AN ANALYSIS OF THE EFFECT OF ALCOHOL CONSUMPTION ON HOUSEHOLD INCOME IN IRELAND COMPARING LIMITED AND FULL INFORMATION METHODS OF ESTIMATION

Gillian Ormond, Rosemary Murphy

ABSTRACT

This paper presents a study of the effects of alcohol consumption on household income in Ireland using the Slán 2007 dataset, accounting for endogeneity and selection bias. Drinkers are categorised into one of three categories; non, moderate and heavy drinkers, based on the recommended weekly drinking levels by the Irish Health Promotion Unit. Previous studies into the effect of alcohol on income have not accounted for the fact that alcohol consumption can be viewed as ordered data. This study accounts for the ordinality of alcohol consumption. Limited and Full Information Methods of Estimation are compared and the effect of alcohol consumption on income is estimated using both methods of estimation. Results from both methods of estimation show that income for moderate drinkers is higher than abstainers or heavy drinkers. The difference in income between moderate drinkers and heavy drinkers is much greater when using the FIML method, with income of heavy drinkers being far less than that of moderate drinkers and substantially less than non-drinkers.

1. INTRODUCTION

This paper investigates the effect of individual alcohol consumption on household income in Ireland while accounting for the potential endogenous relationship between alcohol and income. Much research has been carried out into the effect of alcohol consumption on income, an important feature of the more recent research being that the potential endogenous relationship between alcohol status and income has been controlled for (Hamilton and Hamilton, 1997; Barrett, 2002). 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, alcohol consumption is governed in part by unobserved factors which may also be important determinants of the dependent variable income, implying the possibility that the drinking status variables may be correlated with the error term of the conditional demand equation (Zarkin *et al*, 1998; Hamilton and Hamilton, 1997; Di Pietro & Pedace, 2008; 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 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).

Previous studies such as Hamilton and Hamilton (1997) and Barrett (2002) among others, have estimated the effect of alcohol consumption on income using the Multinomial Logit Ordinary Least Squares (OLS) Two Step Procedure proposed by Lee (1982). This is an extension of the Heckman Probit OLS Two Step Estimate which controls for selection correction and the endogeneity of the choice variable, alcohol consumption, when treated as a polychotomous choice variable. These studies have however assumed that alcohol status is unordered and have estimated the alcohol status equation as such, however alcohol consumption could 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 by OLS. Such estimations can be carried out using Limited Information or Full Information methods of estimation. The Limited Information 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), while Full Information Methods estimate all the equations in a model simultaneously and this joint estimation of equations brings efficiency gains (Greene, 2002).

There have been many studies comparing limited information methods of estimation and full information methods of estimation. Generally findings have been that the full information methods tend to be more efficient primarily due to the fact that with full information methods all the information that is available in the system is used simultaneously as opposed to the limited information methods whereby each equation in the system is estimated individually (Gujarati, 2004). Generally the limited information methods have tended to be used due to the fact that they are computationally easier to estimate (Gujarati, 2004).

Using data from the 2007 Slán Health and Lifestyle Survey, both methods of estimation are used to examine the relationship between alcohol use and income for three categories of drinkers; non, moderate and heavy drinkers accounting for endogeneity and selection bias, and also accounting for the ordinality of alcohol consumption. Both sets of results are compared. As part of this analysis the relationship is between both alcohol and income, a range of socio economic variables is looked at.

The major finding from this analysis is that that alcohol consumption does have an effect on income, with the income of moderate drinkers being the higher compared with both heavy drinkers and abstainers. This paper is organised as follows. The next section compares limited and full information methods of estimation. Section III looks at the empirical model used in the estimation of the effect of alcohol on income. Section IV gives details on the data used and provides the results in terms of the difference in income among the different drinking categories using both limited and full information methods of estimation. Finally section V briefly summarises the major results and the conclusions that can be drawn from them.

