Polish Household Consumption of Tobacco and Alcohol: A Censored System

Shengfei Fu
University of Georgia
Department of Ag & Applied Economics
306 Conner Hall, Athens GA 30602
shengfei@uga.edu

and

Wojciech J. Florkowski
University of Georgia
Department of Ag & Applied Economics
1109 Experiment St., 212 Stuckey Building, Griffin, GA 30223
wojciech@uga.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, San Antonio, Texas, February 6-9 2016

Polish Household Consumption of Tobacco and Alcohol: A Censored System

Abstract

The addictive nature of tobacco and alcohol suggests that it is more appropriate to model these consumption as a system. We expand Heckman's sample selection model into a censored system (or multivariate sample selection model) to analyze household tobacco and alcohol consumption.

We use a pooled cross sectional data of 77,043 observations from Polish Household Survey data in the period of 2005 to 2008 and apply full information maximum likelihood estimation. Empirical investigation indicates that the decisions to smoke and drink as well as their expenditure levels are indeed, respectively, positively correlated. We examine the effects of demographic, socio-economic factors and outmigration, a special issue in Poland, on the consumption decisions and expenditure on tobacco and alcohol. Findings provide insights for the reduction and prevention of tobacco and alcohol use.

Keywords: smoke, drink, tobacco, alcohol, multivariate sample selection model, censored system, Poland, depopulation, worker migration.

Polish Household Consumption of Tobacco and Alcohol: A Censored System

1 Introduction

Tobacco use is one of the main risk factors for a number of chronic diseases, including cancer, lung diseases, and cardiovascular diseases. World Health Organization (WHO) has estimated that tobacco use is currently responsible for the death of about six million people across the world each year (WHO 2016a).World organizations and nations are taking measurements to reduce tobacco use (e.g., the World Health Organization Framework Convention on Tobacco Control, WHO FCTC). Similarly, the consumption of alcohol carries a risk of adverse health and social consequences related to its intoxicating, toxic and dependence-producing properties (WHO 2016b).

Tobacco and alcohol-related harm are two of the priority public health challenges facing Europe. In Poland, 29.4% of the population is estimated to use smoked tobacco in 2013 (WHO 2016c). Per capita (15 years and older) alcohol consumption in Poland is steadily increasing during the period 2000-2010 (WHO 2016d). With the growing attention to tobacco- and alcohol-related social and health problems and public-policy campaigns against tobacco and alcohol use, an analysis of the determinants of household tobacco and alcohol consumption remains important. This study takes advantage of a second-hand survey data collected from a household panel by Poland's Main Statistical Office (GUS).

Earlier studies on individual tobacco and alcohol use have identified a variety of demographic and socio-economic factors as consumption determinants, including income, education level, age, gender, region of residency, and employment status (e.g., Blaylock and Blisard 1992; Jones and Labeaga 2003; Yen 2005). Income is one of the most commonly used variables in studies of cigarettes and alcohol (Yen 2005). Individuals with a higher educational attainment level may be more aware of the risks of tobacco and alcohol consumption than those with less education. Urban residency, compared to rural residency, and employment status may reflect different lifestyle and economic wellbeing. Age is relevant as previous study suggest a life-cycle pattern for smoking (Freeth 1998) and such pattern is likely for alcohol drinking. Also, WHO reports on tobacco and alcohol use clearly reveal different patterns for female drinkers/smokers from their male counterparts (WHO 2016c, 2016d).

In addition, household food consumption literature suggests household size and structure also plays a role in household consumption decisions. The presence of children generally is associated with healthy food choices. And the presence of elders may also indicate difference in consumption pattern. This study, therefore, includes household size as explanatory variable. This measurement is further broken down into the number of adults and the presence as well as the numbers of children and elders, respectively. Specifically, the presence of children and elders are assumed to affect the participation decisions of whether to buy tobacco or alcohol, while the numbers of different family members are assumed to affect the consumed amounts.

A special factor in Poland is worker migration and depopulation, especially after Poland's accession to the EU in 2004, coupled with free job market entry to other EU countries. Migration leads to changes in population structure and exposure to different lifestyle and cultural values, which in return contributes to different consumption features. Previous studies focused on the dampening effect of depopulation on the economic growth; however, less attention has been paid to the dietary welfare of people living in the depopulating regions at a micro or household level. This study investigates determinants of household expenditure on tobacco and alcohol, with special attentions paid to the effect of depopulation associated with the domestic migration, a current issue in Poland. Additionally, the study takes into account the effects of migration to other countries. The study applies quantitative methods to generate measurable effects of individual explanatory factors.

The remainder of this study is organized as follows: Section 2 describes the methodology, including economic theory and econometric modeling. Section 3 introduces data source and variable definitions. Section 4 reports estimation results and goodness of fit. Finally, Section 5 concludes with discussion.

