A double-hurdle approach to modelling tobacco consumption in Italy

Aristei, David; Pieroni, Luca

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Mitglied der

Leibniz-Gemeinschaft

gesis
Leibniz-Institut
für Sozialwissenschaften

A Double-Hurdle Approach to Modelling Tobacco Consumption in Italy

powered by ScholarOne
Manuscript Central™

A Double-Hurdle Approach to Modelling Tobacco Consumption in Italy

David ARISTEI[1] – Luca PIERONI[2]

Abstract

This paper analyses the determinants of tobacco expenditures for a sample of Italian households. A Box-Cox double-hurdle model adjusted for heteroscedasticity is estimated to account separate individual decisions concerning smoking participation and tobacco consumption and to correct for non-normality in the bivariate distribution of the error terms. Nested univariate and bivariate models are found to be excessively restrictive, supporting the adequacy of a generalized specification.

Estimation results show that consumption decisions are significantly affected by income and demographic characteristics. In particular, income positively impacts tobacco expenditure, while participation probability substantially declines as age increases. The existence of significant gender differences in both smoking participation and tobacco consumption patterns is found, while high education and white collar occupation reduce the likelihood to smoke and tobacco expenditure levels. Single adult households have a lower probability of smoking initiation even if, conditional on smoking, they consume more. Finally, complementarity between tobacco and alcohol beverages suggests the necessity of joint public health strategies.

[1] *Department of Economics, Finance and Statistics, University of Perugia and Department of Economic Sciences, University of Verona and.*

[2] *Department of Economics, Finance and Statistics, University of Perugia.*

Corresponding author: Luca Pieroni, Department of Economics, Finance and Statistics, University of Perugia, via Pascoli 20, 06123 Perugia; Tel. +390755855280, Fax +390755855299, e-mail: lpieroni@unipg.it

1. Introduction

In the last years the empirical literature has produced a large body of evidence on the price and non-price determinants of tobacco demand. One strand of literature has adopted an aggregate time series approach to provide empirical support to the rational addiction model proposed by Becker and Murphy (1988) (Chaloupka, 1991; Becker et al., 1994; Bask and Melkersson, 2004). On the other hand, the growing availability of microdata from household expenditure surveys has allowed to model tobacco consumption accounting for zero observations and simultaneously exploiting the richness of survey data information to control for heterogeneous individual (or household) behaviour (Jones, 1989, 1992; Blaylock and Blisard, 1992; Garcia and Labeaga, 1996; Yen, 2005a). From a policy perspective, cross-sectional surveys enables to improve the knowledge of the impacts of socio-demographic variables on tobacco expenditure and help the design of public health programs to achieve smoking-reduction objectives.

While it would be interesting to obtain simultaneous empirical responses concerning addiction, censoring and heterogeneity in tobacco consumption decisions in Italy, the absence of a true panel data does not enable us to account for addictive behaviours while controlling for demographic and socio-economic characteristics[1]. Thus, in this paper we investigate household tobacco expenditures, addressing the issues connected to limited dependent variable models by an approach based on a double-hurdle specification (Cragg, 1971; Jones, 1989; Yen and Jones, 1996; Su and Yen, 2000). Several empirical studies (Blundell and Meghir, 1987; Blaylock and Blisard, 1993;

---

[1] Only few countries give a panel data structure to their household expenditure surveys. Recently, Labeaga (1999) and Jones and Labeaga (2003), using a panel of Spanish households (the Continuous Family Expenditure Survey) have attempted to test rational addiction and simultaneously account for censoring and unobservable heterogeneity.

Garcia and Labeaga, 1996; Yen and Jones, 1996) have shown the inadequacy of the standard Tobit model (Tobin, 1958) in cross-sectional analysis of tobacco consumption, connected with its failure in accounting for differences concerning the generation of zero observations. Since it is implausible that all zero observations in tobacco consumption arise from standard corner solutions generated by a constrained budget, we emphasize the importance of a double-hurdle specification distinguishing between abstentions and corner solutions.