2. LIMITED INFORMATION METHODS & FULL INFORMATION METHODS OF ESTIMATION

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). Alcohol status is estimated for each of the three categories of drinkers; non, moderate and heavy drinkers. From this estimation predicted values for the inverse mills ratio can be generated which are then included as an additional variable in the income equations. Both the income equation and the alcohol status equation is hypothesised to depend on a vector of human capital variables and socio demographic characteristics. All the variables in the income status equation are also included in the alcohol status equation, which accounts for income, and there are also two additional variables that are included in the alcohol status equation only. As part of the analysis into the effect of alcohol consumption on incomes, the relationship between these other socio economic variables with both household income and alcohol status is examined. Simultaneous Equations Models depend 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). 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; 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.

2.1 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 Liklihood Method (LIML) are generally the methods used to estimate simultaneous equations consistently while accounting for endogeneity. Much research has been carried out into comparing Limited Information Methods and Full Information Methods of Estimation (Intriligator *et al* 1996; Adepoju and Olaomi, 2009; Puhani, 2000). Findings from Monte Carlo studies have shown that 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. Its computation via a system of equations means that an error in one equation or in one variable will propagate throughout the whole system in the process of estimation (Intriligator *et al* 1996). In addition to this sensitivity to error, the full information estimators particularly FIML, are computationally more complicated than other estimators and hence more costly to use. Furthermore, both FIML and 3SLS require a much larger sample size than the limited information estimators. In their analysis of small sample properties, Adepoju and Olaomi (2009) and Adepoju (2009) find that FIML is a much better estimator in open ended intervals but poor in relation to closed intervals. Adepoju and Olaomi (2009) in their study in relation to small sample properties, argue that it is important to rank estimators on the merit they have when applied to small samples because in practice researcher's usually work with small samples.

In contrast to the full information method approaches, the limited information approach estimates only one equation at a time, 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, hence is not greatly affected by specification errors (Intriligator *et al*, 1996). Furthermore, it is also generally easily and inexpensively computed (Intriligator *et al*, 1996).

Puhani (2000) in his analysis of different research carried out into comparing the Heckman Limited Information Maximum Liklihood Method with the Full Information Maximum Liklihood Method (FIML) 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.

Similarly Enders and Bandalos (2001) find that FIML is unbiased and more efficient than the other methods.

2.2 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 simultaneous equation models, 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 full information estimators such as FIML or 3SLS estimator, since all the information that is available in the system is not used (Judge *et al*, 1988).

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. 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. 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 1 summarises both the limited information and full information methods of estimation.

Table 1. 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 one 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).

2.3 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 and is based on the entire system of equations of simultaneous equation models. 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 equation and certain restrictions on the error structure. The major disadvantage of FIML, however is that it is difficult and expensive to compute, involving the estimation of rather awkward simultaneous nonlinear equations, which usually must be computed via iteration (Greene, 2002).

3. EMPIRICAL MODEL

Much of the original research into alcohol status used a multinomial logit OLS two step model (Hamilton and Hamilton, 1997; Barrett, 2002) which 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). 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). 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 (Kennedy, 2003). Alcohol consumption could be viewed as ordered data and should be estimated as such (Harris *et al*, 2006). 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 (McElvey and Zavoina, 1975). In this study alcohol status is estimated as an ordered probit from which the Inverse Mills Ratio is derived, and included in the income regression.

In the estimation of the effect of drinking on income, drinking is estimated as ordered data using both approaches. In relation to the limited information method, the Heckman two step method is used. In relation to the Full Information method of estimation, the Full Information Maximum Likelihood method is used to 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).