2 Modeling Approach

2.1 Economic Theory

A qualitative choice model based on a random utility maximization developed by McFadden (1980) provides the theoretical foundation for model specification. Our empirical model is derived by extending the discrete choice model (Pudney, 1989). A household maximizes the random utility function subject to a budget constraint. The household random utility function is given by:

$$V(y,q;\boldsymbol{w}) = d \cdot U(y,q;\boldsymbol{w}) + (1-d) \cdot U^{*}(q;\boldsymbol{w}) \qquad (1)$$

where U is the utility for buyers and U* for non-purchasers, y is the quantity of a commodity with price p, q is a composite commodity for other goods with price normalized to 1, w is a vector of demographic variables, and d is a binary variable that equals one if the household buys the commodity and zero otherwise.

Assume the outcome for tobacco and alcohol consumption, the participation decision, is generated by a binary choice structure:

$$d = 1 \; if \; \boldsymbol{z}'\boldsymbol{\alpha} + u > 0 \qquad (2),$$

$$= 0 \; if \; \boldsymbol{z}'\boldsymbol{\alpha} + u \leq 0$$

where **z** and **α** are vectors of variables and parameters affecting binary purchase decision, and u

is a random error. In cross-sectional demand modeling, zero observations are often treated as the result of economic non-consumption (i.e., corner solution). In some cases, however, zero purchase might be caused by behavioral factors other than prices. Because y does not enter the purchasers' utility function $U^*(q; \boldsymbol{w})$ as described in equation (1) and p > 0, the optimal level is y = 0 for a non-consumer. This optimal zero purchase could be corner solution or the result of opting out of the market. For a buyer, the optimal level of y results from a solution to the constrained utility maximization problem with a fixed budget I:

$$\max_{y,q} \{U(y, q; \boldsymbol{w}) | \, py + q = I\} \qquad (3)$$

Assume that the utility function $U(y, q; \boldsymbol{w})$ is regular strictly quasi-concave and has positive first partial derivatives with respect to y and q. Furthermore, assume an interior solution for y and q. Then, solving Equation (5) yields the notional (latent) demand for milk, y*. Denote as $\boldsymbol{x}$ the vector of income and demographic variables (with corresponding parameter vector **β**) affecting the quantity demanded.

Further, assume latent quantity y* is expressed by the lognormal distribution, which accommodates right-skewness and ensures positive purchase amount:

$$y^* = \boldsymbol{x}'\boldsymbol{\beta} + v \qquad (4)$$

where $\boldsymbol{x}$ and $\boldsymbol{\beta}$ are variables and corresponding parameters affecting quantity decision and v is a random error.

2.2 Econometric Modeling

The occurrence of excessive percentage of zeros in micro-data sets mandates a proper treatment for the censoring of the dependent variables. Such zero observations may occur for three main reasons: infrequency of purchase in survey data with short recording periods, some individuals are out of market for various reasons, and economic non-consumption under current price and individual income.

The particular interpretation given to zero observations can have a crucial bearing on the estimation approach adopted (Madden 2008). Various modeling structures are proposed in existing literature to accommodate the censored data, including the Tobit model, hurdle model, two-part model, and Heckman's sample selection model. More recent development features a sample selection system or censored system in the sense of multiple-goods decisions, which allows correlation within and/or across participation decisions and intensity decisions among multiple goods. Such modeling feature is important for studying the consumption of closely related products, such as the consumption of tobacco and alcohol. A number of censored-system estimation procedures have existed in the literature. These include maximum-likelihood estimators of Amemiya (1974), Wales and Woodland (1983), and Lee and Pitt (1986), and two-step estimators of Heien and Wessells (1990), Shonkwiler and Yen (1999), and Perali and Chavas (2000), as well as an extended full system approach of Stewart and Yen (2004), and Yen (2005).

Due to the additive nature as well as observed culture of drinking and smoking, it is more appropriate to model their consumption as a system (Pierani and Tiezzi 2009). This study uses a censored system which specifies a set of level equations, each exclusively subject to a binary selection rule, and which accommodates error correlations among all equations.

To facilitate the presentation of models, re-write the binary choice rules and level equations, described by Equations (2) and (4), respectively, in a system. Then, each outcome variable y_i is governed by a binary selection rule of whether to consume as follows (observation subscription omitted):

$$\log(\mathrm{y_i}) = \mathbf{x}'\boldsymbol{\beta}_\mathbf{i} + \mathrm{v_i} \quad \text{if } \mathbf{z}'\boldsymbol{\alpha}_\mathbf{i} + \mathrm{u_i} > 0 \qquad (5)$$

$$\mathrm{y_i} = 0 \quad \text{if } \mathbf{z}'\boldsymbol{\alpha}_\mathbf{i} + \mathrm{u_i} \leq 0\ , \quad \mathrm{i} = 1{,}2$$

where **z** and **x** are vectors affecting binary purchase decision and level decision, respectively; $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of parameters; $\boldsymbol{u}_\mathbf{i}$ and $\boldsymbol{v}_{\boldsymbol{i}}$ are random error in the participation and level equation, respectively.