As shown by Arabmazard and Schmidt (1982), maximum likelihood estimates are sensitive to misspecification issues and they will be inconsistent if the bivariate normality assumption is violated. One way to correct for the non-normality of the error terms, providing to generalize the standard double-hurdle model, consists in applying a Box-Cox transformation on the dependent variable (Yen, 1993; Yen and Jones, 2000). In this way we can encompass, besides the standard double-hurdle model, a wide range of specifications, that differ for the distributional assumptions on the error terms, and test the best model to rationalize the data.

Standard Tobit model, standard Double-hurdle model with independence and Box-Cox Tobit model seem to be restrictive specifications, confirming both the violation of the bivariate normality assumption and the existence of separate individual decisions on participation and consumption. Moreover, the independence of the errors for the Box-Cox double-hurdle is not rejected by the data, concluding that the non-normal specification with independent errors is the best model to account for household behaviours on tobacco expenditures.

The remainder of the paper is organized as follows: the next section outlines the theoretical framework upon which the empirical models are based. In Section 3 we

discuss econometric methods. In particular, both the standard double-hurdle model and

the extended specification based on the Box-Cox transformation are presented. In

Section 4 data used in the empirical analysis, taken from 2002 Italian Household Budget

Survey, are discussed and factors influencing participation and consumption equations

are examined. In Section 5 specification and estimation results are presented and

discussed, with specific attention devoted to the nested strategy used to derive the best

specification and to the analysis of the estimated parameters and elasticities. Section 6

offers some concluding comments.

2. Theoretical Framework

Standard consumer choice theory assumes that every individual (or household) is a

potential consumer of all goods. However, for commodities like tobacco this may not be

true and, regardless of price and income levels, some individuals cannot be induced to

smoke. In these cases, zero observations are not the result of economic non-

consumption, but they may be determined by other behavioural factors apart from prices

and income. Following Pudney (1989), alternative models of tobacco demand have been

derived by using discrete random preferences regimes. According to this approach,

smokers are hypothesized to have a different preference structure than non-smokers.

Observed zero expenditure then reflect either the decision to not smoke or a standard

corner solution and hence only potential smokers determine the parameters of tobacco

Engle curve (Blaylock and Blisard, 1993).

Given this setting, individual's utility function takes the following form:

$$U = U(dc_1, c_2, ..., c_n; w) \tag{1}$$

where c_1 is the quantity of tobacco (with price p_1), $c_2,...,c_n$ represent all other goods, w is a vector of demographic variables representing the qualitative characteristics of smoking, and d is binary variable which is equal to one if an individual is an actual or potential smoker and zero otherwise. If d always equals one, every individual is assumed to be a potential smoker and observed zero consumption is a standard corner solution.

Equation (1) can be alternatively rewritten as:

$$U = dU^s(c_1,c_2,...,c_n;w) + (1-d)U^{ns}(c_2,...,c_n;w) \qquad (2)$$

where U^s is the utility function of smokers (actual and potential) and U^{ns} for non-smokers. For a non-smoker, given that c_1 does not enter $U^{ns}(c_2,...,c_n;w)$ and p_1 is in any case positive, the optimal tobacco consumption level is $c_1^* = 0$. For actual and potential smokers, the optimal level of c_1 is determined by solving the following constrained utility maximization:

$$\max_{c_1,...,c_n} \left\{ U^s(c_1,c_2,...,c_n;w) \right\} \quad \text{s.t.} \quad p'c = m \qquad (3)$$

where p is a vector of prices (including p_1) and m is individual's (or household's) budget. Assuming the utility function $U^s(c_1,c_2,...,c_n;w)$ to be continuous, increasing, and quasi-concave, then the notional demand for tobacco can be expressed as a demand function $f(p,m;w)$ and the corresponding expenditure equation can be denoted as $g(m;w)$. As price information was not collected in the ISTAT Italian Household Budget Survey, we assume that all households face the same relative prices (Yen and Jensen, 1996). The notional demand and expenditure have been derived as the results of utility maximization with only the budget constraint, given individual (or household)

characteristics. However, as the quantity and expenditure are also subject to a non-negativity constraint, the optimal level of tobacco expenditure ($e_{c_1}^*$) can be either an interior solution or a corner solution (that is: $e_{c_1}^* = \max\{0, g(m;w)\}$), justifying the use of a double-hurdle specification for modelling tobacco consumption, since it jointly accounts for both abstentions and corner solutions.

