, state that a strong instrument should be strongly statistically correlated with a choice variable, however very often many of the instruments used in IV studies have been invalid and weak. By employing instrumental variables to treat the heterogeneity of health inputs, estimates of the health production function is unbiased (Rosenzweig and Schultz, 1983).

Browning and Meghir (1991) state that an immediate reaction to a conditional demand equation, such as the demand for alcohol, is that it includes variables on the right hand side that may be endogenous for the demand equation. A particular concern in a study such as that of the effects of alcohol on income, and one which was raised in other studies into the effect of alcohol on the financial welfare of an individual (Zarkin *et al*, 1998; Hamilton and Hamilton, 1997; Barrett, 2002), is that the effect of alcohol consumption is governed in part by unobserved factors which may also be important determinants of the dependent variable income and if unobserved factors influence both income and the decision to drink alcohol, then alcohol use will be correlated with the error term in the income equation. If endogeneity exists then the income equation cannot be regressed using OLS, as it would result in biased and inconsistent results (Di Pietro & Pedace, 2008; Hamilton and Hamilton, 1997; Barrett, 2002; Choudhury, 1993) and will also result in the interpretation of the parameter estimates to be difficult in that it would not be possible to say whether or not alcohol consumption has an effect on ones income or whether or not ones income affects alcohol consumption (Zarkin *et al*, 1998; Hamilton and Hamilton, 1997; Barrett, 2002).

2.1.3.1: Testing for Endogeneity

The exogeneity assumption of the instrument is not directly testable (Ziebarth and Grabka, 2009). If endogeneity exists then the Ordinary Least Squares estimator will be biased and inconsistent and so an alternative to OLS, such as the Instrumental Variable (IV) approach, which is the most common way to deal with the problem of endogeneity will have to be used (Knowles and Owen, 1997; Barrow and Sala-i-Martin, 1999; Cho, 1996). If there is no measurement error both OLS and IV will be consistent and will have the same probability limit and OLS will be preferred (Verbeek, 2008).

The Hausman test is a test based on a comparison between two estimators, whereby estimates from OLS and IV are tested to see if differences exist (Kennedy, 2003; Cameron and Trivedi, 2005). The Hausman test, tests for the null hypothesis that the OLS estimator is consistent and fully efficient (Griffiths *et al*, 1993). The test involves estimating the model via both OLS and IV approaches and comparing the resulting vectors. Under the null hypothesis the OLS estimate is consistent in that there is no measurement error and the OLS and IV coefficients will not be systematically different (Griffiths *et al*, 1993). Acceptance of the null hypothesis requires that the difference between the two sets of estimates be small (Griffiths *et al*, 1993).

Many studies into the effect of alcohol consumption on an individual's financial welfare have not used specific tests for endogeneity but have accounted for potential endogeneity of alcohol status by estimating different earnings equations for each category of drinker (Hamilton and Hamilton, 1997; Barrett, 2002).

2.1.4: Selection Bias

In determining the effect of alcohol on income, the possibility of sample selection bias arises which is where individuals self select into different drinking categories, and would result in the outcome differences being potentially explained as a result of pre-existing differences between the groups, as opposed to the actual levels of alcohol consumed (Hamilton and Hamilton, 1997; Barrett, 2002). In such an instance selection into a particular category would be non-random and the unobserved individual characteristics affecting the choice variable also influence the income level (Heckman, 1979).

Hamilton and Hamilton, (1997), when trying to identify the effect of alcohol on earnings, categorise individuals into three categories of drinkers; abstainers, moderate drinkers and heavy drinkers, however if individuals are self selecting into drinking categories, then the samples will not be random (Hamilton and Hamilton, 1997). People may select into particular drinker groups because of their individual characteristics and because they know that by so doing it will not have a negative effect on their income; an example being that only individuals who can cope with heavy drinking without incurring a marked drop in earnings, choose to do so (Hamilton and Hamilton, 1997). In this case selection into drinking categories may not be random in so far as heavy drinkers may have systematically different characteristics from those who are not heavy drinkers (Di Pietro and Pedace, 2008). Such characteristics could exert an influence on not only an individual's probability of being a particular category of drinker but also on their earnings potential (Di Pietro and Pedace, 2008). Where selection into particular categories may occur, choices have to be treated endogenously to get consistent estimates of the income equation coefficients (Zhang, 2004). It is for this reason that a drinking selection equation needs to be considered when estimating the income equation and by including the variables that determine income in the drinking status choice equation, this controls for the effect of income on drinking behaviour, which is similar to what both Hamilton and Hamilton (1997) and Barrett (2002) did in their studies into the effect of alcohol on earnings.

As Heckman (1979) and others demonstrate, this non-randomness, or self selection of alcohol consumption violates the Gauss Markov assumptions and consequently the desired outcomes by OLS yields potentially biased results (Hilmer, 2001). Failure to account for non-random selection in drinking status will lead to biased estimates (Hamilton and Hamilton, 1997; Barrett, 2002).

One way of accounting for potential selection bias is to use the standard two step estimation proposed by Heckman (Berger and Leigh, 1988; Di Pietro and Pedace, 2008) which is set out in Appendix C. The first step being to estimate a person's propensity to drink through probit analysis using information on the observed drinking decision. The probit estimates then generate predicted values for the Inverse Mills Ratio which are then inserted into the corresponding income equations, producing consistent results of the income equations corrected for selection bias (Berger and Leigh, 1988; Di Pietro and Pedace, 2008).

Lee (1982) extends the Heckman Probit OLS Two Step Estimate to a Multinomial Logit OLS Two Step Estimate, to allow for selection correction for polychotomous choices. Hamilton and Hamilton (1997) and Barrett (2002) in their analysis of alcohol status on earnings, group drinkers into three categories and use the multinomial logit two step estimation to carry out the analysis (Hamilton and Hamilton, 1997; Barrett, 2002). Selection bias treats the sector selection alcohol, endogenously in that different characteristics can influence an individual's probability of being in a particular category of drinker and can influence their earnings (Hamilton and Hamilton, 1997; Barrett, 2002). In isolating the effect of alcohol consumption on earnings it is necessary to control for the potential endogeneity of drinking status (Hamilton and Hamilton, 1997; Barrett, 2002). Step one involves using multinomial logit to estimate the alcohol status equation, which generates predicted values for the Inverse Mills Ratio. In estimating the alcohol status equation in step one, there must be at least one instrument that has no effect on income except through its effect on alcohol. Such a variable must be a significant determinant of alcohol yet satisfy the exclusion restriction $Cov(w, \varepsilon_j) = 0$ for all of selection categories (Chiburis and Lokshin, 2007). In the second step an OLS earnings regression is run which includes the values for the Inverse Mills Ratio (Hamilton and Hamilton, 1997; Barrett, 2002). By estimating separate earnings equations including the Inverse Mills Ratio, the endogeneity bias that may arise through simultaneity of drinking status and earnings due to the reverse causation from income and alcohol consumption or unobserved heterogeneity addressed (Hamilton and Hamilton, 1997; Barrett, 2002).

Before alcohol status can be estimated drinkers need to be categorised. Studies have adopted different approaches to the categorisation of drinkers.

2.1.5: Definition and Categorisation of Alcohol Consumption

In the literature there is no definition for alcohol consumption, it is defined through the categorisation of drinkers (Knupfer, 1984; Hamilton and Hamilton, 1997; Barrett, 2002; Kenkel and Ribar, 1994; Zarkin *et al*, 1998).

Knupfer (1984) in measuring the frequency of intoxication finds that those who drink at least eight drinks a day one or more times per week face the highest risk of social disapproval or personal concern over their drinking habits. Some studies have based their categorisation of drinkers on these findings (Hamilton and Hamilton, 1997; Barrett, 2002).

Hamilton and Hamilton (1997) in a study into the effect of alcohol consumption on earnings for males aged between 25 and 59 years in Canada using data from the 1985 General Social survey, define non drinkers as those who drink less often than once a month or not at all over the previous year. Moderate drinkers drink once a month, or everyday, but never consume eight or more drinks on a single day in the previous week. Heavy drinkers are those who drank at least once a week in the previous twelve months and drank eight or more drinks on one or more days in the previous week.

Barrett (2002) in analysing the effect of alcohol consumption on the earnings of males between 25 and 59 years of age in Australia using data from the Australian

National Health survey 1989-1990, defines non-drinkers as individuals who never drink or who did not have a drink in the month prior to the survey, a heavy drinker is defined as someone who drank eight or more standard drinks on at least one day during the reference week and moderate drinkers are anyone who had a drink in the last month and did not drink more than seven standard drinks on any given day during that period.

Kenkel and Ribar (1994) using data from the US National Longitude Survey of Youth define the threshold for heavy drinking, at 6 or more drinks on any one day. Lye and Hirschberg (2004), use the US National Alcohol Abuse and Alcoholism (NIAAA, 1997) definition of moderate drinking as no more than 2 standard drinks per day for a man aged 65 and under, and no more than one standard drink per day for men over the age of 65 and all women. An upper limit for men aged 65 years and older is selected because amounts of muscle tissue decrease with age and therefore the same dose of alcohol produces a higher blood alcohol level (Lye and Hirschberg, 2004). There is no upper limit set for women.

Zarkin *et al* (1998), in their study into the effect of alcohol on wages use data from the 1991 and 1992 sweeps of the US National Household Survey on Drug Abuse, categorise drinkers into one of eight categories. One category for non-drinkers, two for light drinkers, three for moderate drinkers and two for heavy drinkers. Non drinkers are the respondents who did not drink alcohol in the previous 30 days. For the other categories of drinkers, men and women are assessed differently.

Men are categorised as follows:

- Light drinkers — 5 drinks in past 30 days (up to 1 drink per week)
  6-16 drinks in past 30 days (from 1 drink p.w. to 1 drink every other day)
- Moderate drinkers — 17-31 in past 30 days (1 drink every other day up to 1 drink per day)
  32-62 in past 30 days (1 to 2 drinks per day)
  63-93 drinks in past 30 days (2-3 drinks per day)
- Heavy drinkers — 94-124 drinks in past 30 days (3-4 drinks per day)
  125 or more in past 30 days (4+ drinks per day)

For women these amounts were halved.

A US National Survey on Drug Use and Health (SAMHSA 2007) defines binge drinking as having five or more drinks on the same occasion (i.e. at the same time or within a couple of hours of each other) on at least one day in the past 30 days. Heavy use is defined as five or more drinks on the same occasion on each of 5 or more days in the previous 30 days.

Mullahy and Sindelar (1996) using the 1988 Alcohol Survey of National Health Interview Survey (NHIS) which is a stratified, multistage probability sample of the US population, look at the effect of problem drinking on employment and unemployment. Based on the information in their sample, they formulate the indicators of heavy drinking using the 90th and 95th percentile in the distribution of the observed ethanol consumption. They weight each beer, wine and spirits by the amount of ethanol typically found in each type of drink. The following apply: a 12 ounce glass of beer with 0.045 ethanol per ounce, a 4 ounce glass of wine with 0.129 ethanol per ounce, and one ounce of spirits per drink with 0.411 ethanol per ounce. The total amount of ethanol does not vary much across types of drinks using these assumptions. A standard drink of beer, wine, or spirits contains about one half of an ounce of ethanol (Mullahy and Sindelar, 1996).
Table 2.1.2 sets outs the different definitions of non, moderate and heavy drinkers, that are used in the different studies. This allows a comparison to be made between the different definitions.

Table 2.1.2. Categorisation of Non, Moderate and Heavy drinkers

	Non-Drinkers	**Light Drinkers**	**Moderate**	**Heavy drinkers**
Knupfer(1984)				8+ drinks per day one or more times per week.
Hamilton & Hamilton (1997)	less than once month, or not at all over previous in year		Drink once month, or everyday but never 8 or more on a single day in the week prior to the study.	Drank at least once a week in the previous 12 months & drank 8 or more drinks on one or more days in the previous week.
Barrett (2002)	Never drink or did not have a drink in the month prior to the survey		Did drink in last month but not more than 7 drinks on one occasion in reference period	8 or more drinks on at least one day in the a reference/ given week
Kenkel & Ribar (1994)				6 or more on any one day
Zarkin *et al* (1998) **(These amounts are for men, they are halved for women)**	Never had a drink or had no drink in 30 days prior to survey.	5-16 drinks in previous 30 days (up to 1 drink per week, or 1 drink every other day)	17-93 drinks in past 30 days (between 1 drink every other day and 3 drinks per day)	94 or more drinks in past 30 days (3+ drinks per day)

Table 2.1.2. Continued: Categorisation of Non, Moderate and Heavy drinkers

	Non-Drinkers	**Light Drinkers**	**Moderate**	**Heavy drinkers**
Lye & Hirschberg (2004)			Up to 2 drinks per day for men aged 65 or under Up to 1 per day for men and women aged over 65.	
SAMHSA (US, 2007)				5+ drinks on same occasion on each of 5 or more days in the previous 30 days

(Source: Authors own)

Once alcohol consumption is categorised, the effect of alcohol on income can then be investigated.

2.1.6: Estimation of Alcohol Status Equation

Previous studies such as Hamilton and Hamilton (1997) and Barrett (2002) in their estimation of the effect of alcohol on earnings, while accounting for selection bias, have used the two step model as proposed by Lee (1982, 1983). Step One requires the alcohol consumption equation to be estimated by a multinomial logit. This includes a range of socio demographic and personal characteristics which affect an individual’s level of alcohol consumption.

Step One

Many studies have been carried out into the factors that affect ones alcohol consumption (Moore *et al* , 2005; Blow *et al* ,2005; Hamilton & Hamilton, 1997; Auld, 2005; Barrett, 2002; Mullahy & Sindelar, 1996; Balsa and French, 2010) and find that many different factors influence the amount of alcohol one consumes such as gender, age, health among others, and these factors in turn result in a person more likely to be in one category of drinker over another (Hamilton and Hamilton, 1997; Barrett, 2002).

Age has an affect on the levels of alcohol consumed. Moore *et al* (2005), Blow *et al (*2005), Hamilton & Hamilton (1997), Auld (2005), Barrett (2002), Mullahy & Sindelar (1996), Balsa and French (2010) all had similar findings in that on average people drink less as they get older, and as a result are less likely to be heavy drinkers. Interestingly Moore *et al* (2005) in estimating the effects of age and other socio demographic influences on alcohol consumption, find that the decline in alcohol consumption with increasing age was smaller in more recent birth cohorts. Barrett (2002) shows that as men get older they are significantly more likely to be non-drinkers with individuals in the 45-54 age groups having the highest probability of abstaining.

Hamilton & Hamilton (1997) also looked at the age at which people started to consume alcohol. They find that those who started drinking before the age of 18years are more likely to be current heavy drinkers as opposed to being non-drinkers. Similarly some studies find that if an individual smoked at the age of 18 they are more likely to be a consumer of alcohol (Barrett, 2002; Moore *et al*, 2005). This is due to the fact that smoking is a health risk behaviour and reflects an individual's attitude towards risk (Hersch and Viscusi, 1990).

Mullahy & Sindelar (1996) argue that in looking at the drinking behaviour of individuals, different age groups behave very differently which needs to be accounted for. Mullahy & Sindelar (1996) find that young people have a greater than average prevalence for alcoholism. Similarly Moore *et al* (2005) shows that there is a steeper age related decrease in alcohol consumption among men, non-whites, respondents who were married, respondents with less education and smokers and that heavy drinkers tend to reduce their drinking faster than light to moderate drinkers.

Many studies find that in looking at men and women of the same age, men consume greater amounts of alcohol than women and that alcohol abuse/dependence is roughly three times more prevalent among males than females (Fillmore 1994; Moore, 2005; Blow *et al*, 2005; Moore *et al*, 2005; Mullahy & Sindelar, 1996). Similarly in their study into the labour market consequences of heavy and abusive drinking in Uruguay, Balsa and French (2010) observe that men were more likely to drink heavily or to intoxication.

Education plays an important role in a person's level of alcohol consumption (Hamilton & Hamilton, 1997; Barrett, 2002; Balsa and French, 2010; Su and Yen, 2000; Van Ours, 2004). Findings tend to show that higher educated people particularly those with third level degrees, tend to consume moderate amounts of alcohol and are less likely to abstain or be heavy drinkers (Hamilton & Hamilton, 1997). Barrett (2002) discovers this to be the case however observes that for those who didn't attend university, education is not strongly linked to drinker type compared to those who did attend. Van Ours (2004) in looking at the effect of smoking and alcohol consumption on the wages of males in The Netherlands, finds higher education to have a positive impact on alcohol use for men and a much greater impact for women. Su and Yen (2000) discover that in the United States higher education leads to people consuming more wine and less beer and argue that this could be due to the fact that better education may bring more social occasions for wine drinking than casual beer occasions. Contrary to these findings, Balsa and French (2010) find that in Uruguay those with primary education were more likely to drink heavily.

Race can also have an impact on ones level of alcohol consumption, with findings showing that white people tend to consume greater amounts of alcohol, while those who abstain from alcohol tend less often to be white (Mullahy & Sindelar, 1996; Moore *et al.* 2005). Su and Yen (2000) on the other hand find that in the US black people tend to consume more beer.

Being married tends to result in people being less likely to be heavy drinkers and more likely to be moderate drinkers (Barrett, 2002; Auld, 2005; Hamilton and Hamilton, 1997). Van Ours (2004) find that having a partner or children does not have a significant impact on alcohol consumption.

In terms of occupation, Auld (2005) and Barrett (2002) discover that professionals, who work in management and those who work in the service industry are less likely to be abstainers or heavy drinkers. Barrett (2002) also shows that public sector employees are significantly less likely to be heavy drinkers compared with their private sector counterparts.

Generally findings show that price has an impact on the demand for alcohol in so far as higher prices tend to mean people consume less resulting in people more likely to be moderate and less likely to be heavy consumers of alcohol (Auld, 2005; Hamilton and Hamilton, 1997, Department of Health, 2012). McGuiness (1980) finds that in looking at total demand for alcoholic beverages in the UK 1956-1975, that price does have an impact on the demand for alcohol. In 1975 if the price of alcoholic drinks had been 1% higher in real terms, consumption would have been reduced by two thirds of a fluid ounce of alcohol on average for every adult. This change was however accompanied by a diversion of more expenditure to the purchase of alcohol every adult spending on average, an additional 86 pence at 1975 values (McGuiness, 1980).

Hamilton & Hamilton (1997) show that the region or area where a person is from, can also affect whether or not a person is a non, moderate or heavy drinker. The reason given for this is that people's behaviours around the consumption of alcohol are generally based on social influences and what happens in the community around them (Last 1998; Cook and Moore 1999, 2000).

Similar to these findings Barrett (2002) observes that abstainers are more likely to reside in an area where there is a slightly above average fraction of abstainers, similarly moderate drinkers tend to be located in areas where there are more moderate drinkers and likewise with heavy drinkers, they tend to be located in an area with above average proportion of heavy drinkers.

Su and Yen (2000) in their study of alcohol consumption in the US also find that the region where people came from had an impact on their levels of alcohol consumption and that individuals from the Midwest and South consumed less beer and wine than those from other areas. Su and Yen (2000) also find that people living in urban areas were more likely to consume more than those in rural areas. Wang *et al* (1996) observe that households in the South US are less likely to drink alcohol.

Religion can influence how people view alcohol and the levels they consume, with findings being that Catholics seem to behave differently to other religions (Hamilton & Hamilton, 1997; Auld, 2005). Religious attendance seems to result in people being more likely to be a non-drinker versus a moderate or heavy drinker, but that Roman Catholics are different in so far as they have a higher propensity to be a moderate or heavy drinker and those with no religious faith are more likely to be heavy drinkers (Hamilton & Hamilton (1997). Auld (2005) in looking at the effect of alcohol consumption on the wages of males aged between 25 and 59 years in Canada, had similar findings in so far as Catholics tend to drink more and Non-Catholic religious individuals are more likely to abstain from drinking. Van Ours (2004) on the other hand finds that religion does not have a significant impact on alcohol consumption.

Some studies show that a correlation between whether an individual smoked at the age of 18 years and their current alcohol consumption can exist (Barrett, 2002; Moore *et al,* 2005). This measure is viewed as a retrospective measure of an individual's attitude towards risk, the rationale being that smoking is a health risk behaviour and in part reflects an individual's attitude toward risk (Hersch and Viscusi, 1990). Hersch and Viscusi (1990) use contemporaneous smoking behaviour as a proxy for individuals' attitudes towards risk in estimating wage differentials for risk of lost work-day injury. Barrett (2002) looks at smoking in the past as opposed to current smoking because the retrospective measure of smoking is not likely to influence current income however current smoking behaviour is likely to affect current income.

Income affects ones alcohol consumption in that generally people with higher incomes consume more alcohol than those on lower incomes (Britain, 2001). Gallet (2007) in an analysis of 24 countries finds that income elasticity for all alcohol beverages is 0.50, meaning that a 1% increase in consumers incomes leads to a 0.5% increase in alcohol consumption. In Ireland alcohol consumption has a high income elasticity of demand (Davies and Walsh, 1983). It is argued that those who spend a large proportion of their income on alcohol may be more sensitive to price changes (UK, 2008).

Hamilton and Hamilton (1997) and Barrett (2002) in estimating the effect of alcohol on earnings, estimate the drinking status equation along with the earnings equation in order to account for selection bias. The drinking status equation is estimated by multinomial logit, from which the inverse mills ratio can be derived. This is then included as an additional variable in the earnings equation. By including all the variables from the earnings equation, in the alcohol status equation, earnings are accounted for (Hamilton and Hamilton, 1997; Barrett, 2002).

2.1.6.1: The Multinomial Logit Model

The Multinomial Logit model is based on the framework of the random utility model which means that given a choice of alternatives, in this case alcohol consumption, utility is determined by a number of different factors, some of which are specific to the individual and have nothing do with the nature of the choice and some factors specific to the choice and have nothing to do with the individual (Borooah, V, 2001). Multinomial Logit models are conditional which means that the choices between alternatives may depend not just upon the characteristics of the individual making the choice but also upon the attributes of the choice (Borooah, V, 2001).

The Multinomial Logit assumes that a person chooses the quantity of alcohol they consume by comparing the indirect utility or satisfaction provided by each path and chooses the category that provides the maximum utility (Barrett, 2002).

The probability of expressing each potential outcome can be easily expressed and the resulting log-likelihood function can be maximised in a straight forward fashion (Hilmer, 2001). In the multinomial logit choice model, one alternative is selected as the base alternative and other possible choices are then compared to this base alternative with a logit equation (Studenmund, 2005). The Multinomial logit controls for the choice of drinking status and explicitly addresses the endogeneity bias arising through selection bias (Barrett, 2002). Both Hamilton and Hamilton (1997) and Barrett (2002) in their studies include all the variables that determine earnings, in the drinking choice equation as this controls for the effect of earnings on drinking behaviour.

Regressions of earnings on dummy variables indexing frequency of use, do not take account of the possibility that returns to human capital characteristics may vary by drinking status (Hamilton and Hamilton, 1997).

By dividing drinkers into three categories and running separate earnings regressions for each category it is possible to look at the returns to human capital characteristics by drinking category and the consequences of drinking on earnings can be identified (Hamilton and Hamilton, 1997; Barrett, 2002).

Hamilton and Hamilton (1997) and Barrett (2002) state that by estimating the multinomial logit model for alcohol consumption which controls for an individual's self-selection of drinking status, predicted values for the inverse mills ratio are generated which are then inserted into the earnings equations and estimated by OLS regression.

The basis of the Multinomial Logit model in such an estimation of the effect of alcohol on an individual's financial welfare as set out by Hamilton and Hamilton (1997) and Barrett (2002) is that individuals are assumed to select the earnings-drinking status combination that maximises their expected utility. The i^{th} individuals expected utility from an earnings-drinking status combination is modelled by the index function.

$$U_{ij} = z_i \gamma_j + \eta_{ij} \tag{2.1.2}$$

Where:	U	expected utility
	z	vector containing exogenous variables affecting earnings or alcohol consumption
	γ	vector of unknown utility parameters
	η	error term
	i	indexes individuals
	j	indexes drinking status where $j = 1,2,3$

The error terms η_{ij} and u_{ij} in the income equation, represent the impact of unobserved variables on utility levels. The vector z_i contains exogenous variables hypothesised to affect either an individual's earnings or preference for alcohol consumption and thus includes X_i which is a vector of human capital and socio demographic variables that affect earnings. It is not observed directly but an indicator for each individual's choice of drinking status is observed which is denoted by I_i (Hamilton and Hamilton, 1997; Barrett, 2002).

If individual i chooses sector j then

$$I_i = j \quad \text{if} \quad U_{ij} > MaxU_{is} \tag{2.1.3}$$

Where:	I_i	an indicator for each individuals choice of drinking status.
	U_{ij}	utility of individual i receives from consuming alcohol status j
	j	indexes drinking status where $j = 1,2,3$
	s	$1,2,3, s \neq j$

Hamilton and Hamilton (1997) and Barrett (2002) following the formulation of Lee (1983), define the residual for each individual and sector as

$$\varepsilon_{ij} = MaxU_{is} - \eta_{ij} \qquad \textbf{(2.1.4)}$$

Where: ε_{ij} residual for each individual and sector

U_{is} utility of individual i receives from consuming alcohol status s

η_{ij} error term

s $=1,2,3, s \neq j$

then $I_i = j$ if

$$\varepsilon_{ij} < z_i \gamma_j \qquad \textbf{(2.1.5)}$$

Where: I indicator for each individuals choice of drinking status.

ε_{ij} residual for each individual and sector

z vector containing exogenous variables affecting earnings or alcohol consumption

γ vector of unknown utility parameters

i indexes individuals

j indexes drinking status

So that $\ln Y_{ij}$ is observed if and only if $\varepsilon_{ij} < z_i \gamma_j$. Assuming that the $\varepsilon_{ij} (j = 1,2,3)$ error terms are independently and identically Gumbel distributed with the type I generalised extreme value distribution (Bali, 2003, Hamilton and Hamilton, 1997; Barrett, 2002), then equations 2.1.2 through 2.1.5 define a standard multinomial logit model setting out the probability that an individual is likely to be in a particular drinking category.

$$\Pr(I_i = j) = \frac{\exp(z_i \gamma_i)}{\sum_{s=1}^{3} \exp(z_i \gamma_s)} \qquad \textbf{(2.1.6)}$$

Where:	z	vector containing exogenous variables affecting earnings or alcohol consumption
	γ	vector of unknown utility parameters
	I	an indicator for each individuals choice of drinking status.
	i	indexes individuals
	j	indexes drinking status
	s	$s = 1,2,3 \quad s \neq j$

Consider the transformation to normality of the form

$$\varepsilon_{ij}^{\ *} = \phi^{-1}(F_j(\varepsilon)) \qquad \textbf{(2.1.7)}$$

Where:	ε_{ij}	the residual for each individual and sector
	ϕ	the standard normal cumulative distribution function of the standard normal
	i	indexes individuals
	j	indexes drinking status

then $\varepsilon_{ij} < z_i \gamma_j$ if

$$\varepsilon_{ij}^{\ *} < \phi^{-1}(F(z_i \gamma_j)) \qquad \textbf{(2.1.8)}$$

The above multinomial logit can be used to derive the correct earnings specifications which account for selectivity bias, which are estimated in Step two, similar to the method employed in previous studies (Hamilton and Hamilton, 1997; Barrett, 2002). Self Selection implies conditional earnings equations as shown in equation 2.1.13.

2.1.6.2: Assumptions of the Multinomial Logit Model

The multinomial logit makes an assumption known as the independence of irrelevant alternatives (IIA) (Hausman and McFadden, 1984; Small and Hsiao, 1985). The IIA property states that the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set (Hausman and McFadden, 1984; Small and Hsiao, 1985). In essence this means that the relative probability of two existing outcomes is unrelated to the addition or drop of a third outcome, that is, alternative outcomes are irrelevant (Long and Freese, 2005). In terms of drinking categories, the IIA Assumption means that if there are two drinking categories one can choose from, adding another drinking category will not affect the odds of choosing one of the initial categories (Hamilton and Hamilton, 1997; Barrett, 2002)

There are various tests that can be carried out to test the IIA Assumption which involve comparing the estimates from the null model to those from the restricted estimation (Cheng and Long, 2007). Two such tests that can be carried out are the Hausman tests or the Small and Hsiao tests (Long and Freese, 2005).

The Hausman test proposed by Hausman and McFadden (1984) involves estimating the full model with all j outcomes included, with estimates of $\hat{\beta}_F$ (Long and Freese, 2005). A restricted model is then estimated by eliminating one or more outcome categories, with estimates in $\hat{\beta}_R$. The third step involves letting $\hat{\beta}_F^*$ be a subset of $\hat{\beta}_F$ after eliminating coefficients not estimated in the restricted model (Long and Freese, 2005). The test statistic is

$$H = \left(\hat{\beta}_R - \hat{\beta}_F^*\right)' \left[\hat{\mathrm{var}}\left(\hat{\beta}_R\right) - V\hat{a}r\left(\hat{\beta}_F^*\right)\right]^{-1} \left(\hat{\beta}_R - \hat{\beta}_F^*\right) \tag{2.1.9}$$

Where: $\hat{\beta}_R$ = estimates of the restricted model

$\hat{\beta}_F^*$ = estimates of the full model with all j outcomes included

H is asymptotically distributed as chi-squared with degrees of freedom equal to the rows in $\hat{\beta}_R$ if IIA is true. Significant values of H indicate that the Independence of Irrelevant Alternatives (IIA) assumption has been violated (Hausman and McFadden, 1984; Long and Freese, 2005).

With regard to the Small and Hsiao (Small and Hsiao, 1985) test of IIA, the sample is divided into two subsamples of about equal size. The unrestricted Multinomial Logit Model is estimated on both subsamples, where $\hat{\beta}_u^{S1}$ contains estimates from the unrestricted model on the first subsample and $\hat{\beta}_u^{S2}$ it's counterpart for the second subsample (Long and Freese, 2005). A weighted average of the coefficient is computed as

$$\hat{\beta}_u^{S1S2} = \left(\frac{1}{\sqrt{2}}\right)\hat{\beta}_u^{S1} + \left[1 - \left(\frac{1}{\sqrt{2}}\right)\right]\hat{\beta}_u^{S2} \qquad \textbf{(2.1.10)}$$

Where: $\hat{\beta}_u^{S1}$ Estimates from the first subsample

$\hat{\beta}_u^{S2}$ Estimates from the second subsample

A restricted sample is then created from the second subsample by eliminating all cases with a chosen value of the dependent variable. The multinomial logit is estimated using the restricted sample yielding the estimates $\hat{\beta}_r^{S2}$ and the likelihood $L\left(\hat{\beta}_r^{S2}\right)$ (Long and Freese, 2005). The Small and Hsiao statistic is

$$SH = -2\left[L\left(\hat{\beta}_u^{S1S2}\right) - L\left(\hat{\beta}_r^{S2}\right)\right] \qquad \textbf{(2.1.11)}$$

Where: $\hat{\beta}_r^{S2}$ Multinomial logit estimates of the restricted sample

$L\left(\hat{\beta}_r^{S2}\right)$ Likelihood of the Multinomial logit estimates of the restricted sample which is asymptotically distributed as a chi squared with the degrees of freedom equal to K+1, where K is the number of independent variables (Long and Freese, 2005).

Results of the Hausman Tests and Small Hsiao tests are typically inconclusive or contradictory (Long and Freese, 2005).The suest-based Hausman test which is a modification of the Hausman and McFadden test, is a robust procedure implemented in Stata to deal with the issues raised by Long and Freese (Siegel and Lucke, 2009). Long and Freese (2005) recommend the suest-based Hausman test for testing the IIA assumption.

The Hausman test via Suest is comparable to that computed by Hausman, but they use different estimators of the variance of the different estimates (Stata, 2013) Hausman estimates V(b-B) by V(b) -V(B), whereas Suest estimates V(b-B) by V(b) – Cov (b,B) – Cov (B,b) + V(B).

Yan *et al* (2011) employed the suest based Hausman test to test the IIA assumption in their study into motor vehicle-bicycle crashes in Beijing, which they state specifically measures that if one alternative is removed, the before and after estimators under the null hypothesis are consistent.

2.1.7: Estimation of the Income Equation

Given the completion of step one and derivation of the inverse mills ratio, Hamilton and Hamilton (1997) and Barrett (2002) estimate the earnings equation in step two of the model.

Step Two

Assume the potential earnings for individual i with drinking status j is given by equation 2.1.12. Earnings for each individual is hypothesised to depend upon a vector X_i of human capital variables and sociodemographic characteristics and Y_{ij} is observed only if drinking status j is chosen (Hamilton and Hamilton, 1997; Barrett, 2002).

$$\ln Y_{ij} = X_i \beta_j + u_{ij} \qquad \textbf{(2.1.12)}$$

Where:	$\ln Y$	log of earnings
	X	vector of human capital variables & socio-demographic characteristics
	β	coefficients on the observable characteristics
	u	error term
	i	indexes individuals where $i = 1,2,......N$
	j	indexes drinking status where $j = 1,2,3,$

This specification allows labour market returns to individual characteristics to differ by drinking status. By comparing the estimated $\beta's$ across drinker type it is possible to gauge whether the earnings given ones socio-demographic characteristics is greatest for one category of drinker over another (Barrett, 2002).

Self Selection implies conditional earnings as shown in equation 2.1.13.

$$E(\ln Y_{ij} | I_i = j) = X_{ij}\beta_j + E(u_{ij} | \varepsilon^*_{ij} < \phi^{-1}[F(z_i\gamma_j)]) \qquad \textbf{(2.1.13)}$$

Where:	$\ln Y$	log of earnings
	I	indicator for each individuals choice of drinking status.
	X	vector of human capital variables & sociodemographic characteristics
	β	coefficients on the observable characteristics
	ε_{ij}	residual for each individual and sector
	ϕ	standard normal cumulative distribution function of the standard normal
	z	vector containing exogenous variables affecting earnings or alcohol consumption
	γ	vector of unknown utility parameters
	u	error term
	i	indexes individuals
	j	indexes drinking status

Thus $u_{ij}'s$ can be characterised as following a truncated normal distribution which can be accounted for using the standard Heckman selection correction technique (Hamilton and Hamilton, 1997). The earnings equations are estimated using an extension of the generalised two step procedure presented in Lee (1982, 1983). The appropriate specification of the earnings equation conditional on alternative j being chosen is;

$$\ln Y_{ij} = X_{ij}\beta j - \sigma_j p_j \frac{\phi[\Phi^{-1}(F_j(z_i\gamma_j))]}{F_j(z_i\gamma_j)} + v_{ij} \qquad \textbf{(2.1.14)}$$

Where:	$\ln Y$	log of earnings
	X	vector of human capital variables & socio-demographic characteristics
	β	coefficients on the observable characteristics

σ variance of the error term η_{ij}

p correlation coefficient between the unobservables in the earnings and selection equations.

z vector containing exogenous variables affecting earnings or alcohol consumption

γ vector of unknown utility parameters

ϕ probability density function (pdf) of the standard univariate normal distribution respectively.

Φ cumulative distribution function (cdf) of the standard univariate normal distribution respectively.

v error term which has a zero mean and in uncorrelated with μ

i indexes individuals

j indexes drinking status

Estimates from equation 2.1.14 provide information on the expected earnings if an individual were randomly allocated to a given drinking status, as well as predicted income given an individual is a particular drinker type (Hamilton and Hamilton, 1997).

The second term on the right hand side of equation 2.1.14 controls for the truncated mean of the observed residual in the earnings equations arising from individuals selecting their preferred drinking status (Hamilton and Hamilton, 1997). The truncated mean is a generalisation of the Heckman correction term (Inverse Mills Ratio) to the situation where individuals choose over multiple alternatives (Hamilton and Hamilton, 1997).

Using the Multinomial Logit OLS Two Step model as proposed by Lee (1983) to estimate the effect of alcohol consumption on earnings, ensures that selection bias is accounted for (Hamilton and Hamilton, 1997; Barrett, 2002). By estimating separate earnings regressions for each category of drinker endogeneity bias is accounted for (Hamilton and Hamilton, 1997; Barrett, 2002).

2.1.8 Decomposition of Wage Differentials

An often used methodology to study labour market outcomes by groups is to decompose mean differences in log of wages known as the Blinder-Oaxaca decomposition (Jann, 2008). This Blinder-Oaxaca decomposition is a standard technique used to divide the wage differential between two groups into a part that is explained by differences in observable characteristics and a residual that cannot be explained by differences in characteristics (Jann, 2008; Pearlman and Tsao, 2008).

The explained part represents the part of the wage gap that is attributable to differences in group characteristics, that is the differences in wages that exists between groups if all groups had the same characteristics (Jann, 2008). The unexplained part is often used as a measure for discrimination, but it also subsumes the effects of group differences in unobserved predictors (Jann, 2008).

Hamilton and Hamilton (1997) in using the Oaxaca decomposition in their study, state that the unconditional earnings differential, measures the difference in earnings between two workers who have observable characteristics identical to the average person of each drinker type. The earnings differential is unconditional in that the predicted earnings are calculated independently of the workers actual choice of drinking status and hence the earnings differences are independent of selection effects. Hamilton and Hamilton (1997) state that the unexplained term is a pure wage differential and shows whether the returns to a representative set of observed traits vary by drinking status. It measures the differences in household income if observable characteristics are constant (Hamilton and Hamilton, 1997).

The Oaxaca method set out by Barrett (2002) in his study into the effects of alcohol consumption on wages as follows:

$$E\left[\ln y_j \middle| \bar{x}_j\right] - E\left[\ln y_k \middle| \bar{x}_k\right] = \left(\bar{x}_j - \bar{x}_k\right)\left(\frac{\beta_j + \beta_k}{2}\right) + \left(\beta_j - \beta_k\right)\left(\frac{\bar{x}_j + \bar{x}_k}{2}\right) \qquad \textbf{(2.1.15)}$$

Where:	$\ln Y_{ij}$	log of household income
	X	vector of human capital variables & socio-demographic characteristics
	$\left(\bar{x}_j - \bar{x}_k\right)\left(\frac{\beta_j + \beta_k}{2}\right)$	equals the wage gap attributable to differences in characteristics across drinking categories.
	$\left(\beta_j - \beta_k\right)\left(\frac{\bar{x}_j + \bar{x}_k}{2}\right)$	equals differences in productivity in status j versus status k drinkers
	j	indexes drinking status
	$k = 1,2,3$	$k \neq j$

The first term on the right hand side represents the part of the wage gap attributable to the differences in characteristics across drinking categories and is the explained part of the differential (Barrett, 2002).

The second term on the right hand side represents the component of the wage gap due to differences in coefficients and is the unexplained part of the differential (Barrett, 2002). This part tests whether the returns to a representative set of observed traits differ by drinking status and captures the effect of alcohol consumption on household income if observable characteristics are held constant (Hamilton and Hamilton, 1997).

2.1.9: Testing the relevance of instruments and post estimation tests

The significance of each of the instruments can be tested using a Wald test, which calculates a Z Statistic, which is then squared, yielding a Wald Statistic with a chi-squared distribution and will correspond to a two tailed P Value (Agresti, 1996). The Likelihood Ratio Test is another test which can be used to test the significance of coefficients (Gujarati, 2004). The likelihood-ratio test uses the ratio of the maximised value of the likelihood function for the full model over the maximised value of the likelihood function for the simpler model, the full model being that with an additional one or more parameters. The log transformation of the likelihood function yields a chi-squared statistic (Gujarati, 2004). The t and z statistics test whether a given coefficient is significantly different from zero (Gujarati, 2004).

Heteroskedasticity causes standard errors to be biased. OLS assumes that errors are both independently and identically distributed and that the variance of the error term is constant. If heteroskedasticity is present it would lead to bias in test statistics and confidence intervals (Berry and Feldman, 1985). The presence of heteroskedasticity can be tested using the Breusch Pagan test which tests the null hypothesis that the error variances are all equal (Berry and Feldman, 1985). Whites' general test for heteroskedasticity, which is a special case of the Breusch-Pagan test can also be used (Greene, 2000). This tests the error distribution by regressing the squared residuals on all distinct regressors, cross-products, and squares of regressors (Greene, 2000). A possible solution would be to use robust standard errors when heteroskedasticity is present as these relax the assumptions that the errors are both independent and identically distributed, hence robust standard errors tend to be more trustworthy (Berry and Feldman, 1985).

Multicollinearity arises when two or more predictor variables in a model are highly correlated and could cause coefficient estimates of particular variables to be to be incorrect.

2.1.10: Conclusion

This section reviews the literature on the effect of alcohol on income, and in looking at this relationship the issue of endogeneity and selection bias are reviewed in detail and the possible methods of estimation that can be used to account for these. Endogeneity arises when an explanatory variable such as alcohol is determined within the context of the model. Selection bias arises when an individual selects into different categories of drinking resulting in the sample not being random. The multinomial logit OLS two step estimate as proposed by Lee (1982) estimates alcohol consumption as a multinomial logit, from which the Inverse Mills Ratio is derived and included in the income equation. Separate income equations are estimated for each category of drinker. This method of estimation treats the sector choice as endogenous and accounts for selection bias. The different factors that affect both alcohol consumption and income are also assessed. Similar to previous studies alcohol is assumed to be unordered (Hamilton and Hamilton, 1997; Barrett, 2002) and hence is estimated using a multinomial logit model. Alcohol consumption could however be viewed as ordered data and hence should be estimated as such (Harris *et al,* 2006).

2.2: Alternative Methods of Estimation

Previous studies into the effect of alcohol consumption on an individuals financial welfare such as Hamilton and Hamilton (1997) and Barrett (2002) among others, have assumed that alcohol status is unordered and hence have estimated the alcohol status equation using the multinomial logit model. Alcohol consumption could however be viewed as ordered data (Harris *et al,* 2006). If ordinality is ignored then this may lead to a loss of efficiency and an increased risk of getting insignificant results (Harris *et al*, 2006).

2.2.1: Definition and Estimation of Ordered Data

Ordered data is where the variable of interest follows a strict ordering based on the value of the latent variable (Hilmer, 2001). Some polychotomous dependent variables are in a natural order and are expressed in terms of categories (Kennedy, 2003). Measurement through the use of ordered categories is a common practice in marketing and behavioural sciences (Kennedy, 2003). Ordered data avoids a false sense of precision that continuous scales convey (Sprinivasan and Basu, 1989).

Failure to account for the ordinal nature of the dependent variable can result in incorrect results (Greene, 2002). If a dependent variable is ordered, but the ordinality is ignored then this may lead to a loss of efficiency and an increased risk of getting insignificant results (Harris *et al*, 2006). If data is ordered, estimating the data by a multinomial logit or probit model would not be efficient because no account would be taken of the extra information of the ordinal nature of the dependent variable, nor would OLS be appropriate because the coding of the dependent variable reflects only a ranking, the difference between 1 and 2 cannot be treated as equivalent to the difference between a 2 and a 3 (Kennedy, 2003).

An ordered probit model is an econometric model that can be used to deal with ordered categorical variables and is designed to model a discrete dependent variable that has ordered multinomial outcomes (Jones 2005). An ordered probit model can be expressed in terms of an underlying latent variable y* (Jones 2005). The ordered probit assumes that the variable of interest follows a strict ordering based on the value of the latent variable (Hilmer 2001). The ordered probit and logit models have come into fairly wide use as a framework for analysing such responses (Zavoina and McElvey, 1975). Hilmer (2001) states that the estimated thresholds in the ordered probit model should always be significant and if not, then one could conclude that the assumed natural ordering and consequently the ordered probit is an inappropriate specification. A primary difference between the multinomial logit and ordered probit is that due to the assumed natural ordering the latter does not require the Independence of Irrelevant Alternatives (IIA) property, however for the model to be appropriate, the assumed natural ordering must be realistic (Hilmer, 2001).

Wooldridge (2009) says that the ordered probit and logit models have come into fairly wide use as a framework for analysing such responses. The model is built around a latent regression in the same manner as the binomial probit model.

$$y^* = x'\beta + \varepsilon \qquad \textbf{(2.2.1)}$$

Where:

Where:	y	dependent variable
	x	independent variable
	β	coefficient
	ε	error term

y^* is unobserved but what is observed is

$$
\begin{aligned}
y &= 0 \text{ if } y^* \le 0 \\
y &= 1 \text{ if } 0 < y^* \le \mu_1 \\
y &= 2 \text{ if } \mu_1 < y^* \le \mu_2 \\
&. \\
&. \\
&. \\
&= J \text{ if } \mu_{J-1} \le y^*
\end{aligned}
$$

Where: y dependent variable

μ_J known cutoffs

In this equation where $y^* \leq 0$, these respondents are in category 0. Where y^* is greater than 0 but less than μ_1 category 1 is observed and where y^* is greater μ_1 but less than μ_2 category 2 is observed.

In order to address the issue of selection bias when data is of an ordered nature, various extensions of the Heckman two step model have been adopted (Greene and Hensher, 2010). A variety of extensions to the Heckman model (1979) have been developed for ordered choice models, one being to use an ordered probit extension of the Heckman correction (Vella, 1998; Greene and Hensher, 2010). This is where the selection equation is estimated using an ordered probit model, from which an estimate of lambda is computed for each individual in the selected sample. This is then included as an additional regressor in the outcome equation (Vella, 1998).

Many studies have adopted this approach whereby the selection equation is estimated as an ordered probit which allows the inverse mills ratio to be derived (Garen, 1984; Butler *et al*, 1994, 1998; Frazis, 1993; Jimenez and Kugler, 1987; Harmon and Walker,1995; Chiburis and Lokshin, 2007). Langpap and Kerkvliet (2002) in their study into whether the endangered species act in the US has been successful in promoting species recovery, estimated an probit in the first step, from which the inverse mills ratio is derived and the second step is then estimated as an ordered probit.

Chiburis and Lokshin (2007) in their study into the estimation of wages for public, private and informal sectors for male workers in India use an ordered probit selection model. The categorical variable describing the sector individuals work in, is estimated as an ordered probit on the basis of an ordered probit selection rule. They set out the model specification as follows;

Step One – Estimation of the Selection Equation

$$c_i^* = \alpha' s_i + \varepsilon_i \qquad \textbf{(2.2.2)}$$

$c_i = 1$	if	$-\infty < c_i \le \mu_1$
$c_i = 2$	if	$\mu_1 < c_i \le \mu_2$
$c_i = 3$	if	$\mu_2 < c_i \le \infty$

Where:	c	sector category
	α	unknown vector of parameters,
	s	independent variables
	ε	standard normal shock
	μ_J	cutoffs
	i	indexes individuals

The category an individual is in depends on a range of independent variables s (Chiburis and Lokshin, 2007). It is assumed that the independent variables s_i and the categorical variables c_i are observed. It is important that the ordered probit selection model contains a variable that is not an independent variable in the income equation (Chiburis and Lokshin, 2007). There must be at least one instrument in the selection variable s that has no effect on y except through its effect on c. If all the variables in the selection equation are also in the wages equation, then the identification of the coefficient β_j would be weak (Chiburis and Lokshin, 2007). This is due to the fact that additional variables in the first step selection equation are important for identification of the second step estimates which would inflate second step standard errors and unreliable estimates of coefficients (Vella, 1998).

In the first step the selection equation is estimated by an ordered probit of c on s, yielding the consistent estimates $\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, ...\hat{\mu}_J$ (Chiburis and Lokshin, 2007). The probability of observing c=1,2,3 is defined as follows:

$$\begin{aligned}&\Pr(c=1|s)=\Pr(s'\alpha+\varepsilon\le\mu_1)\\&=\Pr(\varepsilon\le\mu_1-s'\alpha)\\&=\Phi(\mu_1-s'\alpha)=1-\Phi(s'\alpha-\mu_1)\end{aligned}$$

$$\begin{aligned}&\Pr(c=2|s)=\Pr(\mu_1<s'\alpha+\varepsilon\le\mu_2)\\&=\Pr(\varepsilon>\mu_1-s'\alpha,\varepsilon\le\mu_2-s'\alpha)\\&=[1-\Phi(\mu_1-s'\alpha)]-\Phi(s'\alpha-\mu_2)\\&=1-[1-\Phi(s'\alpha-\mu_1)]-\Phi(s'\alpha-\mu_2)\\&=\Phi(s'\alpha-\mu_1)-\Phi(s'\alpha-\mu_2)\end{aligned}$$

$$\begin{aligned}&\Pr(c=3|s)=\Pr(s'\alpha+\varepsilon\le\mu_2)\\&=\Pr(\varepsilon\le\mu_2-s'\alpha)\\&=\Phi(\mu_2-s'\alpha)=1-\Phi(s'\alpha-\mu_2)\end{aligned} \qquad \textbf{(2.2.3)}$$

where $\Phi(\cdot)$ is the cumulative normal function.

Defining $\hat{c}_i^* \equiv \hat{s}'\alpha_i$ as a consistent estimator of the Inverse Mills ratio correction term, λ_i can be obtained from the ordered probit equations (Chiburis and Lokshin, 2007; Jimenez and Kugler, 1987; Hamilton and Nickerson, 2001).

$$\hat{\lambda}_i \equiv \frac{\phi(\hat{\mu}_j-\hat{c}_i^*)-\phi(\hat{\mu}_{j+1}-\hat{c}_i^*)}{\Phi(\hat{\mu}_{j+1}-\hat{c}_i^*)-\Phi(\hat{\mu}_j-\hat{c}_i^*)} \qquad \textbf{(2.2.4)}$$

Where:	c	sector category
	j	indexes sector category where $j=c_i$
	μ	cutoffs
	ϕ	probability density function
	Φ	cumulative distribution function

λ_i is included as an omitted variable in the OLS equation estimated in step 2.

Step 2 – Estimation of the Wages Equation

In the second step if the two step estimation, Chiburis and Lokshin (2007) estimate the wages equation while including the selection correction term as an additional regressor.

$$\ln Y_{ij} = X_i \beta_j + u_{ij} \tag{2.2.5}$$

Where:	$\ln Y$	log of wages
	X	vector of human capital variables & socio-demographic characteristics
	β	coefficients on the observable characteristics
	u	error term
	i	indexes individuals where $i = 1,2,......N$
	j	indexes sector category where $j = 1,2,3,$

Chiburis and Lokshin (2007) state that the observed dependent variables y_i is a linear function of some observed independent variables x_i, but the coefficients depend on category c_i

$$y_i = \begin{cases} \beta_1' x_i + u_{i1} & & c_i = 1 \\ \beta_2' x_i + u_{i2} & & c_i = 2 \\ \beta_3' x_i + u_{i3} & & c_i = 3 \\ . & \text{if} & \\ . & & \\ . & & \\ \beta_J' x_i + u_{iJ} & & c_i = J \end{cases}$$

(2.2.6)

Where:	y_i	dependent variable wages for individual i
	x_i	independent variables for individual i
	c_i	sector category for individual i
	u_i	error term
	$j \in \{0,.....,J\}$	

u_{ij} has a mean of 0, has a variance of σ_j^2, and is bivariate normal with ε_i with correlation p_j. It is assumed that the shocks u_{ij} and ε_{ij} are independently and identically distributed across all observations (Chiburis and Lokshin, 2007)

Since only one sector category is observed for each individual and the observations are independent, the correlations between u_{ij} and u_{ik} for $j \neq k$ cannot be identified.

β_j can be consistently estimated with an OLS regression of y on x and $\hat{\lambda}$ by using only the observations i for which $c_i = j$.

$$E[y_i | c_i, s_i, x_i] = \beta_j' x_i + E[u_{ij} | c_i = j, s_i] = \beta_j' x_i + p_j \sigma_j \lambda_i' \qquad \textbf{(2.2.7)}$$

Where:	y	dependent variable wages
	c	sector category
	s	independent variables in selection equation
	x	independent variables in the wages equation
	β	coefficient on observable characteristics in wage equation
	u	error term
	j	indexes sector category
	p	the correlation coefficient between the unobservables in the income and selection equations.
	σ	the standard deviation of the error term
	λ	selection correction term

When regressing y on x over the subsample $\{i : c_i = j\}$, by adding λ as an extra regressor then the estimate of $\hat{\beta}_j$ will be consistent compared with regressing y on x using an OLS regression (Chiburis and Lokshin, 2007; Hamilton and Hamilton, 1997; Barrett, 2002).

By estimating a selection equation as an ordered probit, the ordinality of the variable is accounted for (Harris *et al*, 2006). By deriving the selection correction term, the inverse mills ratio, and including this as an additional regressor in the primary equation, selection bias is accounted for (Vella, 1998; Greene and Hensher, 2010). Possible endogeneity is accounted for by running separate earnings regressions for each category separately (Hamilton and Hamilton, 1997; Barrett, 2002).

2.2.2: Conclusion

Previous studies into the effect of alcohol consumption on income have treated alcohol consumption as a polychotomous choice and used the multinomial logit method of estimation (Hamilton and Hamilton, 1997; Barrett, 2002). This study carries out a similar estimation using a multinomial logit model, however in assessing alcohol consumption a limitation of these previous studies is analysed in that alcohol consumption can be viewed as being ordered data and the fact that previous studies have not accounted for this potential ordinality by estimating alcohol consumption by the multinomial logit model, could lead to less efficient results (Greene, 2002; Harris *et al*, 2006). A variety of extensions to the Heckman model (1979) have been developed for ordered choice models in order to account for selection bias, one being to use an ordered probit extension of the Heckman correction (Vella, 1998; Greene and Hensher, 2010). Chiburis and Lokshin (2007) adopted such an approach whereby the selection equation was estimated as an ordered probit, from which the inverse mills ratio was derived. This was then included as an additional regressor in the primary equation estimated in the second step of the process.

Different methods for consistent estimation of such a model of simultaneous equations exist, most of which fall into one of two categories, limited information methods and full information methods (Gujarati, 2004).

Section 2.3: Estimation of Simultaneous Equations

In the estimation of the effect of alcohol status on income accounting for endogeneity and selection bias, both the alcohol status equation and the income equation need to be estimated (Hamilton and Hamilton, 1997; Barrett, 2002). Limited Information methods or Full Information methods of estimation could be used to estimate such an effect (Chiburis and Lokshin, 2007). This section compares the different methods of estimation that can be adopted and the findings of previous studies in terms of the efficiency of both methods.

Section 2.3.1: Simultaneous Equations

Simultaneous Equations Models depends on more than one equation interacting together to produce the observed data (Gujarati, 2004). Unlike the single equation model in which a dependent variable is a function of independent variables, other dependent variables are among the independent variables in each equation within the simultaneous equation model (Barreto and Howland, 2006). The dependent variables in the system are jointly (or simultaneously) determined by the equations in the system (Barreto and Howland, 2006). Two or more equations together is the structure of the model (Greene, 2002).

In matrix terms the system of equations may be written as

$$[y_1, y_{2,} .....y_M]_t \begin{bmatrix} \gamma_{11} & \gamma_{12} & ...... & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & ...... & \gamma_{2M} \\ & & . & \\ & & . & \\ & & . & \\ \gamma_{M1} & \gamma_{M2} & & \gamma_{MM} \end{bmatrix} + [x_1, x_2, ...x_K]_t \begin{bmatrix} \beta_{11} & \beta_{12} & & \beta_{1M} \\ \beta_{21} & \beta_{22} & & \beta_{2M} \\ & & . & \\ & & . & \\ & & . & \\ \beta_{K1} & \beta_{K2} & & \beta_{KM} \end{bmatrix}$$

$$= [\varepsilon_{1,} \varepsilon_2 ......\varepsilon_M]_t \qquad \textbf{(2.3.1)}$$

In a system of equations to determine y_t in terms of x_t and ε_t, Greene (2002) states that the terms may be written as

$$y_t'\Gamma + x_t'B = \varepsilon_t' \qquad \textbf{(2.3.2)}$$

Where:	y	endogenous dependent variable
	Γ	MxM non singular matrix
	x	exogenous independent variable
	B	KxK parameter matrix
	ε	structural disturbances
	t	used to index observations t =1,…….,T

There are M equations with M endogenous variables and K exogenous variables, where every column is a vector of coefficients in a particular equation and each row applies to a specific variable. In order to determine y_t in terms of x_t and ε_t in the system of equations, the reduced form of the model is used (Greene, 2002). One of the variables in each equation is labelled the dependent variable, hence there will be at least one '1' in each column of Γ (Greene, 2002). The joint determination of the variables in the model is recursive. The first is completely determined by exogenous factors, then given the first, the second is likewise determined and so on (Greene, 2002). Reduced form equations represent each endogenous variable as a function of only exogenous variables (Greene, 2002).

$$y_t = -x_t'B\Gamma^{-1} + \varepsilon_t'\Gamma^{-1} \;=\; x_t' \prod + v_t' \qquad \textbf{(2.3.3)}$$

Where:	x	exogenous variables
	B	KxK parameter matrix
	Γ	MxM non singular matrix
	ε	error term
	$\prod$	KxM reduced form coefficient matrix which equals $-B\Gamma^{-1}$
	v	equals $\varepsilon_t'\Gamma^{-1}$
	t	index observations

If endogeneity exists and regressors are correlated with the error term then the OLS method is inappropriate for the estimation of an equation in a system of simultaneous equations and would lead to biased and inconsistent results (Gujarati, 2004). Two approaches may be adopted in the estimation of simultaneous equation models, namely single equation methods or limited information methods and system methods known as full information methods (Gujarati, 2004).