To facilitate presentation of the log likelihood functions, define diagonal $\mathrm{S} = \mathrm{diag}(\sigma_1, \sigma_2)$ as standard deviation of $\boldsymbol{v}$. Let $\mathbf{R}_{\boldsymbol{uu}} = [\rho_{ij}^{uu}]$, $\mathbf{R}_{\mathbf{vu}} = [\rho_{ij}^{vu}]$, and $\mathbf{R}_{\boldsymbol{vv}} = [\rho_{ij}^{vv}]$ be 2 x 2 correlation matrices among elements of $\boldsymbol{u}$ and $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{u}$, and $\boldsymbol{v}$ and $\boldsymbol{v}$, respectively.

The censored system, which allows error correlations among all equations, assumes the concatenated error vector $[\boldsymbol{u}', \boldsymbol{v}']' \equiv [u_1, u_2, v_1, v_2]'$ is distribted as 4-variate normal with zero mean and covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad (6)$$

where $\Sigma_{11} = \mathrm{E}(\boldsymbol{uu}') = \mathbf{R}_{uu}$, $\Sigma_{21} = \Sigma_{12}' = \mathrm{E}(\boldsymbol{vu}') = \mathbf{S}'\mathbf{R}_{vu}$, and $\Sigma_{22} = \mathrm{E}(\boldsymbol{vv}') = \mathbf{S}'\mathbf{R}_{vv}\mathbf{S}$**.**

Define vectors $\boldsymbol{r} \equiv [r_1, r_2]' \equiv [z_1'\alpha_1, z_2'\alpha_2]'$ and$\boldsymbol{v} \equiv [\log(y_i) - \boldsymbol{x}'\boldsymbol{\beta}_{\boldsymbol{i}}]$. Let $\phi(\boldsymbol{v})$ be the marginal probability density function (pdf) of $\boldsymbol{v} \sim N(0, \Sigma_{22})$ and $\phi(\boldsymbol{u}|\boldsymbol{v})$ be the conditional pdf of$\boldsymbol{u}|\boldsymbol{v} \sim N(\mu_{\boldsymbol{u}|\boldsymbol{v}}, \Sigma_{\boldsymbol{u}|\boldsymbol{v}})$, where $\mu_{\boldsymbol{u}|\boldsymbol{v}} = \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{v}$ and$\Sigma_{\boldsymbol{u}|\boldsymbol{v}} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Then, the likelhood contribution for the positive regime, where both dependent variables are positive, is given by:

$$L_1 = \phi(\boldsymbol{v}) \prod_{j=1}^{2} y_j^{-1} \int_{\boldsymbol{u}>-\boldsymbol{r}}^{+\infty} \phi(\boldsymbol{u}|\boldsymbol{v}) d\boldsymbol{u} = g(\boldsymbol{v}) \prod_{j=1}^{2} y_j^{-1} \Phi_2(\boldsymbol{r} + \mu_{\boldsymbol{u}|\boldsymbol{v}}; \Sigma_{\boldsymbol{u}|\boldsymbol{v}}) \qquad (7)$$

where $\prod_{j=1}^{2} y_j^{-1}$ is the Jacobian of the transformation from $[v_1, v_2]'$ to $[y_1, y_2]'$ and $\Phi_2(\boldsymbol{r} + \mu_{\boldsymbol{u}|\boldsymbol{v}}; \Sigma_{\boldsymbol{u}|\boldsymbol{v}})$ is the bivariate normal cumulative distribution function (cdf) with zero mean, covariance matrix$\Sigma_{\boldsymbol{u}|\boldsymbol{v}}$, and finite upper integration limits$\boldsymbol{r} + \mu_{\boldsymbol{u}|\boldsymbol{v}}$.

The second regime is one in which the values of both variables are zeros (when $z'\alpha_i + u_i \leq 0$, $i = 1, 2$). The likelihood contribution is identical to that of an all-zero regime in the bivariate probit:

$$L_2 = \int_{-\infty}^{u \leq -r} \phi(\boldsymbol{u}, \Sigma_{11}) d\boldsymbol{u} = \Phi_2(-\boldsymbol{r}; \Sigma_{\mathbf{11}}) \qquad (8)$$

where $\phi(\boldsymbol{u}, \Sigma_{11})$ is the marginal pdf of $\boldsymbol{u} \sim N(0, \Sigma_{\mathbf{11}})$. Specifically, $\phi(\boldsymbol{u}, \Sigma_{11}) = (2\pi)^{-1} |\Sigma_{11}|^{-1/2} e^{-\frac{1}{2}\boldsymbol{u}' \boldsymbol{\Sigma_{11}}^{-1} \boldsymbol{u}}$.

For mixed regime, without loss of generality, denote u_i as the error term associated with the non-censored variable and u_j associated with the zero-valued variable. A mixed regime is characterized by:

$$z'\alpha_i + u_i > 0 \qquad \log(y_i) = x'\beta_i + v_i \qquad (9)$$

$$z'\alpha_j + u_j \leq 0 \qquad y_j = 0.$$

Let $\tilde{v} \equiv v_i$ Then $[\boldsymbol{u}', \tilde{v}]'$ is 3-variate normal with zero mean and covariance matrix $\tilde{\Sigma}$, where $\tilde{\Sigma}$ is a 3x3 sub-matrix containing the first three rows and coluns of the error covariance matrix Σ in Equation (6). Partition $\tilde{\Sigma}$ at the third row and column such that $\tilde{\Sigma} = \begin{bmatrix} \Sigma_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{bmatrix}$.