In the next section, alternative empirical models are developed in order to account for the presence of different zero-generating mechanisms in tobacco consumption.

3. Econometric Specification

Recent empirical analyses have shown the inadequacy of the standard Tobit model in cross-sectional analysis of tobacco consumption, stressing the relevance of a double-hurdle approach for microeconomic analysis of tobacco consumption (Jones, 1989; Blaylock and Blisard, 1992; Garcia and Labeaga, 1996; Yen and Jones, 1996; Labeaga, 1999). The main feature of the double-hurdle model is that participation and consumption decisions are assumed to stem from two separate individual choices and the determinants of the two decisions are allowed to differ.

The double-hurdle model, originally proposed by Cragg (1971), assumes that two separate hurdles must be passed before a positive level of consumption can be observed. In the context of tobacco consumption analysis, the first hurdle involves the decision of whether or not to smoke (*participation decision*). It is reasonable to assume that the choice of smoking is not only an economic decision, but also influenced by social and demographic factors which are independent of the quantity consumed. The second hurdle concerns the level of tobacco consumption to choose *(consumption decision)*.

We now turn to the econometric specification of the model. Both hurdles are assumed to be linear in the parameters (α, β), with disturbance terms u and v randomly distributed with a bivariate normal distribution. The matrices z and x include the variables that are assumed to influence participation and consumption decisions, respectively. Formally, following Jones (1989) and Pudney (1989), the bivariate model can be written as:

i) Observed consumption:

$$y_i = d \cdot y_i^{**} \tag{4}$$

ii) Participation equation:

$$w_i = z_i'\alpha + u_i \qquad u_i \quad N(0,1) \tag{5}$$

$$d = \begin{cases} 1 & \text{if } w > 0 \\ 0 & \text{otherwise} \end{cases}$$

iii) Consumption equation:

$$y_i^* = x_i'\beta + v_i, \quad v_i \quad N(0,\sigma^2) \tag{6}$$

$$y_i^{**} = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

A positive level of tobacco consumption y is observed only if the individual (or the household) is a potential smoker ($d = 1$) and actually consumes tobacco (y^{**}). For this reason, in double-hurdle models, differently from Heckman selection model (Heckman, 1979), in which zeros are not affected by the consumption decision, observed zero expenditures are the result of either participation or consumption decisions and potential smokers may have zero tobacco expenditure.

In the present analysis, different double-hurdle models are used to analyze household tobacco expenditure patterns, with particular attention devoted to the specification of the stochastic structure of the model by testing whether the assumptions of bivariate

normality, homoscedasticity and independence of the error terms across participation and consumption equations are acceptable.

To account for the presence of heteroscedasticity, the variance of the error terms specified it as a function of a set of continuous variables:

$$\sigma_i = \exp(z_i' h) \tag{7}$$

where z_i is a vector of continuous variables included in x_i ($z_i \in x_i$) and h is a conformable vector of coefficients (Yen, 1993; Yen and Jensen, 1996; Newman et al., 2003).

The assumption of correlated error terms allows for the possibility that participation and consumption decisions are simultaneously taken. In particular, we assume that u and v are distributed as a bivariate normal:

$$(u,v) \quad BVN(0,\Sigma), \qquad \Sigma = \begin{bmatrix} 1 & \sigma\rho \\ \sigma\rho & \sigma^2 \end{bmatrix} \tag{8}$$

where ρ is the correlation coefficient.