Section 2.3.2: Limited Information Methods

Limited Information Methods or a single equation method is where each equation in the system is estimated individually taking into account any restrictions placed on that equation without worrying about the restrictions placed on other equations in the system (Gujarati, 2004). There are a number of different single equation methods that can be used. OLS is generally inappropriate in the estimation of single equation models due to the frequent presence of endogenous regressors (Gujarati, 2004). The Two Stage Least Squares (2SLS) and the Heckman Two Step Method also known as the Limited Information Maximum Likelihood Methods of Estimation (LIML) are generally the methods used to estimate simultaneous equations consistently while accounting for endogeneity (Gujarati, 2004).

Heckman (1979) proposed a simple practical solution to the problem of sample selection whereby the selection problem is treated as an omitted variable problem. This is an easy to implement method, which is known as the two step or Limited Information Maximum Likelihood (LIML) method (Puhani, 2000). A limitation of Heckman’s two step model is that it is only applicable to binary choice situations. Lee (1982) extends the Heckman Probit OLS Two Step Estimate to a Multinomial Logit OLS Two Step Estimate, to allow for selection correction for polychotomous choices. Step one involves running a multinomial logit for the choice variable, which generates predicted values for the Inverse Mills Ratio (Hamilton and Hamilton, 1997; Barrett, 2002). In the second step an OLS regression is run which includes the values for the Inverse Mills Ratio (Hamilton and Hamilton, 1997; Barrett, 2002).

Two Stage Least Squares is a common approach to consistently estimate simultaneous equations when there are endogenous variables present (Zellner and Theil, 1962). Gujarati (2004) states that the idea behind Two Stage Least Squares (2SLS) is to replace the stochastic endogenous explanatory variable by a linear combination of the predetermined variables in the model and use this combination as the explanatory variable in lieu of the original endogenous variable. The 2SLS method thus resembles the instrumental variable method of estimation in that the linear combination of the predetermined variables serves as an instrument or proxy, for the endogenous regressor (Gujarati, 2004). A feature of 2SLS is that as the sample size increases indefinitely, the estimates converge is closer to the true population values (Puhani, 2000). The estimates may not satisfy small sample properties such as being unbiased and minimum variance (Puhani, 2000).

Section 2.3.3: Monte Carlo Studies

The Monte Carlo approach is defined by Intriligator *et al* (1996) as that of estimating known parameters, which are chosen beforehand using different techniques. It is the process of estimating parameters using a controlled setting, in which the true parameters are known. Agunbiade and Iyaniwura (2010) compare the method as the nearest thing to a controlled laboratory type experiment in econometrics. The Monte Carlo Approach has been applied in determining the choice of alternative estimators in looking at the impact of heteroskedasticity, serial correlation and other violations of basic econometric assumptions on the performance of different estimators in a given study (Agunbiade and Iyaniwura, 2010; Intriligator, 1996).

Van Dijk *et al* (1995) state that Monte Carlo integration methods make use of the following two properties;

1. Generating a large sample of pseudo-random numbers is very easy using a computer procedure. The use of Monte Carlo involves usually a computer procedure for the generation of these random numbers. Pseudo-random numbers are generated on a computer by means of a deterministic

method, thus a sequence of pseudo-random numbers is perfectly reproducible (Van Dijk *et al,* 1995)

2. An integral may be interpreted as the expectation of a random variable. This expected value is estimated using generated random numbers. The accuracy of the estimation procedure is measured using standard results from large sample theory (Van Dijk *et al,* 1995).

Problems handled by the Monte Carlo methods are of two types called probabilistic or deterministic according to whether or not they are directly concerned with the behaviour and outcome of random processes (Koutsoyannis, 1977). The use of this approach to probabilistic problems, involves observing random numbers chosen in such a way that they directly simulate the physical random processes of the original problem, and to infer the desired solution from the behaviour of these random numbers Koutsoyannis, 1977). The idea behind the approach to deterministic problems is to exploit the strength of the theoretician while avoiding its associated weakness by replacing theory experiment whenever the former falters (Koutsoyannis, 1977)

Adepoju (2009) sets out what the Monte Carlo experiment involves:

i. Specifying a “true” model (the explanatory, the coefficients, the sample size and the distribution of the error term)
ii. Generating a data set using (i)
iii. Obtaining estimates for the parameters using the generated samples
iv. Repeating the experiment a numbers of times
v. Evaluating how frequently the estimators accepts or rejects the “true” model in the set of replicates.

Findings from the Monte Carlo approach have generally been that FIML is the most desirable technique in the estimation of simultaneous equations, however it is computationally expensive and is very sensitive to specification and measurement error (Intriligator *et al,* 1996). The full information techniques, specifically 3SLS and FIML, generally provide the most desirable estimators in terms of both bias and mean squared error when the model is correctly specified and the variables are correctly measured (Intriligator *et al*, 1996). FIML, is however extremely sensitive to both specification error and measurement error; a slight misspecification or measurement error can change the results so as to make FIML less desirable than the limited information estimators (Intriligator *et al*, 1996). This can arise given that in the FIML approach to computation through a system of non-linear equations, an error in one equation or in one variable will propagate throughout the whole system in the process of estimation (Intriligator *et al*, 1996). Gujarati (2004) also states that the full information estimators particularly FIML is computationally more complicated than other estimators and hence more costly to use. Furthermore, both FIML and 3SLS require much larger sample size than the limited information estimators (Intriligator *et al* 1996). In analysing the FIML approach Adepoju (2009) finds that FIML is remarkably best in the open ended intervals and remarkably poor at the closed intervals. He states that the ranking of the estimators with respect to the magnitude of the average total bias is invariant to the choice of the upper P_1 or lower P_2, triangular matrices. The three stage least squares (3SLS) ranked best generating the minimum Average Total Absolute Bias (TAB), closely followed by Limited Information Maximum Likelihood (LIML) while the FIML performed poorly.

Likewise Agunbiade and Iyaniwura (2010) have similar findings in their analysis of six different estimation techniques for a just-identified simultaneous three equation econometric model with three multicollinear exogenous variables. The estimation techniques used were Ordinary Least Squares (OLS), Limited Information Maximum Likelihood (LIML), Two-Stage Least Squares (2SLS), Indirect Least Squares (ILS), Three Stage Least Squares (3SLS) and Full Information Maximum Likelihood (FIML). The performances of the estimators are evaluated based on the average or mean values of parameter estimates and total absolute bias of parameter estimates. Agunbiade and Iyaniwura (2010) find that estimates for the three estimators LIML, 2SLS and ILS are virtually identical and these estimators are best for estimating parameters of data plagued by the lower open interval negative level of multicollinearity while FIML and OLS respectively rank highest for estimating parameters of data characterized by closed interval and upper categories level of multicollinearity. In their analysis of small sample properties, Adepoju and Olaomi (2009) also find that FIML is outstandingly best in open ended intervals however poor in relation to closed intervals. Their study was in relation to small sample properties as they argue that it is important to rank estimators on the merit they have when applied to small samples as in practice researcher's usually work with small samples, and the asymptotic properties of the estimates are of little assistance in ones choice of technique.

By contrast to the full information method approaches, the limited information approach estimates only one equation at a time, and confines a misspecification in one equation to that particular equation and confines an error in measurement in one variable to those equations containing that particular variable (Gujarati, 2004). Intriligator *et al* (1996) find that of the possible limited information estimators, the 2SLS estimator generally performs best in terms of both bias and mean squared error and usually more stable than the others; in particular it is not greatly affected by specification errors. Furthermore, it is generally easily and inexpensively computed (Gujarati, 2004). Vandenberghe and Robin (2004) in their study find that the Heckman two-step method imposes a linear form on the outcome equation.

Sherkat (2004) argues that while Monte Carlo simulations have shown that FIML estimation is preferred generally, they also show that the two step estimation provides better estimates when collinearity is present, and OLS estimates are more efficient and less biased when there are multiple violations of the assumptions of the models (Sherkat, 2004). Collinearity across equations and among predictor variables may influence estimates from the FIML model (Sherkat, 2004).

Puhani (2000) in his analysis comparing the Heckman Limited Information Maximum Likelihood Method with the Full Information Maximum Likelihood Method (FIML), analyses other research carried out in relation to this. Puhani (2000) sets out the main conclusion drawn from existing Monte Carlo Studies whereby the relative performance of the estimators is studied in relation to the joint distributions of the error terms u_1 and u_2, the correlations between the error terms, the degree of censoring, and the degree of collinearity between the regressors x_1 and x_2 or between x_1 and the inverse mills ratio (λ). As to the joint distribution of u_1 and u_2, Puhani (2000) states that no clear result emerges when the distributional assumption of joint normality is violated. For the extreme cases Cauchy errors, Puhani (2000) states that Hay, Leu, and Rohrer (1987) and Paarsch (1984) do not identify an estimator which behaves superior to the others. Puhani (2000) also states that Zuehlke and Zeman (1991), who model bivariate t_5 and χ_5^2 errors, do not reach firm conclusions on this issue. The correlation between the error terms, $\mathrm{corr}(u_1, u_2)$, seems to have an affect of the performance of the LIML estimator. Although, Hay, Leu and Rohrer (1987), Manning, Duan and Rogers (1987), and Zuehlke and Zeman (1991) do not reach any strong results, Nelson (1984), Stolzenberg and Relles (1990) and Nawata (1993; 1994) provide evidence that the higher the correlation between u_1 and u_2, the greater the superiority of the FIML (and maybe OLS) estimator over LIML in terms of efficiency. Table 2.3.1 shows Puhani's (2000) summary of the main conclusion drawn from existing Monte Carlo Studies. The column 'Estimators Used' shows what estimators were used specifically in the Monte Carlo study and the column 'Main Results' shows the findings from the Monte Carlo Studies.

Table 2.3.1 Summary of Monte Carlo Studies in Puhani (2000)

Study	Models Analysed	Estima-tors Used	Sample Size	Repetitions	Distribu-tions of u_1, u_2	Variables Changed	Judgement criteria for estimators	Main Results
Nelson (1984)	sample selection model with and without exclusion restrictions	LIML FIML OLS	2,000	n.a.	biv. normal	$R^2(\lambda, x_1) = 0,$ 0.35, 0.641, 0.953, 0.999 Corr (u_1, u_2)= -0.5, 0, 0.25, 0.5, 0.75, 0.95	bias and variance of parameter estimates	Relative efficiency of FIML over LIML rises with higher $R^2(\lambda, x_1)$ and corr (u_1, u_2) OLS dominates LIML only when $R^2(\lambda, x_1)$ is very high and/or corr (u_1, u_2) is small
Paarsch (1984)	sample selection model without exclusion restrictions and identical errors (Tobit)	LIML FIML (Tobit) OLS Powell's LAD	50 100 200	100	normal Laplace Cauchy	degree of censoring 25 and 50%	bias, variance, median, lower and upper quartile of parameter estimates	LIML much less efficient than FIML (Tobit) when errors are normal (or Laplace) FIML (Tobit) performs poorly when errors are Cauchy OLS worst estimator In all cases use of Powell's LAD limited.

(Source: Puhani, 2000)

Table 2.3.1 continued: Summary of Monte Carlo Studies in Puhani (2000)

Study	Models Analysed	Estima-tors Used	Sample Size	Repetitions	Distribu-tions of u_1, u_2	Variables Changed	Judgement criteria for estimators	Main Results
Hay, Leu, and Rohrer (1987)	sample selection model without exclusion restrictions	LIML FIML TPM	300 1,500 3,000		biv. normal logistic/normal Cauchy/Cauchy	corr (u_1, u_2) =0 0.33, 0.66, 0.90, 1.00	mean squared error of fit mean bias of fit mean squared error of parameter estimates	TPM most robust when error distributions are normal or logistic. In the Cauchy case, none of the models can establish a superiority over the others no firm results concerning corr (u_1, u_2)
Manning Duan and Rogers (1987)	Sample selection model with and without exclusion restrictions	LIML FIML TPM Data-Analytic TPM*	1,000	100	biv.normal	corr (u_1, u_2) = 0.5, 0.9 Degree of censoring 25, 50,75%	Mean squared error fit Mean bias of fit	LIML worst when no exclusion restrictions (Data-Analytic TPM and best then) FIML and LIML perform badly when censoring is high No firm results concerning corr (u_1, u_2)

(Source: Puhani, 2000)

Table 2.3.1 continued: Summary of Monte Carlo Studies in Puhani (2000)

Study	**Models Analysed**	**Estima-tors Used**	**Sample Size**	**Repetitions**	**Distribu-tions of u_1, u_2**	**Variables Changed**	**Judgement criteria for estimators**	**Main Results**
Stolzenberg and Relles (1990)	Sample selection model with exclusion restrictions	LIML OLS	500	100	Normal/ normal	corr $(x_1, x_2)^2$ =0, 0.25, 0.5, 0.75 corr $(u_1, u_2)^2$ =0, 0.25,0.5,0.75 Var(u_1) = 1/9, 1, 9 Var(u_2) =0.25, 1,4	Bias and mean absolute error of parameter estimates	No clear relationship between the variances of u_1 and u_2 and the performance of the two estimators High corr $(x_1, x_2)^2$ and high corr $(u_1, u_2)^2$ render LIML superior to OLS in terms of bias, than in OLS in over a third of cases.
Zuehlke and Zeman (1991)	Sample selection model without exclusion restrictions	LIML OLS Lee's robust estimator	100	1,000	biv. Normal biv. t_5 biv. χ_5^2	corr (u_1, u_2) = 0,0.5,1 Degree of censoring 25,50,75%	Bias and mean squared error of parameter estimates	LIML reduces bias, but has very large standard error compared to OLS due to the collinearity of x_1 and λ OLS preferable to LIML, especially when the degree of censoring is high. Lee's robust estimator worst of all No firm results concerning corr (u_1, u_2)

(Source: Puhani, 2000)

Table 2.3.1 continued: Summary of Monte Carlo Studies in Puhani (2000)

Study	**Models Analysed**	**Estima-tors Used**	**Sample Size**	**Repetitions**	**Distribu-tions of u_1, u_2**	**Variables Changed**	**Judgement criteria for estimators**	**Main Results**
Rendtel (1992)	Sample selection model with and without exclusion restrictions	LIML FIML OLS	400	100	Normal/ normal	Additional variables in selection model (i) correlated with y_1 and y_2 (ii)correlated only with y_1 (iii) correlated only with y_2 (iv) correlated with neither y_1 nor y_2	Bias and variance of parameter estimates	Without exclusion restrictions OLS is slightly preferable to FIML and clearly preferable to LIML With exclusion restrictions LIML and especially FIML dominate OLS only if the additional variable in the selection model is only correlated with y_2 (case (iii)); Otherwise (cases (i), (ii) and (iv))exclusion restrictions do not improve the FIML estimator.
Nawata (1993)	Sample selection model with and without exclusion restrictions	LIML OLS	200	500	biv.normal	corr(u_1, u_2) = 0,0.2,0.4,0.6,0.8,1 corr (x_1, x_2) = 0,0.5,0.8,0.9,0.95,1	Bias, variance, median, lower and upper quartile of parameter estimates	LIML less efficient the higher (corr (x_1, x_2)) corr(u_1, u_2) >0.9 renders the LIML estimator very unstable OLS preferable for high corr (x_1, x_2) and high corr(u_1, u_2)

(Source: Puhani, 2000)

Table 2.3.1 continued: Summary of Monte Carlo Studies in Puhani (2000)

Study	Models Analysed	Estima-tors Used	Sample Size	Repetitions	Distribu-tions of u_1, u_2	Variables Changed	Judgement criteria for estimators	Main Results
Nawata (1994)	Sample selection model with and without exclusion restrictions	LIML FIML	200	200	biv.normal	corr (u_1, u_2) = 0,0.4,0.8 corr (x_1, x_2) = 0,0.5,0.8,0.9,0.95,1	Bias, variance, median, lower and upper quartile of parameter estimates.	FIML dominated LIML especially for high corr (x_1, x_2) >0.9 renders the LIML estimator very unstable FIML generally preferable.

(Source: Puhani, 2000)

Puhani (2000) concludes that where collinearity does not exist, Heckmans LIML estimator may be employed, but given the constant progress in computing power the FIML estimator is recommended, as it is usually more efficient than the LIML estimator.

Using a Monte Carlo simulation, Enders and Bandalos (2001) examine the performance of four missing data methods in structural equation models: full information maximum likelihood, list wise deletion, pair wise deletion and similar response pattern imputation. They examine the effects of three independent variables on four outcome measures and find that FIML estimation was superior across all conditions of the design under missing data conditions. FIML is unbiased and more efficient than the other methods (Enders and Bandalos, 2001).

Generally findings have been that the Full Information Maximum Likelihood Method of estimation (FIML) is the most favourable technique in the estimation of simultaneous equations (Puhani, 2000; Intriligator *et al,* 1996; Enders and Bandalos, 2001), primarily due to the fact that this estimator is based on the entire system of equations being estimated jointly (Greene, 2002).

Section 2.3.4: Full Information Methods

Full information methods estimate all the equations in the model simultaneously, taking due account of all restrictions on such equations by the omission or absence of some variables (Gujarati, 2004). Both the Full Information Maximum Likelihood (FIML) and 3 Stage Least Squares (3SLS) estimators are full information methods (Gujarati, 2004). In order to preserve the spirit of simultaneous equation models, ideally one should use the systems method such as the full information maximum likelihood method (FIML) (Gujarati, 2004)

There are two theoretical reasons why in estimating the system, limited information methods or the one-equation-at-a time procedure can be improved upon (Wonnacott and Wonnacott, 1979);

1. Estimation of the first equation in the series of equations does not exploit ones prior information about other equations in the system in particular, the zero restrictions imposed in other equations

2. The estimate of the first equation might be improved further if each possible correlation between the errors in each structural equation is allowed for.

The joint estimation of equations in a simultaneous equation model brings efficiency gains (Greene, 2002). Estimations of the system using limited information methods, has the benefit of computational simplicity but these methods neglect information contained in the other equations (Wonnacott and Wonnacott, 1979). In general the limited information estimator is asymptotically less efficient than the full information estimators such as the FIML or 3SLS estimator, since it does not use all the information that is available in the system (Judge *et al,* 1988). In contrast to many other findings in relation to the comparison of limited and full information methods of estimation, Seaks (1974) finds that consistent single equation methods does better than a systems method such as FIML or 3SLS when a model is simulated over a long period. Seaks (1974) in analysing the work of Klein, finds that in comparing Least Squares, 2SLS, and FIML, the least squares method does poorly relative to the 2SLS and FIML and that 2SLS seems best for simulations over the entire database, while FIML has the edge for one period simulation.

In practice full information methods are not used for a variety of reasons (Gujarati, 2004). Firstly the computational burden is enormous. Secondly methods such as the FIML method lead to solutions that are highly non-linear in the parameters and are therefore often difficult to determine (Gujarati, 2004). Thirdly if there is a specification error in one or more equations of the system, that error is transmitted to the rest of the system and as a result the systems methods become very sensitive to specifications errors (Gujarati, 2004). In practice, therefore single equation methods are often used despite the fact that in estimation of simultaneous equations FIML is the ideal system (Gujarati, 2004)

Table 2.3.2 summarises both the limited information and full information methods of estimation.

Table 2.3.2 Summary of Methods of Estimation

	Limited Information Methods	**Full Information Methods**
Least Squares	Two Stage Least Squares (2SLS)	Three Stage Least Squares (3SLS)
Maximum Likelihood	Limited Information Maximum Likelihood Method (LIML)	Full Information Maximum Likelihood Method (FIML)

(Source: Authors own)

Table 2.3.2 shows the two main Full Information Methods of estimation are the Full Information Maximum Likelihood Method and the Three Stage Least Squares method (Gujarati, 2004).

Section 2.3.4.1: Three Stage Least Squares

Three Stage Least Squares as a systems method, developed by Zellner and Theil (1962), adds a third stage to the two-stage least squares method. It is a full information method, since it exploits all available information as it simultaneously estimates all equations in the system, in contrast to the 2SLS which is a limited information method, and which estimates each equation in the system one at a time (Wonnacott and Wonnacott, 1979). 3SLS generates a set of observed errors $\hat{e}_1, \hat{e}_2, \hat{e}_3$ etc., which is used to estimate the covariance matrix of the errors in the system (Wonnacott and Wonnacott, 1979)

The first two of the three stages of 3SLS are those of 2SLS, the first stage being the estimation of all reduced form coefficients using the least squares estimator, while the second stage is the estimation of all structural coefficients by applying 2SLS to each of the structural equations (Intriligator *et al*, 1996). The third stage is then the generalised least squares estimation of all of the structural coefficients in the system, using a covariance matrix for stochastic disturbance terms of the structural equations that is estimated from the second stage residuals (Intriligator *et al*, 1996). Using the information contained in this covariance matrix has the effect of improving efficiency. In terms of properties of estimators, the 3SLS technique is an improvement over 2SLS, in that while both are consistent, 3SLS is asymptotically more efficient than 2SLS (Intriligator *et al*, 1996). Intriligator *et al (*1996) state that the basis rationale for the use of 3SLS, as opposed to 2SLS, is its use of information on the correlation of the stochastic disturbance terms of the structural equations in order to improve asymptotic efficiency. 3SLS can be viewed as an extension of the method of seemingly unrelated equations to a system of equations in which explanatory endogenous variables are present in some or all of the equations. If there are no explanatory endogenous variables in the system then 3SLS reduces to seemingly unrelated equations (Intriligator *et al*, 1996).

Section 2.3.4.2: Full Information Maximum Likelihood Method

Full Information Maximum Likelihood (FIML) is a technique for estimating systems of simultaneous equations which may be linear or non-linear (Greene, 2002). The full information maximum likelihood estimator is based on the entire system of equations of simultaneous equation models. This estimator treats all equations and all parameters jointly and takes account of the fact that errors may be correlated between equations (Greene, 2002; Pearce, 1986). With the FIML approach the likelihood function for the entire system is maximised by choice of all system parameters, subject to all priori identifying restrictions (Intriligator *et al,* 1996). With normally distributed disturbances, FIML is efficient among all estimators (Greene, 2002), resulting in the estimators being consistent and asymptotically efficient. While FIML has the same asymptotic properties as 3SLS including the same asymptotic covariance matrix, a major advantage of FIML over 3SLS , is that it is possible to use this technique in the estimation of a wide range of a priori information, pertaining not only to each equation individually but also to several equations simultaneously, such as constraints involving coefficients of different structural equations and certain restrictions on the error structure (Greene, 2002). The major disadvantage of FIML however, is that it is difficult and expensive to compute as it can involve the estimation of awkward simultaneous nonlinear equations, which usually must be computed via iteration (Greene, 2002). Very often due to simplicity and asymptotic efficiency, 2SLS is used almost exclusively, when ordinary least squares is not used, for the estimation of simultaneous equation models (Greene, 2002).

Adepoju and Olaomi (2009) state that all simultaneous equation estimation methods have some desirable asymptotic properties and these properties become effective in large samples. Their study looked at the small sample properties of these estimators when the errors are correlated to determine if the properties still hold when available samples are relatively small and the errors are correlated. Adepoju and Olaomi (2009) find that FIML is outstandingly best in the open ended intervals and outstandingly poor in the closed interval.

The Full Information Maximum Likelihood (FIML) estimator is based on the entire system of equations (Greene, 2002). With normally distributed disturbances, FIML is efficient among all estimators. The FIML estimator treats all equations and parameters jointly which are set out by Greene (2002).

Details of the log-likelihood function and how it is maximised, as outlined by Greene (2002) is set out in Appendix D.

2.3.5: Estimation of a regression model with an ordered probit selection equation using the Full Information Maximum Likelihood Estimator

Chiburis and Lokshin (2007) in their study into the estimation of wages for public, private and informal sectors for male workers in India use an ordered probit selection model. The categorical variable describing the sector individuals work in, is estimated as an ordered probit on the basis of an ordered probit selection rule, which is set out in section 2.2.1. They estimate this using both the Limited Information Maximum Likelihood Method of Estimation and the Full Information Maximum Likelihood Method of Estimation.

The Full Information Maximum Likelihood method estimates both the selection equation and the wage equations jointly (Greene, 2002; Chiburis and Lokshin, 2007).

As set out in section 2.2.1 the selection equation is defined as follows:

$$c_i^* = \alpha' s_i + \varepsilon_i \qquad \textbf{(2.3.4)}$$

Where:	c	sector category
	α	an unknown vector of parameters
	s	independent variables
	ε	a standard normal shock
	μ_J	cutoffs
	i	indexes individuals

The wage equation is defined as

$$\ln Y_{ij} = X_i \beta_j + u_{ij} \qquad \textbf{(2.3.5)}$$

Where:	$\ln Y_{ij}$	log of wages
	X	vector of human capital variables & socio-demographic characteristics

	β	coefficients on the observable characteristics
	u_{ij}	error term
	i	indexes individuals where $i = 1,2,......N$
	j	indexes sector category where $j = 1,2,3,$

The vectors of the unknown parameters are β and α .

The FIML estimation consists of finding the parameter values that maximise the likelihood of the data (Chiburis and Lokshin, 2007).

The parameters to be estimated are α , β_1 , β_2,….. β_{J-1} ; μ_1, μ_2, …. μ_J ; ρ_0, ρ_1, …. ρ_{J-1}; σ_0 , σ_1, ….. σ_{J-1}

but β_0 ρ_0 σ_1 do not exist for categories j in which y is missing (Chiburis and Lokshin, 2007).

The likelihood of an observation i in which the category is j and y_i is observed is

$$L^y{}_{ij} \equiv L\left[y_{i,} j \middle| x_i, \beta_j, \sigma_j, \rho_j, \alpha, s_i, \mu_j, \mu_{j+1}\right]$$

$$\equiv L\left[y_i \middle| x_i, \middle| \beta_j, \sigma_j\right] P_r\left[j \middle| y_i, x_i, \beta_j, \sigma_j, \rho_j, \alpha, s_i, \mu_j, \mu j+1\right]$$

$$= \frac{1}{\sigma_j}\phi(t_i)\left[\Phi\left(\frac{\alpha' s_i + \rho_j t_i - \mu_j}{\sqrt{1-\rho_j^2}}\right) - \Phi\left(\frac{\alpha' s_i + \rho_j t_i - \mu_{j+1}}{\sqrt{1-\rho_j^2}}\right)\right] \qquad \textbf{(2.3.6)}$$

Where:	y	wages
	x	vector of human capital variables & socio-demographic characteristics
	β	coefficients on the observable characteristics
	α	is an unknown vector of parameters,

s independent variables

μ_J cutoffs

ρ correlation coefficient

σ standard deviation of the error term

$t_i \equiv (y_i - \beta_j' x_i)/\sigma_j$

ϕ is the standard normal density function

Φ standard normal cumulative distribution function

i indexes individuals where $i = 1,2,......N$

j indexes sector category where $j = 1,2,3,$

If ε, u are standard bivariate normal with correlation ρ, then the conditional distribution of ε given u is normal with mean ρu and variance $1-\rho^2$

If j is a category in which y is unspecified, then the likelihood is simply

$$L_{ij} \equiv \Phi(\alpha' s_i - \mu_j) - \Phi(\alpha' s_i - \mu_{j+1}) \qquad \textbf{(2.3.7)}$$

Taking the logarithm of equations 2.3.11 to get the log likelihood for observation i, and since observations are independent the log likelihood can be added across observations to get the log likelihood for the entire sample (Chiburis and Lokshin, 2007).

$$L \equiv \sum_{i=1}^{n} \begin{cases} \log L_{ici}^{y} & \text{if } y_i \text{ is observed} \\ \log L_{ici}^{\cdot} & \phantom{\text{if }} y_i \text{ is missing} \end{cases} \qquad \textbf{(2.3.8)}$$

Chiburis and Lokshin (2007) in their study state that such an estimation of an ordered selection equation and a wage equation can be done through Limited Information method of estimation such as the Limited Information Maximum Likelihood method or the Full Information method of estimation such as the Full Information Maximum Likelihood method. Many studies have been carried out into comparing the two methods (Puhani, 2000; Intriligator *et al*, 1996; Enders and Bandalos, 2001). Chiburis and Lokshin (2007) in their study find the FIML estimator to be slightly more efficient than the two step estimator when the data exactly meet the model specifications.

2.3.6: Conclusion

Where data is ordered, estimation methods to account for this ordinality while also accounting for endogeneity and selection bias is analysed. Given that simultaneous equations are being estimated in looking at the effect of alcohol on income, the different methods of estimation of simultaneous equations that is, limited information and full information methods are assessed. Many studies find Full Information Methods to be more efficient methods of estimation although they are often not used due to being computationally difficult (Puhani, 2000; Intriligator *et al*, 1996). In an estimation such as the effect of alcohol consumption on income while accounting for the potential endogeneity of the choice variable, the full information maximum likelihood method of estimation would mean that both equations are estimated jointly while accounting for the fact that the errors may be correlated (Gujarati, 2004). In comparison the limited information maximum likelihood method estimates the selection equation initially and then in the second step, estimates the primary equation (Gujarati, 2004).

2.4: Health Status and Health Care Utilisation

Previous research shows that moderate drinkers tend to enjoy a more beneficial health status compared with abstainers and heavy drinkers (Berger *et al*, 1999; Klatsky *et al* 2001; Bau *et al,* 2007), similar to findings in terms of the relationship between alcohol and financial welfare where the financial welfare of moderate drinkers is better than abstainers or heavy drinkers (French and Zarkin, 1995; Heien, 1996; Hamilton and Hamilton, 1997; Barrett, 2002). This section will review the definition of health and the factors that affect both health status and individual levels of health care utilisation.

2.4.1: Definition of Health

The World Health Organisation (1948) defines health as a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity. Health is a resource for everyday life, not the objective of living. Health is a positive concept emphasising social and personal resources, as well as physical capacities (WHO, 1948).

Modern literature on health promotion defines health as having two distinct dimensions; positive health (well being) and negative health (ill health) (Downie *et al*, 1996). The positive dimension of health consists of the qualitative aspects of health and human life in general, and is strongly associated with the concept of "fitness". The negative dimension is determined by the presence or absence of disease, illness, deformity, unwanted states, injury, disability and handicap.

Rivera (2001) states that health is affected by many factors which can be divided into four groups of variables; biological, socioeconomic, lifestyle and medical resources. In Rivera's study (2001) biological variables include age, gender and race; socioeconomic variables include education, income, employment status; lifestyle variables reflect customs or habits of the interviewee which includes the variables describing whether or not a person is a smoker, the levels of alcohol one consumes and whether or not one takes regular exercise. In relation to the fourth category, medical resources, Rivera (2001) did not have this information available for this study.

Health is measured in different ways. Many studies use the measure of Self Assessed Health (Jurges, 2008) while others use the frequency of use of health services as a measure of health (Hernandez-Quevedo and Rubio, 2009).

2.4.2: Self Rated Health

Self Assessed Health is probably the most common measure of health in general purpose surveys and often the only available indicator of the respondent's health (Jurges, 2008). The popularity of this measure for applied economists stems from the fact that it lends itself to the treatment of health as a latent, unidimensional variable (Jurges, 2008). The Self Assessed Health measure is widely used both as an outcome variable in studies of social influences on health (Jurges, 2008; Contoyannis and Jones, 2004; Kiuila and Miesztowski, 2007) and as an explanatory variable in other studies (Wang, 1997; Disney *et al*, 2006). Fayers and Sprangers (2002) state that in relation to the question

> *"What do you think about your health in general? Very good, Good, Fair, Poor, Very Poor?"*

There is widespread agreement that this simple global question provides a useful summary of how patients perceive their overall health status. Fayers and Sprangers (2002) also state that this view is borne out by the large number of studies that have consistently shown, in a wide range of disease areas, that Self Rated Health is a powerful predictor of clinical outcome and mortality.

There can be cross-national variations in self-rated health (Von dem Knesebeck *et al*, 2006). In Ireland and Switzerland only a small proportion of less than 20%, has a poor self-rated health, whereas in Eastern and Southern European countries like Hungary, Poland, or Portugal about 50–60% of the people rate their health less than good. In comparison to other EU countries, Ireland continues to have the highest levels of self-perceived health of any EU country (Ireland, 2012). In

2010, 84% of males and 83% of females rated their health as being good or very good. This is the highest in the EU and compares with an average of 71% and 66% for males and females respectively across the EU (Ireland, 2012).

A major concern with Self Assessed Health is that respondents may have different response styles or different reference points against which they judge their health (Jurges, 2008). This would in turn give rise to a fundamental identification problem, namely to distinguish differences in true health from differences in reporting behaviour. An example of this would be where older respondents tend to have a milder view of their health that is they tend to rate their health as better than otherwise comparable younger respondents (Groot, 2000, Jurges, 2008). This most probably happens because health declines in general with age, so that the perception of what good health is, also changes when individuals get older, which may result in a survey understates the decline in true health (Groot, 2000, Jurges, 2008). An attempt to design a common comparable scale (not only for health) is the anchoring 'vignette' approach (King *et al*, 2004). The Vignettes are short descriptions of persons in different health states, which respondents are asked to judge on the same scale as they are asked to judge their own health. Respondents are explicitly asked to think about the vignette persons as people of their own age and background. The idea behind this is that the respondents put themselves in the shoes of the vignette persons (Jurges, 2008). If responses are consistent across vignettes and self-ratings, it is possible to cover the respondent's individual reference points. Jurges (2008) states that the main problem with this is the fact that no vignettes have yet been developed for general health. If one was to specifically look at the reports of chronic conditions instead of general health this is probably subject to the same measurement error (Jurges, 2008).

2.4.3: Health Care Utilisation

Socio-economic factors such as education, employment and income among others are strong determinants of health status (Rivera, 2001; Behrman & Wolfe, 1989; Berger & Leigh, 1989; Gilleskie & Harrison, 1998; Hartog & Oosterbeek, 1998; Kenkel, 1991, 1995; Leigh, 1998; Yen *et al*, 2010; Lin, 2008). In turn, numerous studies have been carried out into the association between health status and the frequency of use of health services and Hernandez-Quevedo and Rubio (2009) show that need is the most important determinant of health care use in that relative to being in good health, being in very bad health increases the probability of using all types of health services.

The most immediate determinant of utilisation is health status (Gruber & Kiesel, 2010). Health status is highly correlated with health care utilisation. Generally findings have been that the health status variables are strongly associated with both visits to GP's and specialists and that individuals who report a poorer health status are more likely to report greater use of physician services (Dunlop *et al*, 2000; Laroche, 2000). Similarly Sarma *et al* (2006) find that health status has a positive significant effect on utilisation of both GP's and specialists, and that females are more frequent users of health services than men. In Ireland, health status is positively related to the utilisation of GP services (Madden *et al*, 2005).

In examining the use of health care services for individuals with and without a chronic back disorder using the Canadian Community Health Survey 2000-2001, Lim *et al* (2005) conclude that the greater the disability and pain, the higher the utilisation of physicians. People with less severe symptoms and pain were 3.6 times more likely to seek help from a chiropractor than people with no back disorder (Lim *et al*, 2005). Rotermann (2006) describes the use of health care by Canada's senior population with focus on utilisation of general practitioners, hospitalisation, medication and home care. The main findings of this study are that seniors who perceived their health as fair or poor were heavy users of health care services. In addition, the presence of a chronic condition contributes significantly to the use of health services (Rotermann, 2006).

Similarly Finkelstein (2001) investigates the self-reported health status and its influence on health care utilisation, where the fee from claims submitted by physicians defines utilisation of physicians' services. The main findings of this study are that mean expenditure is substantially higher among those who reported worse health status and that self-reported health status was significantly related to the probability of seeing a specialist. Compared to its reference (excellent health status) the proportion of respondents reporting fair health status seeing a specialist was 25% higher (Finkelstein, 2001).

2.4.4: Human Capital Model of the Demand for Health

Michael Grossman's (1972) human capital model of the demand for health has been argued by some to be one of the major theoretical innovations to have emerged from health economics (Wagstaff, 1986). The determinants of health constitute an issue of vital importance to health policy (Gerdtham *et al*, 1999). Becker (1965) outlined a model where households are seen as producers of "commodities" instead of solely consumers of goods and services. Households are assumed to derive utility from the basic commodities they produce by combining their own time with market goods. Hence, the utility associated with a market good is conditional on the time that is allocated to its consumption. Using the key concept of home production elaborated in Becker's work, Grossman (1972) used the theory of human capital to explain the demand for health and health care.

Grossman (1972) constructed a model where individuals use medical care and their own time to produce health (Gerdtham *et al*, 1999). The stock of health capital depreciates over time, however the consumer can produce gross investments in health according to a household production function using medical care and their own time as inputs. In Grossman's formulation, individuals derive utility from the services that health capital yields and from the consumption of other commodities. The model considered a utility function where utility depends on both the flow of healthy days from a stock of health available in a given period, and on the consumption of other commodities, which are produced at home by combining purchased market goods and time (Leibowitz, 2004). A change in health stock in any period is the net result of gross investments in health and the depreciation in health stocks that occurs with age. Greater stocks of human capital were conceptualised as improving the technology of health capital production, yielding greater health outputs for given levels of time and medical inputs (Folland, Goodman, Stano, 2001).

$$I = I(M, TH, E) \qquad \textbf{(2.4.1)}$$

$$B = B(X, TB, E) \qquad \textbf{(2.4.2)}$$

Where:

	I	health investment
	M	market health inputs
	TH	time spent improving health
	B	home good production
	X	market goods necessary for the production of the home good
	TB	time spent in producing the home good
	E	technical efficiency level

Folland, Goodman, Stano (2001) describe the concept of time spent producing health in that investment to health is produced by time spent improving health (TH) & market inputs (M) such as medical services or drugs. A home good is produced (B) with time TB and market purchased goods (X). A person uses money to buy health care inputs (M) or home goods (X) and uses leisure time either for improving health care (TH) or for producing the home good (TB). Variable E in these functions is included to suggest that productivity in producing I or B may vary from person to person. Grossman proposed that this technical efficiency level would be related to the individual's education level, E in that educated people may produce one good or the other more efficiently (Folland, Goodman, Stano, 2001).

Individuals have to allocate their time between time spent enhancing health, leisure time, time lost to illness and time spent working (Folland, Goodman, Stano, 2001).

$$\text{T} = 365 \text{ days} = \text{TH} + \text{TB} + \text{TL} + \text{TW} \qquad \textbf{(2.4.3)}$$

Where:

	T	total time available (365 days per period)
	TH	health enhancing time
	TB	now called leisure time

TL	time lost to illness
TW	working time

Time available for work or leisure = 365-TH-TL = TW +TB

The rate at which a persons stock of health depreciated may increase during some periods of life and decline during others (Folland *et al*, 2001). Eventually as one ages, the depreciation rate is likely to increase, that is, the health of older people is likely to deteriorate faster than the health of younger people (Folland *et al*, 2001). Individuals are assumed to invest in health production until the marginal cost of health production equals the marginal benefits of improved health status. Health status is assumed to affect utility indirectly, through increasing labour income, and directly by the value that individuals place on good health. Grossman (1972) argues that if one can improve their health status they are then in a position to work more, they are absent from work less and are more productive which results in higher income. Grossman (1972) adds to this theory by saying that an increased wage rate increases ones returns from healthy days so workers will therefore tend to increase their optimal capital stock of health.

A second major contribution was to treat education as a factor that increases one's efficiency in producing health and reducing the shadow price of investment at any given age (Leibowitz, 2004). Michael Grossmans (1972) theory of demand entails a central role for education. Under Grossman, better-educated people tend to be economically more efficient producers of health. Educated people have better knowledge around the importance of good health and they know what are the ways to achieve and maintain good health (Folland *et al*, 2001). They also know better how to use medical and other market inputs and their own time to produce health. Education is also seen as a factor allowing a person to improve the efficiency with which, one can produce investments in health (Leibowitz, 2004). It raises the marginal product of the direct inputs. A given investment in health can be generated at less cost for an educated person and therefore they experience a higher rate of return to a given stock of health (Leibowitz, 2004). Grossman argued that better educated people tend to be economically more efficient producers of health status (Leibowitz, 2004). From the demand side educated people tend to recognise the benefits of improved health (Folland *et al*, 2001). They have a greater demand for health relative to other goods and have a grater appreciation for the benefits brought about by being healthy (Folland *et al*, 2001).

In light of the literature that has developed in the intervening 30 years on the effects of health habits on an individual's health, (Leibowitz, 2004) argues that it is important to incorporate the role of non-medical consumption into the health production model. Not only do non-medical commodities compete with health investments for an individual's time and money resources, but other consumption also may directly affect health. Some consumption activities, such as smoking tobacco, may provide current utility, but can be expected to increase the number of unhealthy days in later periods. Leibowitz (2004) also states that with the Grossman model, optimality requires that the marginal cost of the investment (forgoing the utility of consumption in an early period) must equal the present value of the marginal benefits. The optimal level of health capital for any individual is determined by the point at which the marginal cost of investment in health capital is equal to the marginal utility of healthy days (Kiiskinen, 2003).Conversely, the marginal utility of health-depleting consumption must equal the present value of the expected loss of utility in future periods (Kiiskinen, 2003).

Leibowitz (2004) expands this view of health production and treats net investments in health (H) in a given period, time t, as depending not only purchased medical inputs (M) and medical care (h), but choices about time spent on other consumption (tc) and choices about non-medical purchased goods (X). Other household consumption activities may have either positive or negative effects on net health stocks. Similarly time spent in the labor market (tw) may have either positive effects on health or negative effects due to occupational injuries (Leibowitz, 2004). In addition to specifying a role for an individual's choices about allocating time and money to health promoting or health reducing activities, equation 2.4.4 also includes the impact on net health investments of environmental inputs (N) that are beyond an individual's control. This acknowledges that air pollution or high crime levels in an individual's environment will affect the accumulation of health capital (Leibowitz, 2004). Education is also included. Existing health stocks (H_{t-1}) enter into the production of additions to health capital. Thus, the marginal product of all other inputs is likely to be smaller when there are lower levels of the fixed factor, the existing stock of health (Leibowitz, 2004).

$$H_t = H(th, tc, tw, X_t, M_t, N, H_{t-1}, E) \qquad \textbf{(2.4.4)}$$

Where:	H	net investments in health in a given period
	h	medical care in time t
	tc	time spent on other consumption
	tw	time spent in the labour market
	X	choices about non-medical purchased goods
	M	purchased medical inputs
	N	net health investments on environmental inputs
	H_{t-1}	existing health stocks
	E	Education
	t	time

Ziebarth and Grabka (2009) suggest that alcohol may affect the stock of human capital through at least two channels, in that alcohol consumption may influence an individuals productivity and thus wages through a persons health status and in addition to this social network effects could be induced through drinking habits.

The demand-for-health model by Grossman (1972) has become a corner stone in the field of health economics. The model is not, however, undisputed. A key criticism of the model has been that it fails to take into account the uncertainty of the future health status and the uncertainty of the effects of investments in health production and some argue that by ignoring the possibility of stochastic shocks, such as accidents or major illnesses which may result in large and permanent decreases in the level of health capital leads to overestimates of an individual's control of his own health in the long run (Van Doorslaer, 1987; Zweifel and Breyer, 1997).

The Grossman model highlights some of the different variables that affect ones ability to produce health. Other studies have found that many different variables can affect ones health status and in turn ones health care utilisation. These are discussed in section 2.4.5.

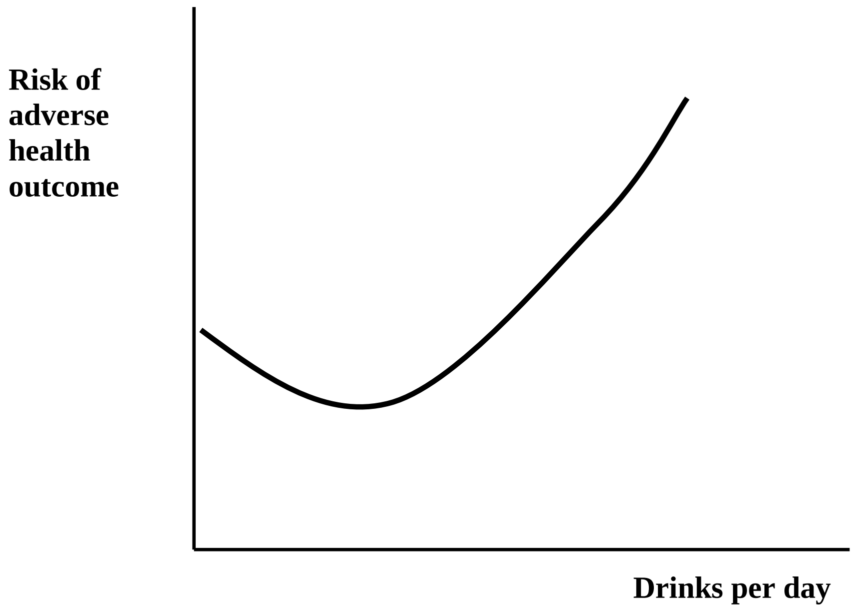
2.4.5: Factors affecting Health Status and Health Care Utilisation

Much research as been carried out into what factors affect ones health status and health care utilisation and findings show that many human capital and socio demographic variables have a significant affect on health and utilisation of health care services (Lin, 2008; Yen *et al*, 2010; Llena-Nozal *et al*, 2004; Gilleskie & Harrison, 1998; Hartog & Oosterbeek, 1998; Kenkel, 1991, 1995; Leigh, 1998; Bassuk, Berkman, & Amick, 2002; Bopp & Minder, 2003; Mackenbach *et al*., 2004; Lin, 2008; Kwan, 2010; Zick and Smith, 1991).

2.4.5.1: Alcohol

The effects of alcohol on ones health status has been the subject of much research and in general findings have tended to be that moderate levels of alcohol consumption are beneficial towards ones health status, compared with abstaining from or consuming heavy amounts of alcohol which has a negative effect of health status (Berger *et al*, 1999; Klatsky *et al* 2001; Bau *et al*, 2007). This gives rise to a U shaped curve or a partial U shaped curve referred to as a J shaped curve, showing a reduced relative risk of given diseases and in general better health for moderate consumers of alcohol compared with abstainers or heavy drinkers (Berger *et al*, 1999; Klatsky *et al* 2001; Bau *et al*, 2007)

Figure 2.4.1 U shaped curve showing relationship between alcohol consumption and risk of adverse health outcome



(Source: Authors own)

Apart from being a drug of dependence, alcohol has been known for many years as a cause of some 60 different types of disease and conditions, including injuries, mental and behavioural disorders, gastrointestinal conditions, cancers, cardiovascular diseases, immunological disorders, lung diseases, skeletal and muscular diseases, reproductive disorders and pre-natal harm, including an increased risk of prematurity and low birth weight (WHO, 2012).

Research carried out into the effects of alcohol consumption on coronary heart disease, show that moderate consumers of alcohol tend to have a lower risk of coronary heart disease than abstainers or heavy drinkers (Shaper *et al*, 1999; Rimm and Moats, 2007; Bryson *et al*, 2006; Klatsky *et al*, 2005.) Becker *et al*, (1996) have similar findings in terms of the effects of alcohol on liver disease. Findings are also similar in relation to the effects of alcohol in terms of the risk of stroke, whereby moderate drinkers are at less of a risk of stroke than abstainers and heavy drinkers have a higher risk of stroke (Berger *et al*, 1999; Mukamel, 2007; Klatsky *et al*, 2001).

Mukamal *et al* (2003) in looking at the effects of alcohol consumption on dementia find that moderate drinkers of 1-6 drinks per week have a lower odds ratio for dementia. The odds of abstainers contracting dementia are about twice as high as the odds of moderate consumers of alcohol. Heavy drinkers who consume 14 drinks or more in the week, have the highest odds of dementia occurring. Leibovici *et al* (1999) and Orgogozo *et al* (1997) find that moderate wine consumption has a protective effect in relation to the risk of Alzheimer's disease.

Walsh and Walsh (2011) find that in Ireland alcohol consumption is a significant influence on the suicide rate among younger males. Its influence on the female suicide rate is not well-established, although there is some evidence that it plays a role in the 15-24 age group. Higher alcohol consumption played a significant role in the very rapid increase in suicide mortality among young Irish males between the late 1980s and the end of the century (Walsh and Walsh, 2011). Walsh and Walsh (2011) argue that in the early twenty first century a combination of falling alcohol consumption and low unemployment led to a marked reduction in suicide rates and suggest that the recent rise in suicide rates may be attributed to the sharp rise in unemployment, especially among males, but it has been moderated by the continuing fall in alcohol consumption.

In terms of mental and physical health, Green (2001) finds that light to moderate drinkers of alcohol appear to be in better health, both mentally and physically, have better functional status and they are also more likely to engage in preventative health care services, compared with abstainers or heavy drinkers. Former drinkers were slightly less well off in terms of health and functional status than lifelong abstainers, which could possibly indicate that they stopped drinking due to ill health or declining health (Green, 2001).

Clearly through looking at previous studies into the effects of alcohol on health, there is a J or U shaped relationship between health status and alcohol consumption, showing that moderate consumers of alcohol tend to have better health status than non-drinkers or heavy drinkers.

In relation to alcohol consumption and the utilisation on health services, Dunlop *et al* (2000) observe that both males and females consuming between 1 and 11 drinks per week are more likely to make use of specialist services compared to those consuming zero drinks per week. In relation to those who had more than 12 drinks per week, females are less likely and males are more likely to attend a specialist when compared with either those who don't drink at all or those who had 1-11 drinks per week (Dunlop *et al* , 2000). Males who do not drink are more likely to have had 6 or more GP visits in the pervious 12 months when compared with male drinkers who have between 1-11 drinks per week and those who have 12 or more drinks per week. Female non-drinkers are more likely to have attended the GP when compared with those who drink between 1-11 drinks per week, however a female who has 12 drinks or more per week is more likely to have attended a GP 6 times or more in the previous year when compared with either non or moderate drinkers (Dunlop *et al* , 2000). In summary males who do not drink are most likely to attend the GP 6 times or more and males who had over 12 drinks per week are most likely to attend a specialist. Females who had over 12 drinks per week are more likely to attend a GP 6 times or more and female who had 1-11 drinks are most likely to attend a specialist (Dunlop *et al*, 2000)

In Ireland between 1995 and 2004 there were 139,962 alcohol-related hospital discharges (HRB, 2007). Males accounted for 75% (105,184) and women for 25% (34,778) of discharges. The number of discharges increased by 92% between 1995 and 2002 (HRB, 2007). The number of alcohol-related discharges peaked in 2002, and had decreased slightly (by 2%) by 2004. Alcohol-related discharges accounted for 874,395 bed days (including day and in-patients) between 1995 and 2004 (HRB, 2007). In 2004, alcohol-related discharges amounted to 117,373 bed days which represented 2.9% of all bed days that year. In 1995, the number of alcohol-related bed days was 55,805, accounting for 1.7% of all bed days (HRB, 2007).

Data from the Irish Hospital Inpatient Enquiry (HIPE) database show that almost 11,500 episodes of care provided in Irish public hospitals in 2005 had a discharge diagnosis relating to alcohol – this accounts for 1.14% of all episodes reported to

HIPE (HSE, 2007). Almost three-quarters of these episodes of care were to male patients (HSE 2007). A pilot study on the role of alcohol in Accident and Emergency Room attendance carried out in 2001 showed that alcohol was a contributory factor for one in four patients attending the A&E department (HSE, 2007). Barry and Skally (2011) found that in 2011 alcohol is associated with approximately 2000 beds being occupied every night in Irish acute hospitals.

Data from the National Psychiatric Inpatient Reporting System's (NPIRS) database showed that of the 22,279 admissions to psychiatric units and hospitals in 2004, 3,217 (14% of all admissions) were for alcohol disorders, the third highest after depressive disorders and schizophrenia (HSE 2007).

2.4.5.2: Education

Many studies have looked at the direct effect of health on education and almost all studies find that education strongly contributes to better health (Behrman & Wolfe, 1989; Berger & Leigh, 1989; Gilleskie & Harrison, 1998; Hartog & Oosterbeek, 1998; Kenkel, 1991, 1995; Leigh, 1998; Bassuk, Berkman, & Amick, 2002; Bopp & Minder, 2003; Mackenbach *et al*, 2004). Grossman (1972) also predicts that education contributes to a more efficient production of health and a decrease in the frequency of consultations. Similarly Kenkel (1995) finds that schooling improves productive efficiency and allows an individual to produce more health from the same set of health inputs. Kenkel (1995) also finds that schooling increases self reported health status and decreases activity limitation and restricted activity days, while controlling for differences in observed health inputs. Increased education can induce individuals to exercise regularly and to get regular health check ups however the affect of education is mainly through its effects on job characteristics (Park and Kang, 2008). Currie and Moretti (2003) find that education has a positive effect on women's choice of pre-natal care and has a negative effect on smoking during pregnancy. Rivera (2001) finds that an individual is more likely to value their health status as being good or positive when they have university qualifications.

Education and health are the two most important investments in human capital that individuals make in so far as both education and health make individuals more productive (Groot and Maassen van den Brink, 2007). Groot and Maassen van den Brink (2007) in analysing the effect of education on health in the Netherlands, test for causality between education and health and test whether the results are affected by scale of reference bias and unobserved heterogeneity, and the results are used to calculate the health returns to education. The study finds that the implied health returns to education are 1.3-5.5%. When a further test for causality between education and health and for the effects of scale of reference bias and unobserved heterogeneity is carried out the results are not affected.

Von dem Knesebeck *et al* (2006) analysed health inequalities in 22 European countries and find that educational inequalities in health are a generalised, though not invariant, phenomenon in European countries. There was two health indicators used, self-rated health and functional limitations, which showed relatively large inequalities observed for Hungary, Poland, and Portugal and small inequalities for Austria, Norway, Sweden, and the United Kingdom. In countries with a small inequalities effect, estimators suggest a trend towards impaired health in lower educational groups. Associations between education and self-rated health are stronger among women than among men in most countries. The study by Von dem Knesebeck *et al* (2006) finds associations between low education and ill health in 13 different age groups in all countries. Among men, significant associations of education with self-rated health until the age of 75 exist. Among women, education is significantly related to self-rated health until the age of 80 with one exception at ages 66–70. Especially among women, associations are strongest for the ages 25–55. In terms of functional limitations, education has a consistent and significant effect until the age of 55 in males and females. Among men, this effect diminishes in the higher age groups, whereas significant associations occur at ages 71–80 among women.

In terms of the effect of education on the utilisation of health care, findings tend to be that those with lower education tend to visit the general practitioner more (Habicht and Kunst, 2005; Jatrana and Crampton, 2009; Morris *et al*, 2005). Similarly Hernandez-Quevedo and Rubio (2009) find that as the level of education increases, the probability of visiting a GP and using hospital emergency services decreases, however they do note that the probability of visiting a specialist doctor increases. Other studies show that members of the population categorised into the lower social classes attend the GP more frequently than those in the higher social segments who have greater knowledge due to a higher level of education (Nolan 1994, McNiece 1999, Cooper *et al*, 1998). In contrast to this Dunlop *et al* (2000) find that in Canada those with a higher level of education were more likely to access GP services during a one year period. Gruber & Kiesel (2010) estimate that men's utilisation of health services increases about 10% for every additional year of education. McNamara *et al* (2013) in their study into health care utilisation in Ireland, use education as a socio-economic indicator in measuring health care utilisation, and find that those with lower levels of education have slightly more GP visits than these with higher levels of education.

2.4.5.3: Age

Grossman (1972) argues that the rate of depreciation for an individual's health capital stock increases with age. Older people tend to report poorer health (Lin, 2008; Yen *et al*, 2010). Yen *et al* (2010) find that not only are older people more likely to be in poor health, but that the magnitudes differ only slightly between smokers and non-smokers. Wilson *et al* (2011) explored differences in health status and health care use between older and younger Aboriginal people in Canada. The term Aboriginal is used to refer to the three broad indigenous groups in Canada that is the First Nations, Inuit and Métis populations. Wilson *et al* (2011) examine differences in health status and health care use between older (55 years and over) and younger (less than 55 years of age) aboriginal people as well as examining the relative importance of age as a determinant of health status/health care use. Findings are that as age increases, self-reported health status worsens with the older ages displaying worse levels of health than the young age group and the young to old age group. 15% of the population aged 18 - 54 years report their health as fair/poor in comparison with 38% of the population aged 55-64 and 45% of those aged 65 years and older. The percentage of the population aged 65 years and older reporting difficulty with activities or chronic illness is significantly higher (69%) as compared to the population aged 55-64 (50 %) and 18-54 years (26%). Only 6% of those aged 18-54 years report three or more chronic conditions in comparison with 26% of those aged 55-64 and 35% of those aged 65 years and older. These are set out in table 2.4.1.

Table 2.4.1: Findings from the study by Wilson *et al* (2011) of differences in health status across age groups of Aboriginal people in Canada

Age group	**% reporting fair/poor health**	**% reporting difficulty with activities or chronic illness**	**% with 3 or more chronic conditions**
18-54 yrs	15%	26%	6%
55-64 yrs	38%	50%	26%
65 yrs plus	45%	69%	35%

(Source: Authors own)

Kiuila and Mieszkowski (2007) use the core interviews of the US Health Interview Survey for the years 1987–1994, to study the effects of socioeconomic status (SES) on mortality and self-reported health, and conclude that general biological deterioration associated with aging is the most significant factor involved in the weakening of the socioeconomic mortality gradient at older ages. Kiuila and Mieszkowski (2007) find that as age increases a larger proportion of persons report themselves as being in poor health. The number of persons in fair and poor health becomes more evenly distributed throughout the income distribution. In contrast, at ages 25–44, those in fair and poor health, are concentrated among low income, poorly educated individuals. Robert and House (2000) indicate that the differences in health outcomes across socioeconomic groups are small at younger ages, but widen throughout middle age and narrow once more at older ages.