Let $\phi(\tilde{v})$ be the marginal pdf of $\tilde{v} \sim N(0, \tilde{\Sigma}_{22})$ and $\phi(\boldsymbol{u}|\tilde{\boldsymbol{v}})$ be the conditional pdf of $\boldsymbol{u}|\tilde{\boldsymbol{v}} \sim N\ (\mu_{\boldsymbol{u}|\tilde{v}}, \Sigma_{\boldsymbol{u}|\tilde{v}})$, where $\mu_{\boldsymbol{u}|\tilde{v}} = \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\boldsymbol{v}}$ and $\Sigma_{\boldsymbol{u}|\tilde{v}} = \Sigma_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}$. Then the likelihood contribution for this regime is:

$$L_3 = y_i^{-1}\phi(\tilde{v}) \int_{u_i > -r_i}^{+\infty} \int_{-\infty}^{u_j \leq -r_j} \phi(u_1, u_2|\tilde{v}) du_2 du_1 = y_i^{-1}\phi(v_i)\Phi_2\big(\boldsymbol{D}\big(\boldsymbol{r} + \mu_{\boldsymbol{u}|\tilde{v}}\big); \boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{u}|\tilde{v}}\boldsymbol{D}\big) \quad (10)$$

where $\boldsymbol{D} = diag(2d_1 - 1, 2d_2 - 1)$, $d_i = 1\ if\ z\alpha_i + u_i > 0$. The sample likelihood function for the censored system is the product of the likelihood contributions L_1, L_2, or L_3 across observations, depending on the regime for each observation.

2.3 Marginal Effects[1]

Economically meaningful measure, marginal effects, are calculated based on conditional means for the joint distribution. The probability of purchase is given by:

$\Pr(y_i > 0) = \Phi(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i})$. (11)

Elasticity for continuous explanatory variable is defined as the change in probability of purchase, corresponding to a one-unit change in z_j. The marginal effects for indicator explanantory variables are the discrete change in purchase probabilties obtained in Equation (11) when the explanatory variable takes value of one versus zero:

$$m_i^{Prob} = \begin{cases} \frac{d\Pr(y_i>0)}{dz_j} = \phi(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i}) \cdot \alpha_{ij}, & if\ z_j\ continuous \\ \Phi\left(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i} \middle| \mathbf{z}_\mathbf{j} = 1\right) - \Phi\left(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i} \middle| \mathbf{z}_\mathbf{j} = 0\right), & if\ z_j\ binary \end{cases} \quad (12)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. The conditional mean of expenditure y_i is (Rosiniski and Yen, 2004):

$$E(y_i|y_i > 0) = \exp\left(\boldsymbol{x}'\boldsymbol{\beta}_{\boldsymbol{i}} + \frac{\sigma_i^2}{2}\right) \cdot \Phi(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}} + \rho_{ii}^{uv}\sigma_i^2)/\Phi(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}}). \quad (13)$$

Multiplying Equations (12) and (13) gets the unconditional mean of y_i:

$$E(y_i) = \exp\left(\boldsymbol{x}'\boldsymbol{\beta}_{\boldsymbol{i}} + \frac{\sigma_i^2}{2}\right) \cdot \Phi(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}} + \rho_{ii}^{vu}\sigma_i^2). \quad (14)$$

Let's consider a variable that enters the level equation as well as the participation equation. In this case, when deriving the semi-elasticity of conditional expected value of y_i with respect to x_j, we have to consider that vector **z** also contains x_j.

Semi-elasticity (discrete change) of the conditional mean is obtained by differentiating (differencing) Equation (14) with respect to variable x_j:

[1] Estimated marginal effects are not reported in the current paper.

$$\mathrm{m}_{\mathrm{i}}^{\mathrm{c}} = \begin{cases} \frac{\mathrm{d}\ \ln E(y_i|y_i > 0)}{\mathrm{dx_j}} = \beta_{\mathrm{ij}} + [\lambda(\boldsymbol{z}'\boldsymbol{\alpha_i} + \rho_{ii}^{vu}\sigma_i) - \lambda(\boldsymbol{z}'\boldsymbol{\alpha_i})]\alpha_{\mathrm{ij}}\ ,\ if\ x_j\ continuous \\ \Delta\ \ln E(y_i|y_i > 0) = \beta_{\mathrm{ij}} + \Delta[\lambda(\boldsymbol{z}'\boldsymbol{\alpha_i} + \rho_{ii}^{vu}\sigma_i) - \lambda(\boldsymbol{z}'\boldsymbol{\alpha_i})]\ ,\ if\ x_j\ binary \end{cases} \quad (15)$$

where α_{ij} and β_{ij} are the parameters of x_j in the participation equation and level equation for dairy product i, respectively; $\Delta[\cdot]$indicates the difference of its argument when x_j takes value of one versus zero. And, the inverse Mill's ratio is $\lambda(\boldsymbol{z}'\boldsymbol{\alpha_i}) \equiv \frac{\phi(\boldsymbol{z}'\boldsymbol{\alpha_i})}{\Phi(\boldsymbol{z}'\boldsymbol{\alpha_i})}$.