3.1 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. Ignoring the ordinal nature of the variable may result in inaccurate results (Maddala, 1983). Harris *et al* (2006) states that the approach to modelling alcohol consumption should be that the propensity of choosing higher levels of alcohol consumption are mapped in an orderly manner.

$$c_i = \alpha' s_i + \varepsilon_i \tag{1}$$

$$\begin{aligned} c_i &= 1 \quad \text{if} \quad -\infty < c_i \le \mu_1 \\ c_i &= 2 \quad \text{if} \quad \mu_1 < c_i \le \mu_2 \\ c_i &= 3 \quad \text{if} \quad \mu_2 < c_i \le \infty \end{aligned}$$

Where

c	category of drinker
α	unknown vector of parameters
s	independent variables
ε	standard normal shock
μ_J	cutoffs
i	indexes individuals

The amount of alcohol individuals consume is affected by a range of independent variables s. It is assumed that the independent variables s_i and the categorical variables c_i are observed. The alcohol status equation comprises of income by including all the variables in the income equation as well as other variables (Hamilton and Hamilton, 1997; Barrett, 2002).

In the ordered probit selection model it is important that the independent variables in the alcohol status equation contain a variable that is not an independent variable in the income equation. There must be at least one instrument for the selection variable s that has no effect on y except through its effect on c. If all the variables in alcohol status equation are also in the income equation then the identification of the coefficient β_j would be weak (Chiburis and Lokshin, 2007). In this study, 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 are included in the alcohol status equation but not in the income equation.

3.2. Income Equation

Assume the potential income for individual i with drinking status j is given by equation 2. Household Income for each individual are hypothesised to depend upon a vector X_i of human capital variables and sociodemographic characteristics and W_{ij} is observed only if drinking status j is chosen.

$$\ln Y_{ij} = X_i \beta_j + u_{ij} \qquad (2)$$

Where	$\ln Y$	log of income
	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 $j = 1,2,3,$

This specification allows household income returns to individual characteristics to differ by drinking status.

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

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

Where:	y_i	dependent variable income 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).

3.3. Estimation of the effect of Alcohol Consumption on Income using the LIML Method

In the first step the alcohol status equation is estimated by an ordered probit of c on s. 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 $c_i = j$.

$$E\left[y_i \middle| c_i, s_i, x_i\right] = \beta_j' x_i + E\left[u_{ij} \middle| c_i = j, s_i\right] = \beta_j' x_i + p_j \sigma_j \lambda_i' \qquad (4)$$

Where:

	y	dependent variable income
	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	correlation coefficient between the unobservables in the income and selection equations.
	σ	standard deviation of the error term
	λ	selection correction term

3.4. Estimation of the effect of Alcohol Consumption on Income using FIML Method

The aim of this study is to estimate the parameter vectors $\beta_1, \beta_2 ... \beta_J$ in the income equation which will show the impact on income for an individual given particular characteristics with a particular drinking category.

y_i could be missing for certain categories of j, since only one category of j is observed for each individual and the observations are independent, hence the correlations between u_{ij} and u_{ik} for $j \neq k$ cannot be identified (Chiburis and Lokshin, 2007)

Estimating the model using Full Information Maximum Likelihood method 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).

Given the parameters, the likelihood of an observation i in which category j and y_i is observed is

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

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

$$= \frac{1}{\sigma_j}\phi(t_i)\left[\Phi\left(\frac{\alpha' s_i + \rho_j t_i - \mu_j}{\sqrt{1-\rho_j^2}}\right) - \Phi\left(\frac{\alpha' s_i + \rho_j t_i - \mu_{j+1}}{\sqrt{1-\rho_j^2}}\right)\right] \tag{5}$$

Where	y	income
	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
	$t_i \equiv \left(y_i - \beta_j' x_i\right)/\sigma_j$	
	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

$$L_{ij} \equiv \Phi\left(\alpha' s_i - \mu_j\right) - \Phi\left(\alpha' s_i - \mu_{j+1}\right) \tag{6}$$

Taking the logarithm of equation 6 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).

$$L \equiv \sum_{i=1}^{n} \begin{cases} \log L_{ici}^{y} & \text{if } y_i \text{ is observed} \\ \log L_{ici} & \quad\; y_i \text{ is missing} \end{cases} \tag{7}$$

4. DATA AND EMPIRICAL RESULTS

4.1 Data

The data to be used in this research will be taken from the 2007 Slán National Health and Lifestyle Survey which is commissioned by Department of Health and Children. This survey is a cross section of the Irish adult population, aged 18 and over and consists of 10,364 people (62% response rate). The selection is a random sample which is proportionately distributed across counties, locality, gender and urban/rural locations, across age groups and social classes.