In Ireland, visiting rates to a GP vary remarkably with age, with an increase in age leading to an increase in visiting rates (Nolan, 1994). In a further study, Nolan (2007) in an analysis of the determinants of GP visiting in Ireland finds that GP visiting is an increasing function of age, although the effect is only significant after age 55 years. Age remains significant even after medical card eligibility and health status are controlled for, reflecting perhaps a greater awareness of good health, or lower opportunity costs, as age increases (Nolan, 2007). A study by McNamara *et al* (2013) find that in Ireland, age itself does not drive health care utilisation but that it is the characteristics that are associated with age that affect health care utilisation.

Jatrana and Crampton (2009) find that those who are of a younger age are associated with increased odds of deferring a doctors visit while Pohlmeier and Ulrich (1994) indicate that the length of treatment by a general practitioner, approximately measured as the number of visits to a physician, strictly increases with age.

2.4.5.4: Occupation

Several reports were commissioned by the Health and Safety Executive in the UK which provide information on the distribution of health and injures by work status (Llena-Nozal *et al*, 2004). In 2001, 2.3 million people in Great Britain suffered from an illness caused by their work or which was aggravated by it, accounting in total for 32.9 million workdays lost at work. The national statistics report shows that the most common type of work related illnesses were musculoskeletal disorders and stress, depression or anxiety, followed by breathing and lung problems and hearing problems. Llena-Nozal *et al* (2004) also find that work related stress varies by occupation and that occupational groups, such as teachers and nurses, have the highest prevalence rates.

Many studies into the effect of occupation on health status find that people with high occupational status have good health and low rates of premature mortality (Macleod *et al*, 2005; Marmot *et al*, 1984). Fischer and Sousa-Poza (2009) in looking at the impact of job satisfaction on the health of persons active in labour market using a national German data set, while accounting for cross-sectional causality problems, use both subjective and objective health measures and also analyse the effect of levels of and changes in, job satisfaction on changes in health. Fischer and Sousa-Poza (2009) find that self reported measures of heath are positively influenced by job satisfaction. In relation to objective measures of health, results show an unambiguously increasing effect of job satisfaction on health. With respect to more specific health problems, job satisfaction decreases the self reported impediment of certain daily activities and lowers the likelihood of medical treatment. In a study of health and gender differences between middle and senior managers in the Canadian Public Service, Tomiak *et al* (1997) find that senior managers tend to engage in healthy lifestyles and enjoy a number of health advantages, that is not smoking, having normal BMI and normal blood pressure, taking part in regular physical activity, reporting a good health status, seeing a doctor less often, and having fewer days incapacitated, in comparison with middle management.

Llena-Nozal *et al* (2004) consider the effect of work choices and changes in labour market status on mental health and look at whether this differs across occupations, with results showing that for females there are large differences from the effect of occupation; the higher the occupation the better the mental health. The quality of the job seems to be very important to females. For males no differential effects with respect to the type of occupation exist. Employment status is important for males and males who are out of the labour force have substantially worse mental health whereas this is not the case for females (Llena-Nozal *et al*, 2004).

Boyce and Oswald (2011) study the effect of an individuals job rank on ones health. A panel data set collected annually between 1991 and 2007 is used, with information on over 1000 individual occupational promotions to follow what happens to the health of those who gain seniority when compared to the health of those who are not promoted. Findings are that there is little evidence that a persons health improves after he or she is promoted. After a person gains seniority at work, the mental health levels of those who become managers typically worsen, and in a way that goes beyond short term change.

In relation to health care utilisation, findings vary with Gruber & Kiesel (2010) showing that women's health care utilisation does not vary significantly between those employed and those retired. Employed men however show a positive aging effect in health care utilisation and in retirement men are characterised by overall lower treatment intensity than employed men. Gruber & Kiesel (2010) also find that the self employed have a lower probability to consult with a doctor. Lin (2008) finds that working status variables are not a significant determinant of self-assessed health. Morris *et al* (2005) shows that students and those in full time education are less likely to go to the GP, whereas individuals looking after the home or family are more likely to visit the GP. McNamara *et al* (2013), find that in Ireland those who are working are slightly less likely to visit their GP in comparison to those who are not working, even when age and the health status are accounted for.

2.4.5.5: Income

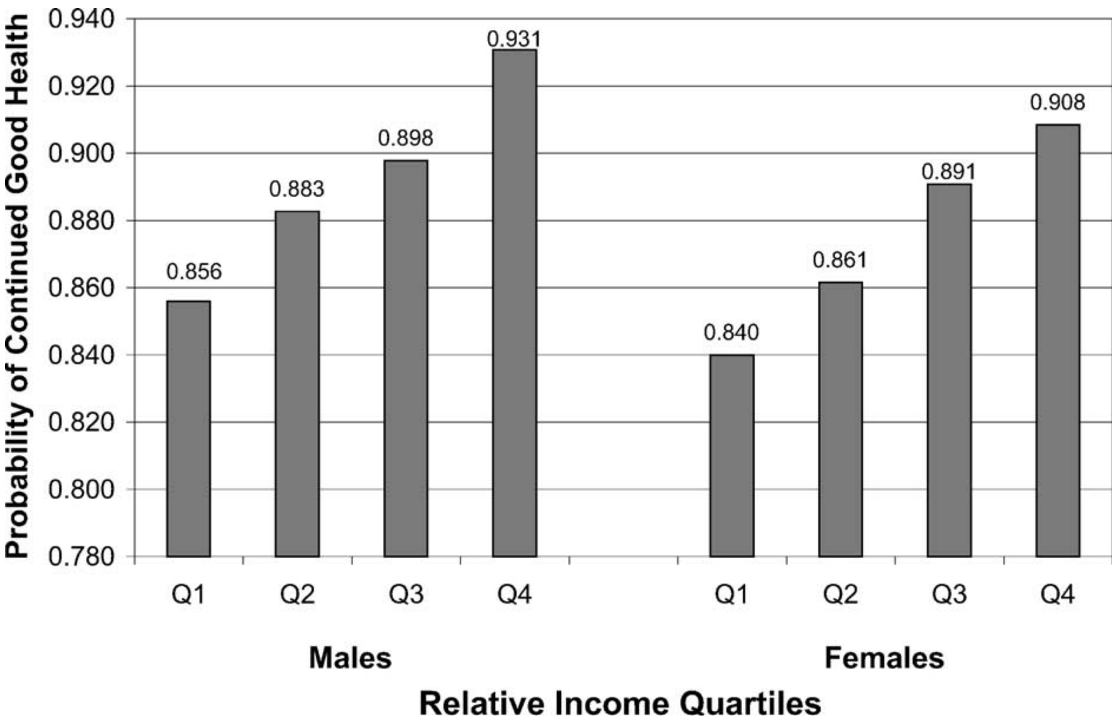
In the analyses of health status, much research has been carried out in the effects of income on ones health status and generally findings have been that those with lower income also reported a lower self-reported health (Yen *et al*, 2010). Similarly Tremblay *et al* (2002) also find that lower levels of household income is associated with worse health, as are smoking, obesity, and lack of frequent exercise. Ettner (1996) estimated the structural impact of income on health using instrumental variable estimates, and shows that income significantly improves both mental and physical health. Battel-Kirk and Perdy (2007) state that economic inequalities cause and exacerbate health inequalities at all levels, locally, regionally, nationally and internationally. As well as a wide gap between the health of rich and poor there is consistent evidence that the risk of poor health increases relative to lower position in the socio-economic scale. They also state societies with more equal distribution of income across the population have higher average life expectancies and better health outcomes than less equal societies. Battel-Kirk and Perdy (2007) give the example of the Scandinavian countries, which have comparatively equitable wealth distribution, there are smaller differences in health than countries which have a wider gap between the rich and the poor.

There is however a difficulty in assessing the direction of causation in the income – health relationship (Smith, 1999). The association between income inequality and health inequality is well established but the causality is less clear (Evans, 2002). The major difficulty in analysing the effect of income on health status is the problem of causality between both variables (Deaton, 2003). While the availability of resources may affect a person's ability to maintain good health, the person's state of health may in turn affect their ability to hold a job and earn income; therefore the endogeneity of income in a regression model needs to be accounted for (Deaton, 2003).

Frijters *et al* (2005) in investigating whether there was an effect of income on health satisfaction, use panel data of East and West Germany in the years following reunification on the basis that reunification was completely unanticipated and resulted in a rapid and exogenous rise in average household incomes in East Germany. Savings increased in real terms overnight, collectively bargained wages were at set levels far exceeding previous levels, many jobs in industry and government were suddenly much higher paid than before and yet there was no obvious immediate change in other health satisfaction producing circumstances. Evidence shows that increased income leads to improved health satisfaction but the quantitative size of the effect is small.

Buckley *et al* (2004) in their study into the effect of income on health status assessing the change in health as opposed to actual health state; model the change or constancy of health status in the subsequent years for which panel data observations are available as a function of the initial year income level and other variables also defined in the initial year. In essence what is being assessed is whether the persons propensity to stay healthy or to move into a poor health state, may be related to his/her household income and other state variables. Findings are that both men and women in the highest income quartile are more likely to remain in good health than those in the lowest.

Figure 2.4.2 Relationship between income and the probability of continued good health



(Source: Buckley et al, 2004)

Buckley *et al* (2004) in their assessment of income on health status, state that the effect in the other direction implies that income is an endogenous variable in the model. Requirements for a suitable variable to serve as an instrument in this case is that it be correlated with income, uncorrelated with the error term in the equation used to model the effects on health, and not itself an explanatory variable in the model. Buckley *et al* (2004) suggest level of education might be thought of as a possible choice since it is obviously highly correlated with income but education is also a candidate for explaining variations in health status among individuals, and so is disqualified if one wants to identify the separate effects of income and education. Buckley *et al* (2004) argue that it is difficult to identify a suitable instrumental variable for the purpose and that the endogeneity of income in a health-on-income regression model is a troublesome problem for anyone using one-time survey data for model estimation.

Establishing the direction of causality poses significant problems in assessing the association between health and income (Kiuila and Mieszkowski, 2007). Income is often viewed as a measure of resources available for the production of gross health-investments but poor health may cause lower incomes (Kiuila and Mieszkowski, 2007). Kiuila and Mieszkowski (2007) are unable to instrument for income in their study and cannot account for the unmeasured heterogeneity of individuals by specifying fixed effects, and the interpretation of the findings showing a positive association between income and health will necessarily remain ambiguous. However, as a partial control for the possible effect of poor health on low income, Kiuila and Mieszkowski (2007) include two variables about a persons employment status, one which shows if a person is employed and the other showing whether or not the person is in the labour force.

Other studies into the causal effect of income on health similarily find that that increased wealth has a very small positive effect on health (Meer *et al*, 2003; Case, 2001).

In terms of health care utilisation, many studies find that higher income is associated with a person more likely to visit a general practitioner (Dunlop *et al*, 2000; Habicht and Kunst, 2005). Hernandez-Quevedo and Rubio (2009) find that while income is positively associated with the probability of an individual contacting a specialist, it is negatively associated with the probability of a GP visit. Similarly, Van Doorslaer *et al* (2006) find that higher income people are more likely to seek specialist care than lower income people. Stewart (1990) argues that this may perhaps be due to the fact that poorer people may be less able to express their need for care and that perhaps those with a higher socio economic status may have different attitudes about the benefits that can be realised by assessing specialist care and may in turn be more motivated to seek opportunities by requesting specific physician visits, hence those on higher income can access and thereby benefit from the health care system more effectively.

2.4.5.6: Race

Generally studies have shown that the black race tends to have poorer health when compared to other races (Thompson, 2011; Thorpe *et al* , 2008). Thompson (2011) assesses racial health by looking at the Black and White distributions of self rated health. Thompson (2011) finds that while the White distribution is certainly more favourable overall than the Black distribution, most of this racial difference occurs in the middle range of health ratings, as opposed to the tails, that is very poor or excellent health. In particular, Black people are less likely to report the highest possible health rating of 5 (18.4% as compared to 23.5% for Whites). Thorpe *et al* (2008) find that the black race is associated with poorer functional status, independent of demographic and health-related factors even within a functionally limited subset of urban community resident women.

Stuber *et al* (2003) in a study into the self assessed mental and physical health among Latinos and blacks find that living in a highly segregated black neighbourhood was positively associated with poor physical health. Respondents living in highly segregated Latino neighbourhoods are less likely to report physical health problems. High racial and ethnic composition is associated with poor physical health among blacks and with poor mental health among Latinos. By contrast individuals living in disproportionately black neighbourhoods are less likely to report poor mental health (Stuber *et al*, 2003).

Morris *et al* (2005) find that non-white people are more likely to consult GP's relative to white people. Habicht and Kunst (2005) find that ethnic differences were generally very small in relation to health care utilisation, with no consistently higher use by one group. Jatrana and Crampton (2009) find that in New Zealand, Maori and Pacific people were more likely to defer a doctors visit compared with New Zealanders, Europeans and others.

2.4.5.7: Gender

Male respondents are more likely to have better self assessed health than females (Lin, 2008; Kwan, 2010). Liu (2008) in looking at the health status of the elderly population in China, focuses specifically on the health differences between men and women who are 60 years of age or older. Three measures of health status were used; one was self reported health status. The second was whether or not a person had a chronic condition and the third measure is activities of daily living which shows the independent living ability of the elderly and is seen as a general indicator of physiological and psychological health of the elderly. Results show that more males report good or very good health than females and that more females report poor or very poor health (Liu, 2008). In terms of the chronic illness of the aged population, Liu (2008) find that chronic disease rate is higher for females than for males in all age groups except the oldest and conclude that elderly women tend to suffer more from chronic diseases than elderly men. In terms of the activities of daily living, the older people are, the higher the disability and that the disability rate for females is higher than that of males. When lifestyle variables and social variables were controlled for, again elderly males report better health status than females (Liu, 2008).

Similarly Lahelma *et al* (1999) find that women have slightly poorer health than men however the magnitude of the differences in ill-health is relatively small, except for mental and somatic symptoms, and disabilities for respondents above 50 years. Female excess of ill health tends to change to one of male excess when more severe ill health among older people is assessed. Lianga *et al* (2003) state that women not only suffer from significantly more serious and chronic health conditions and functional limitations, but also rate their own health status poorer and women also manifest significantly more depressive symptoms and cognitive impairment.

In Ireland although women have a higher life expectancy than men, when life expectancy is expressed as years lived in good health (healthy life years), the difference between women and men is much less significant, indicating that women live longer but with more health problems (Ireland, 2012).

Many studies find that visits to general practitioners and health care utilisation are higher among females (Dunlop *et al*, 2000; Jatrana and Crampton, 2009, Nolan 1994, Tussing 1985). Hernandez-Quevedo and Rubio (2009) in their study into comparing health care utilisation patterns between foreigners and the national population in Spain find that looking at the interactions of age and sex, that 16-34 year old females have a higher probability of attending a general practitioner and other medical services and of being hospitalised than their male counterparts which Hernandez-Quevedo and Rubio (2009) say is probably due to healthy women availing of maternity related services. In Ireland females visit their GP more frequently than males, even when recent childbirth is taken into account (Nolan, 2007). In contrast to this, McNamara *et al* (2013) find that in Ireland there is very little difference between males and females only women are slightly less likely to use outpatient services.

2.4.5.8: Marital Status

There is considerable evidence that married individuals are healthier than single individuals (Rosengren, Wedel and Wilhelmsen, 1989; Zick and Smith, 1991). A possible explanation that has been suggested for this is the effect of marriage in reducing risky behaviours (Umberson, 1987). Consistent with this explanation is the evidence that individuals with spouses are less likely to smoke (Sloan, Smith, and Taylor, 2003).

Divorced, separated and widowed individuals are more likely than married respondents to report fair or poor health (Wilson *et al*, 2011). Similarly, Rivera (2001) finds that divorced or separated people are more likely to assess health negatively. Contrary to much of the findings in other studies, Kiuila and Mieszkowski (2007) show that married people over the age of 44 years report poorer health than people who have never been married, are divorced, separated or widowed. They are unable to find an explanation for this beyond the possibility that unmarried people who survive to old ages are in especially good health. Lin (2008) observe that being single is the most favourable marital status category for people in Taiwan in terms of health status.

Being married, divorced, widowed or living in a common law situation all had the effect of increasing the likelihood of making at least one visit to a GP (Dunlop *et al*, 2000). Jatrana and Crampton (2009) observe that being previously married was significantly associated with increased odds of deferring doctors visits. Similarly in Ireland Madden *et al* (2005) finds that those who are married or separated/divorced have a significantly higher number of GP visits, however McNamara *et al* (2013) find that martial status does not have an effect on the use of health care services in Ireland.

2.4.5.9: Number in Household

A large household size appears to be health protective, with Aboriginal people living in households with 3, 4, 5 or more people having a lower likelihood of being in fair or poor health than those Aboriginal people who live alone (Wilson *et al*, 2011). Similarly, Yen *et al* (2010) show that being in a larger household increases the chance of reporting excellent health. Rivera (2001) finds that if a person is living alone they are more likely to report health negatively.

In terms of health care utilisation, females with children under the age of 12 years residing in the household were more likely to become frequent users of primary care (Dunlop *et al*, 2000)

2.4.5.10: City or Rural

Wilson *et al* (2011) in their study into health status and health care use between older and younger Aboriginal people in Canada, find that Aboriginals who were living in a rural area are more likely to report fair/poor health than those living in urban areas. Contrary to this, Lin (2008) discover that people living in urban areas in Taiwan are more likely to report poorer health.

Findings on the correlation between geographical location of where people live and their likelihood of visiting a general practitioner is varied. Habicht and Kunst (2005) show that those living in rural areas are more likely to use the general practitioner, while contrary to this Dunlop, Coyte & McIsaac (2000) find that residents of urban communities make more visits to specialists as well as general practitioners than rural residents. Carr-Hill *et al* (1996) similarly show that those who are living further away or in a rural setting may be less likely to consult a general practitioner than a patient who is living in an urban area with greater access to a surgery, since the time taken to attend would be greater.

In Ireland, McNamara *et al* (2013) find that people living in Dublin are slightly less likely to visit both primary and secondary care services in comparison to those living in another town or city, or rural area in the Republic of Ireland. They state that the possible explanation for this may be the fact that there is a more acute deprivation concentrated in areas surrounding the city.

2.4.5.11: General Medical Scheme

The General Medical Scheme (GMS) provides care free at the point of use for the most economically deprived section of the population and the elderly (Teljeur *et al*, 2010). 37% of the population is now covered by a medical card under the GMS Scheme (Ireland, 2012). Numbers covered have increased by almost 45% over the decade and by nearly 5% between 2010 and 2011. This is in contrast to the numbers covered by private health insurance which has declined since 2008 (Ireland, 2012). Eligibility for the GMS scheme is determined on a means tested basis for under 70's and was available to all those over 70 from 2001 to 2008, but is subject to a means test since 2008 (HSE, 2013). Conversely the majority of the population pay full fees to access GP's and full costs for prescriptions (Teljeur *et al*, 2010). Although covering only 30% of the population, the GMS scheme accounts for 57% of GP income and is much valued by GP's as it is superannuated and attracts subsidies for staffing (Teljeur *et al*, 2010). Nolan (2007) states that despite the presence of a universal public health system, nearly 50% of the Irish population hold private medical insurance. However, this does not cover the cost of GP consultations (except where large deductibles are exceeded) and is primarily concerned with providing cover for private or semi-private hospital care (Nolan, 2007). The medical card system leads to a clear differential in the economic incentives facing those with a medical card and those without a medical card and one would expect this to lead to significant differences in GP utilisation (Nolan, 2007). Nolan (2007) finds that even after controlling for a variety of demographic, socio-economic and health status characteristics, those with medical cards have a significantly higher number of GP visits per annum.

Madden *et al* (2005) and McNamara *et al* (2013) has similar findings in that medical card eligibility has a consistently positive and significant effect on the utilisation of GP services in Ireland and that medical card patients have a significantly higher number of GP visits than private patients, even after controlling for a variety of demographic, socio-economic and health status characteristics.

2.4.5.12: Private Health Insurance

In 2004 almost half the Irish population were paying for private health insurance, one of the highest levels of coverage in the OECD (Nolan and Nolan 2004), which primarily covers the cost of in-patient and out-patient services in public and private hospitals but does not generally cover the cost of GP services, prescribed medicines or dental, ophthalmic and aural services except where large deductibles are exceeded (Nolan and Nolan, 2003). The numbers of people in Ireland covered by private health insurance has declined since 2008 (Ireland, 2012). The numbers of persons covered by private health insurance has declined to over 2.1 million in 2011 (Ireland, 2012).

Insurance is taken out primarily to ensure speed of access to hospital services and to guard against large medical bills (Harmon and Nolan, 2001).Nolan and Nolan (2003) find that having private medical insurance significantly increases the probability of visiting a GP but is insignificant in determining the frequency of visits. They argue that the result in terms of frequency of visits is not surprising given that private medical insurance in Ireland does not cover the cost of GP visits, except in cases where a large deductible is exceeded. Nolan and Nolan (2003) state that the significance of insurance in determining the contact decision may reflect differences in attitudes towards health care between the two groups with those covered by private medical insurance possibly more risk averse than those without. Nolan and Nolan (2003) also suggest that it is also possible that the GP realises that the patient is not covered by insurance for GP visits and therefore does not recommend follow-up visits. Nolan and Nolan (2003) also find that medical card eligibility has a larger affect than the insurance variable, reflecting the greater importance of medical card eligibility in influencing the decision to visit the GP, as private medical insurance does not cover the cost of GP consultations in Ireland except in cases where a large deductible is exceeded.

Harmon and Nolan (2001) and Hurd and McGarry (1997) either find that those in better health are more likely to be insured or no evidence for adverse selection. Hofter (2006) similarly finds that people with private health insurance tended to be healthier individuals.

2.4.5.13: Lifestyle Choices

Studies show that the greatest current potential for improving health is based on individual's lifestyle choices (Fuchs, 1986; Kenkel, 1995). Fuchs (1986) argue that while low level provision of food, hygiene and basic health care contribute to ones health status it is personal lifestyle choices that cause the greatest variation in health.

The World Health Organisation defines lifestyle as a 'general way of living based on the interplay between living conditions in the wide sense and individual patterns of behaviour as determined by sociocultural factors and personal characteristics'. In terms of lifestyle that affects ones health, Jones and Contoyannis (2004) define lifestyle as a set of behaviours which are considered to influence health a priori and are generally considered to involve a considerable amount of free choice.

In 1965 a study of the health practices of a sample of residents in Alameda county, California, was carried out (Kenkel, 1995). Following this study a number of practices were found to be associated with good health which were, never smoking cigarettes, moderate or no use of alcohol, maintaining proper weight, eating breakfast, not snacking between meals, regular physical activity and getting 7-8 hours sleep regularly (Kenkel, 1995). Each of these seven practices are associated with better health and those who report more of these practices are healthier on average than those who reported a lesser amount (Kenkel, 1995). Some follow up studies carried out, again find that most of the health practices from the Alameda Study were correlated with future health status and eventual mortality rates (Wiley and Camacho, 1980). Kenkel (1995) suggests that the fact that the US National Health Interviews Survey periodically collects data on the Alameda Seven is indicative of the previous studies which prove convincingly the importance of the seven health practices determining health status. Lin (2008) in looking at the effect of lifestyles on health in Taiwan shows that health status is more likely to be assessed poorly if the respondent adopts an unhealthy lifestyle such as smoking cigarettes or consuming alcohol. However in this study, Lin (2008) does not account for the potential endogenous relationship between health status and lifestyles.

Smoking

Cigarette smoking is the single most preventable cause of death in the world today (WHO, 2009). Worldwide it kills one person every 6 seconds, causes one death in 10 among adults, and claims more than 5 million lives annually (Mathers & Loncar, 2006; WHO, 2009). Smoking not only causes premature deaths but also leads to several diseases which may not necessarily kill a person but does affect health, such as chronic bronchitis, mucus hypersecretion, bladder cancer and peptic ulcer disease (Samet, 2001; Yen *et al*, 2010). More deaths are caused each year by tobacco use than by all deaths from human immunodeficiency virus (HIV), illegal drug use, alcohol use, motor vehicle injuries, suicides, and murders combined (Mokdad *et al,* 2004).

Despite such findings Yen *et al* (2010) show that cigarette smoking remains common throughout the world, with many countries having in excess of a quarter of its adult population smoking. China for example is the largest producer of tobacco and estimates suggest that 48.9% of men and 3.2% of women were current smokers in 2003 (Yen *et al,* 2010). In 2006, 53.3% of men were current smokers and 3.7% of women (Yen *et al,* 2010). In Ireland smoking is estimated to be the cause of approximately 7,000 deaths each year, chiefly by illnesses such as lung cancer, heart disease, stroke and emphysema (Department of Health and Children, 2011) and costs to provide health services for smokers are €1 Billion per year (Department of Health and Children, 2011).

Smoking is now identified as a major cause of heart disease, stroke, several different forms of cancer, and a wide variety of other health problems (Doll, 1986; Mattson *et al*, 1987). Manning *et al* (1991) estimate that smoking reduces the life expectancy of a 20 year old by about 4.3 years or 7 minutes per cigarette.

Cancers may begin to occur in people aged in their 30's if they have been smoking for 15-20 years (Holman *et al*, 1988). In Ireland, 90% of lung cancers are caused by smoking and 50% of all smokers will die from smoking related diseases (Department of Health and Children, 2011). In the UK, tobacco consumption is recognised as the single greatest cause of preventable illness and early death with around 102,000 people dying in 2009 from smoking-related diseases including cancers (Peto *et al,* 2012). Overall tobacco smoking is estimated to be responsible for more than a quarter of cancer deaths in the UK, that is, around 43,000 deaths in 2009(Peto *et al*, 2012). In the US, smoking causes an estimated 90% of all lung cancer deaths in men and 80% of all lung cancer deaths in women (America, 2004).

The effects of smoking on self assessed health have been widely studied with findings being that non-smokers are more likely to report good health than smokers (Ho *et al*, 2003; Contoyannis and Jones, 2004; Yen *et al*, 2010; Lin, 2008). Ho *et al* (2003) in their study show that in China those who had never smoked had better perceived health than those who were currently smoking, however those who had been previous smokers and had quit had the worst perceived health, for both genders. In a study of the effect of lifestyle behaviours on the effect Self Assessed Health, Contoyannis and Jones (2004) find that non-smoking has a large positive effect on the probability of reporting excellent or good health. Yen *et al* (2010) using data from the 2006 China Health and Nutrition Survey to look at the effect of cigarette smoking on self assessed health in China while accounting for the endogenous relationship between smoking and health, find that non-smokers had better perceived health than those currently smoking.

Jones (1996) shows that those with poor or fair Self Assessed Health are less likely to have quit smoking than those with better health and that those who experienced serious injury or illness at the end of the period of analysis were more likely to quit smoking.

Jatrana and Crampton (2009) observe that current smokers were significantly associated with increased odds of deferring doctors visits, however Dunlop *et al* (2000) find that smoking was not a significant factor in the number of times an individual visited a General Practitioner and also Dunlop *et al* (2000) find that smoking is not a significant factor in a person visiting a Specialist, but is a significant factor whereby a person went to a specialist 6 or more times in the year previous; in this case a smoker was more likely to have visited a specialist six or more times compared with a non-smoker.

Exercise

According to the World Health Organisation (2010) physical inactivity is now identified as the fourth leading risk factor for global mortality. Physical inactivity levels are rising in many countries with major implications for the prevalence of noncommunicable diseases (NCDs) and the general health of the population worldwide. In 2010, the WHO developed the "Global Recommendations on Physical Activity for Health" with the overall aim of providing national and regional level policy makers with guidance on the dose-response relationship between the frequency, duration, intensity, type and total amount of physical activity needed for the prevention of NCDs. The World Health Organisation (2010) state that overall, strong evidence demonstrates that compared to less active adult men and women, individuals who are more active:

- have lower rates of all-cause mortality, coronary heart disease, high blood pressure, stroke, type 2 diabetes, metabolic syndrome, colon and breast cancer, and depression
- are likely to have less risk of a hip or vertebral fracture
- exhibit a higher level of cardiorespiratory and muscular fitness
- are more likely to achieve weight maintenance, have a healthier body mass and composition

Exercise is primarily a primary preventative behaviour for most chronic diseases (Honda, 2004). Much research has been carried out into the effect of exercise on different diseases or health problems and the general finding are that exercise helps many serious conditions and overall general health (Honda, 2004). Individuals are more likely to value their health status as being good or positive when they do exercise in their leisure time (Rivera, 2001). Paffenbarger (1996) states that physical fitness and exercise can reduce the risk of diseases such as heart disease, non-insulin-dependent diabetes mellitus, some cancers, osteoarthritis and osteoporosis, and obesity.

Ransford and Palisi (1996) examine the relationship between different forms of aerobic exercise (swimming, walking, jogging and dancing) and two measures of health, subjective health and psychological well being. In essence what is being looked at is whether the relationship between exercise and health is more pronounced within age and gender sub-groups. Findings from Ransford and Palisi (1996) study is that exercise and health measures are most strongly correlated among older respondents and that among the older respondents the no-exercising group was much less likely to define their health as good or excellent health. In relation to young people, they are more likely to define their health as good or excellent regardless of exercise involvement (Ransford and Palisi, 1996). In describing and identifying the self-assessed predictors of physical and mental health of nurses, Sveinsdottir and Gunnarsdottir (2008) find that of the nurses who rated their self assessed physical health as good or very good compared with those who rated theirs as poor or very poor, a higher proportion of them reported exercising at least three times a week.

Individuals reporting physical inactivity are significantly more likely to visit a specialist 6 or more times than those reporting physical activity, while physical activity is not significant in terms of visiting a general practitioner (Dunlop *et al*, 2000).

Weight

The World Health Organisation (2011) defines overweight and obesity as abnormal or excessive fat accumulation that presents a risk to health. They state that a crude population measure of obesity is the body mass index (BMI), a person's weight (in kilograms) divided by the square of his or her height (in metres). A person with a BMI of 30 or more is generally considered obese. A person with a BMI equal to or more than 25 is considered overweight (WHO, 2011). The World Health Organisation (2011) state that being overweight and obese are major risk factors for a number of chronic diseases, including diabetes, cardiovascular diseases and cancer. Once considered a problem only in high income countries, overweight and obesity are now dramatically on the rise in low- and middle-income countries, particularly in urban settings. Worldwide obesity has more than doubled since 1980 (WHO, 2011). In 2008, 1.5 billion adults, 20 years and older, were overweight and of these over 200 million men and nearly 300 million women were obese and in 2010, nearly 43 million children under the age of five were overweight (WHO, 2011).

Health Status is more likely to be assessed poorly if the respondents adopt an unhealthy lifestyle such as having a body mass index (BMI) of over 30 (Lin, 2008). In his study, Lin (2008) finds that males who are overweight are more likely to be in poor heath and for females the effects are not significant (Lin, 2008). Contoyannis and Jones (2004) measure obesity on BMI and describe males with a BMI of below 30 and females with a BMI of below 28.6 as not being obese. Contoyannis and Jones (2004) find that not being obese has a positive effect on the probability of reporting excellent or good health. Kenkel (1985) shows that excessive weight is a harmful input in the health production function.

For both overweight and obese people the probability of GP and indirect costs is significantly higher compared with normal weight participants (Wolfenstetter, 2011). In Ireland over half the adult population are now considered overweight or obese (Doherty *et al,* 2012). Using the Slán 2007 survey Doherty *et al* (2012) find that overweight and obesity are significant predictors of GP utilisation and obesity is a significant predictor of inpatient episodes.

Williamson *et al* (1987) in their study into alcohol and weight, finds that, among men alcohol only has a slight effect on weight and among women, drinkers weighed less than non-drinkers. Williamson *et al* (1987) do argue that further studies are needed to understand the causal mechanisms by which alcohol is associated with body weight.

Many variables affect health status and health care utilisation such as alcohol, gender, age, education among others. Table 2.4.2 below summarises these variables.

Table 2.4.2 Summary of the variables found to have various effects on health status and health care utilisation

Variables	**Variables affecting health status**	**Variables affecting health care utilisation**
Alcohol	Yes	Yes
Education	Yes	Yes
Age	Yes	Yes
Occupation	Yes	Yes
Income	Yes	Yes
Race	Yes	Yes
Gender	Yes	Yes
Marital Status	Yes	Yes
Number in Household	Yes	Yes
City or Rural	Yes	Yes
General Medical Scheme	-	Yes
Private Health Insurance	Yes	Yes
Lifestyle Choices		
Smoking	Yes	Yes
Exercise	Yes	Yes
Weight	Yes	Yes

(Source: Authors own)

2.4.6: Endogeneity of health inputs

Endogeneity occurs when an independent variable included in the model is potentially a choice variable, and variables can be jointly determined which as a result leads to correlation between the unobservables and the disturbance term (Chenhall and Moers, 2007). The concept of endogeneity has already been discussed in section 2.1.4 and selection bias in section 2.1.5.

Kenkel (1995) and Rosenzweig and Schultz (1983) state that the observed choices of health inputs such as alcohol, smoking, exercise among others, are the result of an individuals optimising behaviour and hence can be endogenous which could lead to biased estimates of the relationships between health inputs and health outcomes.

Many of the studies such as (Contoyannis and Jones, 2004) that look at the effect of lifestyle variables on health status while accounting for the endogenous relationship between the two, use panel data. Contoyannis and Jones (2004) considering the role of several lifestyle variables in terms of self assessed health, use original data in the study which is then supplemented by follow-up panel data. Advantage is taken of the exogenous variables from the follow up data to model the lifestyle variables from the earlier data, and use lifestyle variables to explain self assessed health.

In looking at the relationship between health and addiction models, typically the instruments used are regional price variations (Clark and Etile, 2002; Leigh and Schembri, 2004; Mityakov and Mroz, 2011). In a study into the effects of smoking on physical functional status, Leigh and Schembri (2004), use the instrument cigarette price. Price per pack data from 50 American States was matched to persons who resided in those states on that basis. Leigh and Schembri (2004) state that cigarette price is logically related to and strongly correlated with smoking; higher prices result in less smoking on average but on the other hand the price of cigarettes is not logically related to an individuals health. Price is commonly used in studies using American data such as Leigh and Schembri (2004) however Irish alcohol and cigarette prices do not vary systematically by region, as there are no regional level taxes (Ireland, 2011).

As well as price, access to health inputs is often used as the basis for identifying the health input demands (Schultz, 2005). Characteristics of parents have also been used as instruments (Strauss and Thomas, 1998). Kenkel (1995) suggests that input prices, the individual's income and individual characteristics related to tastes could be included in the input demand equations but excluded from the health production function. Kenkels (1995) study into the effect of lifestyle variables on health in the US using the 1985 Health Interview Survey, used the variables income, input prices, marital status, employment and occupation dummies specific to the health input demand function and using two stage model estimated the health production function however this yielded much less reasonable results than where the inputs were treated as exogenous (Kenkel, 1995). Contrary to this many other studies find that income should not be excluded from the health equation (Yen *et al*, 2010).

Yen *et al* (2010) state that while endogeneity of smoking and health status is accounted for in the study there are a number of regressors such as drinking and exercise, that may be potentially endogenous, but the lack of viable instruments does not allow further exploration of the potential endogeneity of these variables. However Yen *et al* (2010) find that results of an alternative model without these variables produce few discernable differences in the treatment effects and marginal effects of other explanatory variables, based on the current sample. Similarly Kenkel (1995) states that the difficulty in relation to endogenous health inputs is the lack of suitable instruments for the input demand, and that while two stage models have been identified to address the problem of endogeneity, the lack of suitable instruments mean that the model is not very powerful. Kenkel (1995) finds that the two stage models yielded much less reasonable results than alternative models.

Alcohol Consumption is defined as a health input by Kenkel (1995) and is potentially endogenous. The lack of suitable instruments is a major difficulty in terms of accounting for endogeneity (Kenkel, 1995). In the estimation of alcohol on income, Hamilton and Hamilton (1997) and Barrett (2002) , estimate wage equations by drinker type in order to account for the endogenous relationship between income and alcohol.

2.4.7 Selection Bias of Health Inputs

As was set out in section 2.1.5, alcohol status can cause selection bias to arise whereby individuals self select into certain drinking categories on the basis of individual characteristics (Hamiltion and Hamilton, 1997). Various generalisations of the Heckman (1979) two step estimator which accounts for selection bias have been developed. Where the dependent variable in the selection equation is ordered, methods of estimating such data by an ordered probit estimation in the first step of the two step model, from which an inverse mills ratio can be derived have been set out by Chiburis and Lokshin (2007), Vella (1998) and Greene and Hensher (2010). Where the data in the primary equation is ordered, and where this is to be estimated in step two of the two step model and includes the inverse mills ratio as an additional regressor, this is set out by Greene and Hensher (2010 and Langpap and Kerkvliet (2002).

Step 1 – Estimation of Alcohol Status Equation

The selection equation 2.4.5 similar to that set by Chiburis and Lokshin (2007), Vella (1998) and Greene and Hensher (2010) assumes that the independent variables s_i and the categorical variables c_i are observed. Individuals i are sorted into J categories of 1,2,3 on the basis of an ordered probit selection rule.

$$c_i^* = \alpha' s_i + \varepsilon_i \qquad \textbf{(2.4.5)}$$

Where:	c	category of ordered outcomes
	α	is an unknown vector of parameters
	s	independent variables
	ε	is a standard normal shock
	i	indexes individuals where $i = 1,....,n$
	n	sample observations

The ordered probit of c on s is estimated yielding a consistent estimation of α . There should be at least one additional variable that is unique to the selection equation that is not included in the main equation of interest (Chiburis and Lokshin, 2007; Vella, 1998). Level of choice is based on its c_i value relative to the cut off points which are maximum likelihood estimates from the selection equation. By estimating the selection equation, an estimation of λ_i is then computed for each individual in the sample which will allow a consistent estimate of β_j to be estimated (Chiburis and Lokshin, 2007; Vella, 1998)

$$\hat{\lambda}_i \equiv \frac{\phi(\hat{\mu}_J - \hat{\alpha}' s_i) - \phi(\hat{\mu}_{J+1} - \hat{\alpha}' s_i)}{\Phi(\hat{\mu}_{J+1} - \hat{\alpha}' s_i) - \Phi(\hat{\mu}_J - \hat{\alpha}' s_i)} \qquad \textbf{(2.4.6)}$$

where $j = c_i$

By defining $\hat{c}_i^* \equiv \hat{\alpha}' s_i$

$$= \frac{\phi(\hat{\mu}_J - \hat{c}_i) - \phi(\hat{\mu}_{J+1} - \hat{c}_i)}{\Phi(\hat{\mu}_{J+1} - \hat{c}_i) - \Phi(\hat{\mu}_J - \hat{c}_i)} \qquad \textbf{(2.4.7)}$$

Where:	α	unknown vector of parameters in the selection equation
	s	independent variables in the selection equation
	μ_J	cutoffs
	c	category of ordered outcomes
	ϕ	probability density function
	Φ	cumulative distribution function
	j	indexes outcome category where $j = 1,2,3,$
	i	indexes individuals

Step 2 – Estimation of Health Status Equation

Greene and Hensher (2010) consider a model of estimating educational attainment in step two as an ordered probit. The primary equation is estimated by an ordered probit regression and λ, derived in step one, is also included in this equation as an additional regressor (Greene and Hensher, 2010).

$$h_i * = \beta_i' x + u_i \qquad \textbf{(2.4.8)}$$

$h_i = k$ if $\mu_{k-1} < h_i^* \leq \mu_k$

Where:	h	dependent variable in primary equation
	β	coefficient on the observable characteristics
	x	vector of independent variables
	u	error term
	μ	cutoffs
	k	indexes outcome category
	i	indexes individuals where $i = 1,2,......N$

Greene and Hensher (2010) state that h, x, the level of education attainment and the independent variables, are observed when individuals select into the programme. Corresponding probabilities that each category is observed is given by

$$P[h_i = k] = \Phi(\mu_k - x_i\beta) - \Phi(\mu_{k-1} - x_i\beta) \qquad \textbf{(2.4.9)}$$

The log likelihood for the ordered probit estimation of the primary equation is as follows:

$$\log L = \sum_{z=0} \log \Phi(-\alpha' s) + \sum_{z=1} \sum_{j=0}^{K} m_{ik} \log\left[\Phi_2(\mu_k - \beta' x, \alpha' s, \rho) - \Phi_2(\mu_{k-1} - \beta' x, \alpha' s, \rho)\right]$$

(2.4.10)

where $m_{ik} = 1$ if $h_i = k$

Where:	α	an unknown vector of parameters in the selection equation
	s	independent variables in the selection equation
	β	coefficient on the observable characteristics in primary equation
	x	vector of independent variables in primary equation
	Φ	cumulative distribution function
	μ	cutoffs
	ρ	correlation of the error terms
	k	indexes outcome category
	i	indexes individuals where $i = 1,2,......N$

The selection correction term estimated in Step 1, is included in the estimation of the primary equation of interest, in step two to account for the potential selection bias.

2.4.8: Conclusion

This section reviewed the different definitions of health that exist and the correlation between health status and healthcare utilisation. The most common measure of ones health is Self Rated Health (Jurges, 2008; Kiuila and Miesztowski, 2007) and the effectiveness of this as a measure of health is looked at. Grossman's human capital model of the demand for health provides a greater understanding of the determinants of health. Previous literature into the different factors that affect both health status and health care utilisation is analysed and the possible endogeneity that may exist with some of these factors. The econometric techniques that could be used to look at the effect of a lifestyle variable, such as alcohol on health status and health care utilisation while accounting for endogeneity and selection bias is examined

2.5: Conclusion

This chapter reviewed previous literature in relation to defining and categorising alcohol consumption. Previous studies into the effect of alcohol consumption on income are looked at, and in particular the problem of potential endogeneity and section bias that may arise in such an estimation. Possible econometric techniques to overcome these estimation difficulties are assessed. The various factors that affect both alcohol consumption and income are also reviewed.

Alcohol Status could also be interpreted as ordered data (Harris *et al,* 2006). Previous studies into the effect of alcohol consumption on income have not taken account of this ordinality (Hamilton and Hamilton, 1997; Barrett, 2002), and if ordinality is ignored then this may lead to a loss of efficiency and an increased risk of getting insignificant results (Harris *et al*, 2006). Methods to measure the effect of alcohol status on income, accounting for the ordered nature of alcohol status, and accounting for the potential endogeneity and selection bias are assessed.

In such estimations, limited or full information methods of estimation can be carried out and both these methods are assessed. Generally findings have been that the Full Information methods of estimation are better however they are more computationally challenging (Gujarati, 2004; Puhani, 2000).

The Grossman Model which is an economic model of the determinants of health is reviewed. Grossman (1972) states that health status impacts the human capital model of earnings determination and those with a higher income receive a higher return from investing in health status. Much research has been carried out into the effect of alcohol on health with findings being that the relationship is similar to that of alcohol and income (French and Zarkin, 1995; Heien,1996; Hamilton and Hamilton, 1997; Barrett, 2002). Barrett (2002) goes onto to argue that the potential impact of health status on earnings represents a straightforward extension of the human capital framework of earnings determination and that alcohol consumption can influence ones health status, the consequences of which can influence ones productivity at work which may ultimately be reflected in an individuals earnings.

Previous literature into the affect of alcohol on health is reviewed and the possible econometric techniques that could be adopted to carry out such an estimation. Health status and health care utilisation are defined and the factors affecting both health status and health care utilisation analysed, given that studies show that a correlation between the two exist (Dunlop *et al*, 2000; Laroche, 2000) and hence both could be used as measures of health.

CHAPTER 3

DATA DESCRIPTION

The study into the effect of alcohol consumption on income, on health status and on health care utilisation uses data from the 2007 Slán Survey of the lifestyle, attitudes and nutrition of people living in Ireland. This chapter describes the Slán survey and gives a description if the data available in Slán. The variables used in this study are described with descriptive statistics also provided.

3.1: Data description

In order to identify the impact of alcohol on the household income, this paper uses data from the Slán National Health and Lifestyle Survey. This cross sectional survey is commissioned by the Department of Health and Children in Ireland. The survey and analyses were carried out by the National University of Ireland, Galway along with the Consortium consisting of the Royal College of Surgeons in Ireland, National University of Ireland, Cork and the Economic and Social Research Institute (ESRI). It surveys a cross section of the Irish adult population, aged 18 and over. Surveys have been carried out in 1998, 2002 and 2007 (Morgan *et al*, 2008).

The Slán Survey aims to:

- produce reliable data of a nationally representative cross-section of the Irish population in order to inform the Department of Health and Children in terms of policy and programme planning
- maintain a survey protocol which will enable lifestyle factors to be measured and re-measured which will allow for trends and changes to be identified
- allow direct comparisons to be made with the heath related behaviours of other countries that carry out similar surveys

The 1998 and 2002 surveys were sent to a random sample of people from the electoral register however in relation to the 2007 survey this was no longer the case due to data protection legislation; hence the GeoDirectory was used instead to provide a random sample for the 2007 survey (Morgan *et al*, 2008). The GeoDirectory is a list of all addresses in the Republic of Ireland, complied by An Post, which distinguishes between residential and commercial establishments (Morgan *et al*, 2008). Surveys were conducted by face to face interviews over a period of time. The survey was weighted to match the 2006 Census. The purpose of survey weighting is to compensate for any imbalances in the distribution of characteristics in the completed survey sample compared to the population of interest. This involved a weight being constructed to compensate for the over representation of individuals in smaller households. Calibration of the sample distribution to population totals along the dimensions; age group by gender, age group by marital status, gender by economic status, gender by level of education, occupational category, ethnicity, household size, geographic region. The weighted sample very closely approximated Census 2006 figures for gender, age, marital status and ethnicity. Prior to weighting, the data would have under-represented the groups that are typically hard to reach in surveys such as men and young single adults (Morgan *et al*, 2008). The characteristic of the Slán 2007 sample compared to the characteristics of the population from the 2006 Census is depicted in table 3.1.1.

Table 3.1.1: Characteristics of SLÁN 2007 sample compared to characteristics of population from Census 2006

		NUMBER OF CASES	UNWEIGHTED SAMPLE %	WEIGHTED SAMPLE %	CENSUS 2006
Age group	18-29 years	1907	18	25	26
	30-44 years	3310	32	31	30
	45-64 years	3178	31	29	29
	65 years and over	1969	19	15	15
Gender	Men	4369	42	50	50
	Women	5995	58	50	50
Marital status	Single (including cohabiting)	3,602	35	41	40
	Married	5211	50	48	49
	Separated or divorced	639	6	4	5
	Widowed	912	9	7	6
Country of birth	Ireland	8820	85	83	85
	Northern Ireland	116	1	1	1
	Other UK	644	6	6	5
	Other EU-27	376	4	5	4
	Other Europe	24	0	0	1
	Africa	96	1	1	1
	USA, Canada, South America	67	1	1	1
	Elsewhere or unknown	221	2	3	2
Ethnicity	White or white Irish	9333	90.0	87.0	87.0
	Irish Traveller	31	0.3	0.4	0.4
	Any other white background	634	6.1	8.0	8.0
	Black or black Irish; African	60	0.6	0.7	0.7
	Any other black background	19	0.2	0.1	0.1
	Asian or Asian Irish; Chinese	32	0.3	0.4	0.4
	Any other Asian background	62	0.6	0.9	0.9
	Other including mixed ethnic background	71	0.7	1.0	1.0
	Unknown	122	1.2	1.5	1.5

(Source: Slán 2007 Report)

The 2007 dataset is the largest survey ever to date (Morgan *et al*, 2008). A scientifically representative random sample of 10,364 respondents (a 62% response rate), aged 18 years and over were interviewed in their own homes, by experienced researchers from the Economic and Social Research Institute (ESRI). In terms of the non-response rate 1,868 individuals refused to respond, 735 individuals are other non-responders which includes cases where the respondent was too ill or temporarily away, 273 are classified as other not eligible and includes addresses that were non-residential, 3,714 are non-contact and 2,231 are not eligible both of which are adjusted for the percentage of dwellings that were vacant according to the Census 2006 figures.

In addition to this, over 1,200 people, aged 45 years and over, who participated in the survey also participated in a detailed medical examination and 967, aged 18 to 44 years, agreed to the measurement of their body mass index (BMI)/waist circumference. The survey covers general health, behaviours relating to health such as alcohol consumption, exercise, nutrition, and the use of health services. The two previous surveys were not as extensive and the number of participants was far less with 6,539 respondents in 1998, 5,992 respondents in 2002 compared with 10,364 respondents in 2007. In the 2007 survey, there are nine sections in the questionnaire which cover general health (including reported height and weight), mental health and well being, physical activity, diet and nutrition, smoking, alcohol and other substances, injury, family-social networks & neighbours, and body weight and waist measurement (Morgan *et al*, 2008). The survey has provided vital baseline data on a range of lifestyle related health behaviours.

The Slán dataset includes responses which have some date missing. In this study complete cases only are used.

General findings in the 2007 survey were that self rated health was recorded as excellent or very good by over half of the sample (58%), with very few (3%) reporting their health as poor (Morgan *et al*, 2008). One tenth of respondents (11%) reported a long term illness, health problem or disability that limited their daily activity. One quarter of respondents aged 65 years or over reported a chronic long term condition. The most common chronic illness in the past year was back pain, with 16% reporting this problem (Morgan *et al*, 2008).

Almost three quarters of respondents (74%) had visited a general practitioner in the past year. Older people were more likely to visit the GP. There were no social class differences in attending a GP at least once in the last year. In relation to physical activity over half the respondents (55%) reported being physically active, with 49% having reported being physically active for more than 6 months. Almost a quarter reported some physical activity but not at the level great enough to be considered physically active (Morgan *et al*, 2008).
In relation to alcohol, most men (85%) and women (77%) drank alcohol on some occasions. One quarter (28%) reported excessive drinking in the last year. This was more common in younger respondents (Morgan *et al*, 2008).

3.2: Variables used in the estimation of alcohol status and income

There are many human capital variables and socio demographic variables that affect both alcohol consumption and income as well as some additional variables that can influence levels of alcohol consumption only. Both the dependent and independent variables are described below.

Dependent Variables

Income
In the Slán 2007 survey individual earnings is not measured, it is total income of the household that is reported. Individuals are presented with different income bands and are asked to select which income band is appropriate to for their household in terms of the household's total net income per week. The total net take home pay includes all sources of family income including social benefits. In the 2007 survey there are twenty-four categories of income given ranging from the lowest category of less than €86 per week to the highest of €1,535 or more per week. Number working in household is included as a control variable.

For the purpose of econometric analysis in this paper, the descriptive statistics for income were derived by taking the midpoint of an individual's income category similar to what Barrett (2002) did in his study and for the open upper category, a value of 10% above the lower income limit of the band, was taken (Von Fintel, D., 2007). These are set out in Table 3.2.1.

Analysis of household income is beneficial in so far as there is a huge correlation between the drinking habits of different individuals within families and within households and in looking at the alcohol consumption of a particular member of the household, it is very likely that other members of the household may have similar drinking patterns (O' Farrell, 1995, Cadoret *et al*, 1995)

Table 3.2.1: Analysis of respondents in each income category

Log Income	Income	No. of Respondents	Percent	Cumulative Distribution
3.76	42.95	68	0.79	0.79
4.58	97.51	37	0.43	1.23
4.88	131.63	58	0.68	1.90
5.15	172.43	265	3.09	4.99
5.38	217.02	679	7.92	12.92
5.58	265.07	346	4.04	16.95
5.74	311.06	289	3.37	20.33
5.89	361.41	380	4.43	24.76
6.01	407.48	479	5.59	30.35
6.12	454.86	304	3.55	33.9
6.22	502.70	357	4.17	38.06
6.31	550.04	356	4.15	42.22
6.40	601.85	396	4.62	46.84
6.47	645.48	242	2.82	49.66
6.54	692.29	333	3.89	53.55
6.61	742.48	418	4.88	58.42
6.67	788.40	358	4.18	62.6
6.73	837.15	234	2.73	65.33
6.79	888.91	363	4.24	69.57
6.84	934.49	374	4.36	73.93
6.96	1,053.63	634	7.4	81.33
7.13	1,248.88	424	4.95	86.28
7.27	1,436.55	395	4.61	90.89
7.43	1,685.81	781	9.11	100
Total		8,570	100	

(Source: Authors own)

In looking at the numbers of respondents in each category of income the lowest categories of incomes hold the least numbers of respondents. The largest group of respondents (9.11%) report being in the highest category of income of €1,535 or more per week. The second highest category of respondents report having household income of between €193 and €240 per week.

Drinking Status

In the drinking status equation, drinking status is the dependent variable. Drinking Status consists of three categories, non-drinkers, moderate drinkers and heavy drinkers.

The Irish Health Promotion Unit (HSE, 2008) state that while there are safe levels of drinking, the low risk weekly limits for women is up to 14 standard drinks in a week and for men up to 21 standard drinks in a week and on any one occasion drink no more than 4 standard drinks for women and 6 for men. They define binge drinking as having more than 6 standard drinks at a time. Respondents are categorised based on recommendations from the Irish Health Promotion Unit (HSE, 2008).

Using data from the 2007 dataset moderate drinkers are defined as those who had a drink in the last month or in the week prior to the survey any women who had up to 14 standard drinks and men who had up to 21 standard drinks. Heavy drinkers are women who drank more than 14 drinks in the week prior to the survey and men who drank more than 21 drinks and non-drinkers are those who do not drink or did not have a drink in the month prior to the survey. The dummy variables for the three categories of drinkers are established based on a number of questions in relation to one’s alcohol consumption in the Slán survey.

One of those questions was

'*how long ago did you last have an alcoholic drink?'*

a. During the last week
b. During the last month, but not in the last week
c. Within the last three months, but not in the last month
d. Within the last 12 months, but not in the last 3 months
e. More than 12 months ago
f. Never had alcohol beyond sips or tastes

Those respondents who are part of the non-drinker category can be clearly determined i.e. those who answered c, d, e or f. Those who answered b are moderate drinkers and those who answered 'a' could fall into either the moderate or heavy drinker categories. To categorise these respondents correctly the following question from the Slán survey was used:

'During the past 7 days how many standard drinks of any alcoholic beverage did you have each day?'

This allows the categorisation of respondents who stated that they had a drink in the last week, to be classified as either a moderate or heavy drinker.

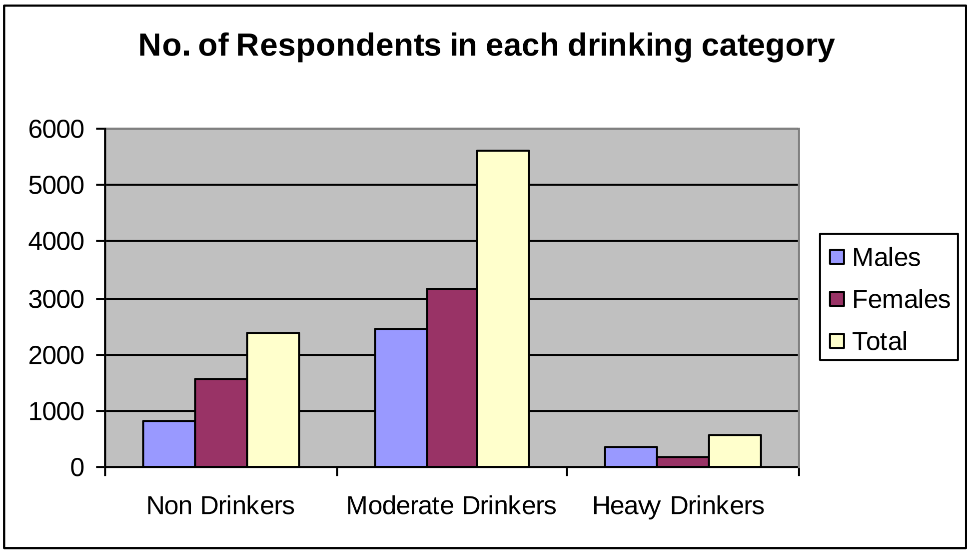
Table 3.2.2 Breakdown of Respondents across drinking categories

	No. of Respondents	% of respondents in each drinking category
Male non-drinker	834	9.73
Female non-drinker	1,557	18.17
Total non-drinkers	**2,391**	**27.9**
Male moderate drinkers	2,458	28.68
Female moderate drinkers	3,168	36.97
Total moderate drinkers	**5,626**	**65.65**
Male heavy drinkers	371	4.33
Female heavy drinkers	182	2.12
Total heavy drinkers	**553**	**6.45**

(Source: Authors own)

The largest group of respondents i.e. 65.65% of respondents to the survey are classified as moderate drinkers, and of those moderate drinkers there are more females than males. 27.9% of respondents are non-drinkers, again the largest group of respondents being female. In terms of heavy drinkers only 6.45% of respondents are in this category, with double the amount of males than females reporting being a heavy drinker. These figures are also depicted in figure 3.2.1.

Figure 3.2.1 Breakdown of Respondents across drinking categories



Category of Drinker

(Source: Authors own)

The Independent Variables

The Slán survey includes a large number of socio-demographic characteristics, a number of which are used as explanatory variables.

Both males and females are included in this study. The survey asks respondents to state whether they are male or female. Similar studies have tended to include males only in their research; however this study will include males and females.

Respondents are asked their age at present in years. Those surveyed are all aged 18 and over which is appropriate given the Intoxicating Liquor Act, 2003 (Ireland, 2003) which states alcohol cannot be sold to anyone under the age of 18 years of age. The age variable is grouped into 6 dummy variables ages 18 to 29 yrs, 30 to 39 yrs, 40 to 49 yrs, 50 to 59yrs, 60 to 69yrs and 70yrs plus. Similar studies such as Barrett (2002) and Hamilton and Hamilton (1997) have also grouped the age variable into dummy variables however they have taken those within the age brackets of between 25 years and 59 years as these studies are looking at the effect of alcohol on an individual's income. In this study because household income is the dependent variable, a wider age group is taken into account.