Semi-elasticity (discrete change) of the unconditional mean with respect to x_j that enters both equations is obtained by differentiating (differencing) Equation (15):

$$\mathrm{m}_{\mathrm{i}}^{\mathrm{u}} = \begin{cases} \frac{\mathrm{d}\ \ln E(y_i)}{\mathrm{dx_j}} = \beta_{\mathrm{ij}} +\ \lambda(\boldsymbol{z}'\boldsymbol{\alpha_i} + \rho_{ii}^{vu}\sigma_i)\alpha_{ij},\ if\ x_j\ is\ continuous \\ \Delta\ \ln E(y_i) = \beta_{\mathrm{ij}} + \Delta[\lambda(\boldsymbol{z}'\boldsymbol{\alpha_i} + \rho_{ii}^{vu}\sigma_i)],\ if\ x_j\ binary. \end{cases} \quad (16)$$

For variables that enter the level equation only, the marginal effects for conditional and unconditional mean under both models are its parameter β_{ij} only.

Individual elasticity or discrete change is averaged over the whole sample to obtain the average marginal effect. Asymptotic standard errors for the average marginal effect estimates are obtained using the delta method (Spanos, 1999).

3 Data and Variable Selection

The data are from the Polish household panel of about 20,000 households annually surveyed by Poland's National Statistics Office (GUS). Despite the attempted panel structure of the survey, fewer than 36% of the households were observed for more than one year. The study uses a pooled cross-sectional sample of 77,043 observations with non-missing values for the period of 2005-2008.

The dependent variables are expenditures in the month preceding survey on tobacco and alcohol. Positive expenditures are logarithm transformed to mitigate deviation from normality and potential heteroskedasticity.

Two variables are reported as measure of depopulation. First, net domestic migration measures the net outflow of population from a region to other regions within Poland. Second, net international migration measures the net outflow of population from a region to other, typically EU countries after Poland's accession to EU in 2004.

Other demographic and socio-economic factors include: household head's gender, age, education level, marital status and employment stability, household location, monthly income, and the numbers of children (age 0-18), adults (age19-60) and elders (age >60). Binary variables, the presence of children and the presence of elders are induced from the numbers of these family members.

Table 1 presents summary of statistics of sample variables. Rural residents account for 37.5 % of all observed households (Village=1). Household income in the month preceding survey averages at 2,781 Polish Zloty (PLN). Nearly three out of five (59.3%) households are headed by male members. And, 67.4% of household heads are married. The proportion of household heads with secondary or higher education is 40.7%. The average household head's age is 51.1 years. In term of employment stability, 26.6% household heads are permanently employed or contract employees. The average household size is 2.98 family members, with the average numbers of children (age 0-18 years), adults (age 19-60 years), and elders (above 60 years of age) broken down into 0.72, 1.80 and 0.45 per household, respectively. Households with the presence of children and elders account for 42.1% and 33.7%, respectively. On average, net migration inflow from a Polish region to another domestic region averages at 1,352 persons over all 16 administrative regions of Poland. And an outflow from a Polish region to a foreign country averages at 1,565 persons. The proportions of households observed in each year are fairly balanced, with 25.1% in 2005 and 2006, and 24.9% in 2007 and 2008, respectively. Lastly, the percent of households who bought tobacco and alcohol in the month preceding survey is 36.3% and 56.2%, respectively. This paper loosely refers positive expenditure on tobacco and alcohol as smoking and drinking, respectively. Average spending on tobacco and alcohol are PLN41.0 and PLN28.5, respectively. Conditional on purchase, households on average spend PLN112.75 per month on tobacco and PLN50.63 per month on alcohol (figures not reported in Table 1).

4 Results

For the censored system (Table 2), parameter estimates are obtained by maximum likelihood estimation. The left panel of Table 2 reports parameter estimates for the participation decisions and the right panel reports on the level decisions. Additionally, Table 2 reports estimated error correlation coefficients among all equations.

The correlation coefficients for the participation and level equations, respectively, are estimated to be positive and statistically significant at 5% level. The correlation coefficient between the decisions to smoke and drink is estimated as high as 0.225 with a *p*-value lower than 1%. The error correlation between the expenditures on tobacco and alcohol is 0.086 (*p*-value <1%). This positive correlation between the behaviors of smoking and drinking are probably due to their addictive nature. The non-zero correlation between the decision to smoke (drink) and the expenditure on tobacco (alcohol) indicates the presence of sample selection. The correlation coefficients across the decision to smoke and the expenditure on alcohol as well as the decision to drink and the expenditure on tobacco are statistically different from zero, too. These results validate the necessity of a system approach.

Higher income is associated with higher probability of drinking, but the effect on the decision to smoke has not been confirmed. A decision to smoke may be made early in life when income as compared to tobacco price is not a major constraint, especially because cigarettes are offered in a wide price range in Poland. Once the smoking habit has become established, the higher expenditure on tobacco in households with higher incomes reflects ability to purchase better quality and quantity. Income does have a statistically significant effect on the probability of increasing the expenditure on alcohol.