The dependent variables used in this study are income and alcohol consumption. Income bands are available for the household's total net income per week. For the purpose of econometric analysis in this paper, the descriptive statistics for income are derived by taking the midpoint of an individuals income category (Barrett, 2002) and for the open upper category, a value of 10% above the lower income limit of the band, was taken (Von Fintel, 2007).

In relation to the drinking status equation drinkers are divided into one of three categories of drinkers; non, moderate and heavy drinkers. Respondents are categorised based on recommendations from the Irish Health Promotion Unit (HSE 2008). Using data from the 2007 Slán 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 ones alcohol consumption in the Slán survey.

The Slán survey includes a large number of socio-demographic characteristics, a number of which are used as explanatory variables and are shown in table two. The drinking status equation contains all the variables that are in the income equation which accounts for the effect of income on alcohol consumption (Hamilton and Hamilton, 1997; Barrett, 2002), along with other variables hypothesised to be unique to the drinking decision.

Table 2. Descriptive Statistics

Variable Definitions	**Drinking Status Equation**	**Income Equation**	**Mean**	**Standard Deviation**
Logincome is the log of household income				
Alcohol Status: Non Drinkers = 1, Moderate Drinker = 2, Heavy Drinkers = 3				
Males = Individuals who are male, 0 = female	X	X	0.427	0.495
Age18-29 = those who are aged is 18 to 29, 0 = otherwise	X	X	0.174	0.379
Age 30-39 = those who are aged is 30 to 39, 0 = otherwise	X	X	0.219	0.414
Age 40-49 = those who are aged is 40 to 49, 0 = otherwise	X	X	0.191	0.393

Age 50-59 = those who are aged is 50 to 59, 0 = otherwise	X	X	0.154	0.361
Age60to69 = those who are aged is 60 to 69, 0 = otherwise *	X	X	0.130	0.336
Age 70plus = those who are aged is 70 plus, 0 = otherwise	X	X	0.132	0.339
Married = Individuals who are married, 0 = otherwise.	X	X	0.506	0.500
Widowed = Individuals who are widowed, 0 = otherwise.	X	X	0.087	0.282
Sep/Div = Individuals who are separated or divorced, 0 = otherwise.	X	X	0.063	0.243
Single/Never Married = Individuals who are single/never married, 0 = otherwise.	X	X	0.280	0.449
Cohabiting = Individuals who are cohabiting, 0 = otherwise*	X	X	0.060	0.240
Edprimary = Individuals who have primary school education only,0 = otherwise *	X	X	0.174	0.379
Educ. Secondary = Individuals who have completed secondary education only, 0=otherwise	X	X	0.440	0.496
Educ. Diploma = Individuals who have a diploma or certificate, 0 = otherwise	X	X	0.185	0.388
Educ. Primary Degree = Individuals who have a primary degree, 0 = otherwise	X	X	0.104	0.306
Educ. Post Graduate = Individuals who have completed a postgraduate /higherdegree, 0 = otherwise	X	X	0.096	0.295
Employee= those whose current employment situation is an employee at work, 0 = otherwise	X	X	0.458	0.498
Selfemployed = those whose current employment situation is self employed or in farming, 0 = otherwise	X	X	0.116	0.320
Disability = those whose current employment situation is unable to work owing to permanent sickness/disability, 0 = otherwise *	X	X	0.040	0.190
State Training/Student = those who are students or on a state training programme, 0= otherwise	X	X	0.037	0.190
Unemployed = those whose current employment situation is unemployed, 0 = otherwise	X	X	0.030	0.169
Homemaker = those whose current employment situation is Homemaker, 0 = otherwise	X	X	0.140	0.347
Retired = those whose current employment situation is wholly retired, 0 = otherwise	X	X	0.170	0.376
Other = those whose current employment situation is classified as other, 0 = otherwise	X	X	0.009	0.097