Of the total respondents in the survey, the largest group of respondents are in the age category of 30 to 39 years, however respondents are relatively evenly distributed across all age categories as can be seen in table 3.2.3.

Table 3.2.3: Breakdown of Respondents across age categories

	No. of Respondents	% of respondents in each category
18 to 29 yrs	1,492	17.41
30 to 39 yrs	1,877	21.90
50 to 59 yrs	1,319	15.39
60 to 69 yrs	1,115	13.01
70 plus yrs	1,133	13.22

(Source: Authors own)

The survey also includes a question on one's present marital status. Five dummy variables are created to represent marital status.

In terms of education the Slán survey categorises education status based on the highest level of education achieved reported by the respondent. In the 2007 survey there are eight different categories and respondents are asked to select which one is relevant to them. In relation to the 2007 survey this study the education variable is grouped into 5 categories similar to those used by Hamilton and Hamilton (1997).

The largest group of respondents report having second level education as being that the highest level of education completed. Those who report having completed diploma or cert is the second highest category of respondents with the lowest being those with a postgraduate qualification. This is depicted in table 3.2.4.

Table 3.2.4: Breakdown of respondents based on highest level of education achieved

	No. of Respondents	% of respondents in each category of education
Primary	1,488	17.36
Secondary	3,775	44.05
Diploma/cert	1,587	18.52
Primary degree	893	10.42
Postgraduate	827	9.65

(Source: Authors own)

In terms of categorising where respondents live, the 2007 survey asks respondents to describe the type of place they live.

> *'What best describes the place where your household is situated as being ...?'*

a) *In open country*
b) *In a village*
c) *In a town (1,500+)*
d) *In a city (other than Dublin)*
e) *In Dublin City or County*

Dummy variables are created to represent the five categories. Data is not available in the 2007 Slán survey in relation to the region where respondents live.

In terms of assessing respondents Health Status five dummy variables are created to indicate how a respondent rates their health. The question in the Slán survey asks people to rate their general health - Excellent, Very Good, Good, Fair, Poor.

In relation to respondents employment situation in the 2007 survey, the information is taken from the question in the survey asking respondents what best describes their usual situation in regard to work. 10 options are then given to people to choose from. In an effort to condense the information some of the categories are merged together in this study and hence the current employment variable is grouped into eight dummy variables. The categories that are merged together are those who report being self employed and being famers, given that both can be defined as employed. Students and those on state training schemes are merged into one category given that both can be classified as training. Overall the employment status variables represent employee's, self-employed including farmers, those on state training schemes, unemployed, homemakers, those on disability, the retired and those in other employment situations.

Given the fact that it is household income that is being assessed, the number of people working in the household is included as a variable. This variable is derived from the question in the Slán survey

> *'How many in your households are currently working, please include all household members who work 15 or more hours per week?'*

The average price of alcohol was included as a variable in the alcohol status equation. The price was derived by dividing the total values of sales in the 2007 by the total volume sold for each type of alcohol in 2007 similar to how it was derived in a study by Hamilton and Hamilton (1997). Given that a constant value for price was derived this was found to be collinear which could not be used. Studies such as Hamilton and Hamilton (1997) that included the price variable had different prices for different types of alcohol and hence were not found to be collinear. Barrett (2002) did not include the price variable.

Race is also included as a variable. Respondents are categorised as Black, White, Asian or Other.

It is important that there is at least one variable in the alcohol status equation that is not in the income equation. If all variables in the alcohol status equation are also in the income equation then the identification of the coefficient in the income equation would be weak (Hamilton and Hamilton, 1997; Barrett, 2002). In this study there are two variables included in the alcohol status equation that are not included in the income equation.

One of these variables is whether or not the respondent is involved in Church activities. A dummy variable for Church Activities is used in this study using the survey data. The question in the Slán survey to which this relates is

> '*do you regularly join in the activities of Church or other religious/parish groups, charitable or voluntary organisations (e.g. collecting for charity, helping the sick, elderly)?'*

Other studies such as Hamilton and Hamilton (1997) found that those who attend religious ceremonies or groups are more likely to be a non-drinker, but that this is not the case with Catholics.

The second variable used is that describing whether or not a respondent previously smoked five or more years ago. Some previous studies have used a dummy variable indicating whether or not an individual smoked at the age of 18 years (Barrett, 2002). This provided a retrospective measure of an individual's attitude towards risk, the rationale being that smoking is a health risk behaviour and in part reflects an individual's attitude toward risk. Hersch and Viscusi (1990) used contemporaneous smoking behaviour as a proxy for individuals' attitudes towards risk in estimating wage differentials for risk of lost work-day injury. Barrett (2002) looked at smoking in the past as opposed to current smoking because he argues that the retrospective measure of smoking is not likely to influence current income however current smoking behaviour is likely to affect current income. In the Slán dataset, information with regard to whether or not the respondent smoked at the age of 18years is not available. Information is available on how long it has been since the respondent last smoked. Based on the responses to this question, a dummy variable is created to categorise those who previously smoked five years ago or more.

3.2.1: Descriptive Statistics

This section includes a more detailed description of the dependent and independent variables from the 2007 Slán Survey. This is set out in Table 3.2.5. Summary statistics of the individual variables have been calculated using the Statistical Package STATA. Most of the variables are presented in the form of dummy variables, where the values 0 and 1 indicate the presence or absence of an attribute.

Description of the dependent and independent variables in the 2002 Slán dataset are provided in Appendix B. Comparison between 2002 and 2007 is not made either between the descriptive statistics or the results as the questions in both surveys are different and there is more information provided in the 2007 survey.

Table 3.2.5 Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Logincome	The log of weekly household income in Euro	6.398	.713	3.76	7.43
Alcohol Status	Non Drinkers = 1, Moderate Drinker = 2, Heavy Drinkers = 3 Based on the standard units alcohol consumed	1.796	.545	1	3
Male	Individuals who are male=1, 0 = female	.427	.495	0	1
Age18to29	Those who are aged is 18 to 29 years =1, 0 = otherwise	.174	.379	0	1
Age30to39	Those who are aged is 30 to 39 years =1, 0 = otherwise	.219	.414	0	1
Age40to49	Those who are aged is 40 to 49 years =1, 0 = otherwise	.191	.393	0	1
Age50to59	Those who are aged is 50 to 59 years =1, 0 = otherwise	.154	.361	0	1
Age60to69 **Base Category**	Those who are aged is 60 to 69 years =1, 0 = otherwise	.130	.336	0	1
Age70plus	Those who are aged is 70plus years =1, 0 = otherwise	.132	.339	0	1
Edprimary **Base Category**	Individuals who have primary school education only =1, 0 = otherwise	.174	.379	0	1
Edsecondary	Individuals who have completed secondary education only =1, 0 = otherwise	.440	.496	0	1

Table 3.2.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Ed diploma/cert	Individuals who have a diploma or certificate =1, 0 = otherwise	.185	.388	0	1
Ed primarydegree	Individuals who have a primary degree =1, 0 = otherwise	.104	.306	0	1
Ed postgraduate	Individuals who have completed a postgraduate /higherdegree =1, 0 = otherwise	.096	.295	0	1
Single/never married	Individuals who are single/never married = 1, 0 = otherwise.	.280	.449	0	1
Cohabiting **Base Category**	Individuals who are cohabiting = 1, 0 = otherwise.	.06	.24	0	1
Separated/Divorced	Individuals who are separated or divorced = 1, 0 = otherwise.	.063	.243	0	1
Married	Individuals who are married = 1, 0 = otherwise.	.506	.500	0	1
Widowed	Individuals who are widowed = 1, 0 = otherwise.	.087	.282	0	1
Opencountry **Base Category**	Individuals living in the open country =1,0= otherwise	.309	.462	0	1
Village	Individuals living in a village =1,0= otherwise	.107	.309	0	1
Town	Individuals living in a town =1,0= otherwise	.242	.429	0	1

Table 3.2.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Cityotherthandublin	Individuals living in a city other than Dublin =1,0= otherwise	.105	.307	0	1
Dublincityorcountry	Individuals living in Dublin city or county =1,0= otherwise	.226	.418	0	1
Healthexcellent	Individuals who classify their health as excellent =1, 0= otherwise	.211	.408	0	1
Healthverygood	Individuals who classify their health as very good =1, 0= otherwise	.358	.480	0	1
Healthgood	Individuals who classify their health as good =1, 0= otherwise	.289	.453	0	1
Healthfair	Individuals who classify their health as fair =1, 0= otherwise	.108	.310	0	1
Healthpoor **Base Category**	Individuals who classify their health as poor =1, 0= otherwise	.032	.175	0	1

Table 3.2.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Employee	Those whose current employment situation is an employee at work = 1, 0 = otherwise	.458	.498	0	1
Selfempl. Incl. farmer	Those whose current employment situation is self employed or in farming = 1, 0 = otherwise	.116	.320	0	1
State training/student	Those on state training scheme or student = 1, 0 = otherwise	.037	.190	0	1
Unemployed	Those whose current employment situation is unemployed = 1, 0 = otherwise	.030	.169	0	1
Disability **Base Category**	Those whose current employment situation is unable to work owing to permanent sickness/disability = 1, 0 = otherwise	.04	.19	0	1
Homemaker	Those whose current employment situation is Homemaker = 1, 0 = otherwise	.140	.347	0	1

Table 3.2.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Retired	Those whose current employment situation is wholly retired =1, 0 = otherwise	.170	.376	0	1
Other	Those whose current employment situation is classified as other = 1, 0 = otherwise	.009	.097	0	1
Num working in Household	No. of people in household working 15 or more hours per week	1.413	1.277	0	20.00
Race White	Individuals of White Race =1, 0 = otherwise	.970	.170	0	1
Race Black	Individuals of Black Race =1, 0 = otherwise	.008	.088	0	1
Race Asian	Individuals of Asian Race =1, 0 = otherwise	.008	.089	0	1
Race other **Base Category**	Individuals of Other Race =1, 0 = otherwise	.005	.077	0	1
Partake in Church activities	Individuals who regularly join in the activities of Church or other religious/parish groups, charitable or voluntary organisations =1, 0= otherwise	.188	.391	0	1
Previous smoker 5+ yrs ago	Individuals who used to smoke five years ago or more =1, 0 = otherwise	.139	.346	0	1

(Source: Authors Own)

3.3: Variables used in the estimation of health status and health care utilisation

There are many human capital and socio demographic variables that affect alcohol consumption, health status and health care utilisation. There are also some additional variables that can influence levels of alcohol consumption but not health status and health care utilisation. Both the dependent and independent variables are described below.

The Dependent Variables

Health Status

In the general health section of the Slán survey, respondents are asked about their health. Self assessed health is one of the most common measures of health in studies (Jurges, 2008). The question posed in the Slán survey to respondents is

> *'In general how would you say your health is…Excellent, Very Good, Good, Fair, or Poor?'*

There is widespread agreement that this simple global question provides a useful summary of how patients perceive their overall health status (Fayers and Sprangers, 2002). Some studies (Wilson *et al,* 2011) also use this single item global measure of health, where an individual is asked to rate their health as excellent, very good, good, fair or poor relative to others their own age, however they then dichotomise these responses into 'excellent/very good/ good' and 'fair/poor'. In this study dummy variables were created to represent each of the five categories of health status and each respondent is categorised into one of the five categories. Table 3.3.1 shows the number of respondents in each category of health status.

Table 3.3.1: Breakdown of Percentage of Respondents reporting different categories of health status

	% Reporting each category of health status
Poor Health Status	3.18
Fair Health Status	10.8
Good Health Status	28.91
Very Good Health Status	35.98
Excellent Health Status	21.13

(Source Authors own)

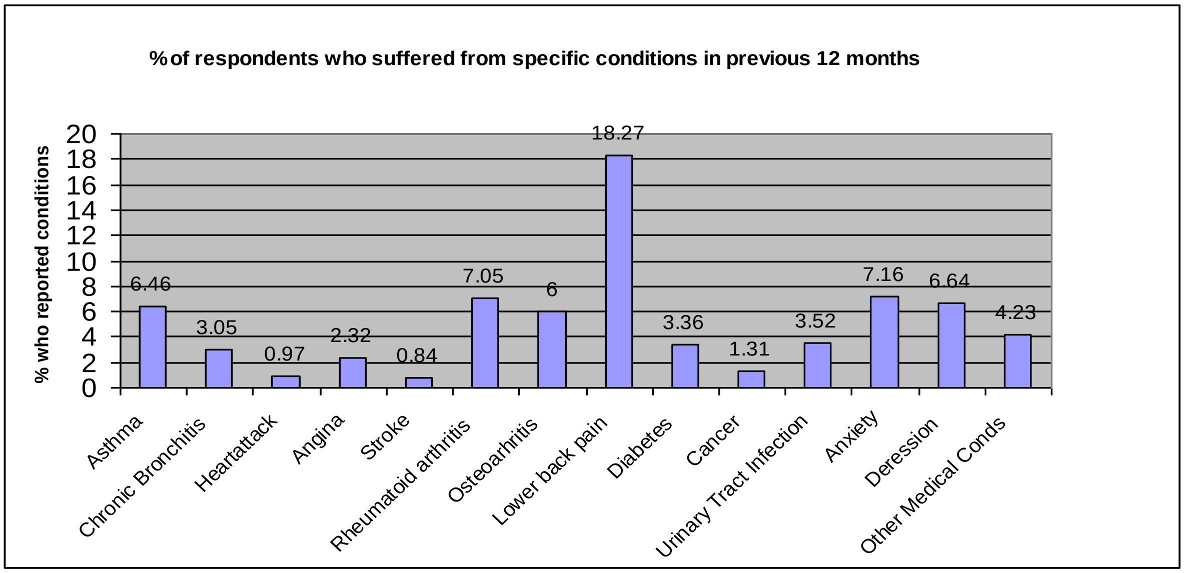
In looking at specific illnesses, the Slán survey asks a question

'Have you had any of the following conditions in the last 12 months?'

a. Asthma,
b .Chronic bronchitis, chronic obstructive lung (pulmonary) disease, emphysema,
c. Heart attack,
d. Angina,
e. Stroke,
f. Rheumatoid arthritis (inflammation of the joints),
g. Osteoarthritis (arthritis, joint degeneration),
h. Lower back pain or other chronic back condition,
i. Diabetes,
j. Cancer (malignant tumour, also including leukaemia & lymphoma),
k. Urinary incontinence, problems in controlling the bladder,
l. Anxiety,
m. Depression
n. Other, specify

Figure 3.3.1 shows the percentage of respondents who reported having each condition in the previous twelve months.

Figure 3.3.1. % of Respondents who reported suffering from specific conditions



(Source: Authors own)

Health Care Utilisation

GP utilisation is used to look at health care utilisation in Ireland. The general health section of the Slán survey asks respondents about the last time they consulted a General Practitioner.

'When was the last time you consulted a GP'

a. In the last 4 weeks

b. Between 1 and 12 months ago

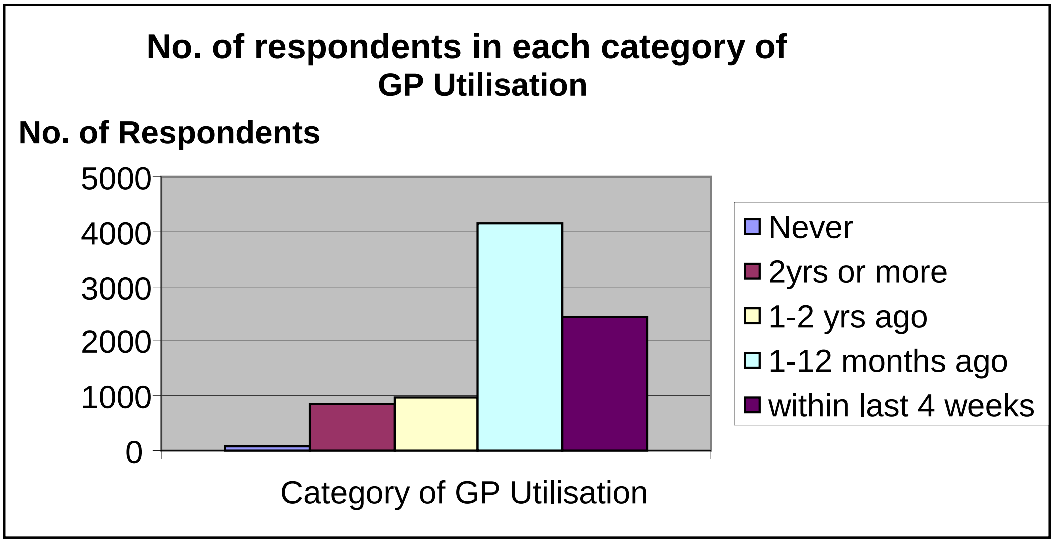
c. Between 1 and 2 years ago

d. More than 2 years ago

e. Never

Figure 3.3.2 depicts the number of respondents in each category of GP utilisation.

Figure 3.3.2 Breakdown of Respondents reporting Different levels of GP Utilisation



(Source: Authors own)

Alcohol Status

The same as in the case with the estimation of alcohol on income, drinkers are categorised into one of three categories of drinkers; non, moderate and heavy drinkers based on recommendations from the Irish Health Promotion Unit (HSE, 2008).

The Independent Variables in the Health Status Equation

In the analysis of individuals' health status, both males and females are included. The survey asks respondents to state whether they are male or female.

Respondents are asked their age at present in years. The age variable is grouped into 6 dummy variables ages 18 to 29 yrs, 30 to 39 yrs, 40 to 49 yrs, 50 to 59yrs, 60 to 69yrs and 70yrs plus. Contoyannis and Jones (2004) and similarly Blaylock and Blisard (1992) control for age by using a continuous variable which give age in years. Wilson *et al* (2011) divided age into three categories: 18-54 years, 55-64 years and 65 years and over.

The survey also includes a question on one's present marital status. Six dummy variables are created to represent marital status similar to other studies such as Wilson *et al* (2011) and Kiuila and Mieszkowski (2007).

In terms of education the 2007 Slán survey categorises education status based on the highest level of education achieved as reported by the respondent. There are eight different categories and respondents are asked to select which one is relevant to them. In this study the education variable is grouped into 5 categories. Contoyannis and Jones (2004) had similar variables in their study.

In relation to respondents employment situation, the information is taken from the question in the survey asking respondents what best describes their usual situation in regard to work. 10 options are then given to people to choose from. In an effort to condense the information some of the categories are merged together and hence the current employment variable is grouped into eight dummy variables. These variables represent employee's, self-employed including farmers, those on state training schemes, unemployed, homemakers, those on disability, the retired and those in other employment situations.

In the Slán survey, income bands are available for the household's total net income per week, per month or per year. The total net take home pay includes all sources of family income i.e. social benefits etc. There are twenty-four categories of income given in the Slán survey ranging from the lowest category of less than €86 euro per week to the highest of €1,535 or more per week. For the purpose of econometric analysis in this paper, the descriptive statistics for income were derived by taking the log of the midpoint of an individual's income category similar to what Barrett (2002) did in his study and for the open upper category, a value of 10% above the lower income limit of the band, was taken (Von Fintel, 2007). Income has been found to have an effect on health status and in general findings have been that those with lower income have lower self reported health (Bradley *et al*, 2000; Yen *et al*, 2010).

In the Slán survey a question is asked about respondent's ethnic or cultural background. The question posed is

'What is your ethnic or cultural background?

a) White or White Irish - Irish, Irish Traveller or any other white background?

b) Black or Black Irish – African or any other black background?

c) Asian or Asian Irish – Chinese or any other Asian background?

d) Other including mixed background?

Four dummy variables are used to represent each of these 4 categories which will allow self assessed health to be analysed by including race similar to what was dine in other studies such as Thorpe *et al (* 2009) and Thompson (2011).

The number of people in the Household is included as a continuous variable.

Whether a person lives in the city or in a rural setting can affect their health status (Lin, 2008; Wilson *et al*, 2011), hence the question in the Slán survey which asks respondents to describe where their household is situated, is used. The Slán survey provides respondents with a choice of five categories to choose from, each describing where their household is situated. From this five dummy variables describing where respondents live are created.

In terms of the lifestyle variables, the Slán survey does include variables describing whether or not one smokes, level of exercise, body weight and level of alcohol consumption. These variables are included in this study. Table 3.3.2 summarises the number of respondents in each of the weight categories and those who are classified as smokers.

Table 3.3.2: Breakdown of Respondents in terms of lifestyle variables – weight & smoking

	No. of Respondents	% of respondents reporting different categories of weight and smoking
Weight		
Weight Right	4,750	55.98%
Weight Too Heavy	2,929	34.52%
Weight Too Light	336	3.96%
Weight Unsure	420	4.95%
Smoking		
Smoker	2298	27.08%

(Source: Authors own)

In terms of describing whether or not respondents eat breakfast and snacks between meals, the questions in the Slán survey ask respondents to describe what they did 'yesterday'. The fact that this does not describe respondent's regular patterns of snacking and eating breakfast, these variables are not included. Respondents are not asked about their general sleeping patterns in the Slán survey.

Section B of the Slán survey relates to Physical Activity. In this study a dummy variable is created to represent whether or not respondents are currently physically active. The World Health Organisation (2010) recommends that for age groups 18-64 years and 65years plus, they should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week or an equivalent combination of moderate- and vigorous-intensity activity. The question in the 2007 Slán data set reports that respondents level of exercise closest to the WHO recommendations is posed as follows:

'Thinking now about regular physical activity, by that I mean: taking part in exercise or sports 2-3 times per week for a minimum of 20 minutes at a time, or more general activities like walking, cycling or dancing 4-5 times per week accumulating to at least 30 minutes per day.
With this in mind could you look at this card and tell me which statement best describes how physically active you have been over the last six months?'

a) I am not regularly physically active and do not intend to be so in the next six months
b) I am not regularly physically active but am thinking about starting to do so in the next six months
c) I do some physical activity but not enough to meet the description of regular physical activity
d) I am regularly physically active but only began in the last six months
e) I am regularly physically active and have been so for longer than six months

The dummy variable showing whether or not a person is physically active is created by categorising respondents who answered d or e as being physically active and those who answered a, b or c are not.

A dummy variable is created to categorise smokers and non smokers. Smokers are anyone who smokes every day or some days. Non smokers are those who do not smoke at all. The question in Slán that is used to define smokers and non smokers is

'Do you smoke every day, some days or not at all?'

Smokers are categorised in a similar manner in numerous studies (Lye and Hirschberg, 2004; Contoyannis and Jones, 2004).

Four dummy variables are created to describe respondent's weight given their age and height. The Slán survey asks

> *Given your age and height, would you say that you are …. About right weight, too heavy, too light or not sure?*

There is no question relating to Body Mass Index in the Slán survey.

A dummy variable is created to describe whether or not respondents have a Medical Card. The question in the Slán survey is

> *Are you covered by a medical card?*
> *Yes – full medical card*
> *Yes – GP only medical card*
> *No*

Those who answered yes to having a full medical card and a GP only medical card are all categorised as having a medical card in this study. Numbers of respondents who have medical cards are depicted in table 3.3.3

Table 3.3.3 Number of Respondents who have Medical Cards

	No. of Respondents	% of respondents with medical cards
Medical Cards		
Numbers without medical card	5,394	63.57%
Numbers with medical card	3,091	36.43%

(Source: Authors own)

As part of the Slán survey, respondents are asked if they have private health insurance. Numbers of respondents are depicted in table 3.3.4.

Table 3.3.4: Number of Respondents who have Private Health Insurance

	No. of Respondents	% of respondents with health insurance
Health Insurance		
Numbers with private health insurance	3,959	46.66%
Numbers without private health insurance	4,526	53.34%

(Source: Authors own)

3.3.1: Descriptive Statistics

This section includes a more detailed description of the dependent and independent variables from the 2007 Slán Survey. This is set out in Table 3.3.5. Summary statistics of the individual variables have been calculated using the Statistical Package STATA. Most of the variables are presented in the form of dummy variables, where the values 0 and 1 indicate the presence or absence of an attribute.

Table 3.3.5 Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Male	Individuals who are male=1, 0 = female	0.427	0.495	0	1
Married	Individuals who are married = 1, 0 = otherwise.	0.506	0.500	0	1
Widowed	Individuals who are widowed = 1, 0 = otherwise.	0.087	0.281	0	1
Sep/div	Individuals who are separated or divorced = 1, 0 = otherwise.	0.063	0.243	0	1
Single/never married	Individuals who are single/never married = 1, 0 = otherwise.	0.280	0.449	0	1
Cohabiting **Base Category**	Individuals who are cohabiting = 1, 0 = otherwise.	0.062	0.242	0	1
Edprimary **Base Category**	Individuals who have primary school education only =1, 0 = otherwise	0.174	0.379	0	1
Educ. Secondary	Individuals who have completed secondary education only =1, 0 = otherwise	0.441	0.497	0	1
Educ. Diploma	Individuals who have a diploma or certificate =1, 0 = otherwise	0.185	0.388	0	1
Educ. Primary Degree	Individuals who have a primary degree =1, 0 = otherwise	0.104	0.306	0	1
Educ. Post Graduate	Individuals who have completed a postgraduate /higherdegree =1, 0 = otherwise	0.096	0.296	0	1

Table 3.3.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Age18-29	Those who are aged is 18 to 29 years =1, 0 = otherwise	0.174	0.379	0	1
Age 30-39	Those who are aged is 30 to 39 years =1, 0 = otherwise	0.219	0.414	0	1
Age 40-49	Those who are aged is 40 to 49 years =1, 0 = otherwise	0.191	0.393	0	1
Age 50-59	Those who are aged is 50 to 59 years =1, 0 = otherwise	0.154	0.361	0	1
Age60to69 **Base Category**	Those who are aged is 60 to 69 years =1, 0 = otherwise	0.130	0.336	0	1
Age 70plus	Those who are aged is 70plus years =1, 0 = otherwise	0.132	0.338	0	1
Employee	Those whose current employment situation is an employee at work = 1, 0 = otherwise	0.458	0.498	0	1
Selfemployed	Those whose current employment situation is self employed or in farming = 1, 0 = otherwise	0.115	0.320	0	1
Disability **Base Category**	Those whose current employment situation is unable to work owing to permanent sickness/disability = 1, 0 = otherwise	0.038	0.192	0	1
State Training/Student	Those who are students or on a state training programme=1, 0= otherwise	0.037	0.189	0	1

Table 3.3.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Unemployed	Those whose current employment situation is unemployed = 1, 0 = otherwise	0.030	0.170	0	1
Homemaker	Those whose current employment situation is Homemaker = 1, 0 = otherwise	0.140	0.347	0	1
Retired	Those whose current employment situation is wholly retired =1, 0 = otherwise	0.169	0.375	0	1
Other	Those whose current employment situation is classified as other = 1, 0 = otherwise	0.010	0.097	0	1
Logincome	The log of weekly household income in Euro	6.398	0.713	3.76	7.43
Race White	Those who are white or white Irish = 1, 0 = otherwise	0.970	0.169	0	1
Race Black	Those who are black or white Irish = 1, 0 = otherwise	0.008	0.089	0	1
Race Asian	Those who are Asian or Asian Irish = 1, 0 = otherwise	0.008	0.089	0	1

Table 3.3.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Race Other **Base Category**	Those who are from another or a mixed background = 1, 0 = otherwise	0.006	0.077	0	1
Total in hh	Total number of people in household	5.811	3.189	0	36
Opencountry **Base Category**	Individuals living in the open country =1,0= otherwise	0.309	0.462	0	1
Village	Individuals living in a village =1,0= otherwise	0.107	0.309	0	1
Town	Individuals living in a town =1,0= otherwise	0.243	0.429	0	1
City other than Dublin	Individuals living in a city other than Dublin =1,0= otherwise	0.106	0.307	0	1
Dublin city	Individuals living in Dublin city or county =1,0= otherwise	0.225	0.418	0	1
Smoker	Individuals who smoke either every day or on somedays =1, 0 = otherwise	0.271	0.444	0	1
Weight right	Individuals who given their age and height, classify their weight as just right = 1, 0 = otherwise	0.560	0.496	0	1

Table 3.3.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Weight too heavy	Individuals who given their age and height, classify their weight as too heavy = 1, 0 = otherwise	0.345	0.475	0	1
Weight too light	Individuals who given their age and height, classify their weight as too light = 1, 0 = otherwise	0.040	0.195	0	1
Weight not sure **Base Category**	Individuals who given their age and height, classify their weight as not sure = 1, 0 = otherwise	0.049	0.217	0	1
Church activities	Individuals who regularly join in the activities of Church or other religious/parish groups, charitable or voluntary organisations =1, 0= otherwise	0.188	0.391	0	1
Health excellent	Individuals with excellent health = 1, 0 = otherwise	0.211	0.408	0	1
Health very good	Individuals with very good health = 1, 0 = otherwise	0.358	0.480	0	1
Health good	Individuals with good health = 1, 0 = otherwise	0.289	0.453	0	1

Table 3.3.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Health fair	Individuals with fair health = 1, 0 = otherwise	0.108	0.310	0	1
Health poor **Base Category**	Individuals with poor health = 1, 0 = otherwise	0.032	0.175	0	1
Medical Card	Individuals who have a medical card = 1, 0 = otherwise	0.360	0.480	0	1
Private Health Insurance	Individuals who have private health insurance =1, 0 = otherwise	0.533	0.500	0	1
Dependent Variables					
Asthma	Individuals who have had asthma in the last 12 months = 1, 0 = otherwise	0.065	0.246	0	1
Chronic bronchitis	Individuals who have had chronic bronchitis, lung disease or emphysema in the last 12 months = 1, 0 = otherwise	0.031	0.172	0	1
Heart Attack	Individuals who have had a heart attack in the last 12 months = 1, 0 = otherwise	0.010	0.098	0	1
Angina	Individuals who have had angina the last 12 months = 1, 0 = otherwise	0.023	0.151	0	1

Table 3.3.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Stroke	Individuals who have had a stroke in the last 12 months = 1, 0 = otherwise	0.008	0.091	0	1
Rheumatoid Arthritis	Individuals who have had rheumatoid arthritis in the last 12 months = 1, 0 = otherwise	0.070	0.256	0	1
Osteo Arthritis	Individuals who have had osteo arthritis in the last 12 months = 1, 0 = otherwise	0.060	0.237	0	1
Lower Back pain	Individuals who have had lower back pain or chronic back pain in the last 12 months = 1, 0 = otherwise	0.182	0.386	0	1
Diabetes	Individuals who have had diabetes in the last 12 months = 1, 0 = otherwise	0.034	0.180	0	1
Cancer	Individuals who have had cancer in the last 12 months = 1, 0 = otherwise	0.013	0.114	0	1

Table 3.3.5 continued: Descriptive Statistics from the Slán Survey 2007

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Urinary	Individuals who have had urinary incontinence, problems controlling the bladder in the last 12 months = 1, 0 = otherwise	.035	0.185	0	1
Anxiety	Individuals who have had anxiety in the last 12 months = 1, 0 = otherwise	.072	0.258	0	1
Depression	Individuals who have had depression in the last 12 months = 1, 0 = otherwise	.066	0.249	0	1
Other medical Conditions	Individuals who have had other medical conditions in the last 12 months = 1, 0 = otherwise	.043	0.202	0	1
Health status	Self Assessed Health Status (1=poor health status to 5=excellent health status)	3.61	1.03	1	5
GP Consultations	Last time an individual visited a GP (1=never and 5=in the last 4 weeks)	3.94	0.95	1	5

(Source: Authors Own)

3.4: Conclusion

The 2007 Slán National Health and Lifestyle Survey is a scientifically representative random sample of 10,364 respondents (Morgan *et al*, 2008). The survey covers general health, behaviours relating to health such as alcohol consumption, exercise, nutrition, and the use of health services. It provides a large amount of data which is used in the study into the effect of alcohol consumption on income, on health status and on health care utilisation and which is described in detail in this chapter.

CHAPTER 4

ESTIMATION OF THE EFECT OF ALCOHOL CONSUMPTION ON HOUSEHOLD INCOME IN IRELAND

This chapter presents an empirical study of the impact of alcohol consumption on household income in Ireland using data from the 2007 Slán Survey of the lifestyle, attitudes and nutrition of people living in Ireland. Drinkers are categorised into non, moderate and heavy consumers of alcohol, and the differences in income between these three groups of individuals is examined.

The relationship between household income and different socio economic variables such as age, gender, health status, marital status, employment situation, the number of people in the household and the province in which people live, is examined. As part of the analysis of alcohol status on household income, the alcohol status equation is estimated initially. This allows the relationship between all these socio economic variables along with the variables describing an individual's involvement in regular church activities and whether a person was a previous smoker, and an individual's alcohol status to be examined.

The drinking status equation is estimated using a multinomial logit model, which is similar to Hamilton and Hamilton (1997) and Barrett (2002). This method allows the predicted values for the inverse mills ratio to be generated. The household income regression is then estimated for each category of drinker using a standard Ordinary Least Squares (OLS) regression which includes the inverse mills ratio. By estimating the income regression using this two step procedure and including the inverse mills ratio, the alcohol sector selection is treated endogenously and selection bias is accounted for (Hamilton and Hamilton, 1997; Barrett, 2002).

4.1: Empirical Techniques

In the analysis of the effect of alcohol consumption on household income in Ireland the endogenous relationship between income and alcohol is accounted for. A two step procedure is used similar to methods adopted by Hamilton and Hamilton (1997) and Barrett (2002) in their studies whereby:

- The first step is to focus on drinking status and the different variables that affect ones drinking habits. The dependent variable is alcohol consumption and this is estimated using a multinomial logit model. In this study similar to what was adopted in previous studies, all the variables in the income equation are included in the alcohol status equation to account for income. In addition to this there are two additional variables included in the income equation which are whether or not one regularly partakes in Church activities and whether the individual was a previous smoker five or more years ago. From this regression the Inverse Mills Ratio is derived.

- The second step is the regression of the income equation, set out below, which includes predicted values for the Inverse Mills Ratio which has been generated through the first step of the regression. The dependent variable in the second regression is household income which is estimated by Ordinary Least Squares (OLS).

$$\ln Y_{ij} = X_{ij}\beta j - \sigma_j p_j \frac{\phi\left[\Phi^{-1}(F_j(z_i\gamma_j))\right]}{F_j(z_i\gamma_j)} + v_{ij} \qquad \textbf{(4.1.1)}$$

Where:

	$\ln Y_{ij}$	log of household income
	X	vector of human capital variables & socio-demographic characteristics
	β	coefficients on the observable characteristics
	σ_j	the standard deviation of the error term η_{ij}

p_j the correlation coefficient between the unobservables in the income and alcohol equations.

z vector containing exogenous variables affecting income or alcohol consumption

γ vector of unknown utility parameters

ϕ probability density function (pdf) of the standard univariate normal distribution respectively.

Φ cumulative distribution function (cdf) of the standard univariate normal distribution respectively.

v_{ij} the error term which has a zero mean and in uncorrelated with μ_{ij}

Estimates from equation 4.1.1 provide information on the expected income of an individual if they were randomly allocated to a given drinking status, as well as predicted income given that a person is a particular drinker type (Hamilton and Hamilton, 1997; Barrett, 2002).

4.2: Results

The alcohol status equation is estimated as a Multinomial Logit Model, from which the inverse mills ratio can be derived. This is then included as an additional variable in the income regression. The results are discussed in section 4.2.1. Results using the 2002 Slán dataset are provided in Appendix E. These are not directly compared to 2006 results as the surveys are different and different variables are used in some instances.

4.2.1 Results from the Multinomial Logit OLS Two Step Estimate using 2007 Slán Survey (Step One & Step Two)

In looking at the effect of alcohol status on income using the Slán 2007 data, a multinomial logit OLS two step estimation is used. All the variables included in the income equation are included in the drinking status equation to control for the effect of income on drinking similar to what Hamilton and Hamilton (1997) and Barrett (2002) did in their study. Two additional variables, unique to the drinking decision are included in the drinking status model. These are whether or not one regularly partakes in Church activities and whether or not a person was previously a smoker five or more years ago.

With a multinomial logit model the parameter estimates are relative to the reference group, in this case moderate drinkers. The coefficients listed in table 4.2.1 indicate the effect that each variable has on the likelihood of an individual being a non or heavy drinker compared with a moderate drinker. The interpretation of the coefficient being that for a one unit change in the independent variable, the logit of the outcome relative to the reference group moderate drinkers, is expected to change by its respective parameter estimates given the variables in the model being held constant. The z statistics is the ratio of the coefficient to the standard error of the respective predictor. The P value shows the probability that the z statistic is observed under the null hypothesis that a particular predictor's regression coefficient is zero, given that the rest of the predictors are in the model, can be rejected.

Table 4.2.1 Results from the Estimation of the Drinking Status Equation using the Multinomial Logit Model

	Non-Drinkers			Heavy Drinkers		
	Coefficient	**Z-stat**		**Coefficient**	**Z-stat**	
Male	-0.432	-6.82*		0.822	7.93*	
age18to29	-0.860	-6.13*		0.172	0.71	
age30to39	-0.490	-4.30*		-0.017	-0.07	
age40to49	-0.463	-4.14*		0.086	0.38	
age50to59	-0.356	-3.24*		0.134	0.60	
age70plus	0.483	4.47*		-0.647	-2.22*	
Ed Secondary	-0.392	-4.89*		-0.001	0.00	
Ed Diploma/Cert	-0.618	-6.15*		-0.215	-1.14	
Ed Primary degree	-0.830	-6.67*		-0.085	-0.42	
Ed Postgraduate	-0.724	-5.87*		-0.405	-1.77	
Singlenevermarried	0.417	3.08*		0.373	1.95	
Separated/divorced	-0.061	-0.35		0.261	1.04	
Married	0.070	0.51		-0.177	-0.87	
Widowed	0.178	1.08		-0.314	-0.88	
Village	-0.283	-2.96*		0.240	1.38	
Town	-0.380	-5.12*		-0.037	-0.26	
City other than Dublin	-0.534	-5.09*		0.551	3.46*	
Dublin city/county	-0.503	-6.38*		0.406	3.02*	
Employee	-0.717	-4.94*		-0.293	-1.21	
Self employed/farmer	-0.590	-3.68*		-0.245	-0.91	
State training/student	-0.923	-4.06*		-0.168	-0.55	
Unemployed	-0.254	-1.23		0.301	1.00	
Homemaker	-0.452	-3.01*		-0.806	-2.55**	
Retired	-0.514	-3.21*		-0.313	-1.02	

Table 4.2.1 continued: Results from the Estimation of the Drinking Status Equation using the Multinomial Logit Model

	Non-Drinkers			Heavy Drinkers		
	Coefficient	Z-stat		Coefficient	Z-stat	
Other	-0.693	-2.43**		-0.779	-1.17	
No. working in h.hold	-0.023	-0.79		-0.006	-0.17	
Race White	-0.626	-2.58*		0.087	0.20	
Race Black	1.327	3.48*		-12.747	-25.34*	
Race Asian	1.732	4.74*		-0.793	-0.71	
Health excellent	-0.736	-4.69*		0.313	0.83	
Health very good	-0.722	-4.81*		0.385	1.04	
Health good	-0.661	-4.43*		0.567	1.55	
Health fair	-0.415	-2.66*		0.606	1.59	
Partake Church activities	0.145	2.06**		-0.640	-3.96*	
Prev smoker 5+yrs	-0.407	-4.87*		0.192	1.34	
_cons	1.975	5.88*		-3.106	-4.80*	

No. of Observations = 7870

Wald Chi2 (70) = 6793.48

Prob > chi2 = 0.0000

Pseudo R^2 = 0.1006

Log Likelihood = -5714.2847

*** indicates significance at 1% level, ** indicates significance at 5% level**

Note: The average price of alcohol was included as a variable in the alcohol status equation. Price was dropped due to collinearity.

The gender variable is statistically significant and results show that males are less likely to be a non- drinker and more likely to be a heavy drinker, which is similar to the findings of previous studies which found that men consume greater amounts of alcohol than women (Fillmore 1994; Moore, 2005; Blow *et al*, 2005; Moore *et al*, 2005; Mullahy & Sindelar, 1996).

Previous studies show that on average people drink less as they got older, and as a result are less likely to be heavy drinkers (Moore *et al*, 2005; Blow *et al*, 2005; Hamilton & Hamilton, 1997; Auld, 2005; Barrett, 2002; Mullahy & Sindelar, 1996). This study has had similar findings in so far as those up to age 59 years are more likely to be a moderate drinker compared with a non drinker, particularly those aged 18-29 years, however for those aged 70 years plus they are more likely to be a non-drinker. For heavy drinkers age is only significant for those over 70 years, and respondents in this age category are less likely to be a heavy drinker which is similar to previous findings (Hamilton and Hamilton, 1997; Barrett 2002).

The results in terms of education show that all education variables are significant for non-drinkers. In particular those with third level education are less likely to be a non-drinker compared with moderate drinkers. Hamilton and Hamilton (1997) and Barrett (2002) find that those with a postgraduate qualification tend to be moderate drinkers as opposed to non or heavy drinkers, and findings in this study are similar.

The variable describing those who are single/never married is the only significant variable describing marital status. A single person or person who never married is more likely to be either a non or a heavy as opposed to a moderate drinker. Previous studies (Barrett, 2002; Auld, 2005; Hamilton and Hamilton, 1997) find that being married is significant in terms of drinking status and that married people are less likely to be heavy drinkers and more likely to be moderate drinkers.

Where one resides has shown to be very significant in terms of ones drinking status. Those who live in a city, either in Dublin or any other city are more likely to be heavy drinkers which is similar to the findings of Su and Yen (2000). In relation to employment status, given all other predictor variables in the model being constant, respondents are more likely to be a moderate drinker as opposed to a non drinker whatever their employment status is. In particular students or those on state training schemes are least likely to be a non drinker.

The Race variables are significant for non-drinkers and for heavy drinkers the variable describing those of black race is significant. A white person is more likely to be a moderate drinker as opposed to a non-drinker, similar to the findings of Mullahy & Sindelar (1996) and Moore *et al (*2005). Asians and Blacks are more likely to be non-drinkers or heavy drinkers.

All the health status variables are significant for non-drinkers. In particular those with excellent, very good or good health, are less likely to be a non drinker which is similar to the findings of previous studies (Berger *et al*, 1999; Klatsky *et al*, 2001; Bau *et al*, 2007) which can be depicted by the U shaped curve showing that the moderate drinkers enjoy better health compared with non or heavy drinkers.

The explanatory variable describing whether or not people regularly partake in Church activities is included in the drinking status equation only. This is a very significant variable across all categories of drinkers. Results show that those involved in Church activities are more likely non drinkers as opposed to moderate or heavy drinkers. This is similar to the findings of Hamilton and Hamilton (1997).

The variable describing those who previously smoked five or more years ago is included in the drinking status equation. This is only significant for non-drinkers and results show that those who previously smoked are more likely to be a moderate drinkers compared with a non-drinker. Barrett (2002) finds that an individual who previously smoked at the age of 18 years is correlated to ones alcohol consumption and he argues that this is the case as it is a retrospective measure of an individual's attitude towards risk.
Results for the income regressions estimated by OLS and corrected for selection bias, using the Slán 2007 dataset, are presented in table 4.2.2. The coefficients listed in table indicate the amount of change one would expect in the dependent

variable, income, for a one unit change in the value of an independent variable, given all the other variables in the model being held constant. The t statistics is used to test whether a given coefficient is significantly different from zero. The P value tests the null hypothesis that a coefficient is zero.

Table 4.2.2: Results from the estimation of the Income Equation by OLS regression accounting for selection bias

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coeffic-ient	**t-stat**		**Coeffic-ient**	**t-stat**		**Coeffic-ient**	**t-stat**	
Male	0.096	2.28**		0.069	4.25*		0.139	1.51	
age18to29	0.204	2.53**		0.105	2.20**		0.285	2.19**	
age30to39	0.244	4.74*		0.119	3.24*		0.075	0.7	
age40to49	0.154	2.83*		0.123	3.45*		0.100	0.94	
age50to59	0.078	1.57		0.112	3.12*		0.048	0.42	
age70plus	-0.073	-1.46		-0.059	-1.48		-0.183	-1.31	
Ed Secondary	0.118	3.17*		0.182	5.53*		0.274	3.47*	
Ed Diploma/Cert	0.216	4.02*		0.306	7.56*		0.432	4.78*	
Ed Primary degree	0.447	6.18*		0.474	10.41*		0.628	6.43*	
Ed Postgraduate	0.427	6.39*		0.549	12.12*		0.675	7.0*	
Singlenevermarried	-0.308	-5.58*		-0.161	-4.71*		-0.103	-1.31	
Separated/divorced	-0.179	-2.68*		-0.287	-7.08*		-0.045	-0.44	
Married	0.099	2.09**		0.168	5.55*		0.311	3.85*	
Widowed	-0.247	-4.09*		-0.185	-4.25*		0.021	0.15	
Village	-0.023	-0.52		-0.016	-0.64		0.052	0.68	
Town	0.030	0.77		-0.068	-2.95*		-0.010	-0.16	
City other than Dublin	0.050	0.9		-0.035	-1.22		0.063	0.69	
Dublin city/county	0.132	2.85*		0.118	5.34*		0.112	1.62	
Employee	0.356	4.76*		0.281	4.65*		0.598	4.99*	
Selfemployed/farmer	0.249	3.18*		0.279	4.58*		0.665	5.07*	
Statetraining/student	0.167	1.21		-0.051	-0.62		0.044	0.23	
Unemployed	-0.188	-1.9		-0.144	-1.98**		0.222	1.5	

Table 4.2.2 continued: Results from the estimation of the Income Equation by OLS regression accounting for selection bias

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coefficient	**t-stat**		**Coefficient**	**t-stat**		**Coefficient**	**t-stat**	
Homemaker	0.189	3.15*		0.186	3.37*		0.266	1.59	
Retired	0.136	2.12**		0.180	2.98*		0.274	1.71	
Other	0.088	0.69		-0.019	-0.18		0.124	0.63	
No. working in house hold	0.105	4.34*		0.134	6.85*		0.127	3.24*	
Race White	0.328	2.49**		0.085	1.43		-0.111	-0.68	
Race Black	-0.261	-1.35		-0.104	-0.71		(omitted)		
Race Asian	0.009	0.04		0.006	0.05		-0.125	-0.59	
Health excellent	0.185	2.58*		0.092	1.57		0.233	1.71	
Health very good	0.162	2.35**		0.032	0.55		0.280	2.05**	
Health good	0.124	1.88		-0.006	-0.11		0.177	1.26	
Health fair	0.077	1.29		-0.053	-0.98		0.170	1.16	
Mills Ratio	0.039	0.26		-0.324	-2.41**		0.149	0.7	
_cons	5.149	34.19*		5.754	28.88*		4.819	8.03*	

Non-Drinkers	**Moderate Drinkers**	**Heavy Drinkers**
No. of obs = 2127	No. of obs = 5216	No. of obs = 527
F(34, 2092) = 64.78	F(34, 5181) = 139.9	F(33,493) =17.67
Prob > F = 00.00	Prob > F = 00.00	Prob > F = 00.00
R Squared = 0.4816	R Squared = 0.4608	R Squared =.5418
Root MSE = .49446	Root MSE = .49009	Root MSE = .49239

*** indicates significance at 1% level, ** indicates significance at 5% level**

The first independent variable considered is the gender variable. Gender is a significant variable in the income regression for both non and moderate drinkers. There is a positive affect on household income for male non-drinkers and moderate drinkers similar to the findings of others (Zhang, 2008). The age variable appears to be much more significant for non and moderate drinkers and not with heavy drinkers. Non and moderate drinkers who are in the younger age categories from 18 to 49 years are more likely to earn more than those who are older, similar to the findings of Hamilton and Hamilton (1997). Barrett (2002) shows that the age profile for moderate drinkers peaks at ages 40-49years, which is also found to be the case in this study. For heavy drinkers the only age variable that is significant is the category 18-29 years and heavy drinkers in this category are likely to have higher incomes than those who are older which again is similar to the findings of Barrett (2002).

Education is a very significant variable in the income regression for all drinker types. In particular those across all drinker types who have a primary degree or a postgraduate degree have higher incomes compared to those with a primary education only which are consistent with previous findings (Barrett, 2002; French & Zarkin, 1995; Heien, 1996). Heavy drinkers who have a postgraduate qualification tend to have the highest income holding the other variables constant.

For all categories of drinkers there is a positive income premium for those who are married. For non and moderate drinkers, there is a negative income premium associated with being single/never married, separated/divorced and widowed compared with those in the base category who are cohabiting. This is similar to previous findings in relation to the income of men (Berger and Leigh, 1988; Schoeni, 1995; Ahituv and Lerman, 2007; Loh, 1996).

In terms of where respondents live, both non-drinkers and moderate drinkers living in Dublin city or county have higher incomes while income of moderate drinkers who live in towns is less, when compared to those living in the country.

For all categories of drinkers, the variables describing those who are employees and self employed have a positive income effect and homemakers who are non or moderate drinkers have a positive income effect compared with those with a disability. Being unemployed is significant for moderate drinkers only and has negative income associated with it compared to those in the base category who have a disability. Retired non and moderate drinkers enjoy a positive household income affect.

The number of people in the household who are working is significant across all drinker types and has a positive correlation with household income. Berger and Leigh (1988) show in their study that race differences in terms of wages are insignificant, and findings using the Slán 2007 dataset are similar in so far as, only the race variable describing white people is significant for non-drinkers. White non-drinkers tend to have higher incomes.

The health variables in the income regressions tend to be more significant for non and heavy drinkers. There is a higher income premium associated with those non and heavy drinkers who report very good and excellent health compared to those with poor health, which is line with Grossmans (1972) argument whereby if one can improve their health status they are then in a position to work more and this then results in ones income increasing.

The Inverse Mills Ratio for non and heavy drinkers is insignificant. It is however significant for moderate drinkers indicating that there is a selection effect into moderate drinking. This is a negative selection effect highlighting that an individual who self selects into the category of a moderate drinker, will have a lower income than an individual with identical observable characteristics drawn at random as a moderate drinker. Hamilton and Hamilton (1987) and Barrett (2002) find that the Mills Ratio for non and moderate drinkers is insignificant however in contrast to this study the Inverse Mills Ratio for heavy drinkers is significant indicating that individuals who self select into heavy drinking earn more on average than an individual with identical observable characteristics drawn at random from the workforce would earn as a heavy drinker.

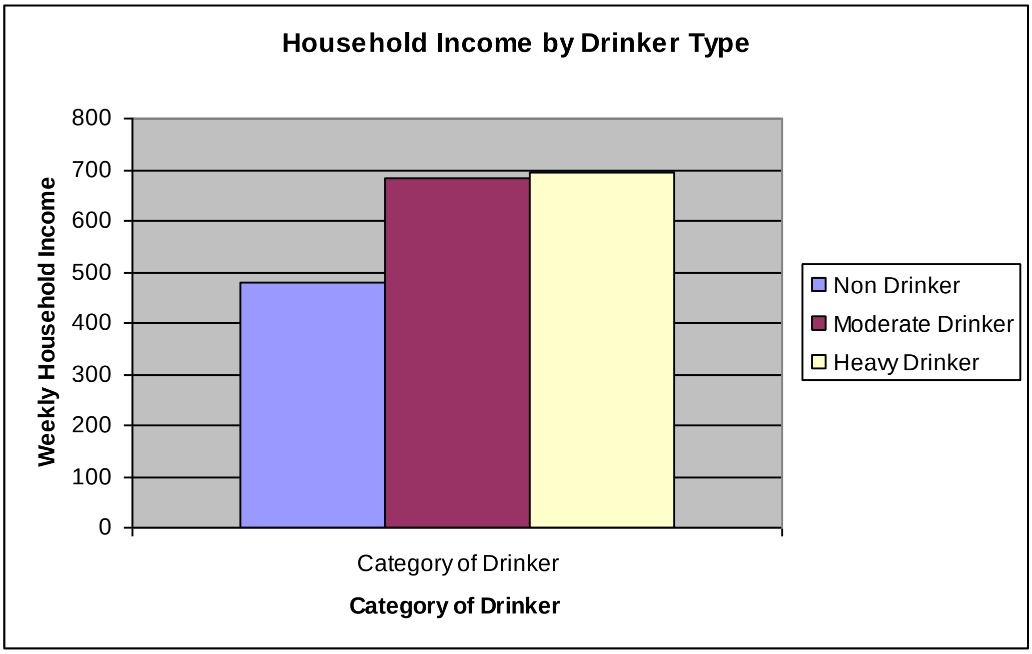
4.2.2 Overall Results of Weekly Income by Drinker Type

Previous studies show a positive association between income and moderate alcohol consumption, compared with income and either non or heavy consumption of alcohol (Hamilton and Hamilton, 1997; Barrett, 2002; French and Zarkin, 1995). Other studies find that while there was a positive correlation between income and alcohol consumption, the drop in income associated with heavy consumption of alcohol compared with moderate consumption, is not found (Zarkin *et al*, 1998; Bastida, 2006).

This study into the effect of alcohol consumption on income in Ireland finds that there is very little difference between the household income of moderate and heavy drinkers, however income of non-drinkers is substantially less.

Weekly household income for non-drinkers is €477.41, compared with €683.36 per week for moderate drinkers and €694.18 for heavy drinkers. This is depicted in figure 4.2.1.

Figure 4.2.1 Weekly Household Income for Non, Moderate & Heavy Drinkers



(Source: Authors own)

Average weekly income for non-drinkers is approximately 31% less than that of moderate and heavy drinkers. There is very little difference between the income of moderate and heavy drinkers. Results using the 2002 Slán dataset are provided in Appendix E.

4.2.3 Decomposition of the Income Differentials

Income is analysed further using the Oaxaca technique which decomposes income into the explained part due to observable characteristics and the unexplained part. The income decompositions are reported in Table 4.2.3.

Table: 4.2.3 Decomposition of Income Differentials between the different categories of drinkers

	Income of Moderate Drinker less Non Drinker	**Income of Heavy Drinker less Non Drinker**	**Income of Heavy Drinker Less Moderate Drinker**
Differential	**Coefficient**	**Coefficient**	**Coefficient**
Prediction 1	6.52	6.542	6.54
Prediction 2	6.16	6.168	6.52
Difference	.3586	.3743	.0157
Decomposition			
Explained	.2801	.2571	-.0425
Unexplained	.0785	.1172	.0582

(Source: Authors own)

The income differentials between each of the categories of drinkers are statistically significant except in terms of the difference between moderate and heavy drinkers. The income decompositions reveal a large income premium for moderate drinkers relative to non drinkers and an even larger income premium for heavy drinkers relative to non-drinkers.

Income differentials show that in relation to the difference in income between non and moderate drinkers, 78% are explained, which means that they are due to differences in endowments. In relation to the difference between non and heavy drinkers, 69 % of the difference is explained by differences in characteristics. Hamilton and Hamilton (1997) refer to the unexplained component as being the pure income differential and isolates the effect of alcohol consumption on income.

4.3 Post Estimation Diagnostics

Testing the Specification of the Model

The Suest-based Hausman test of the Independence of Irrelevant Alternatives (IIA) assumption shows that the null hypothesis, stating that the IIA is valid, can not be rejected, with results showing P values of 1. This means that in looking at the different categories of drinkers, if another drinking category is added to the mix, this will not cause individuals to change their current drinking patterns. Based on this the multinomial logit can be applied.

The significance of each of the variables is assessed using the t and z statistics and results highlight which variables are significant at both the 1% and 5% levels of significance. The Likelihood Ratio test is used to evaluate the relevance of each variable in the model and ensure that each variable is beneficial to the model.

The Wald Test and the F Test show that the models are statistically significant and reject the null at 1% significance level that coefficients of the variables are equal to zero.

Robust standard errors are used in both the alcohol and income equations to which tend to be more trustworthy when heteroskedasticity is present (Berry and Feldman, 1985).

Endogeneity Bias

Selection bias treats the sector selection, in this case alcohol consumption, endogenously. Selection bias of alcohol consumption is accounted for; hence the endogeneity of alcohol consumption is accounted for. Separate Income regressions are then estimated by drinker type which include the Inverse Mills ratio as an additional regressor, which similar to the approach adopted in previous studies (Hamilton and Hamilton, 1997; Barrett, 2002). Barrett (2002) tests for the possible endogeneity of marital status to the drinking decision but finds that endogeneity does not exist. Endogeneity can be tested for relatively easily when suitable instruments are available, however it can be difficult to find suitable instruments Ziebrath and Grabka, 2009). The exogeneity assumption of an instrument is not directly testable and hence the test of analyses of an instrument is mostly a matter of belief (Ziebrath and Grabka, 2009). In this study suitable proxy instruments are not available for marital status and health status variables, however when the drinking status choice equation and the income equations are analysed excluding marital status, the results reported are not sensitive to the treatment of marital status. Similarly where both income and alcohol consumption regressions are run omitting the health status variables, the results are not sensitive to this.

By estimating the effect of alcohol status on income using the Multinomial Logit OLS Two Step Estimate as proposed by Lee (1983), and carried out in similar studies such as Hamilton and Hamilton (1997) and Barrett (2002), selection bias is accounted for.

4.4 Conclusion

This chapter presents an empirical study of the effect of alcohol consumption on household income in Ireland. The relationship between household income and alcohol status with different socio economic variables is examined.