Residency in the rural area is associated with higher propensity of drinking and smoking. However, rural residents’ expenditure on tobacco and alcohol are lower than those of their urban counterparts. The confirmed similarities and differences reflect the similar preferences, but likely difference in lifestyle between rural and urban residents. An earlier study found that there were differences in alcohol and tobacco demand between rural and urban residents (Florkowski and McNamara, 1992).

Households headed by males are associated with higher probabilities of drinking and smoking. However, the expenditures of their households on either good are slightly lower than those of households headed by females. In contrast, household with married household head are less likely to smoke and drink, indicating possibly markedly different lifestyle. Higher household head’s education level decreases the probability of tobacco purchase. This result is consistent with the observations from other countries in Europe, where those with more education are less likely to smoke. However, having received more formal education is associated with the probability of alcohol purchase and may be a result of the changing composition of alcohol consumption in Poland, where more beer and wine consumption has grown relative to spirits in recent years. Conditional on purchase, household heads with higher education level, on average, spend less on both products. This result is consistent with the significance of education in forming healthy consumption choices.

Families with older household heads are more likely to smoke, but less likely to drink.

The result reflects decreasing smoking rates among the young consumers as compared to their parents or grandparents generation. Also, the expenditures of households headed by older consumers are lower on both products. Household head's status of employment does not play a statistically significant role either on the participation or level decisions. This is interesting because some earlier studies reported mixed effects of employment on alcohol and tobacco use (for example, Bilgic and Yen, 2015).

Larger sized household are more likely to buy tobacco but less likely to buy alcohol. A larger family in Poland may consist of multiple generations and more adults, who learn the same habit of smoking, sometimes by easy access to cigarettes. In case of alcohol, larger families tend to spend more than smaller households likely because a larger number of household members also implies more often entertaining.

The presence of elders is associated with higher probability of smoking than the absence of elders. Smoking peaked in Poland a couple of decades ago and older individuals picked the smoking habit much earlier. However, the presence of elderly lowered propensity to drinking. This variable shows a similar effect to the influence of the household head's age on the decision to purchase tobacco and alcohol.

A household with children is less likely to buy alcohol, but interestingly, they are associated with higher probabilities of tobacco purchase. This result seems to contrast to food demand literature where households with children usually make healthier food choices. But the consumption of tobacco and alcohol is different from typical food consumption, for example milk. People might form the habit of drinking or smoking before they have children and this habit persists because of its addictive nature.

Outmigration, a special issue in Poland, because it has intensified as a result of search for job opportunities as the domestic economy adapted its structure to market-driven resource allocation and outside opportunities resulting from the EU accession in 2004. Outmigration measured by domestic and international net outflow, has somewhat mixed effects. Both domestic and international outmigration is associated with lower probability of smoking. Migrating workers are usually young and better educated, in seeking of employment with higher payment. They generally migrate to regions or countries that are economically better developed than Poland. Higher educational attainment and younger age are both associated with lower probability to smoke, while migrants to better developed areas in Poland or EU quite possibly become exposed to different lifestyle and cultural values. The latter are communicated back to their families staying behind affecting the decision to smoke. However, once the migrants decide to smoke and drink alcohol, both domestic and international migrations are associated with higher expenditure on tobacco and alcohol. A large number of migrants to EU countries end up in countries with traditionally much higher beer consumption, for example Germany, Ireland, the Netherlands, than Poland, or wine consumption, for example, France, Italy or Spain. Migrants likely absorb local lifestyle and consumption habits including drinking, which they project to families left behind.

5 Conclusions

Harmful use of tobacco and alcohol is one of the main risk factors for adverse health and social consequences. With the growing attention to tobacco and alcohol-related social and health problems and public-policy campaigns against tobacco and alcohol use, an analysis of the determinants of household tobacco and alcohol consumption remains important. This study takes advantage of household survey data collected by Poland's Main Statistical Office that is not publicly available. The empirical investigation applies a censored system. This multivariate sample selection model addresses the censoring feature of the survey data. It also allows error correlation among all equations to consider sample selection and the possible correlation between tobacco and alcohol use due to their addictive nature.

Our empirical model uses three categories of explanatory variables. Household features include household income, location (rural vs. urban residency), household size and structure. Household head characteristics are age, gender, education level, marital status, and employment status. Lastly, worker outmigration (both domestic and foreign), a special issue in Poland, is investigated.

The empirical estimation indicates that the decisions to smoke and drink and their expenditure levels, respectively, are indeed positively correlated. In the case of tobacco use, rural residency, older household heads, larger household, and the presence of elders are associated with higher probability of tobacco purchase. Married household heads, higher education level, and higher outmigration, both domestically and internationally, are less likely to be associated with buying tobacco. The presence of children, unfortunately, does not play a role in reducing the likelihood of tobacco use and potentially exposes children to second-hand tobacco smoke.