Race White = those who are white or white Irish, 0 = otherwise	X	X	0.970	0.170
Race Black = those who are black or white Irish, 0 = otherwise	X	X	0.008	0.088
Race Asian = those who are Asian or Asian Irish, 0 = otherwise	X	X	0.008	0.089
Race Other = those who are from another or a mixed background, 0 = otherwise *	X	X	0.005	0.077

Table 2 contd. Descriptive Statistics

Variable Definitions	**Drinking Status Equation**	**Income Equation**	**Mean**	**Standard Deviation**
Num working in Household = No. of people in household working 15 or more hours per week	X	X	1.413	1.277
Opencountry = individuals living in the open country,0= otherwise *	X	X	0.309	0.462
Village = individuals living in a village,0= otherwise	X	X	0.107	0.309
Town = individuals living in a town,0= otherwise	X	X	0.242	0.429
City other than Dublin = individuals living in a city other than Dublin,0= otherwise	X	X	0.105	0.307
Dublin city = individuals living in Dublin city or county,0= otherwise	X	X	0.226	0.418
Health excellent = individuals with excellent health, 0 = otherwise	X	X	0.211	0.408
Health very good = individuals with very good health, 0 = otherwise		X	0.358	[illegible]
Health good = individuals with good health, 0 = otherwise	X	X	0.289	0.453
Health Fair = individuals with fair health, 0 = otherwise	X	X	0.108	0.310
Health Poor = individuals with poor health, 0 = otherwise *	X	X	0.032	0.175
Church activities = individuals who regularly join in the activities of Church or other religious/parish groups, charitable or voluntary organisations, 0= otherwise	X		0.188	0.391
Prevsmokerfivemoreyr = Individuals who used to smoke five years ago or more, 0 = otherwise	X		.139	.346

Note: * indicates base category

4.2 Results

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 status on income accounting for selection bias and endogeneity. Alcohol status is estimated as an ordered probit and income as an OLS regression. The results are discussed below.

4.2.1 Results from the LIML Method of Estimation

In the first step of the two step model, alcohol status is estimated by an ordered probit. The Inverse Mills Ratio is generated which is included as an additional variable in the income regression and hence accounts for selection bias. In the second step separate income regressions are estimated for each of the three categories of drinkers.

(a) Estimation of Alcohol status as an Ordered Probit using the LIML Method

Alcohol Status is estimated by an ordered probit in the first step of the two step model that accounts for potential selection bias of alcohol consumption. The Inverse Mills Ratio is generated which is included as an additional variable in the income regression. Results are set out in Table three.

Table 3. 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
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 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**

Alcohol Status estimated by an ordered probit shows that gender is highly significant and that males are less likely than females to report being a non-drinker and are more likely to be drinkers which is in line with the findings of previous studies (Fillmore 1994; 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 less likely to be non-drinkers and are more likely to be heavier drinkers. Those aged 70 years plus are 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. 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). Those living in a city other than Dublin are the least 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).

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 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 more likely to be non-drinkers which is similar to Hamilton and Hamilton's (1997) study, and those who are previous smokers are less likely to be non-drinkers which is similar to the findings of Barrett (2002).

(b) Estimation of the Income Regressions by Drinking Category using the LIML Method

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 four sets out the results of the three income regressions.