The drinking status equation is estimated using data from the 2007 Slán survey, by a multinomial logit model similar to the manner adopted in previous studies (Hamilton and Hamilton, 1997; Barrett, 2002). Predicted values for the Inverse Mills Ratio are then derived and included as additional variables in the income regressions for each category of drinker, which are estimated by OLS regression. By estimating the income regression using this two step procedure and including the inverse mills ratio, selection bias is accounted for. Estimating separate income regressions for each drinking category controls for endogeneity (Hamilton and Hamilton, 1997; Barrett, 2002). Results show that while income of heavy drinkers is more than moderate drinkers the difference is very small. Income of non drinkers is substantially less than both moderate drinkers and heavy drinkers.

Implications of these findings are that there are benefits in terms of household income from the consumption of alcohol. Population based policies aimed at reducing levels of alcohol consumption could result in individuals reducing their alcohol consumption to levels that result in their household income falling.

Harris *et al* (2006) argue that alcohol consumption could be viewed as being ordered data and should be estimated as so. This is something that previous studies into the effect of alcohol consumption on income have not accounted for (Hamilton and Hamilton, 1997; Barrett, 2002). To account for the ordered nature of alcohol consumption, estimation could be carried out by ordered probit as opposed to multinomial logit.

CHAPTER 5

ESTIMATION OF THE EFECT OF ALCOHOL CONSUMPTION AS ORDERED DATA ON HOUSEHOLD INCOME IN IRELAND COMPARING LIMITED & FULL INFORMATION METHODS OF ESTIMATION

This chapter presents an empirical study of the effect of individual alcohol consumption on household income in Ireland while controlling for the potential endogenous relationship between income and alcohol, using an ordered probit model. Previous studies into the effect of alcohol consumption on income such as Hamilton and Hamilton (1997) and Barrett (2002) among others, have assumed that alcohol status is unordered and hence have estimated the alcohol status equation using the multinomial logit model. Alcohol consumption could however be viewed as ordered data (Harris *et al*, 2006). If ordinality is ignored then this may lead to a loss of efficiency and an increased risk of getting insignificant results (Harris *et al*, 2006). Alcohol consumption is estimated as ordered data through the ordered probit model and the income equation is estimated as an OLS regression. Such estimations can be carried out using Limited Information or Full Information methods of estimation. Both methods of estimation are used with the results of both compared.

5.1: Empirical Techniques

The Multinomial logit method adopted in previous studies estimating the effect of alcohol consumption on income (Hamilton and Hamilton, 1997; Barrett, 2002) fails to account for the ordinal nature of a dependent variable (Greene, 2002) and therefore not all the information regarding the particular variable is being examined (Maddala, 1983).

In the estimation of the effect of alcohol consumption on income, alcohol consumption is estimated as ordered data using the limited information and full information approaches similar to that adopted by Chiburis and Lokshin (2007) in their study whereby they estimate a linear regression model with an underlying ordered-probit selection rule. Drinkers are divided into three categories, non-drinkers, moderate drinkers and heavy drinkers based on the recommendations of the Irish Health Promotion Unit (HSE, 2008). Variables used in this study are the same as those used in the estimation of the Multinomial Logit two step OLS estimate as outlined in Chapter 3, Section 3.2.1, page 143.

Alcohol Status Equation

In this model individuals i are sorted into J categories of drinkers 1,2,3 on the basis of an ordered probit selection rule.

$$c_i^* = \alpha' s_i + \varepsilon_i \qquad \textbf{(5.1.1)}$$

Where:	c	category of drinker
	α	is an unknown vector of parameters,
	s	independent variables
	ε	is a standard normal shock
	i	indexes individuals

The amount of alcohol people consume is affected by a range of independent variables. By including all the variables that are included in the income equation in the alcohol status equation, income is accounted for. In addition there are two further variables included in the alcohol status equation that are not included in the income equation in this study. These variables are the variable describing whether or not people regularly partake in Church activities and the variable describing whether or not respondents previously smoked over five years ago. Both these variables have previously been found to have an effect on alcohol consumption but not on income.

Income Equation

The potential household income for individual i with drinking status j is given by equation 5.1.2. Household Income for each individual are hypothesised to depend upon a vector X_i of human capital variables and sociodemographic characteristics and Y_{ij} is observed only if drinking status j is chosen.

$$\ln Y_{ij} = X_i \beta_j + u_{ij} \qquad \textbf{(5.1.2)}$$

Where:	$\ln Y_{ij}$	log of household income
	X	vector of human capital variables & socio-demographic characteristics
	β	coefficients on the observable characteristics
	u_{ij}	error term
	i	indexes individuals where $i = 1,2,......N$
	j	indexes drinking status where $j = 1,2,3,$

Estimating the household income equation allows the $\beta's$ to be estimated across the three categories of drinkers; non, moderate and heavy drinkers, and it is then possible to gauge see whether household income for observed characteristics are greatest for one category of drinker over another.

5.1.1: Estimation of the effect of Alcohol Consumption on Income using the Limited Information Maximum Likelihood Method

Using the LIML method of estimation, the alcohol consumption equation and income equation are estimated separately. In the first step, the alcohol status equation is estimated by an ordered probit of c on s. Since only one category j is observed for each individual and the observations are independent the correlations between u_{ij} and u_{ik} for $j \neq k$ cannot be identified. A consistent estimator of λ_i is derived which is added as an additional regressor in the income equation. The coefficient β_j in the income equation is then estimated with an OLS regression.

5.1.2: Estimation of the effect of Alcohol Consumption on Income using Full Information Maximum Likelihood Method

Using the FIML method of estimation, the alcohol consumption equation and income equation are estimated jointly as opposed to LIML. The parameters to be estimated are

$$\alpha; \beta 1, \beta 2, ...\beta_{J-1;} \mu_1, \mu_2, ...\mu_J; p_1, p_2 ....., p_{J-1}; \sigma_1, \sigma_2, ..., \sigma_{J-1}$$

β_j, p_j, σ_j do not exist for categories drinking status j in which the dependent variable y is missing (Chiburis and Lokshin, 2007).

5.2: Results from both the LIML and FIML Estimations

Both the Limited Information Maximum Likelihood Method of Estimation and the Full Information Maximum Likelihood Method of Estimation is used to measure the effect of alcohol consumption on income. The results of both methods are set out in section 5.2. The results of LIML are discussed in section 5.2.1 and FIML in 5.2.2.

5.2.1: Results from the LIML Method of Estimation

Given that the categories of alcohol consumption are inherently ordered, an ordered probit estimation is used which accounts for this ordinality and hence leads to more accurate results (Greene, 2002). Alcohol Status is estimated by an ordered probit in the first step of the two step model, with non-drinkers being equal to 1, moderate drinkers equal to two and heavy drinkers equal to three. The results of alcohol status estimation are shown in table 5.2.1. The Inverse Mills Ratio is generated which is included as an additional variable in the income regression which accounts for potential selection bias of alcohol consumption. The results of the income regression are shown in table 5.2.2

Table 5.2.1: Results of the Estimation of Alcohol Status as an Ordered Probit using LIML Method

Alcohol Status		
	Coefficient	**Z-Stats**
male	0.345	11.27*
age18to29	0.420	5.89*
age30to39	0.235	3.82*
age40to49	0.236	3.88*
age50to59	0.204	3.35*
age70plus	-0.367	-5.70*
ed secondary	0.231	4.84*
ed diploma/cert	0.281	5.19*
ed primary degree	0.373	6.14*
ed postgraduate	0.277	4.52*
single/never married	-0.073	-1.26
separated/divorced	0.113	1.43
married	-0.030	-0.51
widowed	-0.087	-1.07
village	0.166	3.42*
town	0.162	4.32*
city other than Dublin	0.340	6.66*
Dublin city/county	0.292	7.47*
employee	0.276	3.08*
self employed/farmer	0.234	2.42**
state training/student	0.393	3.42*
unemployed	0.274	2.13**
homemaker	0.116	1.26
retired	0.192	1.94

Table 5.2.1 continued: Results of the Estimation of Alcohol Status as an Ordered Probit using LIML Method

	Coefficient	**Z-Stats**
other	0.209	1.31
No. working in h.hold	0.011	0.78
race white	0.303	2.39**
race black	-0.855	-4.10*
race Asian	-1.019	-4.80*
health excellent	0.454	4.87*
health very good	0.456	5.01*
Health good	0.465	5.15*
health fair	0.338	3.52*
partake Church activities	-0.143	-4.11*
prev smoker 5+yrs	0.208	5.17*
Cut Off 1	0.941	
Cut Off 2	3.24	

No. of Observations = 7870

Wald Chi2(35) = 970.5

Prob > chi2 = 0

Preudo R^2 = 0.0837

Log Likelihood = -5821.0704

*** indicates significance at 1% level, ** indicates significance at 5% level**

Note: Results showing the marginal effects are included in the Appendix F.

Alcohol Status estimated by an ordered probit shows that gender is highly significant and that males are 10% less likely than females to report being a non-drinker and are more likely to be drinkers which is similar to the results from the multinomial logit estimation in Chapter 4 and is in line with the findings of previous studies (Fillmore

1994; Moore, 2005; Blow *et al*, 2005; Moore *et al*, 2005; Mullahy & Sindelar, 1996).

All age categories are significant with results showing that those between the ages of 18 years and 59 years are more likely to be drinkers and in particular those in the category 18-29 years are 12% less likely to be non-drinkers and are more likely to be heavier drinkers. These findings are similar to those using the multinomial logit technique. Those aged 70 years plus are 12% more likely to be non-drinkers which is akin to the findings of Hamilton and Hamilton (1997) and Barrett (2002). None of the marital status variables are significant.

Previous studies show that those with third level qualifications are more likely to be moderate drinkers compared with non or heavy drinkers (Hamilton and Hamilton, 1997; Barrett, 2002). This study finds that all categories describing ones education are very significant with a positive correlation with alcohol consumption. Those who have a primary degree have the largest effect and are more likely to consume higher levels of alcohol consumption compared to those with primary education only. Similarly when alcohol status was estimated using the multinomial logit estimation, those with a primary degree were found to be the least likely to be a non-drinker.

Where ones lives is very significant with results showing that in particular those living in cities are more likely to consume higher amounts of alcohol than those in the country which is similar to the findings of Su and Yen (2000) and similar to those found using the multinomial logit estimation. Those living in a city other than Dublin are almost 10% less likely to be non-drinkers.

The employment variables employee, self employed or a farmer, unemployed, those on state training schemes are all significant and are positively associated with alcohol consumption. Previous studies find that professionals, who work in management and those who work in the service industry are less likely to be abstainers or heavy drinkers (Auld, 2005; Barrett, 2002).

In looking at the individual's race, those of Black and Asian race are more likely to be non-drinkers compared with those in the base category classified as 'other' which is comparable to the findings of Mullahy and Sindelar (1996) and Moore *et al* (2005). Those of Black race are 32% and those of Asian race are 38% more likely to be a non-drinker. Results from the multinomial logit estimation in Chapter 4 also show that those of Black and Asian race are more likely to be a non-drinker compared as opposed to a moderate drinker.

All the health variables are significant and all are strongly correlated to alcohol consumption. In particular those who describe their health as being good, very good or excellent are approximately 13% less likely to be non-drinkers than those in poor health which is analogous to the findings of Berger *et al* (1999), Klatsky *et al* (2001) and Bau *et al* (2007) who show that moderate drinkers enjoy better health than non-drinkers.

There are two explanatory variables specific to the alcohol status equation. One is whether one regularly partakes in Church activity and the other is whether a person was a previous smoker 5 or more years ago. Both are very significant with results showing that those involved in Church activities are almost 5% more likely to be non-drinkers which is similar to Hamilton and Hamilton's (1997) study, and those who are previous smokers are 6% less likely to be non-drinkers which is similar to the findings of Barrett (2002). Results from the multinomial logit estimation in Chapter 4 also find that those involved in Church activities are more likely and previous smokers less likely to be non drinkers.

In step two income regressions are estimated by the three drinking categories. By including the selection correction term potential selection bias is accounted for (Hamilton and Hamilton, 1997: Barrett, 2002). Table 5.2.2 sets out the results of the three income regressions.

Table 5.2.2 Results of the Estimation of Income using LIML Method

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coefficient	**t-stat**		**Coefficient**	**t-stat**		**Coeffi-cient**	**t-stat**	
male	0.093	1.94		0.061	3.23*		0.188	1.96**	
age18to29	0.204	2.73*		0.134	3.08*		0.384	2.42**	
age30to39	0.244	5.00*		0.137	3.99*		0.136	1.12	
age40to49	0.154	2.93*		0.139	4.12*		0.160	1.33	
age50to59	0.077	1.58		0.123	3.51*		0.102	0.81	
age70plus	-0.070	-1.30		-0.059	-1.40		-0.255	-1.67	
ed secondary	0.117	3.22*		0.199	6.49*		0.339	3.35*	
ed diploma/cert	0.217	4.64*		0.342	10.23*		0.523	4.34*	
ed primary degree	0.449	7.11*		0.517	14.07*		0.734	5.38*	
ed postgraduate	0.430	7.53*		0.598	16.99*		0.781	6.27*	
single/never married	-0.312	-6.26*		-0.196	-6.50*		-0.148	-1.87	
separated/divorced	-0.182	-2.67*		-0.294	-7.19*		-0.029	-0.29	
married	0.099	2.09**		0.171	5.66*		0.310	3.9*	
widowed	-0.247	-4.13*		-0.188	-4.31*		0.014	0.10	
village	-0.024	-0.53		-0.010	-0.39		0.084	1.02	
town	0.031	0.88		-0.048	-2.40**		0.034	0.50	
city other than Dublin	0.049	0.85		-0.030	-1.05		0.124	1.16	
Dublin city/county	0.131	2.86*		0.128	6.08*		0.169	1.97**	
employee	0.358	5.25*		0.328	5.99*		0.689	4.95*	
selfemployed/farmer	0.250	3.36*		0.319	5.60*		0.746	5.21*	
statetraining/student	0.168	1.27		0.001	0.01		0.164	0.77	
unemployed	-0.192	-1.90		-0.152	-2.05**		0.279	1.75	

Table 5.2.2 continued: Results of the Estimation of Income using LIML Method

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coefficient	**t-stat**		**Coefficient**	**t-stat**		**Coefficient**	**t-stat**	
homemaker	0.191	3.41*		0.230	4.45*		0.345	2.10**	
retired	0.137	2.30**		0.219	3.89*		0.356	2.04**	
other	0.092	0.72		0.041	0.41		0.223	1.11	
No. working in h.hold	0.105	4.34*		0.135	6.91*		0.132	3.36*	
race white	0.328	2.51**		0.111	1.93		-0.027	-0.15	
race black	-0.257	-1.33		-0.126	-0.86		(omitted)		
race Asian	0.011	0.06		-0.045	-0.35		-0.365	-1.22	
health excellent	0.183	2.56**		0.113	1.97**		0.341	2.01**	
health very good	0.160	2.31**		0.050	0.87		0.385	2.28**	
health good	0.121	1.75		0.002	0.04		0.271	1.62	
health fair	0.074	1.18		-0.056	-1.00		0.226	1.45	
Selection Correction term	0.046	0.30		-0.253	-1.91		0.363	1.29	
_cons	5.143	33.58*		5.601	32.92*		3.967	3.98*	

Non-Drinkers
No. of obs = 2127
F(34, 2092) = 66.54
Prob > F = 00.00
R Squared = 0.4816
Root MSE = .49446

Moderate Drinkers
No. of obs = 5216
F(34, 5181) = 138.91
Prob > F = 00.00
R Squared = 0.4603
Root MSE = .49023

Heavy Drinkers
No. of obs = 527
F(33,493) =17.73
Prob > F = 00.00
R Squared =.5427
Root MSE = .49192

*** indicates significance at 1% level, ** indicates significance at 5% level**

The gender variable is significant for moderate drinkers showing that males who are moderate drinkers are likely to have a slightly higher income compared with females which is comparable to the findings of Zhang (2008). The age variable is particularly significant for non and moderate drinkers showing a positive effect of income up to 70 years. The age category 70 years plus is not significant for any category of drinker. For heavy drinkers the only age category that is significant is 18-29 years and this is strongly positively related to income.

All the education variables are significant across all drinker types. For all types of drinkers those with a primary degree and those with a postgraduate degree have higher incomes as opposed to those with a primary education only which is consistent with the findings of previous studies (Barrett, 2002; French & Zarkin, 1995; Heien, 1996) and consistent with the findings carrying out this estimation using the multinomial logit two step estimation. All the martial status variables are significant for non and moderate drinkers. Compared with those cohabiting incomes for those non and moderate drinkers who are single or never married, separated or divorced and widowed, are likely to be lower. Married people across all drinker types tend to have higher incomes particularly heavy drinkers, which is consistent with previous findings (Berger and Leigh, 1988; Schoeni, 1995; Ahituv and Lerman, 2007; Loh, 1996).

All categories of drinkers living in Dublin city or county tend to have higher incomes and moderate drinkers living in a town tend to have lower incomes compared with those living in the open country. These findings are the same as those from step two of the multinomial logit two step estimation set out in Chapter 4, except for the fact that the variable describing those living in Dublin city or county is not significant for heavy drinkers.

In terms of employment status employees, self employed including farmers, homemakers and those who are retired are all significant across all drinking categories and have a positive correlation with income. Moderate drinkers who are unemployed have lower incomes.

The number of people working in the household variable is very significant and as one would expect is positively correlated to household income, similar to the findings from the multinomial logit two step estimation. The variable describing those of white race is significant for non-drinkers only, with white non-drinkers likely to earn more. Berger and Leigh (1988) find that differences in income between races are insignificant.

In the multinomial logit two step estimation the variable describing those in excellent health is significant for non and moderate drinkers only showing a positive income effect. In this estimation all categories of drinkers with excellent health have higher incomes compared to those in poor health, with heavy drinkers having the highest income. This would correspond to the argument put forward by Grossman (1972). Looking at the health status variables, for moderate drinkers it is only the variable describing those in excellent health that is significant.

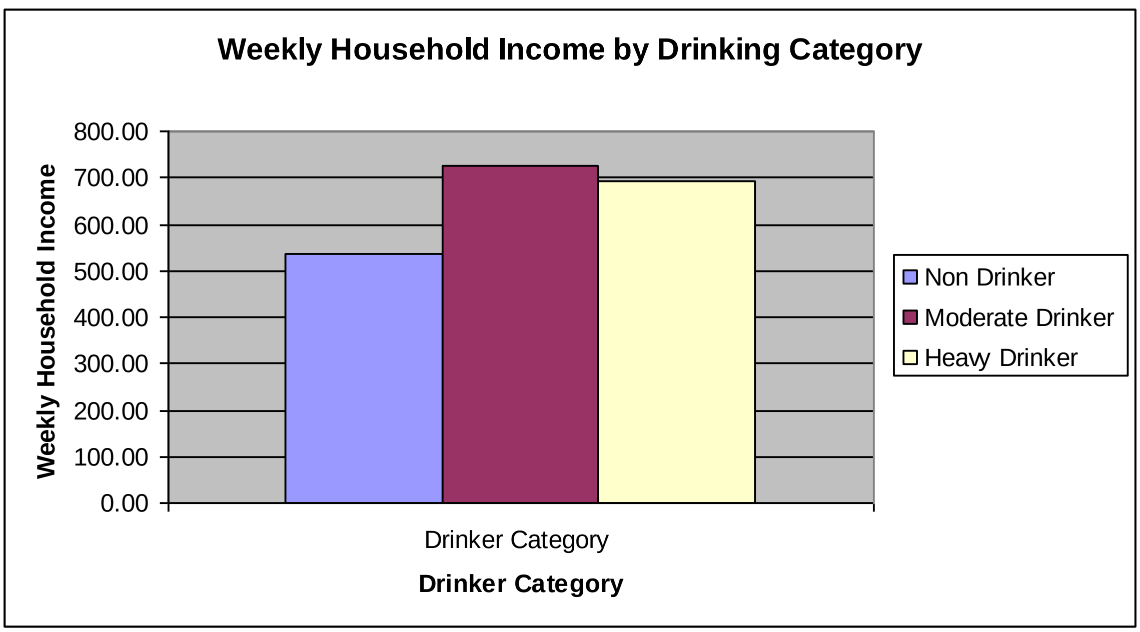
When estimating the effect of alcohol consumption as a Multinomial Logit, the Inverse Mills Ratio is not significant for either non or heavy drinkers, however it is significant for moderate drinkers indicating that there is a selection effect for moderate drinking which is the same as the findings using the multinomial logit two step estimation. Estimating alcohol status as an ordered probit, results show that Inverse Mills Ratio is not significant for any category of drinker showing that there is no selection effect.

Both Hamilton and Hamilton (1987) and Barrett (2002) find that in relation to non and moderate drinkers there is no selection effect however they did find a selection effect in relation to heavy drinking showing that individuals who self select into heavy drinking have higher incomes on average than an heavy drinker with identical observable characteristics drawn at random.

5.2.1.1: Household Income Differentials between the three categories of drinkers

Income regressions are estimated for each of the three categories of drinkers accounting for potential selection bias of alcohol consumption. Many previous studies have had similar findings in that the income of moderate drinkers is more than either abstainers or heavy drinkers (Hamilton and Hamilton, 1997; Barrett, 2002; French and Zarkin, 1995). Findings from this study show that household income for moderate drinkers is highest while income for non-drinkers is lowest. Weekly household income for non drinkers is €535.95, moderate drinkers is €725.45 per week and heavy drinkers is €694.18 per week. This is depicted in figure 5.2.1.

Figure 5.2.1: Weekly Household Income by Drinking Category using the LIML Estimation



(Source: Authors own)

Weekly household income is highest for moderate drinkers and lowest for non-drinkers. Average weekly income for non and moderate drinkers are higher when estimated by Ordered Probit compared with the Multinomial Logit. When using the Ordered Probit OLS Two Step model, income for moderate drinkers is higher than non and heavy drinkers whereas when the Multinomial Logit OLS Two Step model is income for heavy drinkers is slightly more than that of moderate drinkers. With both methods, income for non drinkers is substantially less than either moderate or heavy drinkers.

5.2.1.2: Post Estimation Diagnostics

Testing the Specification of the Model

The significance of each of the variables is assessed using the t and z statistics and results highlight which variables are significant at both the 1% and 5% levels of significance. The Likelihood Ratio test is also used to evaluate the significance of each variable in the model and ensure that each variable is beneficial to the model.
The Wald Test and the F Test show that the models are statistically significant and reject the null that coefficients of the variables are equal to zero.

Robust standard errors are used deal with the presence of heteroskedasticity.

Cut offs not being equal to each other is tested for by testing the null hypothesis cut-off1 less cut-off2 = 0. The null is rejected in all cases showing that the cut offs are not equal to each other.

5.2.2: Results from the FIML Method of Estimation

In looking at the effect of alcohol consumption on household income, the Full Information Maximum Likelihood (FIML) method is used to estimate the linear regression model income with an underlying ordered probit selection rule. This section presents the results of the ordered probit regression for alcohol status and the selection corrected income regressions.

Alcohol Status

Alcohol status is estimated as an ordered probit, with drinkers categorised into one of three categories; non-drinkers, moderate drinkers and heavy drinkers. Table 5.2.3 contains the results of the ordered probit on drinking status using data from the Slán 2007 survey. Results of the alcohol status equation estimated by Full Information Maximum Likelihood method are very similar to the results of the ordered probit estimated in step one of the two step estimation.

Table 5.2.3: Results of the Estimation of Alcohol Status as an Ordered Probit using FIML Method

Alcohol Status		
	Coefficient	**Z-Stats**
Male	0.346	11.27*
age18to29	0.427	5.99*
age30to39	0.240	3.91*
age40to49	0.238	3.93*
age50to59	0.206	3.37*
age70plus	-0.369	-5.73*
edsecondary	0.231	4.84*
eddiplomac~t	0.278	5.14*
edprimaryd~e	0.369	6.09*
edpostgrad~e	0.273	4.46*
singleneve~d	-0.073	-1.28
sepdiv	0.113	1.43
married	-0.031	-0.53
widowed	-0.088	-1.08
village	0.165	3.39*
Town	0.162	4.32*
cityothert~n	0.342	6.69*
dublincity~y	0.293	7.51*
employee	0.277	3.08*
selfemplin~r	0.235	2.43**
statetrain~d	0.394	3.41*
unemployed	0.274	2.13**
homemaker	0.116	1.26

Table 5.2.3 continued: Results of the Estimation of Alcohol Status as an Ordered Probit using FIML Method

	Coefficient	**Z-Stats**
Retired	0.191	1.92**
Other	0.208	1.31
numworkinghh	0.010	0.73
Race White	0.309	2.42**
Race Black	-0.857	-4.09*
Race Asian	-1.015	-4.77*
healthexce~t	0.453	4.86*
healthvery~d	0.455	5.01*
healthgood	0.466	5.15*
healthfair	0.339	3.53*
churchact	-0.124	-3.25*
pr~vemoreyrs	0.223	5.61*

No. of Observations = 7870

Wald Chi2(35) = 970.51

Prob > chi2 = 0

Log Likelihood = -11346.17

*** indicates significance at 1% level, ** indicates significance at 5% level**

Results from estimating the alcohol status equation using the FIML method are very similar to the results from the estimation of alcohol as a two step method. The results show that the gender variable has a very significant effect on alcohol status at the 1% level and being a male has a positive effect. Age across all categories is very significant in terms of alcohol consumption. There is a positive correlation between all ages and alcohol consumption up to age 70 plus years. Those aged 18-29 are most likely to be in the higher drinking category and those over 70 years are likely to be non-drinkers. All categories of Education have a very significant positive correlation with alcohol status with the largest effect being the category of respondents with a primary degree.

Marital Status is not significant in terms of ones alcohol consumption which is in contrast to previous findings such as Barrett (2002), Auld (2005) and Hamilton and Hamilton (1997). Where a respondent is currently living is very significant for all categories and there is a positive correlation between all categories and alcohol status with the largest being for those living in a city other than Dublin.

The explanatory variable describing the respondent's current employment status is significant for all categories except for that of homemakers and those whose employment status is described as other. All have a positive correlation with alcohol status with the largest effect being for those in state training schemes.

The number of people working 15 hours or more per week in the household is not significant. Race is very significant in the alcohol status equation. Those of white race are more likely to consume higher levels of alcohol. A Black or an Asian person is less likely to drink and is likely to be a non-drinker compared to those in the base category classified as being of 'other' race, similar to the findings from studies carried out by Mullahy and Sindelar (1996) and Moore *et al* (2005).

Health Status is strongly related to alcohol consumption. All categories of health status are very significant and all have a strong positive effect, compared to those in poor health.

Findings show that an individual who regularly partakes in Church activities are less likely to consume alcohol. Respondents who previously smoked more than five years ago are more likely to consume alcohol. Barrett (2002) uses the variable whether or not one smoked at the age of 18 and finds that individuals who did are not likely to be current non drinkers.

In general the finding in terms of the effect of independent variables on alcohol status using the Full Information Maximum Likelihood Method of estimation, are similar to those using the ordered probit two step model.

Income Equations corrected for selection bias

The estimation of the income equation for all three categories of drinker allowing for the endogeneity of drinking status is described using the Full Information Maximum Likelihood method. The objective of the analysis is to look at whether or not there is an income premium for the different categories of drinker i.e. does one category of drinker have a higher income than another. Results for the income regressions are presented in table 5.2.4.

Table 5.2.4 Results of the Estimation of Income using FIML Method

	Non Drinkers		Moderate Drinkers		Heavy Drinkers	
	Coefficient	z-stat	Coefficient	z-stat	Coefficient	z-stat
Male	0.091	2.88*	0.047	2.57**	0.139	2.55**
age18to29	0.201	3.44*	0.121	2.91*	0.325	2.67*
age30to39	0.242	5.72*	0.135	4.08*	0.103	0.99
age40to49	0.152	3.31*	0.136	4.20*	0.125	1.23
age50to59	0.076	1.71	0.122	3.63*	0.067	0.63
age70plus	-0.068	-1.58	-0.070	-1.88**	-0.192	-1.67
Edsecondary	0.116	3.88*	0.201	7.15*	0.304	3.89*
eddiplomac~t	0.215	5.45*	0.342	11.08*	0.482	5.20*
edprimaryd~e	0.447	8.48*	0.511	15.08*	0.683	6.83*
edpostgrad~e	0.428	8.32*	0.598	18.02*	0.737	7.57*
singleneve~d	-0.311	-6.39*	-0.193	-6.32*	-0.135	-1.81
Sepdiv	-0.182	-2.74*	-0.297	-7.22*	-0.045	-0.46
Married	0.099	2.12**	0.172	5.63*	0.316	4.08*
Widowed	-0.247	-4.18*	-0.195	-4.48*	0.028	0.20
Village	-0.025	-0.58	-0.013	-0.52	0.061	0.85
Town	0.030	0.95	-0.051	-2.59*	0.012	0.19
cityothert~n	0.047	1.01	-0.042	-1.54	0.075	0.99
dublincity~y	0.129	3.59*	0.119	5.80*	0.126	2.18**
Employee	0.356	5.85*	0.321	5.91*	0.648	5.39*
selfemplin~r	0.249	3.57*	0.315	5.58*	0.709	5.47*
statetrain~d	0.166	1.33	-0.014	-0.19	0.099	0.55
unemployed	-0.194	-2.00**	-0.156	-2.12**	0.238	1.66
homemaker	0.190	3.5*	0.223	4.30*	0.325	2.08**
retired	0.136	2.38**	0.220	3.93*	0.318	2.02**
other	0.090	0.72	0.036	0.36	0.191	1.01
numworkinghh	0.105	4.41*	0.135	6.87*	0.129	3.44*
race white	0.327	2.60*	0.103	1.78	-0.077	-0.48

Table 5.2.4 continued: Results of the Estimation of Income using FIML Method

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coefficient	**z-stat**		**Coefficient**	**z-stat**		**Coefficient**	**z-stat**	
race black	-0.252	-1.46		-0.149	-1.12		(omitted)		
race Asian	0.017	0.11		-0.064	-0.55		-0.212	-1.13	
healthexce~t	0.181	3.00*		0.120	2.25**		0.274	2.08**	
healthvery~d	0.158	2.77*		0.057	1.08		0.317	2.43**	
healthgood	0.118	2.11**		0.009	0.17		0.203	1.59	
Healthfair	0.073	1.28		-0.048	-0.88		0.177	1.32	
_cons	5.142	34.24*		5.527	53.51*		4.565	12.95*	

	Coefficient	**Robust Standard Error**
/cutoff1	0.954	0.182*
/lndelta2	0.833	0.012*
/athrho1	-0.111	0.150
/athrho2	-0.293	0.081*
/athrho3	0.391	0.153*
/lnsigma1	-0.708	0.025*
/lnsigma2	-0.689	0.022*
cutoff1	0.954	0.181
cutoff2	3.254	0.186
rho1	-0.111	0.148
rho2	-0.285	0.075
rho3	0.372	0.137
sigma1	0.492	0.013
Sigma2	0.502	0.011
Sigma3	0.506	0.035

Wald Test if indep. eqn. (rho=0): chi2 (3) = 18.64 Prob > chi2 = 0.0003
No. of Observations = 7870 Wald Chi2 (33) = 970.57
Prob > chi2 = 0 Log Likelihood = -11346.171

*** indicates significance at 1% level, ** indicates significance at 5% level**

Results show that gender has proven to be significant in terms of income across all drinker types. The age variable has a particularly significant effect on income for non and moderate drinkers but this is not the case for heavy drinkers. All ages up to 70 yrs have a positive effect on income; however it is those in the category of 30 to 39 that have the highest age-income profile for non and moderate drinkers. In relation to heavy drinkers the only age category that is significant is that of those aged between 18 years and 29 years having a large positive effect on household income.

As one might expect, the returns to education are extremely significant across all drinker categories, with the highest income being for those with a primary degree and those with a postgraduate qualification which is similar to the findings of others (Barrett, 2002; French & Zarkin, 1995; Heien, 1996).

The significance of the different categories describing marital status varies greatly between the three groups of drinkers. The category single/never married is significant for both non-drinkers and moderate drinkers. Being married is significant for all categories of drinkers there being a positive relationship with income with heavy drinking having the largest effect similar the findings of previous studies (Berger and Leigh, 1988; Schoeni, 1995; Ahituv and Lerman, 2007; Loh, 1996). Being separated, divorced or widowed has a very significant impact on the income of non and moderate drinkers and all with a negative coefficient on these variables, compared with those cohabiting,

In terms of location describing where respondents are living, the only category that is significant for all three categories of drinking is that which describes those who live in Dublin city or county, which has a positive effect on the income compared to those living in the country.

The variable describing ones current employment situation is significant across all drinker types, except for the variables describing state training schemes/and students and those classified as ‘other’. Being employed or self employed as one might expect along with being retired or a homemaker are all very significant effect in terms of income across all drinker types and all have a positive effect on income. Being unemployed is significant and has a negative effect on household income for non drinkers and moderate drinkers but surprisingly being unemployed is not significant in terms of the income of heavy drinkers.

The number working in the household is very significant and has a positive effect on the income of all drinkers.

Race is only significant in terms of the incomes of white people who are non-drinkers. Being of white race and a non-drinker has a positive effect on income. Race is not significant in the income of moderate and heavy drinkers similar to what Berger and Leigh (1988) show in their study.

Most of the variables describing ones health status is significant for non drinkers except for the variable describing health as fair. Excellent health status is the only significant variable for moderate drinkers and health status that is described as being excellent or very good are the only significant variables for heavy drinkers. Where respondents describe their health status excellent there is a positive effect on income, compared to those with poor health.

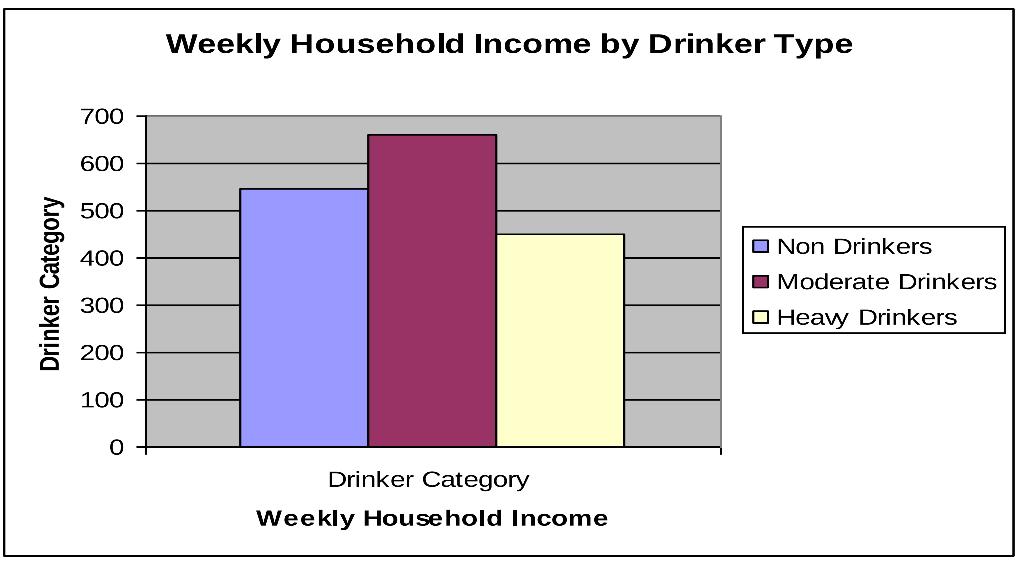
Findings in terms of the effect of independent variables on income using the Full Information Maximum Likelihood Method are also similar to those using the ordered probit two step model.

5.2.2.1: Household Income Differentials between the three categories of drinkers

The income equations are estimated for each of the three categories of drinkers. The log of income is predicted for each of the drinking categories. The Wald test, tests the null hypothesis that there is zero correlation between the error terms in the alcohol equation and the income equation. In this case the null hypothesis is strongly rejected hence there is a need for selection bias correction and an OLS regression would lead to biased results.

The greatest proportion of people are categorised as moderate drinkers. The number of non drinkers is 2,127; the number of moderate drinkers is 5,216 drinkers and heavy drinkers 527. The average log income for non drinkers 6.304 equates to €546.75; for moderate drinkers the mean log of income if 6.492 which equates to €660.10; and for heavy drinkers the mean log of income is 6.109 which is an average of €449.99 income per week. Figure 5.2.2 sets out the percentage differences between the three categories of drinkers.

Figure 5.2.2 Weekly Household Income by Drinking Category using the FIML Estimation



(Source: Authors own)

Results from the analysis using the Full Information Maximum Likelihood Method are similar to the ordered probit two step analysis in so far as moderate drinkers have the highest weekly household income, higher than that of both non drinkers and heavy drinkers. However, a major difference between the results of the two methods is that the income of heavy drinkers is the lowest and substantially so with the FIML method.

This clearly shows that there is an income premium for moderate drinkers compared with non or heavy drinkers, with heavy drinkers having the lowest. This is in contrast to findings using the two step limited information method of estimation, whereby the income of non drinkers is the lowest. Many previous studies into the effects of alcohol consumption on income in other countries similarly find that moderate drinkers earn more than non-drinkers and heavy drinkers even after correcting for endogeneity (Hamilton and Hamilton, 1997; Barrett, 2002).

5.2.2.2: Post Estimation Diagnostics

Testing the Specification of the Model

The significance of each variable is assessed using the z-statistic and results highlight which variables are significant at both the 1% and 5% levels of significance. The Likelihood Ratio test is also used to evaluate the significance of each variable in the model and ensure that each variable is beneficial to the model. The results of both these tests show that each of the variables included in the model has resulted in a statistically significant improvement in the model fit. The Wald Test of independent equations shows that the null stating that the equations are independent and the error terms are not correlated, is rejected showing that correlation between the error terms in the alcohol and income equations exists.

Robust standard errors are used deal with heteroskedasticity.

Cut-off values of the dependent variable estimate the boundaries of each category and differentiate between the different categories. Cut offs not being equal to each other is tested for by testing the null hypothesis cut-off1 less cut-off2 = 0. The null is rejected in all cases showing that the cut offs are not equal to each other.

5.3: Conclusion

This chapter presents an empirical study of the effect of individual alcohol consumption on household income in Ireland while accounting for the potential endogenous relationship between income and alcohol. This is estimated by both the Limited Information Maximum Likelihood Method and the Full Information Maximum Likelihood Method whereby alcohol status is estimated as an ordered probit and income is estimated as an OLS.

Results show that there is an income premium for moderate drinkers compared with non-drinkers and heavy drinkers. With the two step method non drinkers have the lowest income while with the full information method; heavy drinkers have the lowest income. These results are similar to the findings of other studies in so far as moderate drinkers have the highest income more than non-drinkers and heavy drinkers even after correcting for endogeneity (Hamilton and Hamilton, 1997; Barrett, 2002).

The policy implications of these findings are that moderate drinkers have the highest income, more than either abstainers or heavy drinkers and while the misuse of alcohol consumption needs to be targeted, moderate levels of alcohol consumption can have positive affect and this needs to be considered in targeting the development of policies around the misuse of alcohol consumption. With the FIML method of estimation not only do moderate drinkers have the highest income, but heavy drinkers have the lowest and given that previous studies into both methods of estimation find that overall FIML is a better method (Puhani,2000), these results in particular encourage the targeting of heavy drinkers to reduce levels of consumption.

CHAPTER 6

AN INVESTIGATION INTO THE EFFECT OF ALCOHOL CONSUMPTION ON HEALTH STATUS AND HEALTH CARE UTILISATION

This chapter investigates the effect of alcohol consumption on ones health status and health care utilisation in Ireland while accounting for the potential endogenous relationship between alcohol and health. Drinkers are categorised into three categories non, moderate and heavy drinkers.

Techniques similar to those set out by Greene and Hensher (2010) and Langpap and Kerkvliet (2002), Vella (1998) and Chiburis and Lokshin (2007) are followed. The drinking status equation is estimated using an ordered probit model and from this the predicted values for the inverse mills ratio is generated which is then included in the health and health care utilisation equations. Differences in health status for each of these categories of drinkers is examined and the relationship between both alcohol status and health with a host of other personal and socio-economic variables such as age, gender, marital status, employment status and level of education, among others, is also identified. Health care utilisation is also analysed by drinker type. Health care utilisation is used as an alternative measure of health (Lim *et al*, 2005; Rotermann, 2006; Finkelstein, 2001) in this study given that the measure of health status available is self-assessed. The relationship between alcohol status and specific illnesses is also analysed.

6.1: Empirical Techniques

Similar to the technique set out in Chapter 2, Section 2.2.1 page 58, the alcohol status equation is first estimated and from this the inverse mills ratio is derived which accounts for selection bias. This is then included as an additional regressor in both the health status and health care utilisation equations estimated in step two.

6.1.1: Alcohol Status

In this study drinkers are divided into three categories; non-drinkers, moderate drinkers and heavy drinkers and alcohol consumption as set out in equation 6.1.1 is estimated as an ordered probit model. A range of independent variables s describing personal and socio demographic characteristics that affect alcohol consumption are included.

$$c_i^* = \alpha' s_i + \varepsilon_i \qquad \textbf{(6.1.1)}$$

Where:

	c	category of drinker
	α	is an unknown vector of parameters,
	s	independent variables
	ε	is a standard normal shock
	i	indexes individuals where $i = 1,....,n$
	j	indexes drinking status where $j = 1,2,3,$

The independent variables s include all the variables from the health status equation which then accounts for health. In addition the variable describing whether or not one partakes in Church activities is included in the alcohol status equation only as this does have an affect on alcohol consumption but not on health. An estimation of the selection correction term λ_i is then computed for each individual in the sample.

6.1.2: Health Status Estimation

In the second step of the two-step estimation, the health status equation set out in 6.2.1 is estimated for each category of drinker c by an ordered probit regression and λ is also included in this equation as an additional regressor which accounts for selection bias.

$$h^*_{ij} = x_i \beta_j + u_{ij} \qquad \textbf{(6.1.2)}$$

Where:	h	health measure of individual
	x	vector of independent variables
	β	coefficients on the observable characteristics
	u	error term
	i	indexes individuals where $i = 1,2,......N$
	j	indexes drinking status where $j = 1,2,3,$
	k	categories of health status

As opposed to overall health status, specific health conditions are estimated as a probit model as set out by Greene and Hensher (2010) whereby if an individual reports suffering from a specific condition $d_i = 1$ and if not $d_i = 0$. Similarly as in the estimation of health status, the selection correction term from the estimation of alcohol status by ordered probit in step 1 is added as an additional regressor in step two.

6.1.3: Health Care Utilisation Estimation

Health care utilisation is used as an alternative measure of health in this study given that the measure of health status available is self-assessed and both variables are highly correlated (Lim *et al*, 2005; Rotermann, 2006; Finkelstein, 2001). Health Care Utilisation is estimated in the same manner as health status. The dependent variable Health Care Utilisation is based on the number of GP consultations respondents had prior to the survey and is estimated by ordered probit and by drinker type to account for the potential selection bias of alcohol consumption. Alcohol Consumption is estimated initially by ordered probit and the inverse mills ratio is derived from this. The inverse mills ratio is then included as an additional regressor in the health care utilisation equation which is also estimated by ordered probit.

6.1.4: Endogeneity

By estimating the selection correction term, including this as an additional regressor in the health status equation, and then going onto estimate the health status equation by drinker type, means that selection bias and the endogeneity of alcohol status is accounted for.

Lifestyle variables are potentially endogenous in terms of health status (Contoyannis and Jones, 2004; Yen *et* al, 2010; Yen *et* al, 2010). In the Slán dataset respondents are not asked about access to health inputs, characteristics of parents or price variation, variables which could be used as instruments to account for endogeneity. In relation to cigarette prices in Ireland these do not vary by region as regional taxes do not apply. Hence given the lack of viable instruments to account for the potentially endogenous lifestyle variables; smoking, weight, and whether or not one is physically active, means that further exploration of such potential endogenous relationships is not possible.

The model is however also estimated omitting the endogenous lifestyle variables smoking, physical activity and weight to see if the results vary significantly.

Establishing the direction of causality between health and income also poses significant problems (Kiuila and Mieszkowski, 2007). Unable to instrument for income, this study includes variables describing the respondent’s employment status and education which allows for the partial control for the possible effect of poor health on low income (Kiuila and Mieszkowski, 2007; Contoyannis and Jones, 2004).

6.2: Results in the Estimation of the Effect of Alcohol on Health Status

Results form examining the effect of alcohol consumption on individual health status using a two step estimation are set out in this section. Section 6.2.1 presents the results from the ordered probit alcohol status estimation. Section 6.2.2 presents the results for the selection corrected health status equations estimated by the ordered probit model. Section 6.2.3 shows the predicted health status for different categories of drinkers. Section 6.2.5 looks at the effect of drinking on specific health conditions. Section 6.3 presents results for the effects of alcohol consumption on health care utilisation.

6.2.1: Ordered Probit Results for Drinking Status

This paper estimates the effect of alcohol consumption on an individual's health status while accounting for the self selection of individuals into drinking categories. An ordered probit model is used to explain an individual's choice of drinking status. Individuals are categorised into one of three categories of drinkers; non-drinkers, moderate drinkers and heavy drinkers. Non drinkers are reported as category 1, moderate drinkers are category 2 and heavy drinkers are in category 3.The explanatory variables used in the drinking status equation are; gender, age, education, employment situation, log of income, race, marital status, total number in household, where respondents live, smoking, physical activity and weight which are the explanatory variables also assumed to influence health status. By including all of these variables in the drinking status choice equation, the effect of health status on drinking behaviour is controlled for which is similar to what Hamilton and Hamilton (1997) and Barrett (2002) did in their study into the effect of alcohol status on income.

In addition to these variables, a variable describing whether or not one partakes regularly in Church Activities is included as additional exogenous variable in the drinking status equation. This variable is hypothesised to be unique to the drinking decision, as it does not affect health status. This variable is used based on previous studies into what affects alcohol consumption (Berger and Leigh, 1988; Hamilton and Hamilton, 1997; Barrett, 2002). Previous smoking is not included in the alcohol status equation as it is correlated with current smoking which is included as a lifestyle variable that impacts on health. Table 6.2.1 contains the results of the ordered probit estimation of drinking status using data from the 2007 Slán Survey. The coefficients listed in table 6.2.1 indicate the effect each variable has on the probability of an individual being in a higher drinking category. The corresponding z-statistics, testing the null hypothesis of statistical significance of the variables in the alcohol status equation is also given.

Table 6.2.1: Estimation of the Drinking Status equation using the Ordered Probit Model

	Coefficient	**Z-Statistic**
male	0.332	11.23*
Married	-0.012	-0.20
Widowed	-0.061	-0.77
Sep/div	0.108	1.38
Single/never married	-0.014	-0.24
Educ. Secondary	0.232	5.15*
Educ. Diploma	0.287	5.39*
Educ. Primary Degree	0.335	5.48*
Educ. Post Graduate	0.251	4.04*
Age18-29	0.355	5.06*
Age 30-39	0.171	2.84*
Age 40-49	0.209	3.48*
Age 50-59	0.176	2.99*
Age 70plus	-0.294	-4.73*
Employee	0.330	3.98*
Self employed	0.306	3.42*
State Training or student	0.553	5.08*
Unemployed	0.375	3.13*
Homemaker	0.200	2.32**
Retired	0.318	3.47*
Other	0.278	1.85
Logincome	0.188	6.74*
Race White	0.292	2.49**
Race Black	-0.633	-3.38*

Table 6.2.1 continued: Estimation of the Drinking Status equation using the Ordered Probit Model

	Coefficient	**Z-Statistic**
Race Asian	-0.849	-4.26*
Total in hh	-0.005	-0.92
Village	0.160	3.39*
Town	0.154	4.22*
City other than Dublin	0.317	6.44*
Dublin city	0.249	6.56*
Smoker	0.337	10.48*
Weight right	0.173	2.9*
Weight too heavy	0.253	4.12*
Weight too light	0.189	2.06**
Medical Card Holder	-0.001	-0.02
Health Insurance	0.077	2.43**
Church activities	-0.134	-3.97*
Cut Off 1	2.02	
Cut Off 2	4.35	

No. of Observations = 8519

Wald chi2 (37) = 1192.83

Prob > chi2 = 0.0000

Pseudo R^2 = 0.0958

Log Likelihood = -6232.75

*** indicates significance at 1% level, ** indicates significance at 5% level**

Note: The average price of alcohol was included as a variable in the alcohol status equation. The price was derived by dividing the total values of sales in the 2006 by the total volume sold for each type of alcohol in 2006. Price was dropped due to collinearity and the variable Physically Active was also dropped due to collinearity.

Marginal Effects results in Appendix G

The first independent variable considered is gender, whereby the results show that this is a statistically significant variable in the alcohol status equation. Males are more likely to be in a higher drinking category than females which is similar to the findings in previous studies (Fillmore 1994; Moore, 2005; Blow *et al*, 2005; Moore *et al*, 2005; Mullahy & Sindelar, 1996). Females are less likely to be in a higher drinking category than males and are likely to be non or moderate drinkers. Marginal effects show that males are 10.5% less likely to be a non-drinker than female. None of the variables describing marital status are significant in terms of alcohol status.

The explanatory variables describing individual's level of education are all significant and all are positively correlated with alcohol status. Those with a primary degree are more likely to be heavy drinkers or 9.8% less likely to be a non drinker than those with a primary education only, which is different to the findings of Hamilton & Hamilton (1997), who find that higher educated people, those with third level degrees, tend to consume moderate amounts of alcohol and they are less likely to abstain or be heavy drinkers.

The variable age is also significant in terms of alcohol consumption. All ages up to 59 years are positively correlated with alcohol status. In particular those aged 18-29 years are more likely to be in a higher drinking category and are 10.6% less likely to be a non-drinker. Similar to the findings of Hamilton and Hamilton (1997) and Barrett (2002), those aged 70 years or over are less likely to be in a higher drinking category and are 10% more likely to be non-drinkers.

All the variables describing Employment Status are significant except those in the 'Other' category. In particular those on State training schemes or students are more likely to be in a higher category of drinking, with marginal effects showing that they are 14.6% less likely to be non-drinkers. Compared with those with a disability, homemakers are the least likely to be in a higher drinking category and are 6.1% less likely to be a non-drinker.

Log of income is statistically significant. If income was to increase the respondents are more likely to be in a higher drinking category hence showing a positive correlation between income and drinking.

The explanatory variables describing race are all significant. A white person, either white Irish or a person of any other white background is more likely to be in a higher drinking category compared with those of other races. They are 10.2% less likely to be a non-drinker. Both Black's and Asians are less likely to be in a higher drinking category. Blacks are 23.5% more likely to be non-drinkers, and Asians 32% more likely to be non-drinkers. Similarly Mullahy & Sindelar (1996) and Moore *et al (*2005) find that white people tend to consume greater amounts of alcohol and that those who abstain from alcohol tend less often to be white.

Total number of people in the household is not significant in the determination of alcohol consumption.

Where a person lives is a significant variable in terms of alcohol consumption. Those who live in Dublin are 7.7% less likely and those living in a city other than Dublin are 9.3% less likely to be non-drinkers and more likely to be in a higher drinker category when compared with those living in the open country which is similar to the findings of Su and Yen (2000).

In terms of the lifestyle variables, Smoking is highly significant and is positively correlated with alcohol consumption. A smoker is 10% less likely to be a non-drinker compared with a non-smoker which is similar to previous findings (Gulliver *et al*, 1995; Burton and Tiffany, 1997). The variable Physically Active is dropped due to collinearity. The variables describing self assessed weight as being right or too heavy or too light are significant. In particular those who described their weight as too heavy are more likely to consume higher amounts of alcohol and are 8% less likely to be non-drinkers. Previous studies show that alcohol has only a slight effect on weight (Williamson *et al*, *1*987).

Having health insurance is significant and those with health insurance are 2.5% less likely to be a non drinker. Having a medical card is not significant.

The additional explanatory variable that is included in the alcohol status equation but not in the health status equation, is whether respondents regularly partake in Church activities. This is a highly significant variable with a P-value of 0. Church activities are negatively correlated with alcohol consumption. When included in the health status equation is was not found to be significant. Those who regularly partake in Church activities are less likely to be in a higher drinking category, compared with those who do not regularly partake in Church activities, and are in fact 4.4% more likely to be a non-drinker which is similar to what Hamilton and Hamilton (1997) find in their study.

6.2.2: Ordered Probit Results for Health Status

Step two of Selection Model used involves estimating respondents health status separately by drinker type allowing for the self selection of drinking status, by including the predicted values for the selection correction term, which were derived by estimating the ordered probit alcohol consumption in step one.

The Health Status equation is estimated as an ordered probit. Respondents are classified into one of five categories of self assessed health; Poor Health = 1, Fair Health = 2, Good Health = 3, Very Good Health = 4 and Excellent Health = 5.

The objective of this analysis is to test whether different categories of drinkers report higher self assessed health status. Results for the selection corrected health status equations are presented in table 6.2.2.

Table 6.2.2: Estimation of the Health Status Equation corrected for Selection Bias using the Ordered Probit Model

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coefficient	**z-stat**		**Coefficient**	**z-stat**		**Coefficient**	**z-stat**	
Educ Secondary	0.452	4.96*		0.215	3.57*		0.327	0.98	
Educ. Diploma	0.565	4.78*		0.228	3.37*		0.267	0.68	
Educ Prim Dgr	0.689	4.63*		0.333	4.43*		0.284	0.67	
Educ Post Grad	0.479	3.70*		0.388	5.05*		0.177	0.47	
Age18-29	1.303	7.22*		0.434	5.32*		0.701	1.35	
Age 30-39	0.786	6.64*		0.335	4.81*		0.778	2.35**	
Age 40-49	0.679	5.51*		0.257	3.65*		0.808	2.25**	
Age 50-59	0.382	3.66*		0.009	0.13		0.402	1.22	
Age 70plus	-0.485	-4.09*		0.075	0.82		-0.176	-0.35	
Employee	1.593	9.79*		1.126	10.92*		1.439	3.02*	
Self employed	1.418	8.69*		1.172	10.64*		1.231	2.61*	
State Training or Student	1.841	7.31*		1.116	8.62*		1.397	1.94	
Unemployed	1.420	6.6*		1.266	8.9*		1.305	2.37**	
Homemaker	1.209	8.66*		1.068	10.14*		0.861	2.10**	
Retired	1.260	7.93*		0.858	7.29*		1.309	2.34**	
Other	1.157	5.26*		1.366	7.56*		-0.668	-1.01	
Log income	0.299	3.76*		0.118	3.64*		0.249	1.09	
Race White	0.303	1.62		-0.072	-0.58		0.535	1.08	
Race Black	-0.461	-1.48		-0.180	-0.73		(omitted)		
Race Asian	-1.489	-4.37*		-0.118	-0.44		-0.324	-0.32	
Male	0.365	2.87*		-0.136	-3.67*		0.042	0.11	
Married	0.202	1.64		0.118	1.91		-0.015	-0.08	
Widowed	0.311	2.21*		0.039	0.43		0.026	0.07	

Table 6.2.2 continued: Estimation of the Health Status Equation corrected for Selection Bias using the Ordered Probit Model

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coefficient	**z-stat**		**Coefficient**	**z-stat**		**Coefficient**	**z-stat**	
Separated/divorced	0.321	2.08**		0.117	1.40		-0.087	-0.34	
Single/never married	0.180	1.47		0.103	1.70		-0.095	-0.50	
Total in h.hold	-0.003	-0.33		-0.009	-1.53		0.015	0.77	
Village	0.086	0.99		-0.043	-0.81		-0.024	-0.1	
Town	0.086	1.10		-0.097	-2.36**		0.022	0.10	
City not Dublin	0.418	3.02*		0.057	1.04		0.200	0.54	
Dublin city	0.287	2.67*		0.007	0.16		0.074	0.24	
Smoker	0.177	1.36		-0.279	-7.35*		-0.054	-0.14	
Weight right	0.316	2.94*		-0.041	-0.56		0.121	0.39	
Weight too heavy	0.069	0.55		-0.345	-4.50*		-0.174	-0.45	
Weight too light	-0.201	-1.27		-0.309	-2.81*		-0.281	-0.71	
Medical Card Holder	-0.213	-3.39*		-0.220	-5.10*		0.168	1.15	
Health Insurance	0.246	4.20*		0.086	2.43**		0.143	1.02	
Correction sel.	-1.798	-3.80*		-0.408	-1.61		0.830	0.62	
Cut Off 1	0.924			-0.703			3.501		
Cut Off 2	1.912			0.274			4.636		
Cut Off 3	2.902			1.414			5.824		
Cut Off 4	3.972			2.543			6.843		

Non-Drinkers	**Moderate Drinkers**	**Heavy Drinkers**
No. of obs = 2372	No. of obs = 5596	No. of obs = 551
Wald chi2 (37) = 727.83	Wald chi2 (37) = 1098.67	Wald chi2 (36) = 132.65
Prob > chi2 = 00.00	Prob > chi2 = 00.00	Prob > chi2 = 00.00
Pseudo R^2 = 0.1108	Pseudo R^2 = 0.08	Pseudo R^2 = 0.0745
Log likelihood = -3146.91	Log likelihood = -6950.46	Log likelihood = -696.02

*** indicates significance at 1% level, ** indicates significance at 5% level**

Note: Physically Active is dropped in all three categories of drinkers due to collinearity. Race-black is also dropped due to collinearity in the health status equation for heavy drinkers.
Marginal Effects results in Appendix G.