In the case of alcohol purchase, rural residency, higher income, male household head, higher education level, and higher domestic outmigration positively affects the likelihood of drinking. Meanwhile, individuals from larger household size, the presence of children and elders, respectively, and married household heads are less likely to buy alcohol.

The effects of most explanatory variables on expenditures are similar in either equation. Rural residency, male and older household head negatively affects the amount of purchase. Higher income, married household head and outmigration are associated with higher expenditure on tobacco and alcohol. The effects of education and the numbers of adults and elders are mixed across tobacco and alcohol expenditure. Higher education level is associated with higher expenditure on tobacco, but lower spending on alcohol. The numbers of adults and elders in a family are positively associated with alcohol expenditure. Both outmigration measures are associated with lower likelihood of smoking. Households in regions with higher domestic outmigration are more likely to buy alcohol. Domestic outmigration does not have a statistically significant effect on alcohol purchase. In regard of expenditure levels, both outmigration measures are associated with higher expenditure. This might reflect possible changes in lifestyle and, thus, changes in consumption pattern due to exposure to different lifestyle and culture. Overall, this study's findings reveal determinants of household consumption of tobacco and alcohol identifying household features that are likely to be associated with large consumed amounts. Such households or individuals from such households are a potential target for policy aiming at reduction of harmful effects of alcohol and tobacco consumption and interventions.

References

Amemiya, T. 1985. Advanced Econometrics. Cambridge: Harvard University Press.

Bettman, J.R. 1979. "Memory Factors in Consumer Choice: A Review." *Journal of Marketing* 43(Spring 1979): 37-53.

Bilgic, A. and S. T. Yen. 2015. "Household Alcohol and Tobacco Expenditures in turkey: A Sample-selection System Approach." *Contemporary Economic Policy* 33(3): 571-585.

Blaylock, J.R., and W.N. Blisard. 1992. "U.S. Cigarette Consumption: The Case of Low-Income Women." *American Journal of Agricultural Economics* 74(1992): 698–705.

Florkowski, W. J. and K. T. McNamara. 1992. "Policy Implications of Alcohol and Tobacco Consumption in Poland." *Journal of Policy Modeling* 14(1): 93-98.

Freeth, S. *Smoking-Related Behaviour and Attitudes 1997:A Report on Research Using the Omnibus Survey Produced on Behalf of the Department of Health.* London: Office for National Statistics, 1998.

Heien, D. and C.R. Wessells. 1990. "Demand systems Estimation with Microdata: A censored Regression Approach." *Journal of Business & Economic Statistics* 8(3): 365-371.

Jones,A.M.,and J.M.Labeaga.2003. "Individual Heterogeneity and Censoring in Panel Data Estimates of Tobacco Expenditure." *Journal of Applied Econometrics* 18(2003):157–77.

Madden, D. 2008. "Sample selection versus two-part models revisited: the case of female smoking and drinking." *Journal of Health Economics* 27 (2008): 300-307.

McFadden, D. 1980. "Econometric Models for Probabilistic Choice among Products." *Journal of Business* 53(3) Part 2: Interfaces Between Marketing and Economics 13-29.

Perali, F., and J.P. Chavas. 2000. "Estimation of Censored Demand Equations from Large Cross-Section Data." *American Journal of Agricultural Economics* 82:1022–37.

Pierani, P. and S. Tiezzi. 2009. "Addiction and Interaction between Alcohol and Tobacco Consumption." *Empirical Economics* 37(1): 1-23.

Pudney, S. 1989. "Modelling Individual Choice: The Econometrics of Corners, Kinks, and Holes." Cambridge, UK: Blackwell Publishers.

Rosinski, J., and S.T. Yen. 2004. "A Note on the Conditional Moments of Limited Dependent Variable Models with a Transformed Dependent Variable." Unpublished, Dept. Agr. Econ., The University of Tennessee, Knoxville, 2004.

Shonkwiler, J.S., and S.T. Yen. 1999. "Two-Step Estimation of a Censored System of Equations." *American Journal of Agricultural Economics* 81:972–82.

Spanos, A. 1999. Probability Theory and Statistical Inference: Econometric Modeling with Observational Data. Cambridge, UK: Cambridge University Press, 1999.

Stewart, H. and S.T. Yen. 2004. "Changing household characteristics and the away-from-home food market: a censored equation system approach." *Food Policy* 29(6): 643–658.

WHO. 2016a. *Health topics: Tobacco*. Online available at http://www.who.int/topics/tobacco/en/. Accessed Jan 2016.

WHO. 2016b. *Health topics: Alcohol*. Online available at http://www.who.int/topics/alcohol_drinking/en/ . Accessed Jan 2016.

WHO. 2016c. *WHO Report on the Global Tobacco Epidemic, 2015 – Country profile: Poland.* http://www.who.int/tobacco/surveillance/policy/country_profile/pol.pdf?ua=1. Accessed Jan 2016.

WHO. 2016d. *WHO Global status report on alcohol and health 2014 – country profile: Poland.* Online available at http://www.who.int/substance_abuse/publications/global_alcohol_report/profiles/pol.pdf?ua=1. Accessed Jan 2016.