Table 4. Results of the Estimation of Income using LIML Method

	Non Drinkers			**Moderate Drinkers**			**Heavy Drinkers**		
	Coefficien	**t-stat**		**Coefficien**	**t-stat**		**Coefficie**	**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	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	
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	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 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	
Inverse Mills Ratio	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	Moderate Drinkers	Heavy Drinkers
No. of obs = 2127	No. of obs = 5216	No. of obs = 527
F(34, 2092) = 66.54	F(34, 5181) = 138.91	F(33,493) =17.73
Prob > F = 00.00	Prob > F = 00.00	Prob > F = 00.00
R Squared = 0.4816	R Squared = 0.4603	R Squared =.5427
Root MSE = .49446	Root MSE = .49023	Root MSE = .49192

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

Findings from the income regressions show that 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 *et al* (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. All the marital 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. 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.

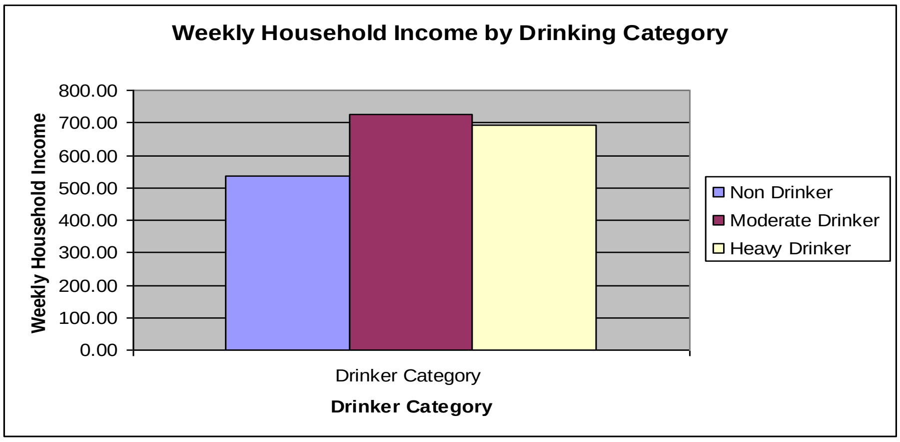
The number of people working in the household variable is very significant and as one would expect is positively correlated to household income. The variable describing those of white race is significant for non-drinkers only, with white non-drinkers likely to earn more. Berger and Leigh (1988) find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

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.

(c) Income Differentials between the three categories of drinkers using LIML

Income regressions are estimated for each of the three categories of drinkers accounting for potential endogeneity bias of alcohol consumption. Many previous studies have had similar findings in that the income of moderate drinkers is highest (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 1.

Figure 1. Weekly Household Income by Drinking Category using the LIML Estimation



(Source: Authors own)

When using the Ordered Probit OLS Two Step model, income for moderate drinkers is higher than non and heavy drinkers and income for non drinkers is substantially less than either moderate or heavy drinkers.

4.2.2 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. Tables setting out the results of the alcohol status equation and the income equation are set out and discussed.

(a) Estimation of Alcohol status as an Ordered Probit using the FIML Method

In estimating the effect of alcohol consumption on income by the Full Information Maximum Likelihood (FIML) method the linear regression model income is estimated with an underlying ordered probit selection rule. Results are set out in Table five.

Table 5. 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
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.

(b) Estimation of the Income Regressions by Drinking Category using the FIML Method

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 Six.

Table 6. 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	-	-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	-	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
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*

Table 6 contd. Results of the Estimation of Income using FIML Method

	Coefficient	Robust Standard Error
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**

The income equations estimated using the Full Information Maximum Likelihood by drinker type accounting for selection bias and endogeneity, 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.