Education is a very significant variable in terms of health status for both non and moderate drinkers but not for heavy drinkers. Both non and moderate drinkers with third level education are more likely to report having a better health status compared with those who have a primary level of education only. Non drinkers who have a primary degree are 2.8% less likely to report poor health status. Previous studies also find that education strongly contributes to better health (Grossman, 1972; Behrman & Wolfe, 1989; Berger & Leigh, 1989; Gilleskie & Harrison, 1998; Hartog & Oosterbeek, 1998; Kenkel, 1991, 1995; Leigh, 1998).

Similar to the findings of previous literature where older people report poorer health (Lin, 2008; Yen *et al*, 2010; Wilson *et al*, 2011), this study finds that all ages is significant in terms of the health status of non-drinkers. In particular non-drinkers in the age category 18-29 years are very likely to report excellent health status and are 4% less likely to report poor health status. Those aged 70 years or over are 4.2% more likely to report poorer health status.

Moderate drinkers aged 18-49 years are more likely to report a higher category of Health Status and are approximately 6% less likely to be report poor health status. For heavy drinkers the only age ranges that are significant in terms of health status are 30-49 years and heavy drinkers in this age group are likely to report a higher category of health status and are 9% less likely to report poor health.

The employment status variables are in general statistically significant in the determination of health status across all categories of drinkers. In particular non-drinkers who are students or are in state training schemes are likely to report excellent health and are 3.3% less likely to report poor health compared to those with a disability in the base category. All variables describing employment status are positively correlated with health status holding other variables constant for all categories of drinkers.

Log of income is a very significant variable in the health status equation for both non and moderate drinkers and is a positive value for both categories of drinkers. Particularly in relation to non drinkers they are likely to report a higher category of self assessed health the higher their income. These findings are generally consistent with previous findings which are that those with lower income also reported a lower self-reported health (Buckley *et al*, 2004; Bradley *et al*, 2000; Yen *et al*, 2010). These studies look at individual income as opposed to household income. Similarly Tremblay *et al* (2002) shows this is also the case in relation to household income.

Race is not significant in the determination of health status of moderate and heavy drinkers. For non drinkers the only race variable that is significant is that describing those of Asian race and for this variable the coefficient is negative showing that those of Asian race who are non-drinkers are 31% more likely to report a poor category of self assessed health. Previous studies have had very varied results in relation to the effect of race on health status. Many studies show that the black race tends to have poorer health when compared to other races (Thompson, 2011; Stuber *et al*, 2003). In contrast Habicht and Kunst (2005) find that ethnic differences were generally very small, with no consistently higher use by one group.

Gender is a significant determinant of health status of non and moderate drinkers. Male non-drinkers are more likely to report a higher category of health status while female moderate drinkers are more likely to report a higher category of health status. Male non drinkers are just over 2% less likely than females to report poor health status. Lin (2008), Kwan (2010), Liu (2008), Lahelma *et al (*1999), Lianga *et al (*2003) all find that males report better health than females.

In terms of marital status, non drinkers who are widowed and separated or divorced are likely to report a higher health status and are approximately 1.7% less likely to report poor health compared to those cohabiting which is in contrast to previous findings such as Wilson *et al* (2011). Being single/never married is not significant. For both moderate and heavy drinkers, marital status is not a significant determinant of health status. Numerous studies previously find that married individuals are healthier than single individuals (Rosengren, Wedel and Wilhelmsen, 1989; Zick and Smith, 1991).

Number of people living in the household is not significant for any category of drinker.

Where one lives can also affect ones health status. The health status of non-drinkers is affected by the variables describing those who live in a city either Dublin or any other city, with non-drinkers who live in a city being more likely to report a higher category of health status, and those in a city other than Dublin are likely to be higher than those in another city. In relation to moderate drinkers it is the variable describing those who live in a town that is a significant determinant of health status whereby those moderate drinkers who live in a town are likely to report a lower category of health status compared with those in the open country and are in fact 0.2% likely to report poor health status. Variables describing where respondents live are not significant for heavy drinkers. These findings show that the health status varies across the urban/rural divide for the different drinking categories. Previous studies also had very varied findings in relation to the health status of people depending on where they lived. Such studies have tended to look at the rural versus urban as opposed to break this down further into cities, towns, villages and open country. Wilson *et al* (2011) show that those living in rural areas are more likely to report fair/poor health than those living in urban areas whereas contrary to this Lin (2008) shows that people living in urban areas in Taiwan are more likely to report poorer health.

In terms of the Lifestyle variables, smoking is only significant in the health status equation for moderate drinkers. Moderate drinkers who smoke are likely to report being in the lower categories of health status i.e. they are 0.7% more likely to report poor health. Previous literature finds that smoking has a negative effect on health status (Mathers & Loncar, 2006; WHO, 2009; Samet, 2001; Yen *et al*, 2010; Manning *et al*, 1991; Holman *et al*, 1988; Ho *et al*, 2003; Jones, 2004, Jones, 2006)

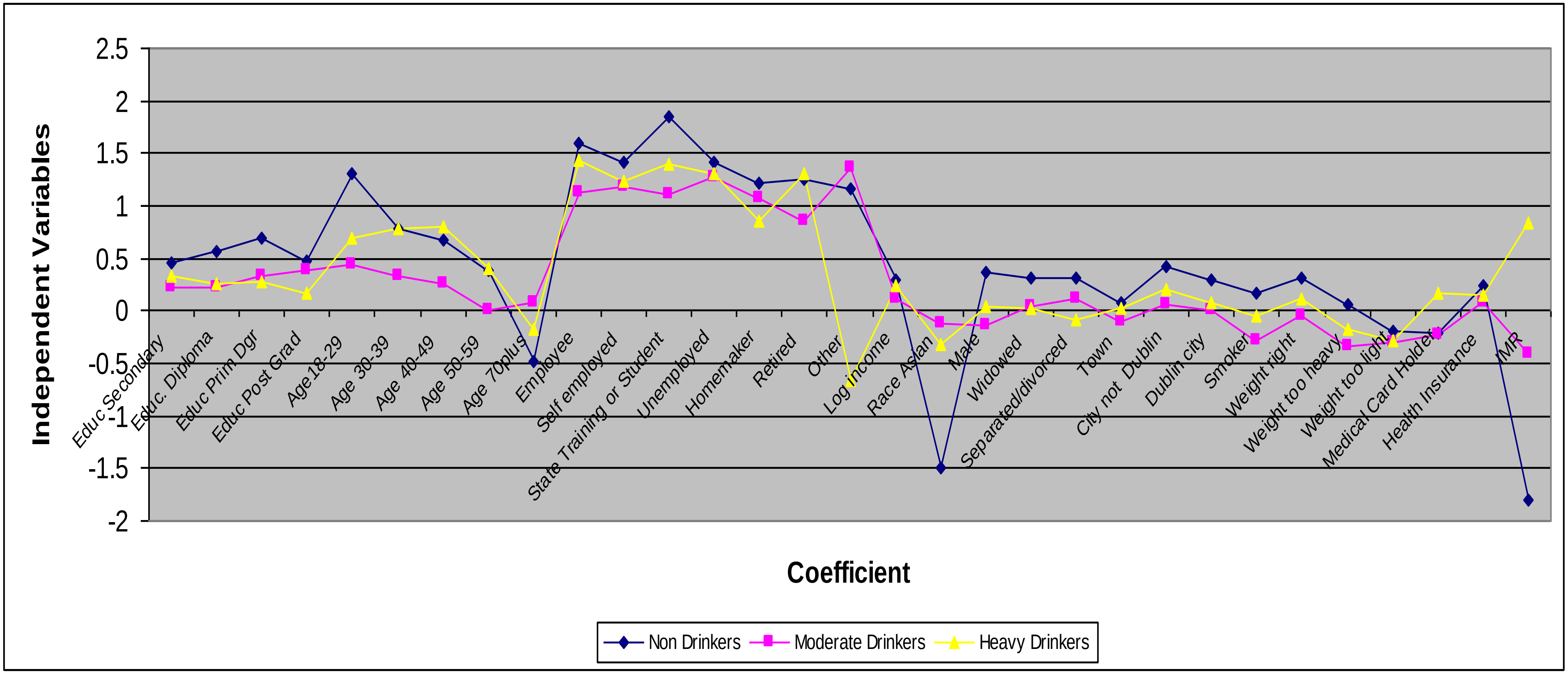
The variable describing whether respondents are physically active is dropped due to collinearity for all drinkers. Non drinkers who describe their weight as 'just right' tend to report higher categories of health status and are 2.3% less likely to report poor health. Moderate drinkers who describe their weight as 'too heavy' or 'too light' tend to report a lower category of health status and are approximately 1% more likely to report poor health. Both these variables are significant in the health status equation of moderate drinkers. Previous findings are similar in that they found that those who are overweight tend to have a poorer health status, particularly in relation to males (Lin, 2008). Contoyannis and Jones (2004) also find that those who are not obese have a higher reporting of excellent or good health.

Having a medical card is significant for non and moderate drinkers. Both non and moderate drinkers who have a medical card are likely to report having lower categories of health. Similarly health insurance is also significant for non and moderate drinkers. Moderate drinkers with health insurance are likely to report having poorer health while non drinkers are 1.7% less likely to report poor health. Harmon and Nolan (2001) and Hurd and McGarry (1997) finds that those in better health are more likely to be insured or at least there is no evidence for adverse selection. Hofter (2006) also finds that people with private health insurance tend to be healthier individuals.

The selection correction terms are significant for non-drinkers but not significant for moderate and heavy drinkers. For non-drinkers the coefficient is negative which indicates that individuals who self select into being a non-drinker, is 12.4% more likely to have poorer health status on average than what an individual with identical observable characteristics drawn at random would have as a non-drinker. Individuals who decide or have a preference to be a non-drinker also tend to be individuals with a poor health status.

The results of the ordered probit estimation of health status are also shown in figure 6.2.1 which graphs the coefficients for each variable for each of the three categories of drinkers. Variables that have been found to be insignificant across all drinker types have been dropped from this graph.

Figure 6.2.1: Coefficients of the individual variables from the Estimation of Health Status Equation



(Source: Authors own)

6.2.3: Overall Health Status by Drinker Type

In relation to all three categories of drinkers, the majority of respondents report good or very good health status. Table 6.2.3 shows the percentage breakdown of the self assessed health for each category of drinker.

Table 6.2.3 Results of Health Status by Drinker Type

	Non Drinkers
Poor Health status	6.05%
Excellent Health Status	16.45%
Fair Health Status	17.04%
Good Health Status	30.18%
Very Good Health Status	30.28%
	Moderate Drinkers
Poor Health status	1.89%
Fair Health Status	8.40%
Excellent Health Status	23.19%
Good Health Status	28.73%
Very Good Health Status	37.78%
	Heavy Drinkers
Poor Health status	1.71%
Fair Health Status	9.85%
Excellent Health Status	22.07%
Good Health Status	32.14%
Very Good Health Status	34.23%

(Source: Authors own)

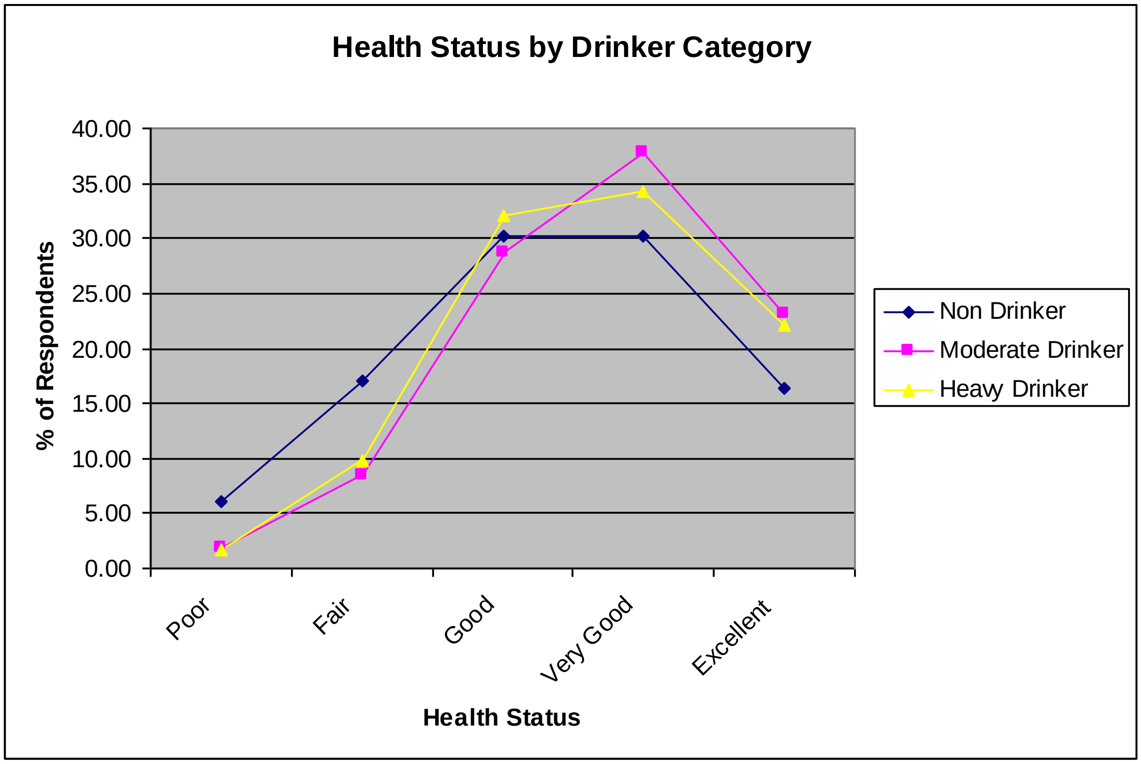
Overall the findings show that more non-drinkers report poor health than either moderate or heavy drinkers, which is in contrast to the findings of Kenkel (1995) who found heavy drinking to be a harmful input in the health production function.

The Self Assessed Health Status of Moderate and Heavy drinkers are very similar in this study. The majority of respondents in both categories of drinkers report good or very good health with the combined % of moderate drinkers who report these two categories of health status, being almost the same as heavy drinkers at approximately 66.4%, where as non-drinkers who report good and very good health is approximately 60.5%. In terms of the numbers who report excellent health, again there is very little difference between that of moderate and heavy drinkers, with marginally more moderate drinkers reporting excellent health, however in comparison far less non-drinkers (16.45%) report excellent health status.

In looking at excellent and very good health combined, this amounts to 46.73% for non drinkers, 61% for moderate drinkers and 56.3% for heavy drinkers showing that a greater proportion of moderate drinkers report very good or excellent health compared with non drinkers and heavy drinkers.

These results are also depicted in figure 6.2.2.

Figure 6.2.2: Health Status by Drinker Category



(Source: Authors own)

In general these findings show that the health status of non drinkers is worse than either moderate drinkers or heavy drinkers, and there is not much of a difference between the later two groups, except that slightly more moderate drinkers report very good/excellent health. These findings are similar to those of Contoyannis and Jones (2004) who find that prudent alcohol consumption is not significant in the determination of self assessed health status, and to the findings of other studies which show that moderate consumer's of alcohol tend to have better health (Berger *et al*, 1999; Klatsky *et al*, 2001; Bau *et al*, 2007).

6.2.4: Post Estimation Diagnostics

Testing the Specification of the Model

The significance of each of the variables is assessed using the z-statistic and results highlight which variables are significant at both the 1% and 5% levels of significance, and the Likelihood Ratio Test is also used to evaluate the significance of variables in the model and to ensure that each instrument is beneficial to the model. The likelihood ratio test statistic for the null hypothesis that coefficients on the Church Activity dummy variable are jointly equal to 0 gives a chi-squared of 15.02 with 1 degree of freedom and p value showing a 1% significance; hence the hypothesis is decisively rejected.

In relation to the health status equation, results show that each of the variables included in the model has resulted in a statistically significant improvement in the model fit. In relation to the variables describing Race, there was an LR chi-squared value of 9.1 with 3 degrees of freedom and a p-value showing a 5% significance. In relation to the dummy variables describing location where respondents live, there was an LR chi-squared value of 15.84 with 4 degrees of freedom and a p-value showing 1% significance.

The Wald Test shows that the models are statistically significant and rejects the null that coefficients of the variables are equal to zero.

Due to the lack of suitable instruments for the potentially endogenous lifestyle variables smoking, physical activity and weight; this study was unable to account for this possible endogeneity. Similar to Kenkel (1995) the alcohol status equation and the health status equations are estimated omitting these and here is no real difference in the self assessed health status of the different drinking categories compared to when the lifestyle variables were included. Results of health status omitting lifestyle variables are set out in Appendix G Table G.3.

Robust Standard Errors are used in the estimation of both alcohol status and health status equations to eliminate the potential problem of heteroskedasticity.

The variable physical activity was omitted from both the alcohol status and health status equations due to collinearity. The variable describing those respondents who are of black race was dropped due to collinearity in the health status equation for heavy drinkers.

The null hypothesis cut-off1 less cut-off2 = 0, is tested in both the alcohol status and health status ordered probit models. The null is rejected in all cases showing that the cut offs are not equal to each other and hence that categories should not be merged.

6.2.5: Effect of drinking on specific health conditions

The General Health Section of the Slán survey also contains questions in relation to whether or not people have suffered from specific illnesses in the 12 months prior to the survey. An analysis of specific illnesses is carried out to see if the findings in terms of general health status are similar in relation to specific illnesses.

In relation to heavy drinkers, in looking at the effect of drinking on specific illnesses in many cases the dataset was too small to provide accurate results, hence in this case drinkers consists of both moderate drinkers and heavy drinkers. The results of the effect of alcohol consumption on overall health status are similar for moderate and heavy drinkers and so the two are grouped together for the purpose of looking at specific illnesses by drinker type. Previous studies into the effects of alcohol on income have done something similar in that drinkers and non-drinkers are analysed (Berger and Leigh, 1988).

Similar to the manner set out by Greene and Hensher (2010) in Appendix G, initially drinking is estimated by probit regression. Results show that males, those in the age categories 18 to 29 years, those on state training schemes, those with higher household income, those of white race and those who live in cities are all more likely to be drinkers. Smokers and those who classify their weight as being too heavy are also likely to be drinkers. A single person or person who never married and those of black or Asian race are likely to be a non-drinker. Respondents who are regularly involved in Church activities are less likely to be drinkers.

The probit regression for drinkers allows an inverse mills ratio to be derived to allow for potential selection bias of drinking. A probit is then run for each specific illness for both drinkers and non-drinkers. This regression includes the inverse mills ratio derived from the drinking regression and this controls for potential selection bias.

Asthma

In looking at the effect of drinking on asthma, findings show that most of the variables that are significant in terms of health status are not in the asthma regression. Employment is significant for non-drinkers and results show that the unemployed, self employed and home makers are least likely to suffer from the condition. The inverse mills ratio is not significant for either drinkers or non-drinkers showing that a selection effect does not arise in relation to asthma.

Chronic Bronchitis

In relation to Chronic Bronchitis the age variable 18 to 29 years is significant for drinkers and ages 30-39 years are significant for both drinkers and non-drinkers. Respondents in these age groups are less likely to have suffered from chronic bronchitis.

Many of the employment status variables are significant. In particular both drinkers and non-drinkers who are employees are less likely to suffer from the condition. Drinkers who are on training schemes and the self employed are also less likely to suffer from the condition. Drinkers who are of a white race are not likely to suffer from chronic bronchitis. There is not a selection effect in terms of drinking status.

Heart Attack

Many variables were dropped due to collinearity in the regressions on the condition heart attack. In looking at the variables that are significant, findings show that drinkers who have a postgraduate level of education, those living in Dublin and smokers are more likely to have suffered a heart attack in the previous year.

The variable smoking is very significant. The Inverse Mills ratio is significant for drinkers. The negative coefficient shows that drinkers have unobservable characteristics associated with being less likely to suffer from a heart attack.

Angina

In relation to the condition angina, drinkers who live in cities tend to be more likely to suffer from the condition. The number of people living in the household is also significant in terms of angina in drinkers.

For non-drinkers those aged 40 to 49 years are less likely to suffer from the condition, while males compared with females and those living in cities and towns are more likely to report having suffered from angina. The selection effect is not significant for both drinkers and non-drinkers.

Stroke

In looking at those who suffered a stroke in the previous twelve months before the survey, in respect of drinkers, many of the age variables are significant showing that those up to aged 49 years are less likely to have reported having suffered from a stroke. Also drinkers who are retired or self employed are less likely to report having suffered from a stroke. Many variables were dropped due to collinearity.

Rheumatoid arthritis

Age is a significant variable in terms of this condition. In particular both drinkers and non-drinkers in the younger age groups are less likely to have suffered from the illness. The employment variables are significant for drinkers, and drinkers in all categories of employment are not very likely to suffer the condition.

The variable that describes those that live in a city other than Dublin is significant and shows that that those that those in this category are less likely to report having suffered from the rheumatoid arthritis. For non-drinkers ones weight is significant and those who are over weight are more likely to suffer from the condition. A selection effect does not arise.

Osteoarthritis

In looking at osteoarthritis the education variables are very significant for drinkers. In particular those who have received third level education are more likely to report osteoarthritis. Drinkers with higher income, who smoke and are overweight, are more likely, while drinkers who describe themselves as home makers and single or never married are less likely to suffer the condition. Those in the younger age categories up to 39 yrs for non-drinkers and 49 years for drinkers are not likely to be suffering the condition. Non-drinkers who classify their weight as just right or too heavy are also more likely to report suffering the condition.

A selection effect is present for drinkers. This is negative showing that individuals who drink are less likely to suffer from osteoarthritis on average compared with an individual with identical observable characteristics drawn at random.

Lower back pain

For both drinkers and non-drinkers weight is a very significant variable, whereby those who are too heavy or too light are more likely to suffer from lower backpain. Similarly those who are married or single/never married as well as non-drinkers who are widowed are not likely to suffer from the condition.

All the employment variables are significant for drinkers and are negatively correlated with lower back pain, and similarly for non-drinkers those who are unemployed or home makers are particularly unlikely to suffer from it. Drinkers in the age group 50-59 years and those with a diploma in education, are more likely to report having suffered from lower back pain along with smokers. All the race variables are significant and negatively correlated with lower back pain. Non drinkers who have received a postgraduate qualification are more likely to suffer from lower pack pain.

Diabetes

For all respondents, both drinkers and non-drinkers in the younger age categories up to 49 years are less likely to report suffering from diabetes. All the employment variables are significant and are negatively correlated with the condition. Those with higher household income are less likely to suffer from diabetes. Specifically for drinkers all the weight variables are significant and positively correlated with diabetes. For non-drinkers education is significant and those with a diploma or degree tend not to suffer from condition and similarly male non-drinkers are less likely to suffer from it than females as well as those living in Dublin.

There is a selection effect for non-drinkers which is positive showing that individuals who do not drink are more likely to suffer from diabetes compared with a non-drinker drawn at random.

Cancer

Both drinkers and non-drinkers under 59 years are less likely to suffer from cancer, however drinkers over 70 years of age are likely to do so. Employment is significant for all respondents, all having a negative correlation with the illness. Education is significant for drinkers except for the variable describing those who have a diploma. Respondents with an education tend not to suffer the condition, particularly those with a degree or postgraduate qualification. Drinkers who are male and those who reside in Dublin are less likely to have suffered from cancer in the previous year. For both drinkers and non-drinkers, findings show that smokers were less likely to have been diagnosed with cancer in the previous twelve months. There is no selection effect.

Urinary tract infection

For drinkers many of the education variables are significant showing in particular that those with a second level education and those with a primary degree are not likely to have the condition. All the age variables are significant with those up to the age 59yrs having a negative coefficient and those over 70 years a positive one, confirming that drinkers who are 70years or over are more likely to have suffered the urinary tract problems in the previous year. Employment status variables and household income are significant for drinkers all with negative coefficients. Female drinkers are more likely to suffer a urinary tract infection. Total number in the household is also significant showing that as the number in the household increases, drinkers are slightly more likely to have suffered this condition. A selection effect does not exist. For non-drinkers none of the variables are significant.

Anxiety

In relation to drinkers, the employment variables are significant, all with negative coefficients. In particular those classified as employees, unemployed and self employed have very high negative coefficients showing that drinkers in each of these categories are less likely to suffer from anxiety. Drinkers who are separated or divorced and those residing in Dublin are more likely to suffer from anxiety. Similarly smokers and those whose weight is either too heavy or too light are likely to suffer from the condition.

Nondrinkers who are males and those who have a third level education are more likely to suffer from anxiety. The variables showing age categories 40years upwards are significant for non-drinkers, with those in the age group 40 to 59 years more likely to have suffered anxiety in the previous 12 months. Those aged 70 years or over are less likely to have suffered the condition. Smoking is a very significant in terms of anxiety in non-drinkers, with results showing that smokers are more likely to suffer from the condition.

Respondents who are too heavy or too light are also more likely to suffer from anxiety. Inverse mills ratio is significant here with a negative coefficient showing that non-drinkers are less likely to suffer from anxiety compared with an individual with identical characteristics drawn at random.

Depression

Male drinkers and non-drinkers are all less likely to suffer from depression compared with females. Similarly employment variables are significant for both groups with the coefficients on the employment variables being negative.

Smoking is very significant for drinkers, and shows that drinkers who smoke are more likely to have suffered from depression. Education is very significant for non-drinkers. Those with second level education and those with a postgraduate qualification are significant, showing that respondents in these categories are less likely to have suffered depression in the previous year. Household income is also very significant for non-drinkers showing that higher income increases the likelihood of depression. Non-drinkers whose weight is described as being too light are more likely to have suffered depression. Mills ratio is significant for non-drinkers, with a positive coefficient which suggests that individuals who do not drink are more likely to suffer from depression on average compared with an individual with identical observable characteristics drawn at random.

Overall Percentage of Drinkers & Non-Drinkers who reported specific illnesses

In looking at the overall percentage of both drinkers and non drinkers who have suffered from the specific conditions in the twelve months prior to the survey, in all cases a higher percentage of non-drinkers report having had the conditions. These results are depicted in table 6.2.4

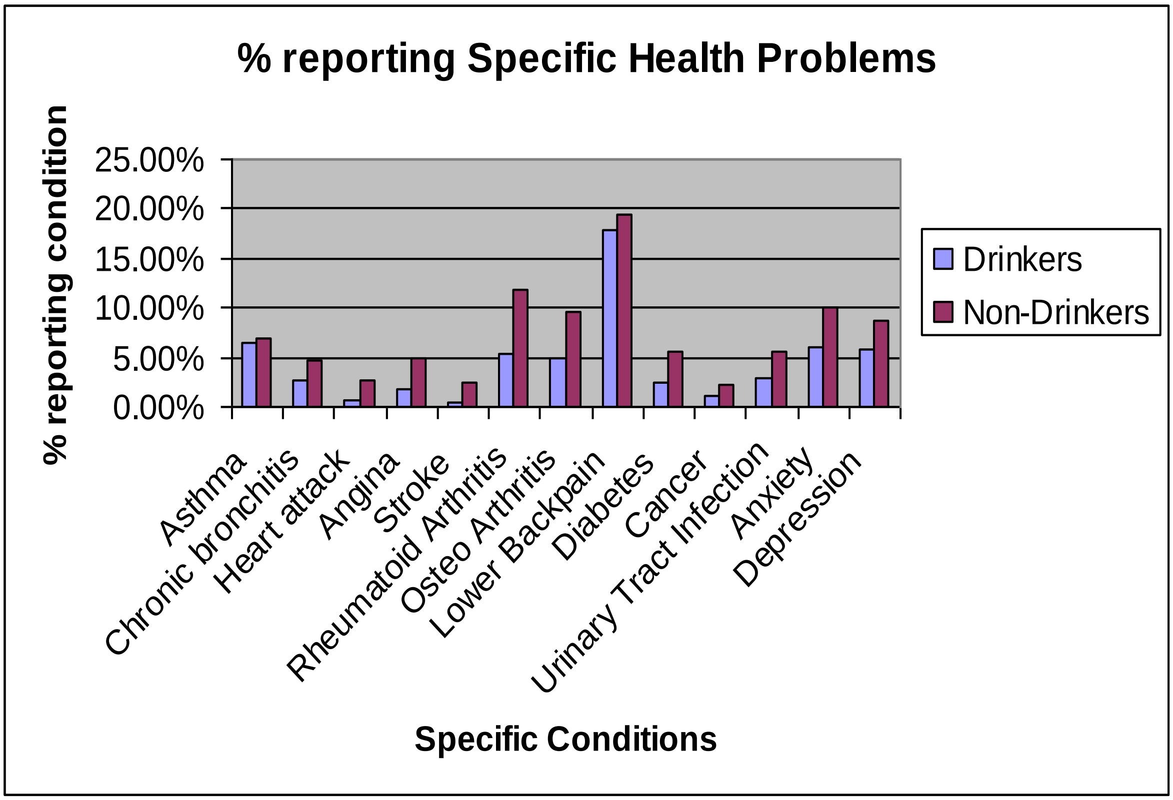
Table 6.2.4 Drinkers & Non-Drinkers who have reported suffering from specific health problems

	Observations	Drinkers	Non-Drinkers
Asthma	6094 Drinkers 2346 Non-drinkers	6.4 %	6.87 %
Chronic bronchitis	5916 Drinkers 2307 Non-drinkers	2.57 %	4.64 %
Heart attack	5531 Drinkers 1444 Non-drinkers	0.78 %	2.63 %
Angina	5835 Drinkers 1866 Non-drinkers	1.8 %	4.88 %
Stroke	5531 Drinkers 1690 Non-drinkers	0.54 %	2.49 %
Rheumatoid Arthritis	6094 Drinkers 2333 Non-drinkers	5.31 %	11.87 %
Osteo Arthritis	6094 Drinkers 2333 Non-drinkers	4.85 %	9.68 %
Lower Backpain	6147 Drinkers 2372 Non-drinkers	17.87 %	19.34 %
Diabetes	6120 Drinkers 2372 Non-drinkers	2.5 %	5.53 %
Cancer	5886 Drinkers 2057 Non-drinkers	1.11 %	2.29 %
Urinary Tract Infection	6121 Drinkers 2267 Non-drinkers	2.84 %	5.52 %
Anxiety	6147 Drinkers 2333 Non-drinkers	6.10 %	10.02 %
Depression	6147 Drinkers 2372 Non-drinkers	5.81 %	8.77 %

(Source Authors own)

These results are also depicted graph format in figure 6.2.3.

Figure 6.2.3: % of Drinkers and Non-Drinkers who have reported suffering from specific health problems



(Source Authors own)

In the Slán questionnaire respondents were provided with a list of thirteen medical conditions and were asked if they had suffered from any of the conditions in the twelve months prior to the survey. In all cases a higher percentage of non-drinkers suffered from each of these conditions compared with drinkers. This is similar to the findings into the effect of alcohol consumption on health status whereby less non-drinkers report having good or excellent health status compared with either moderate or heavy drinkers.

6.3: Results of the Estimation of the Effect of Alcohol on Health Care Utilisation - Consultations with the GP

Using the Slán 2007 dataset, Health Care Utilisation is estimated by an ordered probit accounting for the potential selection bias of drinking status. Health Care Utilisation is measured by looking at the number of times a person consulted with the General Practitioner. Respondents are classified into one of five categories of self assessed health; Never attended a GP = 1, attended a GP more than 2 years ago = 2, attended a GP between 1 & 2 years ago = 3, attended a GP between 1 & 12 months ago = 4 and attended a GP within the last 4 weeks = 5. Potential Selection bias is accounted for by including the selection correction terms derived from the alcohol status estimation. Results from the estimation of both alcohol status and health care utilisation regressions are shown in Appendix I.

Looking at the results of the health care utilisation in tables I3 and I4 in Appendix I, health status is a very significant variable in looking at health care utilisation which is the same as the findings of Gruber and Kiesel (2010). A non or moderate drinker who reports excellent health is approximately 5% more likely and a heavy drinker 7% more likely to report never having consulted a GP. A person with excellent, very good or good health is likely to be in a lower category in terms of attending the GP (Dunlop *et al*, 2000).

The education variables are not significant for either drinkers or non drinkers in terms of their use of GP services. This is in contrast to the previous findings whereby those with a diploma or primary degree are more likely to access GP services (Dunlop *et al*, 2000).

Moderate Drinkers in the lower age brackets up to 49 years are less likely to access GP services regularly, which is similar to the findings of Nolan (2007) and Jatrana and Crampton (2009) who find that GP visiting is an increasing function of age. Age is not significant for non and heavy drinkers. All the variables describing ones employment status are very significant for non and moderate drinkers and in general respondents in all of the different categories are not likely to report high GP utilisation.

Unlike previous studies such as Dunlop *et al* (2000) and Habicht and Kunst (2005) who show that as a person's income increases a person is more likely to visit a general practitioner more frequently and Hernandez-Quevedo and Rubio (2009) who find the opposite to be the case, this study finds that income is not significant for any category of drinking in terms of health care utilisation.

The variable describing those of Asian Race is significant for moderate drinkers. Moderate drinkers of Asian Race are less likely to visit the GP frequently and are almost 2.6% more likely to report never having gone to a GP. Males who are moderate drinkers are less likely to be in a higher category of GP utilisation compared with female moderate drinkers. These findings are similar to other studies such as Dunlop *et al* (2000) and Jatrana and Crampton (2009). Non- Drinkers who are married are more likely to access GP services more often.

The total number in household is significant for non drinkers, showing that as the number in the household increases a person is less likely to visit the General Practitioner as often.

Being a medical card holder is very significant for all categories of drinkers. A medical card holder is more likely to visit the GP more regularly than a person who does not have a medical card which is very similar to previous findings (Nolan, 2007). Similarly the variable describing those with private health insurance is significant across all drinker types and a person with private health insurance is likely to visit the GP regularly. Nolan and Nolan (2003) find the same in that having private medical insurance significantly increases the probability of visiting a GP.

The selection correction term is not significant for any category of drinker.

6.3.1: Overall Health Care Utilisation by Drinker Type

Table 6.3.1 sets out the results showing the level of GP consultation by the three drinker types.

Table 6.3.1 Results showing the level of GP Consultations by Drinker Type

Last time consulted GP	Non-Drinker	Moderate Drinker	Heavy Drinker
Never	1%	1%	1%
more than 2 years ago	8%	10%	15%
1-2 years ago	8%	13%	15%
between 1 and 12 months	45%	51%	47%
in last 4 weeks	38%	25%	23%

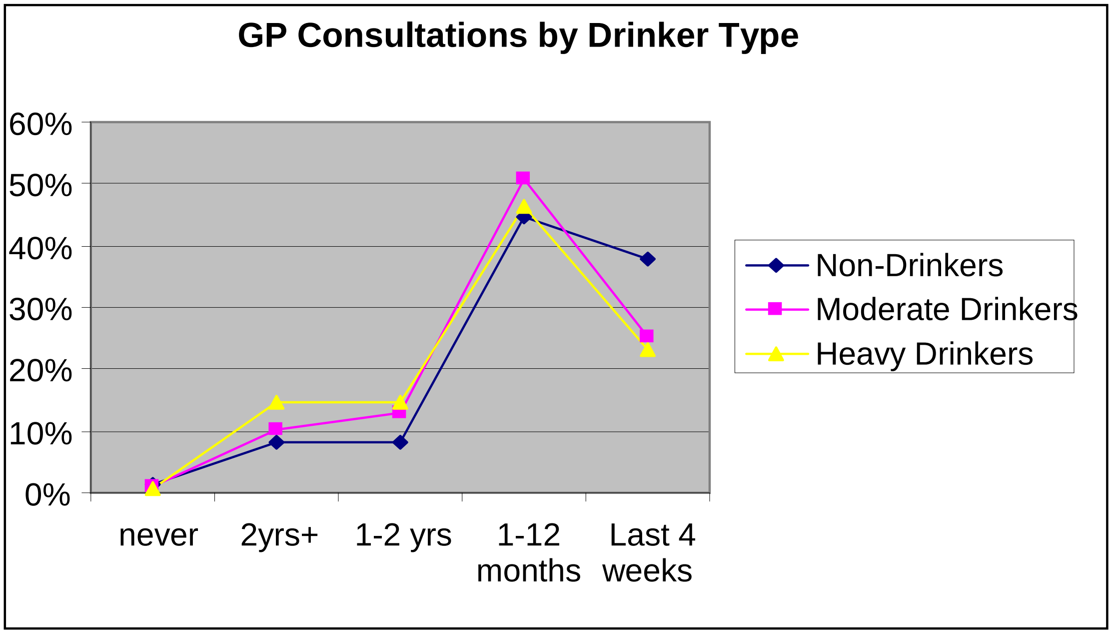
(Source Authors own)

Results show that more non-drinkers consulted the GP in the 4 weeks prior to the survey than either moderate or heavy drinkers. Both moderate and heavy drinkers utilised the GP approximately the same amount in the 4 weeks.

In the year prior to the survey, non drinkers utilised GP services the most, with moderate drinkers utilising services more than heavy drinkers. In looking at those who visited the GP one year ago or more, heavy drinkers have the highest % of visits. Dunlop *et al* (2000) shows that males who do not drink are more likely to have had 6 or more GP visits in the pervious 12 months whereas a female moderate drinkers is least likely to have attended a GP 6 times or more in the previous year when compared with either non or heavy drinkers.

Figure 6.3.1: Results showing the level of GP Consultations by Drinker Type

% of consultations



Last time consulted a GP

(Source Authors own)

Generally usage of GP Services is very similar across drinker types however more non drinkers attended the GP in the 4 weeks prior to the survey compared with both moderate and heavy drinkers, with 38% of non-drinkers attending, 25% of moderate drinkers and 23% of heavy drinkers attending. This is similar to the findings of Dunlop *et al* (2000), who find that male non-drinkers attend the GP more than drinkers, and female non-drinkers attend the GP more than those who drink moderately but not more than those who are heavy drinkers. These findings are in line with the findings on health status in section 6.2 whereby more non drinkers tend to report poorer health.

6.3.2: Post Estimation Diagnostics

The significance of each of the variables is assessed using the z statistic and results highlight which variables are significant at both the 1% and 5% levels of significance. The Likelihood Ratio test is also used to evaluate the significance of each variable in the model and ensure that each variable is beneficial to the model and results shows that coefficients of the variables are equal to zero.

Robust Standard Errors are used both in the alcohol consumption and the health care utilisation equations deal with potential heteroskedasticity.

The variable physical activity is omitted from both the alcohol status and health care utilisation equations due to collinearity. The variable describing those respondents who are from a Black Race was dropped due to collinearity in the health status equation for heavy drinkers.

Due to the lack of suitable instruments for the potentially endogenous lifestyle variables; smoking, physical activity and weight, this study is unable to account for this possible endogeneity. When these variables are omitted from the alcohol status equation and the health care utilisation equations, there is no real difference in the health care utilisation of the different drinking categories compared to when the lifestyle variables were included.

The cut offs in the health care utilisation ordered probit model are tested for being equal to each other. The null hypothesis that cut-off1 less cut-off2 = 0 is rejected showing that the cut offs are not equal to each other hence none of the individual categories of either drinkers or categories in terms of health care usage could be merged with each other.

6.4: Conclusion

This chapter presents an empirical study of the effect of alcohol consumption on individual health status and health care utilisation in Ireland while accounting for the potential endogenous relationship between alcohol and health. Drinkers are categorised into three categories non, moderate and heavy drinkers. The drinking status equation is estimated using an ordered probit model, from which the predicted values for the inverse mills ratio is generated which is then included in the health status and health care utilisation equations. This accounts for the possible selection bias of alcohol.

Overall the findings show that more non-drinkers report poor health than either moderate or heavy drinkers. There is a very small difference between the health status of moderate and heavy drinkers, however slightly more moderate drinkers do report very good/excellent health. Previous studies have found a U or J Shaped curve which depicts that moderate consumers of alcohol tend to have better health when compared with abstainers or heavy drinkers (Berger *et al*, 1999; Klatsky *et al*, 2001; Bau *et al*, 2007). The findings of this study are similar in that moderate drinkers report having the best health status however a substantial drop in the health status of heavy drinkers is not evident.

In looking at the overall percentage of both drinkers and non drinkers who have suffered from the specific conditions in the twelve months prior to the survey, in all cases a higher % of non-drinkers reported having had the conditions.

In relation to health care utilisation, results show that more non-drinkers consulted the GP in the 4 weeks prior to the survey than either moderate or heavy drinkers. Both moderate and heavy drinkers utilised the GP approximately the same amount in the 4 weeks. In the year prior to the survey non drinkers utilised the GP services slightly more than both moderate and heavy drinkers.

Implications of these findings are that moderate drinkers report having the best health status similar to previous findings (Berger *et al*, 1999; Klatsky *et al* 2001; Bau *et al*, 2007). The difference compared with heavy drinkers is small and the difference between moderate drinkers and non drinkers is more substantial. This again highlights some of the positive effects of moderate levels of alcohol consumption, and reiterates the benefit of considering target based policies in order to combat the problem of the misuse of alcohol.

CHAPTER 7
CONCLUSION

This chapter sets out the aim of this thesis and how the study is presented overall. Findings of this research using Irish data from the 2007 Slán survey are summarised and policy implications arising are addressed. This thesis is primarily a technical-econometric study. The results however, relate to a core health policy issue, which is subject to widespread public debate. Various stakeholders and the media will, therefore, be interested in the implications of the findings for the implementation of polices around the misuse of alcohol consumption.

7.1 Chapter Summary

The main aim of this thesis is to examine if differences in income exist for different categories of drinkers in particular non, moderate and heavy drinkers. The impact of alcohol consumption on health and health care utilisation is also examined. These two questions will examine if the correlation between income and alcohol consumption is similar in terms of sign and magnitude to the correlation between health status and alcohol consumption. Barrett (2002) among others has identified a correlation between these two sets of relationships for other countries.

Throughout this study potential endogeneity and selection bias of alcohol consumption is considered and accounted for, which would otherwise lead to biased estimates. Potential endogeneity of alcohol status is accounted for by running separate income regressions along with separate health status and health care utilisation regressions by drinker type similar to the methods adopted in previous studies such as Hamilton and Hamilton (1997) and Barrett (2002). The issue of selection bias is addressed by using various extensions of the Heckman Probit Two Step Estimation.

As part of this analysis the relationship between other socio demographic and personal characteristic variables with alcohol consumption, income, health status and health care utilisation is also examined.

This study provides details of the effects of heavy drinking on income, health and health care utilisation in Ireland. It also provides details of the specific relationship between many personal and socio demographic characteristics on alcohol consumption. This study will encourage the use of a target based approach as opposed to a population based approach in the efforts to reduce alcohol consumption. This would lead to more specific policy formation, targeting particular segments of the population rather than the population as a whole.

Chapter 2 reviews previous studies into the effect of alcohol on income and studies into the effect of alcohol on health and health care utilisation. This chapter identifies how individuals can be categorised into different categories of drinkers. The issue of endogeneity is examined and possible ways to account for potential endogeneity is also looked at. Selection bias is assessed and the possible selection bias that may arise in terms of alcohol consumption. Econometric techniques that account for such bias are analysed. The different factors that affect alcohol consumption, income, health status and health care utilisation are reviewed. This chapter also looks at interpreting alcohol consumption as ordinal data and how estimation can be carried out while accounting for endogeneity and selection bias. Limited Information Methods are compared with Full Information Methods of estimation and a review of previous studies comparing the two is carried out.

Chapter 3 describes the Slán National Health and Lifestyle survey which is used in this study. Each of the variables used in the study are described as set out in the Slán survey. A detailed description of the dependent variables income, drinking status, health status and health care utilisation is given setting out the questions in the Slán survey which provide the data along with the number of respondents and details as to the responses given. Detailed descriptions are also provided in relation to the independent variables. Standard Deviations and the mean values are set out for both the dependent and independent variables along with the minimum and maximum value in relation to each variable.

Chapter 4 presents a study on the effect of alcohol on household income in Ireland. Similar studies that were previously carried out in relation to other countries are reviewed. The issues arising in such an estimation, primarily the endogeneity and selection bias of alcohol consumption, are assessed along with possible methods that could be used to deal with such difficulties. The Lee Multinomial Logit Ordinary Least Squares (OLS) Two Step Estimate is used which involves the estimation of the alcohol status equation in step one from which the inverse mills ratio is derived which is then included as an additional regressor in the income equation in step two. This estimation allows the relationship between household income and alcohol status with different personal and socio economic variables to be examined. Results show that while heavy drinkers have a higher income than moderate drinkers the difference is very small. Income of non drinkers is substantially less than both moderate drinkers and heavy drinkers.

Chapter 5 considers the ordinal nature of alcohol consumption and the potential implications of not accounting for this. Methods of estimating the effect of alcohol on income treating alcohol as an ordinal variable while still accounting for endogeneity and selection bias are reviewed.

Limited Information Methods of Estimation and Full Information Methods of Estimation are also reviewed with both methods used in the estimation of the effect of alcohol on income. Results from the two step method show that non drinkers have the lowest income while with the full information method heavy drinkers have the lowest income. The commonality between the results from the two methods is that moderate drinkers have the highest weekly household income.

Chapter 6 presents an empirical study of the effects of alcohol consumption on health status and health care utilisation while accounting for the potential endogeneity and selection bias of alcohol. Alcohol consumption is estimated as an ordered probit in the first step of the two step procedure, which allows the inverse mills ratio to be estimated. The health status and health care utilisation equations are then estimated as an ordered probit including the inverse mills ratio as an additional regressor. Differences in health status and health care utilisation for each of these categories of drinkers is examined and the relationship between both alcohol status, health status and health care utilisation with a host of other personal and socio-economic variables such as age, gender, marital status, employment status and level of education, among others, is also assessed. The relationship between specific illnesses and alcohol status is also examined.

7.2 Findings Overall

This thesis identifies that moderate drinkers have the highest income in terms of household income.

Estimating the effects of alcohol consumption on income whether considering alcohol as an ordered or unordered variable shows that income of non drinkers is less than moderate drinkers which is similar to findings of previous studies (French and Zarkin, 1995; French and Zarkin, 1998; Hamilton and Hamilton, 1997; Barrett, 2002). Estimating the alcohol status equation as a multinomial logit shows heavy drinkers have the highest income however the difference between the income of moderate and heavy drinkers is very little. This is in contrast to many previous studies such as Hamilton and Hamilton (1997) and Barrett (2002) where it was found that moderate drinkers earn the highest amount however French and Zarkin (1998) and Bastida (2006) find no evidence of a drop in earnings associated with heavy drinking.

Taking account of the ordered nature of the alcohol status variable and estimating alcohol status as an ordered probit, income of moderate drinkers is higher than heavy drinkers. Using the Full Information Maximum Likelihood Method of estimation and accounting for the ordered nature of alcohol consumption, moderate drinkers have a higher income than heavy drinkers, however the difference between income of moderate drinkers and heavy drinkers is much greater when using the FIML method, with the income of heavy drinkers being far less than moderate drinkers and substantially less than non-drinkers. This is similar to the findings of Barrett (2002) who also found that there was a significant earnings penalty for heavy drinkers relative to abstainers. Table 7.2.1 depicts the findings in terms of the weekly household income by category of drinker for each of the different methods of estimation.

Table 7.2.1: Weekly household income by drinking type

	Multinomial Logit Two Step Estimation	**Ordered Probit Two Step Estimation**	**FIML estimation treating alcohol status as ordered**
Non Drinkers	€477.41	€535.95	€546.75
Moderate Drinkers	€683.36	€725.45	€660.10
Heavy Drinkers	€694.18	€694.18	€449.99

(Source: Authors own)

Overall it appears that treating alcohol as an unordered variable, moderate and heavy drinkers have higher household income than non-drinkers. However previous research into the effect of alcohol consumption on income (Hamilton and Hamilton, 1997; Barrett, 2002) did not account for the fact that alcohol consumption can be viewed as being ordered data and not accounting for this may lead to a loss of efficiency and a greater risk of insignificant results (Harris *et al*, 2006). This is a clear limitation of previous research. This study estimates the effect of alcohol consumption on income treating alcohol as ordered data using both the Limited Information Methods and Full Information Methods of Estimation. Generally findings show that Full Information Methods of estimation are more favourable techniques in the estimation of simultaneous equations (Puhani, 2000; Intriligator *et al*, 1996; Enders and Bandalos, 2001). While results differ between the FIML method and the two-step method; both methods find that moderate drinkers are the best off in terms of income. In looking at the results of the Full Information Methods

In terms of the impact of alcohol consumption on health status, this thesis finds that more non drinkers report poor health than either moderate or heavy drinkers. Findings in terms of the health status of moderate and heavy drinkers are very similar, with majority of respondents in both categories reporting good or very good health. In looking at the two highest categories excellent and very good health combined, more moderate drinkers report being in this category than heavy drinkers.

While many previous studies have had similar findings in that moderate drinkers have better health status compared with non drinkers, in contrast to this study they also found that heavy drinkers have poorer health status compared with moderate drinkers resulting in a J or U shaped curve showing a reduced relative risk of given diseases and general better health for moderate drinkers (Berger *et al*, 1999; Klatsky *et al* 2001; Bau *et al*, 2007). This study does not find that a J or U shaped curve exists in Ireland given that heavy drinkers do not suffer a fall in health status compared with moderate drinkers.

Evidence from looking at health care utilisation shows that compared with moderate and heavy drinkers, more non-drinkers consulted the GP both in the 4 weeks and in the year prior to the survey. This would correspond to the fact that more non drinkers report having poor health. This was similar to the findings of Dunlop *et al* (2000) who found that males who do not drink were most likely to attend the GP 6 times or more when compared with male drinkers who have between 1-11 drinks per week and those who have 12 or more drinks per week. In relation to females Dunlop *et al* (2000) found that female non-drinkers were more likely to have attended the GP when compared with those who drink between 1-11 drinks per week, however a female who has 12 drinks or more per week is more likely to have attended a GP 6 times or more in the previous year when compared with either non or moderate drinkers. Both moderate and heavy drinkers utilised the GP approximately the same amount. Table 7.2.2 summaries the findings in terms of health status and health care utilisation of each category of drinker.

Table 7.2.2: Health Status and Health Care Utilisation by drinker type

	Health Status **% Reporting excellent or very good health**	**Health Care Utilisation** **% who consulted GP in 4 weeks prior to the survey**	**Health Care Utilisation** **% who consulted GP in the year prior to the survey**
Non Drinkers	46.6%	38%	83%
Moderate Drinkers	61%	25%	76%
Heavy Drinkers	56.5%	23%	70%

(Source: Authors own)

7.3 Policy Implications

Confidence Intervals at 95% are constructed from the estimation of alcohol on income using the FIML method of estimation. As opposed to just looking at the mean income value, the confidence interval provides a range of values which is likely to contain the populations' income. This thesis can be 95% confident that the true estimate income of non-drinkers lies between €539.15 and €550.04 per week; the true estimate income of moderate drinkers lies between €651.97 and €665.14 per week; and the true estimate income of heavy drinkers lies between €441.42 and €454.86 per week. These are set out in table 7.3.1 below.

Table 7.3.1: Confidence Intervals at 95% showing the true estimate of income for each category of drinker

Income for categories of drinkers:	Average Income	95% Confidence Interval for log income		95% Confidence Interval for weekly household income	
Non-drinker	546.75	6.29	6.31	539.15	550.04
Moderate drinker	660.10	6.48	6.50	651.97	665.14
Heavy drinker	449.99	6.09	6.12	441.42	454.86

(Source: Authors own)

This thesis suggests that non drinkers have a lower weekly household income and lower health status. Moderate drinkers appear to be better off than heavy drinkers in terms of income and slightly better off than heavy drinkers in terms of health status.

There are very varied ideas around target versus population based approaches to dealing with the problem of misuse of alcohol consumption. The WHO (2007) state that there is a large body of evidence showing that not only do alcohol policies and interventions targeted at vulnerable populations prevent alcohol related harm but that policies targeted at the population at large can also have a protective effect on the population as a whole. The WHO (2007) state that while there are advantages from both approaches in terms of reducing alcohol related harm, in some countries support for population approaches has declined in favour of targeted interventions. McCambridge *et al* (2013) suggest however that the alcohol industry actors do not have identical commercial interests and policy preferences and while they favour targeted interventions that focus on a problematic minority of drinkers and emphasising the role of individual responsibility, policy making is not always rational and purely informed by evidence. For this reason McCambridge *et al* (2013) state that policy is subject to a wide range of influences and this complexity warrants dedicated investigations.

Adams and White (2005) in their study of the impact of the population approach to prevention, study situations whereby such an approach may put the health of individuals at risk. They state that the population approach means that risk is reduced for all members of the population, irrespective of their baseline risk, in order to try and maximise the benefit of preventative interventions to public health. Adams and White (2005) argue that this makes the unrealistic simplistic assumption that monotonic relationships exist between specific risk factor exposure and associated risk of morbidity and mortality. Adams and White (2005) have found that in particular where there is a J Shaped relationship as between exposure and risk that population based strategies are not always beneficial. They state that while with a population approach only a small number of people are likely to be negatively affected the negative effect on such individuals should not be ignored. They highlight that there are ethical issues associated with such an approach especially where clearly identifiable groups of individuals can be predicted to be harmed, rather than helped by an intervention and while the population interventions may outweigh any negative effect seen, the effect on individuals concerned should not be overlooked and that more discussion is required on how these individuals should be protected from population interventions. Adams and White (2005) argue that such an approach which can harm some of the population does not meet the requirements of the principal of the non-maleficence of the Hippocratic Oath ‘first do no harm’.

At a Symposium on Moderate Alcohol Consumption; Health Risks and Benefits in Cambridge, Massachusetts in 2006, Professor Smallwood highlighted that many health professionals do not accept that there are health benefits to moderate drinking and simply see minimising harm at population level as the all important issue despite there being clear evidence that there may be benefits to moderate levels of alcohol consumption. Professor Smallwood argues that no society has yet solved the riddle of how to achieve an agreed balance whereby the social and health benefits of moderate drinking can be enjoyed by the majority, while harm caused by and affecting the few is minimised.

There is a substantial body of evidence to show that there are benefits to moderate levels of alcohol consumption. Previous research has been found to show that moderate consumers of alcohol enjoy higher incomes compared with non or heavy consumers of alcohol (French and Zarkin, 1995; Zarkin *et al*, 1998; Hamilton and Hamilton, 1997; Barrett, 2002). Similarly previous research shows that moderate consumers of alcohol enjoy better health compared with non or heavy drinkers (Berger *et al*, 1999; Klatsky *et al* 2001; Bau *et al*, 2007). By adopting a population based policy approach to reducing alcohol consumption, while this may benefit the majority of individuals in society there may be a small number of individuals who are moderate consumers of alcohol, who will be at harm or disadvantaged from such an approach and hence an ethical issue arises from such policy measures (Adams and White, 2005).

In the recommendations from the Steering Group on National Substance Misuse Strategy in February 2012 (Ireland 2012), no reference is made to the potential benefits of moderate levels of alcohol consumption; the majority of recommendations are around the supply side of alcohol and are population based, examples being of further taxation on alcohol and introducing a social responsibility levy among others. This study provides a greater insight into alcohol consumption in Ireland and findings show that there are benefits to moderate levels of alcohol consumption in Ireland, none of which have been considered by the Steering Group in their recommendations around policy, which has been the argument by the Alcohol Beverage Federation of Ireland (ABFI, 2012). Clearly the adoption of many of the policy recommendations as set out by the Steering Group (Ireland, 2012) will have a negative impact on some individuals who are currently moderate consumers of alcohol and such policy approaches may result in them reducing their levels of consumption further which may cause them harm.

It is recommended that the approach to policy around the misuse of alcohol consumption is looked at again in the context of the tailoring of policies to particular groups of individuals rather than providing a one size fits all approach. The at risk individuals should be targeted which would ensure that when people drink they do so in as safe a manner as possible and selective enforcement of policies around the misuse of alcohol should be looked at to ensure that all individuals are protected from harm.

There are several directions in which it may be fruitful to extend this research. This analysis focuses on the effect of alcohol consumption on household income, health status and health care utilisation. It would also be interesting to look at the area of substance misuse as opposed to alcohol consumption on its own, and include for example smoking and illegal use of drugs, and estimate the effect of such variables simultaneously on both income and health. This would be particularly interesting in Ireland given that alcohol policy is included as part of the National Substance Misuse Strategy since 2009 (Ireland, 2012).

Another area of future research would be to look at policy approaches in Ireland and other countries in relation to other lifestyle variables such as smoking and obesity, and whether target based approaches have been adopted and an analysis of the success of these approaches carried out.

The Slán survey used in this study is a cross sectional study. Carlson and Morrison (2009) describe a cross sectional study as an observational study in which exposure and outcome are determined simultaneously for each subject. They argue that cross sectional designs require shorter time commitment and fewer resources to conduct, but can have limitations. One limitation of cross sectional designs is, given that exposure and outcome are simultaneously assessed, there is generally no evidence that exposure caused the outcome and causality can be unclear. Another issue with cross sectional designs outlined by Carlson and Morrison (2009) is that cross sectional studies evaluate prevalent rather than incident outcomes and thus excludes people who develop the outcome but die before the study. The measured association is between exposure and having the outcome as opposed to exposure and developing the outcome. A third limitation identified by Carlson and Morrison (2009) is that the reader needs to assess if alternative explanations for study results have been appropriately ruled out. Given the limitations of cross sectional studies, an interesting area for further analysis would be to carry out similar research but for each of the years of the Slán survey, using the same variables which would allow a comparison of the cross sections to be made hence giving some degree of a time element.

Bibliography

Academy of Medical Science (2004). *Calling Time; the nations drinking as a major health issue.* London: Academy of Medical Sciences.

Adams, J. and White, M. (2005). When the population approach to prevention puts the health of individuals at risk. *International Journal of Epidemiology* 34(1): 40-43.