Yen, S. 2005. "A Multivariate Sample-selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations." *American Journal of Agricultural Economics* 87(2) (May, 2005): 453-466.

Table 1. Summary of Descriptive Statistics of Sample Variables

Variable	Description/Unit	Mean	Std Dev
	Dependent Variables		
Smoke	1, if a household buys tobacco, 0 otherwise	0.363	0.481
Drink	1, if a household buys alcohol, 0 otherwise	0.562	0.496
Tobacco	Expenditure on tobacco in the month preceding survey, in PLN	40.984	78.917
Alcohol	Expenditure on alcohol in the month preceding survey, in PLN	28.470	60.472
	Demographic, Socio-Economic Factors / Explanatory Variables		
Village	1, if a household residents in village, 0 otherwise	0.375	0.484
Income	Household income in the month preceding survey, in 1000 Polish Zloty (PLN)	2.781	2.205
Male	1, if the household head is male, 0 otherwise	0.593	0.491
Married	if the household head is married, 0 otherwise	0.674	0.469
HighEduc	1, if the household head has secondary or higher education, 0 otherwise	0.407	0.491
Age	Household head's age, in years	51.146	15.210
Employed	1 if household head is permanently employed or contract employee, 0 otherwise	0.266	0.442
Hhsize	Number of family members in a household	2.981	1.531
NKid	Number of children (under 18)	0.723	1.040
N1960	Number of adults 60 or under 60 years old	1.804	1.191
N60above	Number of elders above 60	0.453	0.696
DKid	1 if children are present in a household, 0 otherwise	0.421	0.494
DElder	1 if elders (above 60) are present in a household, 0 otherwise	0.337	0.473
OUTD	Net migration domestically to other regions in Poland, in 1000	-1.352	5.714
OUTF	Net migration international to other countries, in 1000	1.565	2.108
YR05	Baseline, 1 if observed in 2005, 0 otherwise	0.251	0.434
YR06	1 if observed in 2006, 0 otherwise	0.251	0.434
YR07	1 if observed in 2007, 0 otherwise	0.249	0.432
YR08	1 if observed in 2008, 0 otherwise	0.249	0.432

Note: N=77,043

Table 2. Maximum-likelihood Estimates for Censored System of Tobacco and Alcohol Consumption

	Binary Decision Of Smoking	**Binary Decision of Drinking**		**Expenditure on Tobacco**	**Expenditure on Alcohol**
	Coeff. (Std. Err.)	Coeff. (Std. Err.)		Coeff. (Std. Err.)	Coeff. (Std. Err.)
Constant	-0.563(0.015)**	-0.379(0.022)**	Constant	5.779(0.038)**	4.230(0.041)**
Village	0.053(0.010)**	0.298(0.010)**	Village	-0.029(0.018)*	-0.018(0.015)
Income	-0.003(0.003)	0.138(0.012)**	Income	0.049(0.007)**	0.051(0.008)**
Male	0.0003(0.001)	0.003(0.001)**	Male	-0.006(0.002)**	-0.005(0.001)**
Married	-0.007(0.003)**	-0.006(0.003)**	Married	0.003(0.004)	0.013(0.003)**
HighEduc	-0.075(0.012)**	0.135(0.014)**	HighEduc	0.174(0.021)**	-0.111(0.018)**
Age	0.218(0.011)**	-0.035(0.011)**	Age	-0.207(0.018)**	-0.108(0.015)**
Employed	-0.004(0.013)	0.009(0.014)	Employed	0.021(0.023)	-0.005(0.019)
Hhsize	0.472(0.004)**	-0.789(0.052)**	N1960	-0.009(0.011)	0.089(0.012)**
DKid	0.405(0.012)**	-0.811(0.055)**	N60above	-0.031(0.012)**	0.112(0.013)**
DElder	0.384(0.005)**	-0.874(0.052)**			
OutD	-0.034(0.013)**	0.021(0.013)*	OutD	0.119(0.022)**	0.044(0.017)**
OutF	-0.056(0.013)**	0.007(0.014)	OutF	0.165(0.023)**	0.048(0.018)**
YR06	-0.141(0.014)**	0.045(0.016)**	YR06	0.147(0.022)**	-0.131(0.018)**
YR07	-0.211(0.018)**	-0.216(0.022)**	YR07	0.329(0.02)**	0.119(0.017)**
YR08	-0.328(0.001)**	0.801(0.052)**	YR08	-0.107(0.008)**	0.003(0.007)

Correlation Coefficient Estimates

	Coeff. (Std. Err.)		Coeff. (Std. Err.)
Rho.Smoke.Drink	0.225(0.009)**	Rho.Smoke.Tobacco	-0.951(0.132)**
Rho.Tobacco.Alcohol	0.086(0.009)**	Rho.Smoke.Alcohol	-0.033(0.009)**
Sigma.Tobacco	1.605(0.010)**	Rho.Drink.Tobacco	-0.143(0.011)**
Sigma.Alcohol	1.408(0.012)**	Rho.Drink.Alcohol	-0.833(0.091)**

** Significant at 5%.

* Significant at 10%.