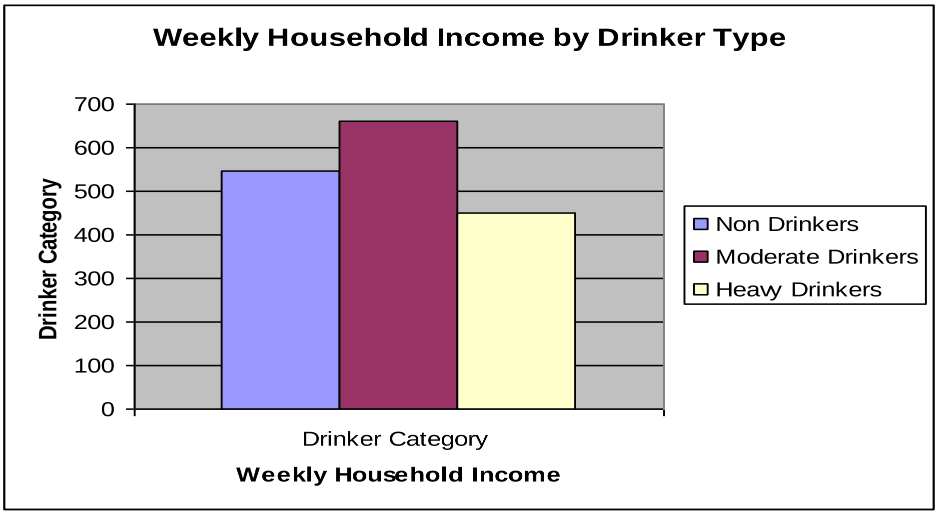
Findings in terms of the effect of independent variables on income using the Full Information Maximum Likelihood Method are also similar to those using the ordered probit two step model.

(c) Income Differentials between the three categories of drinkers using FIML

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, 5,216, are categorised as moderate drinkers; 2,127 non drinkers and 527 as heavy drinkers. The average income for non drinkers is €546.75; for moderate drinkers €660.10 and for heavy drinkers €449.99 per week. Figure 2 sets out the percentage differences between the three categories of drinkers.

Figure 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.

4.3 Post Estimation Diagnostics used in both LIML and FIML

The significance of each of the instruments in both the alcohol status and income regressions is assessed using the z statistic. The Wald test and 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. The null hypothesis that the cut offs are not equal to each other, cut-off1 less cut-off2 = 0, is tested for. The null is rejected in all cases showing that the cut offs are not equal to each other.

5. CONCLUSION

This paper examines the effects of alcohol consumption on household income in Ireland using data from the 2007 Slán survey. The analysis looks at the different estimation methods which could be used in accounting for the endogenous relationship between alcohol status and income. 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. Limited and Full Information Methods are compared and both methods are used to examine the relationship between these two variables.

The major finding from this analysis is that in Ireland, drinking does affect income. 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, income is again higher for moderate drinkers when compared with heavy drinkers, however the difference between the income of moderate drinkers and heavy drinkers is much greater when using the FIML method, with income of heavy drinkers being far less than moderate drinkers and substantially less than non-drinkers. Table seven depicts the findings in terms of the weekly household income by category of drinker for each of the different methods of estimation.

Table 7. Weekly household income by drinking category

	Ordered Probit Two Step Estimation	FIML estimation treating alcohol status as ordered
Non Drinkers	€535.95	€546.75
Moderate Drinkers	€725.45	€660.10
Heavy Drinkers	€694.18	€449.99

(Source: Authors own)

Overall it appears that treating alcohol consumption as ordered data, while results differ between the FIML method and the two-step method; both methods find that moderate drinkers are the better off in terms of household income which is similar to previous findings (Zarkin *et al*, 1998; Lye and Hirschberg, 2004; Hamilton and Hamilton, 1997; Barrett, 2002) despite many of these studies treating alcohol consumption as an unordered variable. Generally previous studies find 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. Confidence Intervals at 95% are constructed from the estimation of alcohol on income using the FIML method of estimation showing 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 eight below.

Table 8. 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)

Despite such evidence, the majority of recent recommendations around alcohol policy in Ireland set out by the Steering Group on National Substance Misuse Strategy in February 2012 (Department of Health, 2012), are population based and no reference is made to the potential benefits of moderate levels of alcohol consumption with the majority of recommendations being around the supply side of alcohol. These results show that similar to findings in relation to other countries, moderate consumers of alcohol in Ireland do enjoy higher household income and by adopting such a population based approach to policy, many individuals in this group will be worse off as a result. Adams and White (2005) state that such an approach to policy brings about an ethical issue in that while it may benefit the majority of individuals there may be a small number of individuals, namely moderate consumers of alcohol, who will be at harm or disadvantaged from such an approach.