Adepoju, A. (2009). Comparative assessment of simultaneous equation techniques to correlated random deviates. *European Journal of Scientific Research* 28(2): 253-265.

Adepoju, A. and Olaomi, J. (2009). Ranking of simultaneous equation techniques to small sample properties and correlated random deviates. *Journal of Mathematics and Statistics* 5(4): 260-266.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: Wiley and Sons Inc.

Agunbiade, D. and Iyaniwura, J. (2010). Estimation under Multicollinearity: A Comparative Approach Using Monte Carlo Methods. *Journal of Mathematics and Statistics* 6(2): 183-192.

Ahituv, A and Lerman, R (2007). How Do Marital Status, Work Effort and Wage Rate Interact? *Demography* 44(3): 623-647.

Alcohol Beverage Federation of Ireland (2012). *The National Substance Misuse Strategy Minority report by the Alcohol Beverage Federation of Ireland*. Dublin: ABFI, 2012.

UK (2008). Alcohol: Tax, Price and Public Health, Institute of alcohol studies. [Online]. (URL: www.alcoholpolict.net) (Accessed 3rd December 2012).

American College of Sports Medicine (2000). *Guidelines for Exercise Testing and Prescription.* 6th Edition. Baltimore: Lippincott Williams & Wilkins; *2000.*

American Diabetes Association. [Online] (URL: www.diabetes.org) (Accessed December 2012).

An Garda Síochána (2006). *Annual Report of An Garda Síochána 2006*. Dublin: An Garda Síochána.

An Garda Síochána (2009) *Annual Report of An Garda Síochána 2009*. Dublin: An Garda Síochána.

Anderson, P. and Baumberg, B. (2006). *Alcohol in Europe. A public health perspective*. London, Institute of Alcohol Studies.

Arts, Sport, Tourism, Community, Rural and Gaelteacht Affairs (2007). *The Relationship between alcohol misuse and the drinks industry sponsorship of sporting activities.* A report to the joint committee.

Auld, C. (2005) Smoking, Drinking and Income. *Journal of Human Resources* (2):505-518.

Badley, E., Wang, P., Cott, C. and Gignac, M. (2000). *Determinants of changes in self-reported health and outcomes associated with those changes.* Arthritis Community Research and Evaluation Unit Working Paper 00-05.

Bali, T.(2003) The Generalised Extreme Value Distribution. *Economics Letters* 79(3): 423-427.

Balsa, A. and French, M. (2010). Alcohol Use and the Labor Market in Uruguay, *Health Economics* 19: 833-854.

Barr, N. (2004). *Problems and definition of measurement: In Economics of the welfare state.* New York: Oxford University Press.

Barreto, H. and Howland, F, (2006). *Introductory Econometrics: using Monte Carlo simulation with Microsoft Excel.* Cambridge University press.

Barrett, G. (2002). The Effect of Alcohol Consumption on Earnings. *Economic Record* 78(1): 79-96.

Barrow, R. and Sala-i-Martin, X. (1999). *Economic Growth.* The MIT Press, Cambridge.

Bassuk, S., Berkman, L. and Amick, B. (2002). Socioeconomic status and mortality among the elderly: Findings from four US communities. *American Journal of Epidemiology* 155(6): 520–533.

Bastida, E. and Soydemir, G. (2006). Alcohol use and Earnings, Findings from a community based study. *Eastern Economic Journal* 32(4):617-628.

Battel-Kirk, B. and Purdy, J. (2007). *Health Inequalities on the island of Ireland; the facts, the causes, the remedies.* Public Health Alliance Ireland for the island of Ireland, Belfast and Dublin.

Bau, P., Bau, C., Rosito, G., Manfroi, W. and Fuchs, F. (2007) Alcohol Consumption, Cardiovascular Health and Endothelial Function Markets. *Alcohol* 41(7).

Baum, C., Schaffer, M. and Sillman, S. (2003). Instrumental Variables and GMM: Estimation and testing. *Stata Journal* 3: 1-31.

Becker, U., Deis, A., Sorensen, T., Gronbaek, M., Borch-Johnsen K., Muller, C., Schnohr, P. and Jensen, G. (1996). Prediction of risk of liver disease by alcohol intake, sex, and age: a prospective population study. *Hepatology* 23(5): 1025-9.

Behrman, J. and Wolfe, B. (1989). Does more schooling make women better nourished and healthier? Adult sibling random and fixed effects estimates for Nicaragua. *Journal of Human Resources* 24: 644–663.

Berger, K., Ajani, U., Kase, C., Gaziano, J., Buring, J., Glynn, R. and Hennekens, C. (1999). Light to moderate alcohol consumption and the risk of stroke among US male physicians. *The New England Journal of Medicine* 341: 1557-1564.

Berger, M. and Leigh, J. (1988). The Effect of alcohol use on wages. *Journal of Applied Economics* 20: 1343-1351.

Berger, M. and Leigh, J. (1989). Schooling, self-selection, and health. *Journal of Political Economy* 97: 433–455.

Berry, W. and Feldman, S. (1985). *Multiple Regression in Practice*. Beverly Hills, CA: Sage Publications, Inc.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structured Estimates, *Journal of Human Resources* 8: 436-455.

Blow, P., Colleran, C., Oslin, D., Owen, F. and Slaymaker, V. (2005). Treatment outcomes for alcohol dependence among middle-aged and older adults. *Addictive Behaviors* 30(7): 1431-1436.

Bopp, M. and Minder, C. (2003). Mortality by education in German speaking Switzerland, 1990–1997: Results from the Swiss national cohort. International *Journal of Epidemiology* 32(3): 346–354.

Boyce, C. and Oswald, A. (2011). Do people become healthier after being promoted?. *Health Economics* 21(5): 580-596.

Breusch, T. and Pagan, A. (1979). A simple test for Heteroskedasticity and Random Coefficient Variation. *Econometrica* 47(5): 1287-1294.

British Medical Association (BMA) (2009). *The human cost of alcohol misuse: doctors speak out.* [Online]. (URL: http://bma.org.uk). (Accessed 30th April 2013)

Brookhart, A; Rassen, J.; Schneeweiss, S. (2010). Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Safety 2010 June; 19(6): 537-554

Brown, W. (1950). *Introducing Econometrics*. St. Paul: West Pub. Co.

Browne, R., Pinkerton, R. and Tuttle, M. (1987) Respiratory infections in smokers. *American Family Physician* 36: 133-140.

Browning M. and Meghir, C. (1991). The effects of male and female labour supply on commodity demands. *Econometrica* 59: 925-951.

Bryson, C., Mukamal, K., Mittleman, M.; Fried, L., Calvin, H., Kitzman, D. and Siscovick, D. (2006). The Association of Alcohol Consumption and Incident Heart Failure. *Journal of American College of Cardiology* 48(2): 305-311.

Buckley, N., Denton, F., Robb, A. and Spencer, B. (2004). The transition from good to poor health: an econometric study of the older population. Journal *of Health Economics* 23: 1013–1034.

Burton, S. and Tiffany, S. (1997). The effects of alcohol consumption on craving to smoke. *Addiction* 92: 15-26.

Butler, J., Finegan, T. and Siegfried, J. (1994). Does More Calculus Improve Student Learning in Intermediate Micro and Macro Economic Theory? *American Economic Review* 84(2): 206-210.

Butler, J., Finegan, T. and Siegfried, J. (1998). Does More Calculus Improve Student Learning in Intermediate Micro- and Macroeconomic Theory? *Journal of Applied Econometrics* 13: 185-202.

Byrne, S. (2011) *Costs to society of problem alcohol use in Ireland*. Dublin: Health Service Executive.

Cadoret, R., Yates, W., Troughton, E., Woodworth, G. and Stewart, M. (1995) Adoption study demonstrating two genetic pathways to drug abuse. *Archives of General Psychiatry* 52: 42-52.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: methods and applications.* Cambridge University Press.

Carr-Hill, R., Rice, N. and Roland, M. (1996). Socioeconomic determinants of rates of consultation in general practice based on fourth national morbidity survey of general practices. *British Medical Journal* 312: 1008-1012.

Carlson, M and Morrison, R.S (2009). Study Design, Precision, and Validity in Observational Studies. *Journal of Palliative Medicine,* 2009: 12 (1): 77-82

Case, A. (2001). *Does money protect health status? Evidence from South African pensions.* NBER Working Paper no. 8495. Cambridge: National Bureau of Economic Research.

Chahine, T., Subramanian, S. and Levy, J. (2011). Socio demographic and geographic variability in smoking in the US: A multilevel analysis of the 2006-2007 Current Population Survey, Tobacco Use Supplement. *Social Science and Medicine* 73: 752-758.

Cheng, S. and Long, J. (2007). Testing for IIA in the multinomial logit model. *Sociological Methods and Research* 35(4):583-600.

Chenhall, R and Moers, F. (2007) The Issue of endogeneity within Theory-Based, Quantitative Management Accounting Research. *European Accounting Review* 16(1): 173-195.

Chiburis, R. and Lokshin, M. (2007). Maximum Likelihood and two-step estimation of an ordered probit selection model. *The Stata Journal* 7(2): 167-182.

Cho, D. (1996). An alternative interpretation of conditional convergence results. *Journal of Money, Credit and Banking* 28: 669-681.

Choudhury, S. (1993). Reassessing the Male-Female wage differential: A fixed Effects Approach. *Southern Economic Journal* 60(2): 327-340.

Clark, A. and Etile, F. (2002). Do health changes affect smoking? Evidence from British panel data. *Journal of Health Economics* 21: 533-562.

Colditz, G., Giovannucci, E. , Rimm, E. *et al* (1991). Alcohol intake in relation to diet and obesity in women and men. *American Journal of Clinical Nutrition* 54: 49-55.

Connors, G., Donovan, D. and DiClemente, C. (2001) *Substance abuse treatment and the stages of change: Selecting and planning interventions*. Guilford Press New York.

Contoyannis, P. and Jones, A.M (2004). Socio-economic status, health and lifestyle. *Journal of Health Economics* 23: 965-995

Cook P. and Moore, M. (1999) *Alcohol*. NBER Working Papers 6905, National Bureau of Economic Research, Inc.

Cook, P. and Moore, M. (2000). *Environment and Persistence in youthful drinking patterns. Risky behaviour among youth: An Economic Analysis.* Chicago: University of Chicago Press.

Cooper, H., Smaje, C. and Arber, S. (1998). Use of health services by children and young people according to ethnicity and social class: secondary analysis of a national survey. *British Medical Journal* 317: 1047-1051.

Currie, J. and Moretti, E. (2003). Mothers Education and the intergenerational transmission of human capital: evidence from college openings. *Quarterly Journal of Economics* 118: 1495-1532.

Davies, P. and Walsh, D. (1983). *Alcohol problems and alcohol control in Europe*. Washington, DC: Taylor and Francis.

Deaton, A. and Lubotsky, D. (2003). Mortality, inequality and race in American cities and states. Social Science Medicine 56(6): 1139-1153.

Di Pietro, G. and Pedace, L. (2008). Changes in the returns to education in Argentina. *Journal of Applied Economics* XI(2): 259-279.

Disney, R., Emmerson, C. and Wakefield, M. (2006). Ill-health and retirement in Britain: a panel data based analysis. *Journal of Economics* 25(4): 621-649.

Dobson R. (2008). News: cutting price of alcohol is associated with higher mortality, study shows. *British Medical Journal* 337:a1504.

Doherty, E., Dee, A. and O' Neill, C. (2012). Estimating the Amount of Overweight and Obesity in Ireland Related Health-Care Use in the Republic of Ireland using Slán Data. *The Economic and Social Review* 43 (2): 227–250.

Doll, R. (1986). Tobacco: an overview of health effects, *IARC Scientific Publications* 74: 11-22.

Dougherty, C. (2007). *Introduction to Econometrics.* Edition: 3. Oxford University Press, 2007.

Downie, R., Tannahill, C. and Tannahill, A. (1996). *Health promotion: models and values*. 2nd Edition. Oxford University Press.

Dunlop, S., Coyte, P. and McIsaac, W. (2000). Socio economic status and the utilisation of physicians services: Results from the Canadian National Population Health Survey. *Social Science and Medicine* 51:123-133.

Durbin, J. (1954) Errors in Variables. *Review of the International Statistical Institute* 22: 23-32.

El-Sayed, M., Ali, N., El-Sayed, A. (2005). Interaction between alcohol and exercise: physiological and haematological implications. *Sports Medicine* 35(3): 257-269.

Enders, C.K. & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal* 8(3), 430-457.

Ettner, S. (1996). New evidence on the relationship between income and health. *Journal of Health Economics* 15: 67-85.

European Commission (2007). *Attitudes towards Alcohol.* Special Eurobarometer 272b / Wave 66.2-TNS Opinion & Social. European Commission at the request by Directorate General SANCO. [Online]. (URL: www.ec.euroopa.eu). (Accessed13th March 2003).

Evans, R. (2002). *Interpreting and addressing inequalities in health: from Black to Acheson to Blair to . . . ?.* London: Office of Health Economic; 2002. Seventh Annual Lecture.

Fayers, P. and Sprangers, M. (2002). Understanding self-related health. *The Lancet* 359: 187-188.

Fillmore, K., Golding, J., Leino, V., Ager, C. and Ferrer, H. (1994) Societal-Level Predictors of Groups' Drinking Patterns: A Research Synthesis from the Collaborative Alcohol-Relation Longitudinal Project. *American Journal of Public Health* 84(2): 247-253.

Finkelstein, M (2001). Do Factors other than Need Determine Utilization of Physicians' Services in Ontario? *Canadian Medical Association Journal* 5: 165.

Fischer, J. and Sousa-Poza, A. (2009). Does job satisfaction improve the health of workers? New evidence using panel data objective measures of health. *Health Economics* 18: 71-89.

Fletcher, G., Balady, G., Amsterdam. E. *et al* (2001). *Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association.* Circulation. 2001; 104(14)*:* 1694–*1740*. [Online] (URL http://www.ncbi.nlm.nih.gov/pubmed/11581152) (Accessed: 11th March 2013).

Foley A (2010) *The economic contribution of the drinks industry*. Dublin: Drinks Industry Group of Ireland.

Folland, S., Goodman, A. and Stano, M. (2001). *The Economics of Health and Health Care*. Third Edition. Prentice Hall.

Frazis, H. (1993). Selection Bias and the Degree Effect. *Journal of Human Resources* 28: 538-554.

French, M. and Zarkin, G. (1995). Is moderate alcohol use related to wages? Evidence from four worksites. *Journal of Health Economics* 14: 319-344.

French, M., Popovici, I. and Maclean, J. (2009). Do Alcohol Consumers Exercise More? Findings from a national survey. *American Journal of Health Promotion* 24(1): 2-10.

Frijters, P., Haisken-DeNew, J. and Shields, M. (2005). The causal effect of income on health: Evidence from German reunification. *Journal of Health Economics* 24: 997-1017.

Gallet, C. (2007). The demand for alcohol: a meta-analysis of elasticities. *The Australian Journal of Agricultural and Resource Economics* 51:121–135.

Garen, J., 1984. The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica* 52(5): 1199-1218.

Gerdtham, U., Johannesson, M., Lundberg, l. and Isacson, D. (1999). The demand for health: results from new measures of health capital. *European Journal of Political Economy* 15:501–521.

Gilleskie, D. and Harrison, A. (1998). The effect of endogenous health inputs on the relationship between health and education. *Economics of Education Review* (17) 279–295.

Gohlann, S., Requate, T., Schmidt, C. and Tauchmann, H. (2008). *Tobacco and Alcohol: Complements or Substitutes? A Structural Model Approach*. C.E.P.R. Discussion Papers 6780.

Green, C. and Polen, M. (2001). The health and health behaviours of people who do not drink alcohol. *American Journal of Preventative Medicine* 21(4): 298-305.

Greene, W. (2000). *Econometric Analysis*. 4th Edition. Upper Saddle River, New Jersey: Practice Hall.

Greene, W. (2002). *Econometric Analysis*. 5th Edition. Upper Saddle River, New Jersey: Practice Hall.

Greene, W. and Hensher, D. (2010). *Modeling Ordered Choices: A Primer.* Cambridge University Press.

Griffiths, W., Carter Hill, R. and Judge, G. (1993) *Learning and practicing econometrics*. Wiley, New York.

Groot, W. and Maassen van den Brink, H. (2007). The Health Effects of Education. *Economics of Education Review* 26: 186-200.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, Mar/Apr72, 80(2): 223-255.

Grossman, M. (2004). The demand for health, 30 years later: A very personal retrospective and prospective reflection. *Journal of Health Economics 23*(4):629–636.

Grossman, Michael (2008). The Relationship Between Health and Schooling. *Eastern Economic Journal*, 2008, 34, (281–292)

Gruber, S. and Kiesel, M. (2010). Inequality in health care utilisation in Germany? Theoretical and empirical evidence for specialist consultation. *Journal of Public Health* 18: 351-365.

Gruchow, H., Sobocinski, K., Barboriak, J. and Scheller, J. (1985). Alcohol consumption, nutrient intake and relative body weight among US adults. *American Journal of Clinical Nutrition* 42: 289-295.

Gujarati, D. (1995). *Basic Econometrics*. McGraw-Hill, Third Edition.

Gujarati, D. (2004). *Basic Econometrics*. McGraw-Hill, Fourth Edition.

Gulliver, S. (1995). Interrelationships of smoking and alcohol dependence, use and urges to use. *Journal of Studies on Alcohol* 56: 202-206.

Gulliver, S., Kamholz, B. and Helstrom, A. (1995). Smoking Cessation and Alcohol Abstinence: What to the data tell us? *Alcohol Research & Health* 29(3): 208-212.

Habicht, J. and Kunst, A. (2005). Social inequalities in health care services utilisation after eight years of health care reforms: a cross sectional study of Estonia, 1999. *Social Science and Medicine* 60(2005) 777-787.

Hamilton, V. and Hamilton, B. (1997). Alcohol and Earnings; does drinking yield a wage premium? *Canadian Journal of Economics* 30(1): 135-151.

Harmon, C. and Nolan, B. (2001). Health Insurance and Health Services Utilisation in Ireland. *Health Economics*, 10 (2): 135-145.

Harris, M., Ramful, P. and Shao, Z.(2006). An ordered generalised extreme value model with application to alcohol consumption in Australia. *Journal of Health Economics* 25: 782–801.

Hartog, J. and Oosterbeek, H. (1998). Health, wealth and happiness: Why pursue a higher education? *Economics of Education Review* 17: 245–256.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica* 52(5): 1219-1240.

Hay, J. and Olsen, R. (1984). Let them Eat Cake: A note on Comparing Alternative Models of the Demand for Medical Care. *Journal of Business & Economic Statistics* 2(3): 279-282.

Hay, J., Leu, R. and Rohrer, P. (1987) Ordinary Least Squares and Sample Selection Models of Health-Care Demand for Medical Care. *Journal of Business & Economic Statistics* 5: 499-506.

Health Promotion Unit (1996). *National Alcohol Policy*. Dublin Stationery: Office: Department of Health.

Health Research Board, Ireland (2012). *Alcohol: Public Knowledge, Attitudes and Behaviours*. Dublin: Health Research Board, Ispos MRBI.

Health Service Executive (2007). *Report of the HSE Working Group on Residential Treatment & Rehabilitation (Substance users).* Dublin: Health Service Executive.

Health Service Executive (2013). Schemes – Medical Cards. [Online] (URL: http://www.hse.ie/eng/services/list/1/schemes/mc/Medical_Cards.html). (Accessed: 29th July 2013).

Health Service Executive (2008). (URL:http://www.hse.ie/eng/services/news/Image_Bank/Alcohol_Awareness_Posters.pdf) (Accessed: 2008).

Heckman, J (1979).Sample Selection Bias as a specification error. *Econometrica* 47(1): 153-161.

Heien, D. (1996). Do Drinkers Earn Less? *Southern Economic Journal* Vol. 63: 60-68.

Hernandez-Quevedo, C. and Rubio, D. (2009). A comparison of the health status and health care utilization patterns between foreigners and the national population in Spain: New evidence from the Spanish National Health Survey. *Social Science & Medicine* 69: 370–378.

Hersch, J. and Viscusi, W. (1990). Cigarette smoking, seatbelt use and differences in wage-risk tradeoffs. *Journal of Human Resources* 25(2): 202-27.

Hilmer, M. (2001). A comparison of alternative specifications of the college attendance equation with an extension to two-stage selectivity-correction models. *Economics of Education Review* 20: 263–278.

Ho, S., Lam, T., Fielding, R and Janus, E. (2003). Smoking and perceived health in Hong King Chinese. *Social Science Medicine* 57(9): 1761-1770

Hofter, R. (2006). Private health insurance and utilization of health services in Chile. *Applied Economics* 38: 423–439.

Hogan, R. and Perrucci, C. (2007). Black women: Truly disadvantaged in the transition from employment to retirement income. *Social Science Research* 36(3): 1184-1199.

Holder, H. (1998). *The cost offsets of alcoholism treatment.* In M. Galanter (ed.), Recent developments in alcoholism. Vol. 14: The Consequences of Alcoholism, pp. 361-374. New York: Plenum Publishing Corporation.

Holman, C., Armstrong, B., Arias, L., Martin, C., Hatton, W., Hayward, L., Salmon, M., Shean, R. and Waddell, V. (1988). *The quantification of Drug Caused Morbidity and Mortality in Australia 1988.* Commonwealth Department of Community Services and Health Canberra.

Honda, K. (2004). Factors Underlying Variation in Receipt of Physician Advice on Diet and Exercise: Applications of the behavioral model of health care utilization. *American Journal of Health Promotion* 18(5): 370-377.

Hope, A. (2007) *Alcohol Consumption in Ireland 1986-2006.* Report for the Health Service Executive – Alcohol Implementation Group. Dublin: Health Service Executive.

Hope, A. (2008). *Alcohol Related Harm in Ireland.* Health Service Executive Report – Alcohol Implementation Group.

Hope, A., Gill, A., Costello, G., Sheehan, J., Brazil, E. and Reid, V. (2005) *Alcohol and injuries in the accident and emergency department - a national perspective*. Dublin: Department of Health and Children.

Hsieh, S., Hu, G., Chaung, Y., Chen, C. and Hu, Y. (2010). The Effects and Safety of Exercise Training in Subjects with Chronic Heart Failure – Do elder subjects gain similar benefits? *International Journal of Gerontology* 4: 165-170.

Humphries, K. and Van Doorslaer, E. (2000). Income-related health inequality in Canada . *Social Science & Medicine* 50: 663 – 671.

Hunkeler, E., Hung, Y., Rice, D. and Weisner, C. (2001). Alcohol consumption Patterns and Health Care Costs in an HMO. *Drug and Alcohol Dependence* 64: 181-190

Hurd, M. and McGarry, K. (1997). Medical Insurance and the Use of Health Care Services by the Elderly. *Journal of Health Economics*, 16 (2): 129-154.

Idson, T. and Feaster, D. (1990). A Selectivity model of employer-size wage differentials. *Journal of Labor Economics* 8, 99-122.

Intriligator, M., Bodkin, R. and Hsiao, C. (1996). *Econometric Models Techniques and Applications*. 2nd Edition. Upper Saddle River, NJ: Prentice Hall, 1996.

Ireland (2002). *Strategic task force on alcohol: interim report.* Dublin: Department of Health and Children, 2002.

Ireland (2004). *Strategic task force on alcohol; Second Report.* Dublin: Department of Health and Children, 2004.

Ireland (2011). Department of Health and Children. *Smoking key statistics and points*. [Online] (URL: http://www.dohc.ie/issues/smoking_ban/smokekey.html). (Accessed: 10th December 2012).

Ireland (2012) Irish Government News Services. *Health in Ireland Key Trends 2012*. [Online] (URL: *http://www.merrionstreet.ie/index.php/2012/12/health-in-ireland-key-trends-2012/*). (Accessed 22nd May 2013).

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal* 8(4): 453-479.

Jatrana, S. and Crampton, P. (2009). Primary Health care in new Zealand. *Health Policy* 93: 1-10.

Jiminez, E. and Kugler, B. (1987). The Earnings Impact of Training Duration in a Developing County: An Ordered Probit Selection Model of Colombia's Servicio Nacional de Aprendisaje. *Journal of Human Resources* 22(2): 228.247.

Jones, A.M (1996). Smoking Cessation and Health: A response. *Journal of Health Economics* 15(6): 775-759

Jones, A. (2005). *Applied Econometrics for Health Economists – A Practical Guide.* Second Edition. Department of Economics and Related Studies, University of York.

Judge, G., Hill, R., Griffiths, W., Lutkepohl, H. and Lee, T. (1988). *Introduction to Theory and Practice of Econometrics*. Wiley and Sons.

Jurges, H. (2008). Self-assessed health, reference levels and mortality. Applied *Economics* 40: 569-582.

Kahn, H., Tatham, L., Rodriguez, C. Calle, E., Thun, M. and Heath, C. (1997) Stable behaviors associated with adults' 10-year change in body mass index and the likelihood of gain at the waist. *American Journal of Public Health* 87(5): 747-754.

Kenkel, D. (1991). Health behavior, health knowledge, and schooling. *Journal of Political Economy* 99: 287–305.

Kenkel, D. (1995). Should you eat breakfast? Estimates from health production functions. *Health Economics* 4:15–29.

Kenkel, D. and Ribar, D. (1994). *Alcohol consumption and young adults socioeconomic status.* Brookings Papers on Economic Activity, Microeconomics 119-61.

Kennedy, P. (2003). *A guide to econometrics*. MIT Press. Fifth Edition.

Kiiskinen, U. (2003). *A Health Production Approach to the Economic Analysis of Health Promotion.* Publications of the National Public Health Institute.

King, G., Murray, C., Salomon, J. and Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* 98: 191-207.

Kiuila, O. and Mieszkowski, P. (2007). The Effects of Income, Education and Age on Health. *Health Economics* 16: 781-798.

Klatsky, A., Armstrong, M., Friedman, G. and Sidney, S. (2001). Alcohol drinking and risk of hospitalisation for ischemic stroke. *American Journal of Cardiology* 88: 703-706.

Klatsky, A.; Chartier, D.; Udaltsova, N.; Gronningen, S.; Brar, S.; Friedman, G. and Lundstrom, R. (2005). Alcohol Drinking and Risk of Hospitalisation for heart failure with and without associated coronary Artery disease. *The American Journal of Cardiology* 96: 346-351.

Klesges, R., Mealer, C. and Klesges, L. (1994). Effects of alcohol intake on resting energy expenditure in young women social drinkers. *American Journal of Clinical Nutrition* 59(4): 805-809.

Knowles, S. and Owen, P. (1997). Education and health in an effective-labour empirical growth model. *The Economic Record* 73: 314-328.

Knupfer, G (1984). The risks of drunkenness (or, ebrietas resurrecta). A comparison of frequent intoxication indices and of population sub-groups as to problem risks. *British Journal of Addiction*, 79(2): 185-196.

Koski, A., Sirén, R., Vuoir, E. and Poikolainen, K. (2007). Alcohol tax cuts and increase in alcohol-positive sudden deaths – a time-series intervention analysis. *Addiction* 102: 362-368.

Koski, A., Sirén, R., Vuoir, E. and Poikolainen. (2007). Alcohol tax cuts and increase in alcohol-positive sudden deaths – a time-series intervention analysis. *Addiction* (102): 362-368.

Koutsoyannis, A (1977). *Theory of Econometrics: An introductory Exposition of Econometric Methods*. MacMillan.

Kwan, Y. (2010). Life Satisfaction and Self-Assessed Health Among Adolescents in Hong Kong. *Journal of Happiness Studies* 11: 383-393.

Lahelma, E., Martikainen, P., Rahkonen, O. and Silventoinen, K. (1999). Gender Differences in illhealth in Finland: patterns magnitude and change. *Social Science & medicine* 48: 7-19.

Langpap, C. and Kerkvliet, J. (2002). *Success or Failure? Ordered Probit Approaches to measuring the effectiveness of the endangered species Act.* Paper presented at Annual Meeting of the American Agricultural Economics Association 2002.

Laroche, M. (2000). Health Status and Health Services Utilization of Canada's Immigrant and Non-Immigrant Populations. *Canadian Public Policy* 26(1): 51-75

Last, J. (1998). *Public Health and Human Ecology*. Prentice Hall, London.

Lee, L. (1982). Some Approaches to the Correction of Selectivity Bias. *The Review of Economic Studies*, Vol. 49: 355-372.

Lee, Lung-Fei (1983). Generalised Econometric Models with Selectivity. Econometrica 51(2): 507-512.

Leibovici, D.; Ritchie, K.; Ledesert, B. and Touchon, J. (1999). The effects of wine and tobacco consumption on cognitive performance in the elderly: A longitudinal study of relative risk. *International Journal of Epidemiology* 28:77–81, 1999.

Leibowitz, A. (2004), The demand for health and health concerns after 30 years. *Journal of Health Economics* 23: 663–671.

Leigh, J. (1998). Parents' schooling and the correlation between education and frailty. *Economics of Education Review* 17: 349–358.

Leigh, P. and Schembri, M. (2004). Instrumental variables technique: cigarette price provided better estimate of effects of smoking on SF-12. *Journal of Clinical Epidemiology* 57, 284-293.

Lianga, J., Bennetta, J., Sugisawac, H., Kobayashid, E. and Fukayad, T. (2003). Gender differences in old age mortality roles of health behaviour and baseline health status. *Journal of Clinical Epidemiology* 56: 572–582.

Lim, K., Jacobs and Klarenbach (2006). A Population Based Analysis of Health care Utilization of Persons With Back Disorders. *SPINE* 31(2): 212-218.

Lin, S. (2008). An analysis of lifestyles and health in Taiwan. *Applied Economics* Letters 15(5): 399-404.

Little, H. (2000). Behavioural Mechanisms Underlying the Link Between Smoking and Drinking. *Alcohol Research & Health* 24(4): 215-224.

Liu, L. (2008). Gender Difference of Health Status in Aging Population of Urban China. *The Chinese Economy* 41(1): 58-71.

Living in Britain Survey 2001.
[Online] (URL: http://webarchive.nationalarchives.gov.uk). (Accessed 14th March 2013).

Llena-Nozal, A., Lindeboom, M. and Portrait, F. (2004). The effect of work on mental health: does occupation matter? *Health Economics* 13: 1045-1062.

Loh, E. (1996) Productivity Differences and the marriage premium for white males. *Journal of Human Resources* 31: 566-89.

Long, J. (1997). *Regression models for categorical and limited dependent variables.* A volume in the Sage Series for Advanced Quantitative Techniques.
Thousand Oaks, CA: Sage Publications.

Long, J. and Freese, J. (2005). *Regression Models For Categorical Dependent Variables Using Stata*. 2nd Edition, College Station, TX: Stata Press.

Loprest, P. (1992). Gender Differences in wage growth and job mobility. *American Economic Review* 82 (2): 526-532.

Lye, J. and Hirschberg, J. (2004). Alcohol consumption, smoking and wages. *Applied Economics* 36: 1807-1817.

Mackenbach, J., Huisman, M., Andersen, O. *et al*. (2004). Inequalities in lung cancer mortality by the educational level in 10 European populations. *European Journal of Cancer*, 40(1), 126–135.

MacDonald, Z and Shields, M. (2001). The Impact of Alcohol Consumption on Occupational Attainment in England. Economica 2001; 68(271): 427-453

Macleod, J., Smith, G., Metcalfe, C. and Hart, C. (2005). Is subjective social status a more important determinant of health than objective social status? Evidence from a prospective observational study of Scottish men. *Social Science and Medicine* 61: 1916–1929.

Madalozzo, R. (2008). An analysis of income differentials by marital status. *Estuos Economicos (Estud. Econ.)*. 38(2) São Paulo 2008.

Maddala, G. (1983) *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, New York.

Madden, D., Nolan, A. and Nolan, B. (2005). GP reimbursement and visiting behaviour in Ireland. *Health Economics* 14: 1047-1060.

Manning, W., Duan, N. and Rogers, W. (1987). Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models. *Journal of Econometrics* 35: 59-82.

Manning, W., Keeler, E., Newhouse, J., Sloss, E. and Wasserman, J. (1991). *The Costs of Poor Health Habits*. Cambridge: Harvard University Press.

Marmot, M. Shipley, M. and Rose, G. (1984). Inequalities in death – specific explanations of a general pattern. *Lancet* 1: 1003–1006.

Martin, J., Barry, J and Skally, M. (2011) Alcohol Attributable Hospitalisations and Costs in Ireland, 2000-2004. *Irish Medical Journal*. 104: 140-4.

Martin, J., Barry, J., Goggin, D., Morgan, K., Ward, M. and O' Suilleabhain, T. (2010). Alcohol – attributable mortality in Ireland. *Alcohol* 45: 379-86.

Mathers, C.D, and Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*, 3(11): 442

Mattsom, M., Pollack, E. and Cullen, J. (1987). What are the odds that smoking will kill you? *American Journal of Public Health* 77: 425-431.

McCambridge, J., Hawkins, B. and Holden, C. (2013). *Industry Use of Evidence to Influence Alcohol Policy: A Case Study of Submissions to the 2008 Scottish Government Consultation.* PLoS Medicine 10(4): e1001431.

McGreggor, P., McKee, P. and O' Neill, C. (2006). GP Utilisation in Northern Ireland: Exploiting the Gatekeeper Function. *The Economic and Social Review* 37(1): 71-90.

McGuiness, T. (1980) An Economic Analysis of total demand for alcoholic beverages in the UK 1956-1975. *Journal of Industrial Economics* 29: 85-109.

McKelvey, R. and Zavoina, W (1975). A Statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4:103-120.

McNamara, A., Normand, C. and Whelan, B. (2013). *Patterns and Determinants of Health Care Utilisation in Ireland*. The Irish Longitudinal Study on Ageing (TILDA). Dublin.

McNamee, D. (1996). A Change in Lifestyle May Prevent Cancer. *The Lancet* 348, 1436.

McNiece, R. and Majeed A. (1999). Socioeconomic differences in general practice consultation rates in patients aged 65 and over: prospective cohort study. *British Medical Journal* 319**:** 26-28.

Meer, J., Miller, D. and Rosen, H. (2003). Exploring the health-wealth nexus. *Journal of Health Economics* 22: 713–730.

Menard, S. (1995). *Applied Logistic Regression Analysis.* Sage Publications. Series: Quantitative Applications in the Social Sciences, No. 106.

Merikangas, K. (1990) The genetic epidemiology of alcoholism. *Psychological Medicine* 20: 11-22.

Miche, E., Roelleke, E., Zoller, B., Wirtz, U., Schneider, M., Heurst, M., Amelang, M. and Radzewitz, A. (2009) A longitudinal study of quality of life in patients with chronic heart failure following an exercise program. *European Journal of Cardiovascular Nursing* 8: 281-287.

Milbourne, R., Otto, G. and Voss G. (2003) Public Investment and Economic Growth. *Applied Economics* 35: 527-540.

Mityakov, S. and Mroz, T. (2011). *Economic Theory as a Guide for the Specification and Interpretation of Empirical Health Production Function.* Presented at 2011 North American summer meeting of the Econometric Society.

Miyoshi, K. (2008) Female wage differentials in Japan. *Japan and the World Economy* 20: 479-496.

Mokdad A., Marks, J., Stroup, D. and Gerberding, J. (2004). Actual Causes of Death in the United States. *JAMA: Journal of the American Medical Association* 291(10):1238–45.

Mongan, D. (2010) The burden of alcohol-related morbidity on hospital services. *Drugnet Ireland* 5: 9-12.

Mongan, D., Reynolds, S., Fanagan, S. and Long, J. (2007) *Health-related consequences of problem alcohol use.* Overview 6. Dublin: Health Research Board.

Moore, A.; Gould, R.; Reuben, D.; Greendale, G.; Carter, K.; Zhou, K. and Karlamangia, A. (2005). Longitudinal Patterns and Predictors of Alcohol Consumption in the United States. *American Journal of Public Health* 95(3): 458-465.

Morgan, K., McGee, H., Watson, D., Perry, I., Barry, M., Shelley, E., Harrington, J., Molcho, M., Layte, R., Tully, N., Van Lente, E., Ward, M., Lutomski, J., Conroy, R., Brugha, R (2008). SLÁN 2007: Survey of Lifestyle, Attitudes & Nutrition in Ireland. Main Report. Dublin: Department of Health and Children.

Morris, S., Sutton, M. and Gravelle, H. (2005). Inequity and inequality in the use of health care in England: an empirical investigation. *Social Science & Medicine* 60: 1251–1266.

Mukamal, K. (2007) Alcohol Intake and Noncoronary Cardiovascular Disease *Annals of Epidemiology* 17(5), Supplement, Pages S8-S12.

Mukamal, K. J., Kuller, L., Fitzpatrick, A., Longstreth, W., Mittleman, M. and Siscovick, D. (2003b). Prospective Study of Alcohol Consumption and Risk of Dementia in Older Adults. *Journal of the American Medical Association* 289: 1405-1413.

Mullahy, J. (1999). It'll only hurt a second? Microeconomic Determinants of Who gets Flu Shots. *Health Economics* 8: 9-24.

Mullahy, J. and Sindelar, J. (1991) Gender differences in labour market effects of alcoholism. *American Economic Review* 81: 161-165.

Mullahy, J. and Sindelar. J. (1993) Alcoholism, work and income, *Journal of labour Economics* 11: 494-520.

Mullahy, J. and Sindelar, J. (1996) Employment, unemployment and problem drinking. *Journal of Health Economics* (1996) 409-434.

Napari, Sami (2009). Gender Differences in early-career wage growth. *Labour Economics* 16: 140-148.

National Suicide Research Foundation (2010). *National Registry of deliberate self harm Annual Report 2010.* Annual Report. Cork.

Nawata, K. (1993) A note on the Estimation of Models with Sample Selection Biases. *Economics Letters* 42: 15-24.

Nawata, K. (1994) Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two Step Estimator. *Economics Letters* 45: 33-40.

Nelson, F. (1984) Efficiency of the Two-Step Estimator for Models with Endogenous Sample Selection. *Journal of Econometrics* 24: 181-196.

Nelson, T. and Wechsler, H. (2001). Alcohol and college athletes. *Medicine and Science in Sports and Exercise* 33(1): 43-47.

Nolan B. (1994). General Practitioner Utilisation in Ireland: The role of Socio-Economic Factors. *Social Science and Medicine* 38(5): 711-716.

Nolan, A. and Nolan, B. (2003). *A Cross-Sectional Analysis of the Utilisation of GP Services in Ireland: 1987-2001.* Working Paper No.1 of Research Programme on Health Services, Health Inequalities and Health and Social Gain: ESRI, ISSC and University of Ulster.

Nolan, A. and Nolan, B. (2004). *Ireland's Health Care System: Some Issues and Challenges*. Dublin: Economic and Social Research Institute.

Nolan, Anne (2007). A Dynamic Analysis of GP Visiting in Ireland: 1995-2001. *Health Economics* 16: 129-143.

Nunes, J. (2008). Health, Social Insurance and Income. *International Advances in Economic Research* 14: 329-335.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labour Markets. *International Economic Review* 14: 693-709.

OECD (2011) *OECD Health Data*. Paris: OECD.

O'Farrell, T. (1995). *Marital and family therapy*. In R. K. Hester & W. R. Miller (Eds.), Handbook of alcoholism treatment approaches: Effective approaches. 2nd edition, pp. 195-220. Boston: Allyn and Bacon.

Office of the Surgeon General (US); Office on Smoking and Health (US). *The Health Consequences of Smoking: A Report of the Surgeon General.* Atlanta (GA): Centers for Disease Control and Prevention (US); 2004. [Online]. (URL: http://www.ncbi.nlm.nih.gov/books/NBK44695/). (Accessed: 27th March 2013).

Ogloblin, C. and Brock, G. (2003). Smoking in Russia: 'The Marlboro Man' Rides but Without 'Virginia Slims' for Now. *Comparative Economic Studies* 45: 87-103.

Orgogozo, J., Dartigues, J., Lafont, S., Letenneur, L., Commenges, D., Salamon, R., Renaud, S. and Breteler, M. (1997). Wine consumption and dementia in the elderly: a prospective community study in the Bordeaux area. *Revue Neurologique (Paris)* 153(3): 185–192.

Paarsch, H. (1984) A Monte Carlo Comparison of Estimators for Censored Regression Models. *Journal of Econometrics* 24: 197-213.

Paffenbarger, R. (1996). Physical Activity and Fitness for Health and Longevity. *Research Quarterly for Exercise and Sports* 67(3): 11-30.

Park, C. and Kang, C. (2008). Does education induce healthy lifestyle? *Journal of Health Economics* 27: 1516-1531.

Pate, R., Heath, G., Dowda, M. and Trost, S. (1996). Associations between Physical Activity and Other Health Behaviors in a Representative sample of US Adolescents. *American Journal of Public Health* 86(11): 1577-1581.

Pate, R., Pratt, M., Blair S. *et al*. (1995) Physical activity and public health: a recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA 273*: 402–*407.*

Pearce, D. (1986). *Dictionary of Modern Economics*. 3rd Edition. London: Macmillan.

Pearlman, A. and Tsao, T. (2009). *Decomposition of the Black-White Wage Differential in the Physician Market*. Levy Economics Institute, Economics Working Paper.

Peto, R., Lopez, A., Boreham, J. and Thun, M. (2012). Mortality from smoking in developed countries 1950-2005 (or later). [Online]
(URL: http://www.ctsu.ox.ac.uk/~tobacco/index.htm). (Accessed: 13th March 2013)

Pohlmeier, W. and Ulrich, V. (1994). An econometric model of the two-part decision-making process in the demand for health care. *Journal of Human Resources* 30(2): 339-360.

Protection of Young Persons Employment Act 1996. Ireland 1996.

Puhani, Patrick (2000). The Heckman Correction for Sample Selection and its Critique. *Journal of Economic Surveys* 14(1): 53-68.

Rabinovich, L., Brutscher, P., De Vries, H., Tiessen, J., Clift, J. and Reding, A. (2009). *The affordability of alcoholic beverages in the European Union. Understanding the link between alcohol affordability, consumption and harms.* Prepared for the European commission. Cambridge, RAND Europe.

Ransford, E. and Palisi, B. (1996). Aerobic exercise, subjective health and psychological well-being within age and gender subgroups. *Social Science and Medicine* 42(11): 1555-1559.

Reimers, C.W. (1983). Labor market discrimination against Hispanic and black men. *The Review of Economics and Statistics* 65 (4):570-579.

Rendtel, U. (1992). *On the choice of a selection model when estimating when estimating regression with selectivity*. DIW – Discussion Paper, 53, Berlin: German Institute Economic Research.

Ressler, R. and Waters, M. (2000). Female earnings and the divorce rate: a simultaneous equations model. *Applied Economics* 32: 1889-1898.

Rice, D., Conell, C., Weisner, C., Hunkeler, E., Fireman, B. and Hu, T. (2000). Alcohol Drinking Patterns and Medical Care Use in an HMO Setting. *The Journal of Behavioural Health Services and Research* 27(1): 3-16.

Rimm, E. and Moats, C. (2007) Alcohol and Coronary Heart Disease: Drinking Patterns and Mediators of Effect. *Annals of Epidemiology* 17(5): S3-S7.

Rivera, Berta (2001). The effects of public health spending on self-assessed health status: an ordered probit model. *Applied Economics* 33: 1313-1319.

Robert, S. and House, J. (2000). *Socioeconomic inequalities in health: integrating individual, community and societal level theory and research.* In: Albrecht G, Fitzpatrick R, Scrimshaw S (eds). The handbook of social studies in health and medicine. Thousand Oaks: Sage Publications 115-135.

Roberts, L. and Linney, K. (2000) *Alcohol problems and couples: Drinking in an intimate relational context.* In K. Schmaling & T. G. Sher (Eds.), The psychology of couples and illness, (pp. 269-310). Washington, DC: American Psychological Association.

Rosengren, A., Wedel, H. and Wilhelmsen, L. (1989). Marital Status and Mortality in Middle Ages Swedish Men. *American Journal of Epidemiology* 129 (1): 54-64.

Rotermann, M. (2006). Seniors' Health Care Use. *Statistics Canada Catalogue* 82-003 Supplement to Health Reports, Volume 16.

Sarma, S. and Simpson. (2006). A Microeconometric Analysis of Canadian Health Care Utilization. *Health Economics 15*: 219-239.

Savitz, D., Harris, R. and Brownson, R. (1993). Methods in chronic disease epidemiology in *Chronic Disease Epidemiology and Control* by Brownson, R., Remington, P. and Davis, J. eds, Washington, DC: American Public Health Association.

Schoeni, R. (1995). Marital Status and Earnings in Developed Countries. *Journal of Population Economics* 8:351-359.

Schultz, P. (2005). *Productive Benefits of Health: Evidence from Low-Income Countries.* IZA Discussion Paper No. 1482; Institute for the Study of Labor (IZA)

Seaks, Terry G. (1974). Simulations with Econometric Models and Alternative Methods of Estimation. *Southern Economic Journal*, 1-9.

Sewell, K. (2002). *International Alcohol Policies: A Selected Literature Review*. Scottish Executive Central Research Unit. Health and Community Research Programme. Research Findings No. 15, 2002.

Shao-Hsun, K. and Huffman, W. (2007). Binge Drinking and labour market success: a longitudinal study on young people. *Journal of Population Economics* 20: 35-54.

Shaper, A. (1988). Alcohol and Mortality in British Men: Explaining the U-Shaped Curve. *Lancet* (December 3) 1267-73.

Shaper**,** G. and Wannamethee, G. (1999). Types of alcoholic drink and risks of major coronary heart disease events and all cause mortality. *American Journal of Public Health* 89(5): 685-690.

Sherkat (2004). Religious intermarriage in the United States: trends, patterns, and predictors. *Social Science Research* 33: 606-625.

Siegel, M. and Lucke, M. (2009). *What determines the choice of transfer channel for migrant remittances? The Case of Moldova*. Working Papers. Kiel Institute for the World Economy.

Silles, Mary A (2009) The causal effect of education on health: Evidence from the United Kingdom. *Economics of Education Review* 28: 122–128.

Simon, H. (1992). Can You Run Away from Cancer? *Harvard Health Letter*, 17(5).

Sloan, F., Smith, K. and Taylor, D. (2003). *The Smoking Puzzle: Information, Risk Perception and Choice.* Cambridge, MA: Harvard Press.

Small, K. and Hsiao, C. (1985). Multinomial Logit specification tests. *International Econometric Review* 26(3): 619-627.

Smith, J. (1999). Healthy bodies and thick wallets: the dual relation between health and economic status. *Journal of Economic Perspectives* 13 (2): 145–166.

Srinivasan, V. and Basu, A. (1989). The Metric Quality of Ordered Categorical Data. *Marketing Science* 8(3): 205-230.

Stata Base Reference Manual Release 13 (2013). A Stata Press Publication StataCorp LP, College Station, Texas.

Stewart, M. (1983). On Least Squares estimation when dependent variable is grouped. *Review of Economic Studies* 50: 737-53.

Stewart, M. (1990). Access to Health Care for economically disadvantaged Canadians: a model. *Canadian Journal of Public Health* 81: 450-455.

Stolzenberg, R. and Relles, D. (1990). Theory Testing in a World of Constrained Research Design, The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental Research. *Sociological Methods and Research* 18 (4): 395-415.

Strauss, J. and Thomas, D. (1998). Health, Nutrition and Economic Development. *Journal of Economic Literature* 36(2): 766-817.

Stubet, J., Galea, S., Ahern, J., Blaney, S. and Fuller, C. (2003). The Association between Multiple Domains of Discrimination and Self-assessed Health: A multilevel Analysis of Latinos and Blacks in four low-Income New York City Neighbourhoods. *Health Services Research* 38(6): 1735-59.

Studenmund, A. (2005). *Introducing Econometrics: a practical guide*. Pearson Education, 2005.

Su, S. and Yen, S. (2000). A censored system of cigarette and alcohol consumption. *Applied Economics* 32: 729-737.

Substance Abuse and Mental Health Services Administration (SAMHSA) (2007). *US National Survey on Drug Use and Health.* [Online]. (URL: http://www.samhsa.gov/data/NSDUH.aspx). (Accessed 20th March 2013).

Sveinsdottir, H. and Gunnarsdottir, H. (2008). Predictors of self-assessed physical and mental health of Icelandic nurses: Results from a national survey. *International Journal of Nursing Studies* 45: 1479-1489.

Teljeur, C., O' Dowd, T., Thomas, S. and Kelly, A. (2010). The distribution of GP's in Ireland in relation to deprivation. *Health and Place* 16: 1077-1083.

Thompson, O. (2011). Racial disparities in the cognition-health relationship. *Journal of Health Economics* 30: 328–339.

Thorpe, R., Kasper, J., Szanton, S., Frick, K., Fried, L. and Simonsick, E. (2008). Relationship of race and poverty to lower extremity function and decline: Findings from the women's health and aging study. *Social Science and Medicine* 66: 811-821.

Tremblay, S., Ross, N. and Berthelot, J. (2002). *Regional Socio-economic context and health.* Supplement to Health Reports, 13: 1-12. Ottawa: Statistics Canada, Catalogue 82-002.

Trutz, H. and Pratschke, J. (2010) *Risk and Protection Factors for Substance Use among Young People: A comparative study of early school-leavers and school-attending students.* Dublin: National Advisory Committee on Drugs.

Trutz, H. and Pratschke, J. (2010) *Risk and Protection Factors for Substance Use among Young People: A comparative study of early school-leavers and school-attending students.* Stationery Office Dublin: National Advisory Committee on Drugs.

Tussing, A. (1985). *Irish Medical Care Resources: An Economic Analysis*. General Research Series Paper 126. Dublin: Economic and Social Research Institute.

Umberson, D. (1987). Family Status and Health Behaviours: Social Control as a Dimension of Social Integration. *Journal of Health and Social Behaviour* 28(3), 306-19.

United States National Institute on Alcohol Abuse and Alcoholism Report (1997) Vol. 21. [Online]. (URL: http://www.niaaa.nih.gov/publications/journals-and-reports/alcohol-research/archives). (Accessed: 27th March 2013).

US Public Health Service (1996). *Physical Activity and Health: A Report of the Surgeon General.* Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. The Presidents Council on Physical Fitness and Sports. [Online].
(URL: http://www.cdc.gov/nccdphp/sgr/pdf/execsumm.pdf).(Accessed: 28th March 2013).

Van Dijk, H., Monfort, A. and Brown, B. (1995). *Econometric Inference Using Simultaneous Techniques*. Wiley.

Van Doorslaer, E. (1987). *Health, Knowledge and the Demand for Medical Care: an econometric analysis.* Assen, The Netherlands: Van Gorcum.

Van Doorslaer, E. and Jones, A. (2003). Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics* 22: 61–87.

Van Doorslaer, E., Koolman, X. and Masseria, C. (2006). Inequalities in Access to Medical Care by Income in Developed Countries. *Canadian Medical Association Journal* 174 (2): 177-183.

Van Ours, J. (2004). A Pint a day raises a man's pay; but smoking blows that gain away. *Journal of Health Economics* 23: 863-886.

Vandenberghe, V. and Robin S. (2004). Evaluating the effectiveness of private education across countries: a comparison of methods. *Labour Economics* 11(4): 487-506.

Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *Journal of Human Resources* 33(1): 127-169.

Verbeek, M. (2008). *A Guide to Modern Econometrics*. Wiley and Sons, Third Edition.

Virgil, 70BC-19BC. [Online]. (URL: www.searchquotes.com). (Accessed 2nd August 2013).

Von Den Knesebeck, O., Verde, P. and Dragano, N. (2006). Education and Health in 22 European Countries. *Social Science & Medicine* 63(5): 1344-1351.

Von Fintel, D. (2007). Dealing with Earnings Bracket Responses in Household Survey – How Sharp are Midpoint Imputations. *South African Journal of Economics* 75(2): 293-312**.**

Wagenaar, A., Salois, M. and Komro, K. (2009). Effects of beverage alcohol price and tax levels on drinking: a meta-analysis of 1003 estimates from 112 studies. *Addiction* 104:179–190.

Wagstaff, A. (1986). The demand for health: a simplified Grossman model. *Bulletin of Economic Research* 38(1):93-95.

Walker, I. (1995). Estimates of the Economic Return to Schooling in the UK. *American Economic Review* 85 (5): 1278-1296.

Walsh, B. and Walsh, D. (2011). Suicide in Ireland: The influence of Alcohol and Unemployment. *The Economic and Social Review* 42(1): 27-47.

Wang, J., Gao, X., Wailes, E. and Cramer, G. (1996). US consumer demand for alcoholic beverages: cross section estimation of demographic and economic effects. *Review of Agricultural Economics* 18: 477-88.

Wang, W. (1997) Semi-parametric estimation of the effect of health on labour force participation of married women. *Applied Economics* 29: 325–9.

Wannamethee S. and Shaper, A. (2003). Alcohol, body weight, and weight gain in middle aged men. *American Journal of Clinical Nutrition* 77(5): 1312-1317.

Welch, G. and Clark, J. (1995). The Other Arthritis. *American Fitness* 13(5): 24-28.

Williamson, D., Forman, M., Binkin, N., Gentry, E., Remington, P. and Trowbridge, F. (1987). Alcohol and Body Weight in the United States Adults. *American Journal of Public Health* 77(10): 1324-1330.

Wilson, K., Rosenberg, M. and Abonyi, S. (2011). Aboriginal peoples, health and healing approaches: The effects of age and place on health. *Social Science & Medicine* 72: 355-364.

Wolfenstetter, S. (2011). Future Direct and indirect costs of obesity and influence of gaining weight: Results from the MONICA/KORA cohort studies, 1995-2005. *Economics and Human Biology* Article in Press.

Wonnacott, T. and Wonnacott, R. (1979). *Econometrics.* Wiley and Sons. Second Edition.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge and London.

Wooldridge, J. (2009) *Introductory Econometrics A Modern Approach*. Southwestern Cenagage Learning Inc. Fourth Edition.

World Health Organisation (1948). *Constitution of the World Health Organisation*. Adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States and entered into force on 7 April 1948.

World Health Organisation (1980). *Problems Related to Alcohol Consumption: report of a WHO expert committee.* Geneva: World Health Organisation.

World Health Organization (1986). *Ottawa Charter for Health Promotion,* Adopted at the First International Conference on Health Promotion, Ottawa, 21 November 1986.

World Health Organisation (WHO) (2000). *European Alcohol Action Plan 2000-2005.* Copenhagen: World Health Organisation Regional Office.

World Health Organisation (1991). Health21: The health for all policy framework for the WHO European Region. European Health for All Series, No.6. Copenhagen, WHO Regional Office for Europe.

World Health Organisation (2004). *Global Status Report on Alcohol 2004.* Geneva: World Health Organisation, Department of Mental Health and Substance Misuse.

World Health Organisation (2006). *Constitution of the World Health Organization - Basic Documents*, Forty-fifth edition, Supplement, October 2006.

World Health Organisation (2007). *WHO Expert Committee Report on Problems Related to Alcohol Consumption*. Technical Report Series 944. Geneva, World Health Organisation.

World Health Organisation (2009). *Report on the Global Tobacco Epidemic, 2009: Implementing smoke free environments*. World Health Organisation.

World Health Organisation (2010). *Global Strategy on Diet, Physical Activity and Health*. Geneva: World Health Organisation.

World Health Organisation (2011). *Global Status Report on Alcohol and Health* Geneva, World Health Organisation.

World Health Organisation (2012). *Alcohol in the European Union: Consumption, Harm and Policy Approaches.* Geneva: World Health Organisation.

Wu, D. (1973). Alternative tests of Independence Between Stochastic Regressors and Disturbances. *Econometrica Econometric Society* 41(4): 733-50.

Yan, X., Ma, M., Huang, H., Abdel-Aty, M. And Wu, C. (2011). Motor vehicle-bicycle crashes in Beijing: irregular manoeuvres, crash patterns, and injury severity. *Accident Analysis and Prevention* 43(5): 1751-8.

Yang, S. (2007). Racial disparities in training, pay-rise attainment, and income. *Research in Social Stratification and Mobility* 25(4): 323-335.

Yen, Steven T., Douglass Shaw, W., Yuan, Y. (2010). Cigarette Smoking and self reported health in China. *China Economic Review*, 2010, Vol. 21 (4): 532-543.

Zarkin, G., French, M., Mroz, T. and Bray, J. (1998). Alcohol use and wages: new results from the National Household Survey on Drug Abuse. *Journal of Health Economics* 17: 53-68.

Zellner, A. and Theil, H. (1962). Three Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrics* 54-78.

Zhang, H. (2004) Self Selection and Wage Differentials in Urban China: A Polychotomous Model with Selectivity. Cambridge, Department of Urban Studies and Planning, Massachusetts Institute of Technology.

Zhang, Y., Hannum, E. and Wang, M. (2008). Gender-Based employment and Income Differences in Urban China: Considering the Contributions of Marriage and Parenthood. *Social Forces* 86(4): 1529-1560.

Zick, C. and Smith, K. (1991). Marital Transitions, Poverty, and Gender Differences in Mortality. *Journal of Marriage and the Family* 53 (2): 327-36.

Ziebarth, N. And Grabka, M. (2009). In Vino-Pecunia? The Association Between Beverage Specific Drinking Behaviours and Wages. Journal of LAbor Research 30(3): 219-244.

Zietz, J. and Zhao, X. (2009). The response of household incomes to stock price and GDP growth by income quantile. *Applied Economics* 41: 1501-1512.

Zuehlke, T. and Zeman, A. (1991). A comparison of Two Stage Estimators of Censored Regression Models. *The Reviews of Economics and Statistics* 72: 185-188.

Zweifel, P. and Breyer, F. (1997). *Health Economics*. Oxford University Press.

APPENDIX

Appendix A: Alcohol Consumption per capita & costs of alcohol consumption in Ireland

Table A1: OECD Indicators of consumption of alcohol per capita

OECD Country	Litres per capita (15 years and over)	Change in consumption levels 1980–2009 %
France	12.3	-37%
Portugal	12.2	-18%
Austria	12.2	-16%
CzechRepublic	12.1	+3%
Estonia	12.0	n.a
Luxembourg	11.8	-14%
Hungary	11.8	-21%
Slovenia	11.5	n.a
Russian Fed.	11.5	+45%
Ireland	**11.3**	**+18%**
United Kingdom	10.2	+9%
Poland	10.2	-11%
Switzerland	10.1	-25%
Denmark	10.1	-14
Australia	10.1	-22%
Spain	10.0	-46%
Finland	10.0	+27%
Germany	9.7	-32%
Belgium	9.7	-28%
Netherlands	9.4	-18%
New Zealand	9.3	-21%
Greece	9.2	-19%

Table A1 continued: OECD Indicators of consumption of alcohol per capita

OECD Country	Litres per capita (15 years and over)	Change in consumption levels 1980–2009 %
OECD	**9.1**	**-9%**
Slovak Republic	9.0	-38%
Korea	8.9	n.a
United States	8.8	-15%
Chile	8.6	-21%
Canada	8.2	-23%
Italy	8.0	-52%
Sweden	7.4	+10%
Japan	7.4	+4%
Iceland	7.3	+70%
South Africa	7.2	+17%
Norway	6.7	+12%
Brazil	6.2	+188%
Mexico	5.9	+74%
China	4.4	+159%
Israel	2.5	-11%
Turkey	1.5	-17%
India	0.7	+47%
Indonesia	0.1	-25%

(Source: Steering Group Report on a National Substance Misuse Strategy 2012)

Table A2: Cost of Alcohol related problems in Ireland in 2001 & 2003

	2001 Euro million	2003 Euro million
Health Care costs	279	433
Costs of Road Accidents	315	322
Cost of alcohol related crime	100	147.5
Loss of output due to alcohol related absences from work	1,034	1,050
Alcohol related transfer payments	404	523.3
Taxes not received on lost outputs	234	210
TOTAL	**2,366**	**2,652.8**

(Source: Department of Health and Children, 2004)

Table A3: Overall cost of harmful use of alcohol in Ireland in 2007

	€ million	% of total cost
Cost to the healthcare system of alcohol-related Illnesses	1,200	32
Cost of alcohol-related road accidents	526	14
Cost of alcohol-related crime	1,189	32
Cost of output lost due to alcohol-related absence from work	330	9
Cost of alcohol-related accidents at work	197	5
Cost of alcohol-related suicides	167	5
Cost of alcohol-related premature mortality	110	3
Total	**3,710**	**100**

(Source: Steering Group Report on a National Substance Misuse Strategy 2012)

Appendix B: Descriptive Statistics from the Slán Survey 2002

Table B1: Descriptive Statistics from the Slán Survey 2002

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Alcohol Status	Those categorised as Non Drinkers = 1, Moderate Drinker = 2, Heavy Drinkers = 3.	1.97	.66	1	3
Log Income	The log of weekly household income in Euro	5.94	.85	3.48	7.55
male	Males =1, 0 = otherwise	0.40	.49	0	1
age1829	Those who are aged is 18 to 29 years =1, 0 = otherwise	0.18	.39	0	1
age3039	Those who are aged is 30 to 39 years =1, 0 = otherwise	0.22	.42	0	1
age4049	Those who are aged is 40 to 49 years =1, 0 = otherwise	0.25	.43	0	1
age5059	Those who are aged is 50 to 59 years =1, 0 = otherwise	.14	.34	0	1

Table B1 continued: Descriptive Statistics from the Slán Survey 2002

Variable	Variable Description	Mean	Standard Deviation	Min	Max
age6069 **Base Category**	Those who are aged is 60 to 69 years =1, 0 = otherwise	.09	.29	0	1
age70plus	Those who are aged is 70 plus years =1, 0 = otherwise	.12	.32	0	1
Ednoschooling **Base Category**	Individuals who have no schooling =1, 0 = otherwise	.001	.04	0	1
Edprimary	Individuals who have primary school education only =1, 0 = otherwise	.15	.36	0	1
Edsecondarysome	Individuals who have some secondary education =1, 0 = otherwise	.21	.40	0	1
Edsecondarycompl	Individuals who have completed secondary education =1, 0 = otherwise	.23	.42	0	1
Edthirdsome	Individuals who have some third level education =1, 0 = otherwise	.11	.32	0	1
Edthirdcompl	Individuals who have completed third level education =1, 0 = otherwise	.21	.41	0	1
Cohabiting **Base Category**	Individuals who are cohabiting = 1, 0 = otherwise.	.04	.19	0	1

Table B1 continued: Descriptive Statistics from the Slán Survey 2002

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Married	Individuals who are married = 1, 0 = otherwise.	.55	.50	0	1
Widowed	Individuals who are widowed = 1, 0 = otherwise.	.07	.25	0	1
Separated/Divorced	Individuals who are separated/Divorced = 1, 0 = otherwise.	.04	.20	0	1
singlenevermarried	Individuals who are single/never married = 1, 0 = otherwise.	.29	.45	0	1
Ulster **Base Category**	Individuals from the province Ulster=1, 0= otherwise	.06	.24	0	1
munster	Individuals from the province Munster=1,0= otherwise	.24	.43	0	1
leinster	Individuals from the province Leinster =1,0= otherwise	.49	.50	0	1
connaught	Individuals from the province Connaught =1, 0= otherwise	.11	.31	0	1
Healthpoor **Base Category**	Individuals who classify their health as being poor =1, 0= otherwise	.02	.14	0	1

Table B1 continued: Descriptive Statistics from the Slán Survey 2002

Variable	**Variable Description**	**Mean**	**Standard Deviation**	**Min**	**Max**
Health Excellent	Individuals who classify their health as being excellent =1, 0= otherwise	.16	.37	0	1
Healthvgood	Individuals who classify their health as being very good =1, 0= otherwise	.37	.48	0	1
Healthgood	Individuals who classify their health as being good =1, 0= otherwise	.33	.47	0	1
Healthfair	Individuals who classify their health as being fair =1, 0= otherwise	.11	.31	0	1
Disability **Base Category**	Those whose current employment situation is unable to work owing to permanent sickness/disability = 1, 0 = otherwise	.04	.19	0	1
Homemaker	Those whose current employment situation is Homemaker = 1, 0 = otherwise	.14	.35	0	1
Unemployed	Those whose current employment situation is unemployed =1, 0 = otherwise	.03	.18	0	1
Student	Those whose current employment situation is at school/student =1, 0 = otherwise	.02	.16	0	1
Retired	Those whose current employment situation is wholly retired =1, 0 = otherwise	.13	.33	0	1

Table B1 continued: Descriptive Statistics from the Slán Survey 2002

Variable	Variable Description	Mean	Standard Deviation	Min	Max
Seekingwork **Base Category**	Those whose current employment situation is seeking work for the first time =1, 0 = otherwise	.004	.07	0	1
Other	Those whose current employment situation is classified as other =1, 0 = otherwise	.034	.18	0	1
Employeeatwork	Those whose current employment situation is an employee at work =1, 0 = otherwise	.44	.50	0	1
Selfemployed	Those whose current employment situation is self employed =1, 0 = otherwise	.11	.32	0	1
hh16yrs+	Number of people in each household aged 16years or over	2.55	1.74	0	60
Partakechurchact	Individuals who regularly join in the activities of Church or other religious/parish groups, charitable or voluntary organisations =1, 0= otherwise	.22	.41	0	1
Religcath	Individuals who belong to a religion and are Roman Catholic=1, 0= otherwise	.83	.37	0	1
Avge price Alcohol	The average price in 2002.	1.87	0	1.87	1.87

(Source: Authors Own)

Appendix C: Heckman Selection Two Step Estimation Model

The Heckman (1979) standard two step estimation as a way of accounting for potential bias, estimates the probit equation by maximum likelihood to obtain estimates of γ in the first step (Greene, 2002). From this estimation a predicted value for the Inverse Mills Ratio can be generated for each observation in the selected sample. The primary equation is then estimated including the inverse mills ratio as an additional regressor (Greene, 2002).

The conventional sample selection model is set out as follows whereby the primary equation is eqn. (1) and the selection equation is equation (2).

$$y_i = x_i'\beta + \varepsilon_i \qquad \textbf{(C1)}$$

$$d_i^* = z_i'\gamma + v_i \qquad \textbf{(C2)}$$

$$d_i = 1 \text{ if } d_i^* > 0, \qquad d_i = 0 \text{ otherwise} \qquad \textbf{(C3)}$$

$$y_i = y_i^* * d_i \qquad \textbf{(C4)}$$

$$(\varepsilon_i, v_i) \sim N(0,0),(1, \rho\sigma_\varepsilon, 1)$$

Where:

	y_i	The equation of primary interest
	d_i^*	Reduced form for the latent variable capturing the self selection
	x_i	vector of variables
	z_i	vector of variables
	β	unknown parameters
	γ	unknown parameters
	ε_i & v_i	zero mean error terms with $\mathrm{E}[\varepsilon_i \mid v_i] \neq 0$

d_i^* is the latent variable with associated indicator function d_i reflecting whether the primary dependent variable is observed and where the relationships between d_i and d_i^*, and y_i and y_i^* are shown in equations 3 and 4.

The sample rule is that y_i is observed only when d_i^* is greater than zero. There is also an assumption that ε_i and v_i have a bivariate normal distribution with zero means and correlation ρ. The error term in the selection equation is assumed to be jointly normally distributed with the error term in the primary equation, and contains any unmeasured characteristics in the selection equation. This misspecification is overcome through the inclusion of a correction term that accounts for the selection bias.

$$E\left[y^*{}_i \middle| x_i, z_i, d_i = 1\right] = \beta' x_i + (\rho\sigma_\varepsilon)\left[\frac{\phi(z_i' \gamma)}{\Phi(z_i' \gamma)}\right] = \beta' x_i + (\rho\sigma_\varepsilon)\lambda_i \qquad \textbf{(C5)}$$

$$= \beta' x_i + \theta\,\lambda_i$$

Where:	y	The equation of primary interest
	x	vector of variables
	z	vector of variables
	d^*	Reduced form for the latent variable capturing the self selection
	β	unknown parameters
	σ	standard deviation
	ρ	correlation of the error terms ε_i and each of the u_{ij} terms
	γ	unknown parameters
	ϕ	probability density function
	Φ	cumulative distribution function
	λ_i	Inverse Mills Ratio
	i	indexes individuals

The term $\left[\frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)}\right]$ is the Inverse Mills Ratio which is denoted by λ_i.

In the second step, the primary equation (1) is estimated with the inverse mills ratio included as an additional regressor and estimates of β are obtained.

Appendix D: Full Information Maximum Likelihood Estimation

Details of the log-likelihood function and how it is formulated, as outlined by Greene (2002) is set out below. To formulate the appropriate log-likelihood function, the reduced form equation is depicted below.

$$Y = X \prod + V \tag{D1}$$

Where:	Y	Endogenous Variables
	X	Exogenous Variables
	$\prod$	K x M reduced form coefficient matrix
	V	Matrix of all reduced form disturbances and $V = E\Gamma^{-1}$

Each row of V is assumed to be multivariate normally distributed, with $E[v_t | X] = 0$ and covariance matrix $E[v_t v_t' | X] = \Omega$ (Greene, 2002)

The Log-likelihood is

$$\ln L = -\frac{T}{2}\left[M \ln(2\pi) + \ln|\Omega| + tr(\Omega^{-1}W)\right] \tag{D2}$$

Where:	M	No. of equations
	π	Coefficient
	Ω	$E[v_t \| v_t' X]$
	W_{ij}	$= \frac{1}{T}(y - X\pi_i^0)'(y - X\pi_j^0)$
	tr	tracing coefficients
	T	t=….T estimators of the parameters
	π_j^0	j th column of π
	i	indexes individuals

This function is maximised subject to all the restrictions imposed by the structure (Greene, 2002). The following substitutions are then made

$$\prod = -B\Gamma^{-1} \text{ and } \Omega = (\Gamma^{-1})'\sum\Gamma^{-1}$$

$$\text{So that } \Omega^{-1} = \Gamma\sum^{-1}\Gamma'$$

Thus

$$\ln L = -\frac{T}{2}\left[M\ln(2\pi) + \ln\left|(\Gamma^{-1})'\sum\Gamma^{-1}\right| + tr\left\{\frac{1}{T}\left[\Gamma\sum^{-1}\Gamma'(Y + XB\Gamma^{-1})'(Y + XB\Gamma^{-1})\right]\right\}\right]$$

(D3)

Which can be simplified firstly as

$$-\frac{T}{2}\ln\left|(\Gamma^{-1})'\sum\Gamma^{-1}\right| = -\frac{T}{2}\ln\left|\sum\right| + T\ln\left|\Gamma\right| \quad \textbf{(D4)}$$

Secondly as $\Gamma'(Y + XB\Gamma^{-1})' = \Gamma'Y' + B'X'$

By permuting Γ from the beginning to the end of the trace and collecting terms;

$$tr(\Omega^{-1}W) = tr\left[\frac{\sum^{-1}(Y\Gamma + XB)'(Y\Gamma + XB)}{T}\right] \quad \textbf{(D5)}$$

Therefore, the log likelihood is as set out by Greene (2002)

$$\ln L = -\frac{T}{2}\left[M\ln(2\pi) - 2\ln\left|\Gamma\right| + tr(\Sigma^{-1}S) + \ln\left|\Sigma\right|\right] \quad \textbf{(D6)}$$

Where:	M	No. of equations
	π	Coefficient
	Γ	an MxM non singular matrix

$$\Sigma = \Gamma'\Omega\Gamma$$

$$S = \frac{1}{T}(Y\Gamma + XB)'(Y\Gamma + XB)$$

In maximising the $\ln L$, the Full Information Maximum Likelihood estimator is produced, which as Greene (2002) states is asymptotically efficient among estimators of the simultaneous equations model.

Appendix E: Results from Multinomial Logit OLS Two Step Estimate of the effect of alcohol consumption on income using 2002 Slán Survey

The table below sets out the results of the alcohol status equation estimated in the first step of the Multinomial Logit OLS two step estimation using the Slán 2002 dataset.

Table E1: Results from the estimation of the Drinking Status using the 2002 Slán dataset

	Non Drinkers			Heavy Drinkers		
	Coefficient	**Z-Stats**		**Coefficient**	**Z-Stats**	
male	-0.296	-3.59*		0.524	6.32*	
Age1829	-0.776	-4.25*		1.908	8.04*	
Age3039	-0.521	-3.63*		1.390	6.20*	
Age4049	-0.703	-5.11*		0.778	3.47*	
Age5059	-0.363	-2.60*		0.611	2.64*	
Age70plus	0.167	1.21		-0.912	-2.75*	
Ed primary	0.686	4.48*		0.223	1.17	
Ed secondary some	0.486	3.36*		0.302	2.07**	
Ed secondary complete	0.238	1.66		-0.158	-1.11	
Ed third some	-0.024	-0.14		0.024	0.15	
Ed third complete	-0.288	-1.87		-0.529	-3.74*	
Married	-0.064	-0.32		-0.434	-2.60*	
Widowed	0.371	1.57		-0.169	-0.51	
Separated/divorced	0.072	0.28		0.296	1.27	
Singlenever/married	0.450	2.22**		0.068	0.41	
munster	0.057	0.52		-0.122	-1.00	
leinster	-0.281	-2.78*		0.037	0.35	
connaught	0.042	0.32		-0.016	-0.10	
homemaker	0.323	2.36**		0.191	1.04	
unemployed	0.121	0.54		0.350	1.58	

Table E1 continued: Results from the Drinking Status using the 2002 Slán dataset

	Non Drinkers			Heavy Drinkers		
	Coefficient	**Z-Stats**		**Coefficient**	**Z-Stats**	
student	0.118	0.40		0.610	2.44**	
retired	-0.324	-2.17**		-0.108	-0.37	
Seekingwork	-0.073	-0.11		0.167	0.32	
Empl other	0.354	1.07		-0.286	-0.7	
Employee at work	-0.185	-1.48		0.284	1.96**	
Selfemployed	-0.107	-0.71		-0.070	-0.39	
No. in hh 16yrs+	0.023	0.77		0.026	1.03	
Healthexcellent	-0.619	-3.19*		0.359	1.13	
healthvgood	-0.766	-4.20*		0.476	1.53	
healthgood	-0.625	-3.50*		0.664	2.14**	
healthfair	-0.232	-1.22		0.623	1.88**	
Religion Catholic	0.182	1.78		0.183	1.76**	
Partake Church activities	0.284	3.42*		-0.365	-3.51*	
Avge Price Alcohol	(omitted)			(omitted)		
_cons	-0.241	-0.76		-2.923	-6.66*	

No. of Observations = 5472

Wald Chi2(66) = 1087.65

Prob > chi2 = 0.0000

Pseudo R^2 = 0.1199

Log Likelihood = -4755.315

*** indicates significance at 1% level, ** indicates significance at 5% level**

Note: The average price of alcohol was included as a variable in the alcohol status equation. The price was derived by dividing the total values of sales in the 2002 by the total volume sold for each type of alcohol in 2002. Price was dropped due to collinearity.

The first independent variable considered is gender which is used to indicate if the survey respondent is male or not. Gender appears to be a very significant variable for both non and heavy drinkers showing that males are less likely to be a non-drinker and more likely to be moderate or a heavy drinker, which is similar to previous findings (Fillmore 1994; Moore, 2005; Blow *et al,* 2005; Moore *et al*, 2005; Mullahy & Sindelar, 1996).

Ones age is very significant in the determination of alcohol status. Those up to age 59 years are less likely to be non drinkers and are more likely to be moderate or heavy drinkers. In particular those between ages 18 and 39 are more likely to be heavy drinkers. Those over 70 years are more likely to be non-drinkers. These findings are similar to those of other studies whereby increasing age decreases the probability of a person being a heavy drinker (Hamilton and Hamilton, 1997; Barrett, 2002). As people get older, in particular those over 65 years of age, they are less likely to be a heavy drinker. Barrett (2002) states that this relationship between age and drinking status reflects an important life-cycle pattern of drinking behaviour – young people, other things being equal, are more likely to drink heavily and that likelihood decreases as they age.

In terms of marital status a single/never married person is more likely to be a non drinker. Being married is associated with a greater probability of being a moderate drinker, and a lower probability of being either a non or heavy drinker. Barrett (2002) finds marital status to have the same effect in his study.

In relation to the education variables those with a primary level of education are likely to be non-drinkers. Those who have some secondary level of education are likely to be either non or heavy drinkers as opposed to moderate drinkers. Those who have completed third level education are more likely to be moderate drinkers as opposed to non and heavy drinkers which is similar to the findings of Hamilton and Hamilton (1997) who have found that higher education has a negative effect on the propensity for individuals to be either non-drinkers or heavy drinkers as opposed to moderate drinkers.

Those from Leinster are more likely to be moderate drinkers as opposed to non-drinkers.

Some of the employment status variables are significant. A homemaker is more likely, whereas a retired person is less likely to be a non-drinker. Students are likely to be heavy drinkers as are employees at work.

Previous studies on the effect of alcohol on health, such as Berger *et al*, 1999; Klatsky *et al*, 2001; Bau *et al*, 2007 among others are that in general moderate levels of alcohol consumption is beneficial towards ones health status compared with abstaining from or consuming heavy amounts of alcohol, which has a negative effect on health status. The results of this study show that the variable health status is very significant in terms of the non-drinker status equation and findings are that those who of good, very good or excellent health are less likely to be a non drinker, which would appear to be in similar to previous findings. This is similar to the findings from the 2007 data.

The drinking status choice model also includes explanatory variables to indicate whether or not one is Catholic and also whether or not one partakes in regular Church activities. Both these variables are included in the drinking status choice model only and not in the income equation because they are hypothesised to affect the drinking decision only. In particular given its significance the variable describing involvement in Church activities ensures a good selection correction term for the wage equations, given that the variable is significant at the 1% level. Those who do partake in such activities are less likely to be heavy drinkers and more likely to be non drinkers. Similarly, Hamilton and Hamilton (1997) find that religious attendance has a positive impact on the propensity to be a non-drinker; however they found that this effect is virtually negated for Catholics, while results in this study show that those who are catholic are more likely to be either a non or heavy drinker compared with a moderate drinker.

The table below sets out the results of the earning regressions which include the inverse mills ratio as an additional variable, using the 2002 dataset.

Table E2: Results from the estimation of the Earnings Equation with selectivity corrections using the 2002 Slán dataset

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
	Coeffic-ient	t-stat		Coefficient	t-stat		Coefficien t	t-stat	
male	0.145	2.20**		0.052	1.85		0.301	3.07*	
Age1829	0.308	1.59		0.075	0.95		0.574	1.62	
Age3039	0.289	2.39**		0.197	3.54*		0.613	2.23**	
Age4049	0.276	2.36**		0.315	5.37*		0.512	2.47**	
Age5059	0.124	1.49		0.171	2.88*		0.395	2.17**	
Age70plus	0.039	0.62		-0.114	-2.06**		-0.037	-0.16	
Ed primary	-0.217	-1.96**		-0.530	-7.71*		-0.553	-4.89*	
Ed secondary some	-0.060	-0.61		-0.327	-5.38*		-0.277	-2.89*	
Ed secondary complete	0.178	1.97**		-0.052	-1.05		-0.085	-0.89	
Ed third some	0.443	4.06*		0.090	1.57		0.219	2.20**	
Ed third complete	0.532	5.23*		0.351	5.85*		0.160	1.37	
Married	0.177	1.39		0.226	3.21*		-0.110	-1.01	
Widowed	-0.255	-1.77		-0.156	-1.77		-0.196	-1.01	
Separated/divorced	-0.261	-1.85		-0.384	-4.49*		-0.314	-2.57**	
Singlenever/married	-0.267	-1.88		-0.319	-4.53*		-0.392	-4.86*	
munster	0.037	0.72		0.012	0.31		-0.025	-0.34	
leinster	0.197	3.29*		0.153	3.90*		0.107	1.68	
connaught	0.065	1.08		-0.058	-1.22		-0.178	-1.84	
homemaker	0.038	0.55		-0.054	-0.87		0.121	1.04	
unemployed	0.017	0.14		-0.291	-3.50*		-0.212	-1.49	
Student	-0.062	-0.33		-0.292	-2.04**		-0.045	-0.26	
Retired	0.146	2.11**		0.206	3.21*		0.319	2.11**	
Seekingwork	1.283	4.01*		0.324	1.15		0.359	1.72	
Empl other	0.190	0.97		0.140	1.26		0.087	0.25	
Employee at work	0.534	7.36*		0.411	8.83*		0.472	4.72*	
Selfemployed	0.241	2.85*		0.209	3.54*		0.301	2.75*	
No. in hh 16yrs+	0.013	1.16		0.003	0.23		0.062	2.57**	

Table E2 continued: Results from the estimation of the Earnings Equation with selectivity corrections using the 2002 Slán dataset

	Non Drinkers		Moderate Drinkers		Heavy Drinkers	
	Coefficient	**t-stat**	**Coefficient**	**t-stat**	**Coefficient**	**t-stat**
Healthexcellent	0.102	0.85	0.252	2.66*	0.046	0.22
healthvgood	0.143	1.19	0.228	2.50**	0.085	0.40
healthgood	0.099	0.88	0.093	1.09	0.098	0.43
healthfair	-0.082	-0.91	0.057	0.67	-0.100	-0.45
Mills Ratio	-0.319	-1.18	0.536	2.48**	0.156	0.47
_cons	5.301	18.04*	5.023	23.58*	4.932	6.04*

Non-Drinkers		**Moderate Drinkers**		**Heavy Drinkers**	
Number of obs	= 1278	Number of obs	= 3079	Number of obs	= 1115
F(32, 1245)	= 27.01	F(32, 3046)	= 56.73	F(32, 1082)	=17.33
Prob > F	= 00.00	Prob > F	= 00.00	Prob > F	= 00.00
R Squared	= 0.3557	R Squared	= 0.3165	R Squared	=.2475
Root MSE	= .6508	Root MSE	= .69109	Root MSE	= .7465

*** indicates significance at 1% level, ** indicates significance at 5% level**

The first independent variable considered in the earnings equations estimates is gender. Gender is proven to be significant for non and heavy drinkers with male non and heavy drinkers having higher earnings compared with females. This is similar to the findings of many other studies such as Mullahy and Sindelar, 1996; Zhang, 2008 among others, who have also found that gender is very significant in terms of earnings and that men tend to earn more than women.

In terms of the age variable, age groups 30-49 is significant across all drinker types and all are positively correlated with income. Similar to Barrett's (2002) findings the age-earnings profile for moderate drinkers has a concave shape, peaking at the ages 40 to 49 years. Those over 70 years are likely to earn less. Findings are similar for heavy drinkers.

The returns to education tend to be fairly uniform across all drinker types with a significant earnings premium for those who have completed a third level education. These results are in line with the findings of Hamilton and Hamilton (1987) and Barrett (2002). Those with lower levels of education i.e. those with a primary or some second level education only, tend to have lower earnings across all drinker types.

The independent variables indicating marital status show that moderate drinkers who are married tend to earn more. Moderate and heavy drinkers who are separated/divorced and those who are single/never married earn less which is in line with previous findings whereby married people earn more (Berger and Leigh, 1988; Mullahy and Sindelar, 1996; Schoeni, 1995; Ahituv and Lerman, 2007; Loh, 1996).

In terms of the provinces from which the respondents are living, Leinster is significant for non and moderate drinkers. Those living in Leinster have a positive earnings premium.

In relation to the variables categorising each respondents current employment status, for employees currently working or the self employed there is a positive earnings premium across all drinker types but more so for non- and heavy drinkers. Being retired is also significant across all drinker types with retired people earning more. A moderate drinker who is a student earns less. Unemployed moderate drinkers earn less.

The number of people in the household over 16 years is significant for heavy drinkers and is slightly positively correlated to earnings. Health Status does not appear to be significant in terms of its effect on earnings across all drinker types, except for the variables describing those in excellent and very good health for moderate drinkers. Moderate drinkers who described their health status as being very good or excellent all have a positive earnings premium. Grossman (1972) argued that a person with improved health status is in a position to work more and earn more.

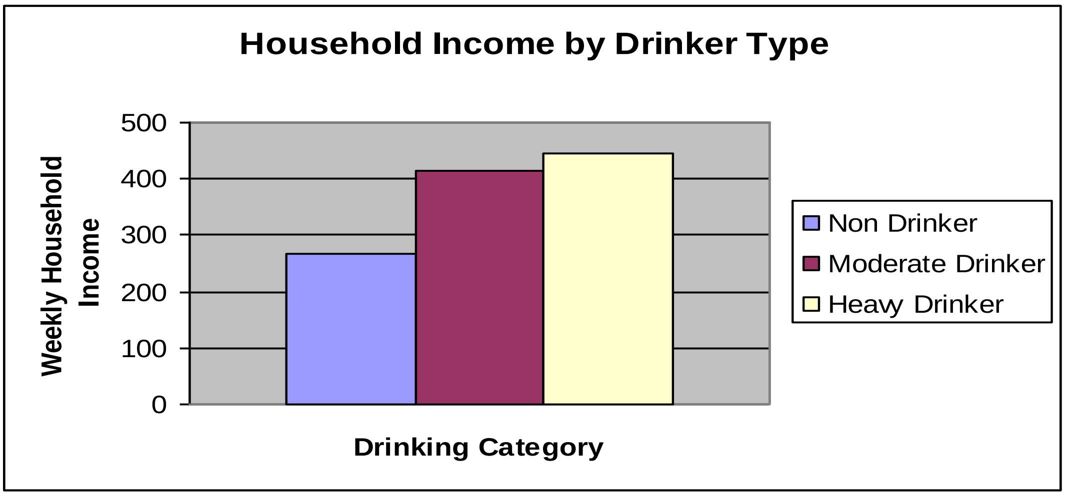
The earnings equations included the Mills Ratio which corrects for endogeneity bias arising from individuals self selecting into their drinking status. The coefficient on the Mills Ratio is only significant for moderate drinkers. The positive coefficient for moderate drinkers indicates that when an individual self selects into moderate drinking, earnings will be more than an individual with identical observable characteristics drawn at random as a moderate drinker.

The coefficient on the Mills Ratio for non drinkers and heavy drinkers is not statistically significant, indicating that selection bias does not exist. Hamilton and Hamilton (1987) and Barrett (2002) find that the Mills Ratio for non and moderate drinkers is insignificant however in contrast to this study they find that in relation to heavy drinkers the Mills Ratio is significant indicating that individuals who self select into heavy drinking earn more on average than an individual with identical observable characteristics drawn at random from the workforce would earn as a heavy drinker.

Results of Weekly Income by Drinker Type 2002 Dataset

Findings show that log of income for non drinkers is 5.59 which converts to €267.74 per week, log of income for moderate drinkers is 6.024 which is €413.23 per week and log of income for heavy drinkers is 6.099 which is €445.41 per week. Overall findings show that, similar to the findings using the 2007 dataset, there is very little difference between the household earnings of moderate and heavy drinkers, however non-drinkers earn substantially less.

Figure E1: Household Income by Drinker Type



(Source: Authors own)

Clearly these findings show that household income is highest for heavy drinkers. The difference between income for heavy and moderate drinkers is not large however there is a big difference between the income of moderate and heavy drinkers compared with non-drinkers, whereby non-drinkers earn substantially less.

Post Estimation Diagnostics

Testing the Specification of the Model

The Multinomial Logit Model has a strong assumption of independence of irrelevant alternatives (IIA). The IIA property states that the ratio of probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set.

To validate this assumption, a Suest-based Hausman test of IIA assumption was employed by using the mlogtest in Stata. Results were a P- Value of 1.0 for non-drinkers, 1.00 for moderate drinkers and 0.96 for heavy drinkers, showing that the null hypothesis, stating that the Independence of Irrelevant Alternatives is valid, was not rejected. This means that when an individual chooses an amount of alcohol to consume i.e. be in a particular drinker category, if another drinking category is added to the mix, this will not cause them to change their current drinking patterns. Based on this the multinomial logit can be applied.

The significance of each of the instruments is assessed using the z statistic. The Likelihood Ratio test is also used to evaluate the relevance of each instrument in the model and ensure that each instrument is beneficial to the model.

Robust standard errors are used in this study to account for heteroskedasticity.

In the alcohol status equation the variable average price of alcohol was dropped due to collinearity.

Appendix F: Marginal Effects of Ordered Probit Estimation of Alcohol Status in the estimation of the effect of alcohol on income using 2007 Slán survey

Table F1: Marginal Effects of Ordered Probit Estimation of Alcohol Status

Marginal effects after oprobit y= Pr(alcoholstatus==1) 0.25263

Variable	dy/dx	Z stat
male	-.107	-11.58*
Age18to29	-.121	-6.62*
Age30to39	-.072	-4.02*
Age40to49	-.072	-4.11*
Age50to59	-.062	-3.54*
Age70plus	.127	5.36*
edsecondary	-.073	-4.90*
eddiplomac~t	-.084	-5.56*
edprimaryd~e	-.107	-6.97*
edpostgrad~e	-.082	-4.94*
singleneve~d	.024	1.25
sepdiv	-.035	-1.48
married	.009	0.51
widowed	.028	1.05
village	-.051	-3.59*
town	-.050	-4.45*
cityothert~n	-.098	-7.48*
dublincity~y	-.088	-7.96*
employee	-.088	-3.10*
selfemplin~r	-.070	-2.60*
statetrain~d	-.109	-4.04*
unemployed	-.079	-2.38**
homemaker	-.036	-1.29
retired	-.059	-2.04**

Table F1: continued marginal Effects of Ordered Probit Estimation of Alcohol Status

Variable	dy/dx	Z stat
other	-.062	-1.43
numworkinghh	-.003	-0.78
Race White	-.105	-2.23**
Race Black	.322	3.94*
Race Asian	.385	4.83*
healthexce~t	-.131	-5.45*
healthvery~d	-.139	-5.29*
healthgood	-.138	-5.60*
healthfair	-.098	-3.94*
churchact	.047	4.01*
pr~vemoreyrs	-.063	-5.47*

(*) dy/dx is for discrete change of dummy variable from 0 to 1

*** indicates significance at 1% level, ** indicates significance at 5% level**

Appendix G: Marginal Effects in the Estimation into the effect of Alcohol Consumption on Health Status

In the estimation of the effect of alcohol on health status, both alcohol and health status are estimated as an ordered probit. The marginal effects are set out below.

Table G1: Marginal Effects of Ordered Probit Estimation of Alcohol Status using 2007 Slán survey

Marginal effects after oprobit y= Pr(alcohol status==1) = 0.25808474

Variable	dy/dx	Z stat
male*	-0.105	-11.53*
married*	0.004	0.20
widowed*	0.020	0.76
sepdiv*	-0.034	-1.43
single~d*	0.004	0.24
edseco~y*	-0.074	-5.21*
eddipl~t*	-0.087	-5.78*
edprim~e*	-0.098	-6.12*
edpost~e*	-0.075	-4.37*
Age18~29*	-0.106	-5.57*
Age30~39*	-0.054	-2.94*
Age40~49*	-0.065	-3.65*
Age50~59*	-0.054	-3.13*
Age70p~s*	0.101	4.50*
employee*	-0.106	-4.04*
selfem~r*	-0.091	-3.77*
statet~d*	-0.146	-6.59*
unempl~d*	-0.106	-3.67*
homema~r*	-0.061	-2.45**
retired*	-0.095	-3.78*
other*	-0.081	-2.08**
loginc~e	-0.061	-6.74*

Table G1 continued: Marginal Effects of Ordered Probit Estimation of Alcohol Status using 2007 Slán survey

Variable	dy/dx	Z stat
racewh~e*	-0.102	-2.34**
racebl~k*	0.235	3.15*
raceas~n*	0.321	4.12*
totali~h	0.002	0.92
village*	-0.049	-3.55*
town*	-0.049	-4.34*
cityot~n*	-0.093	-7.15*
dublin~y*	-0.077	-6.91*
smoker*	-0.103	-11.16*
We~right*	-0.056	-2.88*
weight~y*	-0.080	-4.24*
We~light*	-0.057	-2.21**
medcar~r*	0.000	0.02
health~e*	-0.025	-2.42**
church~t*	0.044	3.87*

(*) dy/dx is for discrete change of dummy variable from 0 to 1

*** indicates significance at 1% level, ** indicates significance at 5% level**

Table G2: Marginal Effects of Ordered Probit Regression of Health Status by Drinker Type

	Non Drinkers		Moderate Drinkers		Heavy Drinkers	
Variable	**dy/dx**	**z-stat**	**dy/dx**	**z-stat**	**dy/dx**	**z-stat**
edseco~y*	-0.030	-4.72*	-0.005	-3.40*	-0.006	-0.89
eddipl~t*	-0.027	-6.10*	-0.004	-3.66*	-0.004	-0.80
edprim~e*	-0.028	-7.00*	-0.005	-5.01*	-0.004	-0.82
edpost~e*	-0.023	-5.16*	-0.006	-5.68*	-0.003	-0.56
Age18~29*	-0.040	-9.32*	-0.007	-5.55*	-0.010	-1.36
Age30~39*	-0.035	-7.84*	-0.006	-4.86*	-0.009	-1.97*
Age40~49*	-0.031	-7.05*	-0.005	-3.91*	-0.009	-1.94
Age50~59*	-0.021	-4.40*	0.000	-0.13	-0.005	-1.37
Age70p~s*	0.042	3.25*	-0.002	-0.88	0.004	0.29
employee*	-0.083	-8.03*	-0.032	-6.42*	-0.042	-1.65
selfem~r*	-0.041	-9.75*	-0.011	-8.13*	-0.010	-2.31**
statet~d*	-0.033	-9.82*	-0.009	-7.96*	-0.008	-2.31**
unempl~d*	-0.033	-9.85*	-0.009	-7.95*	-0.008	-2.35**
homema~r*	-0.048	-9.04*	-0.011	-7.96*	-0.007	-2.33**
retired*	-0.061	-7.80*	-0.010	-7.46*	-0.009	-2.32**
other*	-0.030	-9.35*	-0.008	-7.83*	0.028	0.55
loginc~e	-0.021	-3.62*	-0.003	-3.42*	-0.005	-1.02
racewh~e*	-0.027	-1.29	0.001	0.63	-0.019	-0.63
raceblack	0.048	1.06	0.005	0.60	Omitted	
raceas~n*	0.312	2.54**	0.003	0.38	0.009	0.22
male*	-0.023	-3.03*	0.003	3.39*	-0.001	-0.11
married*	-0.014	-1.63	-0.003	-1.84	0.000	0.08
widowed*	-0.018	-2.65*	-0.001	-0.45	0.000	-0.07
sepdiv*	-0.017	-2.75*	-0.002	-1.57	0.002	0.31
single~d*	-0.011	-1.58	-0.002	-1.76	0.002	0.49
totali~h	0.000	0.33	0.000	1.52	0.000	-0.73
village*	-0.006	-1.06	0.001	0.78	0.000	0.09
town*	-0.006	-1.15	0.002	2.16**	0.000	-0.10
cityot~n*	-0.021	-4.11*	-0.001	-1.09	-0.003	-0.61
dublin~y*	-0.017	-3.09*	0.000	-0.16	-0.001	-0.25

Table G2 continued: Marginal Effects of Ordered Probit Regression of Health Status by Drinker Type

	Non Drinkers		Moderate Drinkers		Heavy Drinkers	
Variable	**dy/dx**	**z-stat**	**dy/dx**	**z-stat**	**dy/dx**	**z-stat**
smoker*	-0.011	-1.50	0.007	5.21*	0.001	0.14
We~right*	-0.023	-2.69*	0.001	0.56	-0.002	-0.38
weight~y*	-0.005	-0.57	0.009	3.68*	0.003	0.42
We~light*	0.016	1.08	0.010	2.04**	0.007	0.52
medcar~r*	0.015	3.29*	0.005	4.10*	-0.003	-1.17
health~e*	-0.017	-4.11*	-0.002	-2.32**	-0.003	-0.92
Mills Ratio	0.124	3.66*	0.009	1.59	-0.015	-0.60

Non-Drinkers y = Pr(healthstatusoprobit==1) (predict) = .03059334
Moderate Drinkers y = Pr(healthstatusoprobit==1) (predict) = .0079029
Heavy Drinkers y = Pr(healthstatusoprobit==1) (predict) = .00661565

(*) dy/dx is for discrete change of dummy variable from 0 to 1

*** indicates significance at 1% level, ** indicates significance at 5% level**

Table G3: Results of Health Status by Drinker Type when Lifestyle Variables are omitted

	Non Drinkers
Poor Health status	6.11%
Excellent Health Status	16.35%
Fair Health Status	17.08%
Good Health Status	30.12%
Very Good Health Status	30.34%
	Moderate Drinkers
Poor Health status	1.90%
Fair Health Status	8.34%
Excellent Health Status	23.19%
Good Health Status	28.66%
Very Good Health Status	37.87%
	Heavy Drinkers
Poor Health status	1.69%
Fair Health Status	9.81%
Excellent Health Status	22.05%
Good Health Status	32.03%
Very Good Health Status	34.42%

(Source: Authors own)

Appendix H: Sample Selected Probit Model

Greene and Hensher (2010) set out a sample selected probit model as follows, whereby the selection equation and the outcome equation are probit models.

The selection equation is as follows:

$$c^* = \alpha' s + \varepsilon \qquad \textbf{(H1)}$$
$$c = 1[c^* > 0]$$

Where:	c	dependent variable
	α	is an unknown vector of parameters,
	s	independent variables
	ε	error term

Inverse mills ratio is constructed for each individual i, from an estimate of the probit selection equation

$$\lambda_i = \phi(\alpha' s_i) / \Phi(\alpha' s_i) \qquad \textbf{(H2)}$$

h, x observed when $c = 1$

$(\varepsilon, u) \sim N[(0,0), (1, \rho, 1)]$

In the second step the inverse mills ratio is added to the outcome equation and estimated as follows:

$$h^* = \beta' x + u \qquad \textbf{(H3)}$$
$$h = 1[h^* > 0]$$

Where:	h	dependent variable in primary equation
	β	coefficient on the observable characteristics
	x	vector of independent variables
	u	error term

The log likelihood for the probit estimation of the primary equation is as follows:

$$\log L = \sum_{c=0} \log \Phi(-\alpha' s) + \sum_{c=1,h=1} \log \Phi_2(-\beta' x, \alpha' s, -\rho) + \sum_{c=1,h=1} \log \Phi_2(\beta' x, \alpha' s, \rho)$$

(H4)

Where: α is an unknown vector of parameters in the selection equation

s independent variables in the selection equation

β coefficient on the observable characteristics in primary equation

x vector of independent variables in primary equation

Φ cumulative distribution function

ρ correlation of the error terms

i indexes individuals where $i = 1,2,......N$

h dependent variable in primary equation

c dependent variable in selection equation

Appendix I: Results in the Estimation of Health Care Utilisation

In the estimation of the effect of alcohol on health care utilisation, both alcohol and health status are estimated as an ordered probit. The results of the ordered probit estimates of the alcohol status equation estimated in step one along with the marginal effects are set out below. The results of the health care utilisation estimation and the marginal effects estimated in step two are also set out below.

Table I1: Results of the Ordered Probit Regression of Alcohol Status using 2007 Slán survey

Variable	Coefficient	z-stat
healthexce~t	0.464	4.89*
healthvery~d	0.449	4.86*
healthgood	0.451	4.94*
healthfair	0.309	3.22*
edsecondary	0.209	4.58*
eddiplomac~t	0.269	5.02*
edprimaryd~e	0.316	5.13*
edpostgrad~e	0.229	3.65*
Age18to29	0.344	4.84*
Age30to39	0.161	2.64*
Age40to49	0.206	3.41*
Age50to59	0.189	3.20*
Age70plus	-0.285	-4.54*
employee	0.203	2.35**
selfemplin~r	0.180	1.94
statetrain~d	0.420	3.72*
unemployed	0.245	2.01**
homemaker	0.085	0.95
retired	0.216	2.29**
other	0.155	1.00
logincome	0.182	6.47*
racewhite	0.310	2.61*
raceblack	-0.629	-3.33*
raceasian	-0.865	-4.27*
male	0.340	11.41*

Table I1 continued: Results of the Ordered Probit Regression of Alcohol Status using 2007 Slán survey

Variable	Coefficient	z-stat
married	-0.009	-0.15
widowed	-0.061	-0.76
sepdiv	0.106	1.35
singleneve~d	-0.009	-0.15
totalinhh	-0.005	-0.93
village	0.161	3.42*
town	0.158	4.31*
cityothert~n	0.314	6.37*
dublincity~y	0.253	6.60*
smoker	0.344	10.62*
physically~e	(omitted)	
weightright	0.152	2.53**
weighttooh~y	0.243	3.93*
weighttool~t	0.196	2.12**
medcardhol~r	0.013	0.31
healthinsu~e	0.072	2.25**
churchact	-0.143	-4.24*
/cut1	2.275	
/cut2	4.609	

Number of obs = 8455
LR Chi2 (41) = 1207.38
Prob > Chi2 = 0
Pseudo R2 = .0982
Log Likelihood = -6162.84

*** indicates significance at 1% level, ** indicates significance at 5% level**

Table I2: Marginal Effects of Ordered Probit Regression of Alcohol Status in the estimation of the effect of alcohol consumption on health care utilisation using 2007 Slán survey

Marginal effects after oprobit y= Pr(alcohol status==1) = .2570

Variable	Coefficient	z-stat
healthexce~t	-0.135	-5.49*
healthvery~d	-0.138	-5.13*
healthgood	-0.135	-5.36*
healthfair	-0.091	-3.56*
edsecondary	-0.067	-4.62*
eddiplomac~t	-0.082	-5.36*
edprimaryd~e	-0.093	-5.69*
edpostgrad~e	-0.069	-3.92*
Age18to29	-0.102	-5.31*
Age30to39	-0.050	-2.73*
Age40to49	-0.063	-3.57*
Age50to59	-0.058	-3.37*
Age70plus	0.097	4.31*
employee	-0.065	-2.36**
selfemplin~r	-0.055	-2.04**
statetrain~d	-0.117	-4.46*
unemployed	-0.073	-2.21**
homemaker	-0.027	-0.97
retired	-0.066	-2.42*
other	-0.047	-1.06
logincome	-0.059	-6.47*
racewhite	-0.108	-2.45*
raceblack	0.233	3.10*
raceasian	0.326	4.13*

Table I2 continued: Marginal Effects of Ordered Probit Regression of Alcohol Status in the estimation of the effect of alcohol consumption on health care utilisation using 2007 Slán survey

Variable	Coefficient	z-stat
male	-0.108	-0.11
married	0.003	0.15
widowed	0.020	0.75
sepdiv	-0.033	-1.40
singleneve~d	0.003	0.15
totalinhh	0.002	0.93
village	-0.050	-3.58*
town	-0.050	-4.43*
cityothert~n	-0.092	-7.06*
dublincity~y	-0.077	-6.96*
smoker	-0.105	-11.33*
physically~e	omitted	
weightright	-0.049	-2.52**
weighttooh~y	-0.076	-4.05*
weighttool~t	-0.059	-2.28**
medcardhol~r	-0.004	-0.31
healthinsu~e	-0.023	-2.25**
churchact	0.047	4.13*
/cut1		
/cut2		

(*) dy/dx is for discrete change of dummy variable from 0 to 1

*** indicates significance at 1% level, ** indicates significance at 5% level**

Table I3: Ordered Probit Regression of Health Care Utilisation by Drinker Type

	Non Drinkers		Moderate Drinkers		Heavy Drinkers	
Variable	**Coefficient**	**z-stat**	**Coefficient**	**z-stat**	**Coefficient**	**z-stat**
healthexce~t	-1.048	-5.64*	-1.201	-7.67*	-1.699	-2.33**
healthvery~d	-0.926	-5.16*	-1.039	-6.74*	-1.500	-2.13**
healthgood	-0.723	-4.12*	-0.832	-5.43*	-1.280	-1.81
healthfair	-0.298	-1.95	-0.379	-2.43**	-1.082	-1.89
edsecondary	0.079	0.90	0.021	0.35	0.086	0.28
eddiplomac~t	0.199	1.70	0.053	0.78	-0.104	-0.27
edprimaryd~e	0.239	1.65	0.004	0.06	0.043	0.10
edpostgrad~e	0.084	0.62	0.064	0.85	0.267	0.74
Age18to29	0.252	1.39	-0.172	-2.02**	-0.261	-0.50
Age30to39	0.022	0.18	-0.178	-2.48**	-0.152	-0.47
Age40to49	-0.165	-1.34	-0.246	-3.42*	-0.171	-0.46
Age50to59	-0.125	-1.12	-0.131	-1.85	0.047	0.14
Age70plus	-0.063	-0.55	0.132	1.49	0.595	1.17
employee	-0.480	-3.21*	-0.513	-4.83*	0.144	0.41
selfemplin~r	-0.684	-4.35*	-0.491	-4.40*	0.218	0.64
statetrain~d	-0.893	-3.75*	-0.476	-3.59*	0.212	0.36
unemployed	-0.604	-3.00*	-0.352	-2.46**	-0.141	-0.34
homemaker	-0.631	-4.57*	-0.426	-3.78*	0.574	1.70
retired	-0.450	-2.99*	-0.340	-2.78*	0.336	0.76
other	-0.650	-2.49**	-0.574	-3.08*	-0.230	-0.21
logincome	0.107	1.37	0.014	0.42	-0.024	-0.09
racewhite	0.174	0.90	0.086	0.63	0.005	0.01
raceblack	-0.027	-0.09	0.145	0.59	(omitted)	
raceasian	-0.263	-0.75	-0.641	-2.03**	-1.734	-1.48
male	-0.208	-1.61	-0.354	-9.27*	-0.334	-0.78
married	0.316	2.39**	0.034	0.55	0.010	0.05
widowed	0.247	1.65	-0.056	-0.58	0.206	0.59
sepdiv	0.170	1.05	-0.042	-0.51	0.067	0.25
singleneve~d	0.015	0.12	-0.050	-0.82	-0.088	-0.52
totalinhh	-0.024	-2.36**	-0.009	-1.43	0.001	0.04
village	0.014	0.15	-0.077	-1.38	-0.255	-1.00

Table I3 continued: Ordered Probit Regression of Health Care Utilisation by Drinker Type

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
Variable	**Coefficient**	**z-stat**		**Coefficient**	**z-stat**		**Coefficient**	**z-stat**	
town	0.076	0.93		-0.080	-1.86		-0.062	-0.25	
cityothert~n	0.061	0.46		-0.018	-0.31		0.023	0.05	
dublincity~y	0.050	0.46		-0.024	-0.53		0.047	0.14	
smoker	0.066	0.50		-0.016	-0.41		-0.044	-0.10	
physically~e	(omitted)			(omitted)			(omitted)		
weightright	0.056	0.53		-0.013	-0.17		-0.417	-1.02	
weighttooh~y	0.100	0.81		0.058	0.73		-0.188	-0.40	
weighttool~t	0.037	0.22		-0.164	-1.49		0.053	0.11	
medcardhol~r	0.477	6.92*		0.332	7.16*		0.677	4.81*	
healthinsu~e	0.141	2.34**		0.143	3.88*		0.335	2.31**	
mills_alco~1	-0.196	-0.42		-0.041	-0.16		-0.109	-0.07	
/cut1	-2.622			-3.918			-4.623		
/cut2	-1.608			-2.753			-3.027		
/cut3	-1.180			-2.200			-2.455		
/cut4	0.221			-0.688			-1.003		

Non-Drinkers	**Moderate Drinkers**	**Heavy Drinkers**
Number of obs = 2345	Number of obs = 5563	Number of obs = 547
LR Chi2 (41) = 432.23	LR Chi2 (41) = 764.92	LR Chi2 (40) = 138.29
Prob > Chi2 = 0	Prob > Chi2 = 0	Prob > Chi2 = 0
Pseudo R2 = .084	Pseudo R2 = .06	Pseudo R2 = .098
Log Likelihood = -2569.54	Log Likelihood = -6435.06	Log Likelihood = -638.62

*** indicates significance at 1% level, ** indicates significance at 5% level**

Table I4: Marginal Effects of Ordered Probit Regression of Health Care Utilisation by Drinker Type

	Non Drinkers			Moderate Drinkers			Heavy Drinkers		
Variable	**dy/dx**	**z-stat**		**dy/dx**	**z-stat**		**dy/dx**	**z-stat**	
healthexce~t	0.049	2.69*		0.055	3.67*		0.072	0.87	
healthvery~d	0.029	2.83*		0.030	3.85*		0.035	0.89	
healthgood	0.020	2.50**		0.026	3.17*		0.027	0.80	
healthfair	0.007	1.48		0.010	1.69		0.033	0.76	
edsecondary	-0.001	-0.90		0.000	-0.35		-0.001	-0.28	
eddiplomac~t	-0.003	-1.91		-0.001	-0.8		0.001	0.24	
edprimaryd~e	-0.003	-2.00**		0.000	-0.06		0.000	-0.10	
edpostgrad~e	-0.001	-0.68		-0.001	-0.89		-0.002	-0.93	
age18to29	-0.004	-1.70		0.003	1.71		0.003	0.40	
age30to39	0.000	-0.18		0.004	2.14**		0.001	0.40	
age40to49	0.003	1.14		0.005	2.73*		0.002	0.39	
age50to59	0.002	0.99		0.003	1.63		0.000	-0.14	
age70plus	0.001	0.52		-0.002	-1.69		-0.002	-1.40	
employee	0.011	2.27**		0.009	3.92*		-0.001	-0.40	
selfemplin~r	0.025	2.41**		0.014	2.79*		-0.001	-0.76	
statetrain~d	0.046	1.84		0.015	2.21**		-0.001	-0.46	
unemployed	0.022	1.67		0.009	1.68		0.001	0.28	
homemaker	0.019	2.69*		0.011	2.54**		-0.002	-1.58	
retired	0.011	2.12**		0.008	2.03**		-0.002	-0.98	
other	0.026	1.37		0.021	1.78		0.003	0.16	
logincome	-0.002	-1.31		0.000	-0.42		0.000	0.09	
racewhite	-0.004	-0.75		-0.002	-0.57		0.000	-0.01	
raceblack	0.000	0.08		-0.002	-0.71		0.143	0.54	
raceasian	0.006	0.56		0.025	1.13		0.002	0.76	
male	0.004	1.43		0.007	6.06*		0.000	-0.05	
married	-0.006	-2.14**		-0.001	-0.55		-0.001	-0.72	
widowed	-0.004	-1.90**		0.001	0.55		-0.001	-0.27	
sepdiv	-0.003	-1.24		0.001	0.48		0.001	0.49	
singleneve~d	0.000	-0.12		0.001	0.79		0.000	-0.04	
totalinhh	0.000	2.15**		0.000	1.40		0.072	0.87	

Table I4 continued: Marginal Effects of Ordered Probit Regression of Health Care Utilisation by Drinker Type

	Non Drinkers		Moderate Drinkers		Heavy Drinkers	
Variable	**dy/dx**	**z-stat**	**dy/dx**	**z-stat**	**dy/dx**	**z-stat**
village	0.000	-0.15	0.001	1.26	0.003	0.67
town	-0.001	-0.97	0.001	1.73	0.001	0.23
cityothert~n	-0.001	-0.49	0.000	0.31	0.000	-0.05
dublincity~y	-0.001	-0.47	0.000	0.52	0.000	-0.14
smoker	-0.001	-0.52	0.000	0.40	0.000	0.10
physically~e	omitted					
weightright	0.000	-0.47	0.000	0.17	0.003	0.91
weighttooh~y	-0.002	-0.84	-0.001	-0.76	0.002	0.36
weighttool~t	-0.001	-0.23	0.004	1.23	0.000	-0.12
medcardhol~r	-0.009	-4.19*	-0.005	-5.76*	-0.004	-1.78
healthinsu~e	-0.002	-2.19**	-0.003	-3.42*	-0.003	-1.48
mills_alco~1	0.003	0.42	0.001	0.16	0.001	0.07

Non-Drinkers y = Pr(gpconsult==1) (predict) = .0063
Moderate Drinkers y = Pr(gpconsult==1) (predict) = .00624
Heavy Drinkers y = Pr(gpconsult==1) (predict) = .00268

*** indicates significance at 1% level, ** indicates significance at 5% level**

Appendix J: Testing the cut off points in Ordered Probit models

---------- Forwarded message ----------
From: **Stata Technical Support** <tech-support@stata.com>
Date: Mon, Feb 13, 2012 at 8:31 PM
Subject: Re: Ordered Probit Cut Off's
To: Gillian Ormond <gillianormond@gmail.com>

Dear Gillian,

Testing the significance of cut points may not be meaningful because there are different ways to specify equivalent ordered probit models that will result in changes in the cut points. You can see from the following FAQ

http://www.stata.com/support/faqs/stat/oprobit.html

that including and intercept in the model will reverse the sign of the first cut point while other cut points can be positive. This means that you could even have an important cut point equal to zero (by chance)or statistically no different from zero with a very good model in the statistical sense. Having said that, below I provide an example where I test if a couple of cut points coefficients are equal to zero using the -test- command. Be careful with the corresponding interpretation.

```
. webuse fullauto, clear
  . oprobit rep77 foreign length mpg
  . test _b[/cut1]=0
  . test _b[/cut2]=0
```

And this is the output:

```
. . webuse fullauto, clear
(Automobile Models)

.   . oprobit rep77 foreign length mpg



Iteration 0:   log likelihood = -89.895098
Iteration 1:   log likelihood = -78.106316
Iteration 2:   log likelihood = -78.020086
Iteration 3:   log likelihood = -78.020025
Iteration 4:   log likelihood = -78.020025
```

Continued: Testing the cut off points in Ordered Probit models

```
Ordered probit regression                    Number of obs   =         66
                                             LR chi2(3)      =      23.75
                                             Prob > chi2     =     0.0000
Log likelihood = -78.020025                  Pseudo R2       =     0.1321
------------------------------------------------------------------------------
       rep77 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     foreign |   1.704861   .4246796     4.01   0.000     .8725037    2.537217
      length |   .0468675    .012648     3.71   0.000      .022078    .0716571
         mpg |   .1304559   .0378628     3.45   0.001     .0562463    .2046656
-------------+----------------------------------------------------------------
       /cut1 |    10.1589   3.076754                      4.128577    16.18923
       /cut2 |   11.21003   3.107527                      5.119389    17.30067
       /cut3 |   12.54561   3.155233                      6.361467    18.72975
       /cut4 |   13.98059   3.218793                      7.671874    20.28931
------------------------------------------------------------------------------
.   . test _b[/cut1]=0

 ( 1)  [cut1]_cons = 0

           chi2(  1) =   10.90
         Prob > chi2 =    0.0010

.   . test _b[/cut2]=0

 ( 1)  [cut2]_cons = 0

           chi2(  1) =   13.01
         Prob > chi2 =    0.0003
```

For more details and examples type -help test-.

Let me know if you have further questions.

Kind regards,
Miguel Dorta

****

Miguel Dorta
Technical Support Representative
tech-support@stata.com
StataCorp LP
4905 Lakeway Drive
College Station, TX 77845

****