BIBLIOGRAPHY

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.

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.

Auld, C. (2005) Smoking, Drinking and Income. *Journal of Human Resources* (2):505-518.

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.

Bau, P., Bau, C., Rosito, G., Manfroi, W. and Fuchs, F. (2007) Alcohol Consumption, Cardiovascular Health and Endothelial Function Markets. *Alcohol* 41(7).

Bau, P., Bau, C., Rosito, G., Manfroi, W. and Fuchs, F. (2007) Alcohol Consumption, Cardiovascular Health and Endothelial Function Markets. *Alcohol* 41(7).

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.

Berry, W. and Feldman, S. (1985). *Multiple Regression in Practice*. Beverly Hills, CA: Sage Publications, Inc.

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.

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. http://www.stata-journal.com/sjpdf.html?articlenum=st0123

Department of Health, Ireland (2012). *Steering Group Report on a National Substance Misuse Strategy*. February 2012. http://health.gov.ie/blog/publications/steering-group-report-on-a-national-substance-misuse-strategy-february-2012/

Di Pietro, G. and Pedace, L. (2008). Changes in the returns to education in Argentina. *Journal of Applied Economics* XI(2): 259-279.

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.

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.

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. and Zarkin, G. (1995). Is moderate alcohol use related to wages? Evidence from four worksites. *Journal of Health Economics* 14: 319-344.

Greene, W. (2002). *Econometric Analysis*. 5th Edition. Upper Saddle River, New Jersey: Practice Hall.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, Mar/Apr72, 80(2): 223-255.

Gujarati, D. (2004). *Basic Econometrics*. McGraw-Hill, Fourth Edition.

Hamilton, V. and Hamilton, B. (1997). Alcohol and Earnings; does drinking yield a wage premium? *Canadian Journal of Economics* 30(1): 135-151.

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.

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.

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.

Intriligator, M., Bodkin, R. and Hsiao, C. (1996). *Econometric Models Techniques and Applications*. 2nd Edition. Upper Saddle River, NJ: Prentice Hall, 1996.

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.

Kennedy, P. (2003). *A guide to econometrics*. MIT Press. Fifth Edition.

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.

Lee, L. (1982). Some Approaches to the Correction of Selectivity Bias. *The Review of Economic Studies*, Vol. 49: 355-372.

Loh, E. (1996) Productivity Differences and the marriage premium for white males. *Journal of Human Resources* 31: 566-89.

Lye, J. and Hirschberg, J. (2004). Alcohol consumption, smoking and wages. *Applied Economics* 36: 1807-1817.

Maddala, G. (1983) *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, New York.

McKelvey, R. and Zavoina, W (1975). A Statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4:103-120.

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.

Mullahy, J. and Sindelar, J. (1996) Employment, unemployment and problem drinking. *Journal of Health Economics* (1996) 409-434.

Pearce, D. (1986). *Dictionary of Modern Economics*. 3rd Edition. London: Macmillan.

Puhani, Patrick (2000). The Heckman Correction for Sample Selection and its Critique. *Journal of Economic Surveys* 14(1): 53-68.

Schoeni, R. (1995). Marital Status and Earnings in Developed Countries. *Journal of Population Economics* 8:351-359.

Sherkat (2004). Religious intermarriage in the United States: trends, patterns, and predictors. *Social Science Research* 33: 606-625.

Su, S. and Yen, S. (2000). A censored system of cigarette and alcohol consumption. *Applied Economics* 32: 729-737.

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**.**

Wonnacott, T. and Wonnacott, R. (1979). *Econometrics.* Wiley and Sons. Second Edition.

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.

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.

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. http://mumford.albany.edu/chinanet/events/past_conferences/shanghai2005/zhanghongliang_en.pdf

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.