Denoting zero consumption as 0 and positive consumption as +, the likelihood function for the full double-hurdle model with heteroscedasticity correction and dependence between u and v can be written as:

a) Heteroscedastic Double-hurdle model with dependent error terms

$$L = \prod_0 [1 - \Phi(z_i'\alpha, x_i'\beta, \rho)] \prod_+ \left[\Phi\left(\left[z_i'\alpha + \frac{\rho}{\sigma_i}(y_i - x_i'\beta) \right] \middle/ \sqrt{1-\rho^2} \right) \frac{1}{\sigma_i} \phi((y_i - x_i'\beta)/\sigma_i) \right] \tag{9}$$

where Φ denotes the standard normal CDF (univariate or multivariate) and ϕ is the univariate standard normal PDF.

In empirical applications it is common to assume that u and v are independent (i.e. $\rho = 0$). Under this hypothesis we obtain the original formulation proposed by Cragg (1971).

b) Heteroscedastic Double-hurdle model with independent error terms

$$L = \prod_0 [1 - \Phi(z_i'\alpha)\Phi(x_i'\beta)/\sigma_i] \prod_+ \left[\Phi(z_i'\alpha)\frac{1}{\sigma_i}\phi((y_i - x_i'\beta)/\sigma_i) \right] \tag{10}$$

This model hypothesizes that the participation and consumption decisions are made separately and that there is a feedback effect from the level of consumption to the participation decision (Atkinson et al., 1984; Deaton and Irish, 1984; Blaylock and Blisard, 1993). In the present analysis we do not make any *a priori* assumption on the correlation structure of the error terms; differently from other studies (Yen and Jensen, 1996; Moffatt, 2005; Newman et al., 2003), the independence of error terms is not assumed as a maintained hypothesis, but its significance is tested by means of a Wald test. Moreover, it is worth noticing that the standard Tobit model is a nested version of the Cragg model, when $\alpha_j = 0$ for $j \neq 0$ and $\alpha_0 = \infty$, that is when $p(u > -z_i'\alpha) = 0$ (i.e. $\Phi(z_i'\alpha) = 1$), and a likelihood ratio test can be used to distinguish between the Tobit and the independent double-hurdle model.

A limitation of the standard double-hurdle specification is that it is built on the assumption of bivariate normality of the error terms. If the normality assumption is violated the maximum likelihood estimates of the model will be inconsistent. This may be particularly relevant when the model is applied to a dependent variable with a highly skewed distribution, as is often the case with survey data on tobacco expenditures.

As suggested by Yen (1993) and Jones and Yen (2000), one way to correct for the non-normality of the error terms consists in applying a Box-Cox transformation to the dependent variable, which gives:

$$y_i^T = \frac{y_i^\lambda - 1}{\lambda}, \qquad \text{with } 0 < \lambda \leq 1 \tag{11}$$

where λ is an unknown parameter. It is worth noticing that the Box-Cox transformation includes, as special cases, a straightforward linear transformation ($\lambda=1$) and the logarithmic transformation ($\lambda \to 0$), but normally we would expect the parameter λ to lie between these two extremes.

The Box-Cox double-hurdle model implies the following relationship between the transformed dependent variable and the latent variables, w and y^*:

$$y_i^T = \begin{cases} y_i^* & \text{if } y_i^* > -1/\lambda \text{ and } w_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where w_i and y_i^* are defined as in (7) and (8).

This specification relaxes the normality assumption on the conditional distribution of y_i and still allows stochastic dependence between the error terms of participation and consumption equations. The likelihood functions for dependent and independent double-hurdle models with heteroscedasticity correction can be written as:

c) Box-Cox Heteroscedastic Double-hurdle model with dependent error terms

$$L = \prod_0 \left[1 - \Phi\left(z_i'\alpha, \frac{x_i'\beta + 1/\lambda}{\sigma_i}, \rho \right) \right] \times$$
$$\times \prod_+ \left[\Phi\left(\left\{ z_i'\alpha + \frac{\rho}{\sigma_i}\left[(y_i^\lambda - 1)/\lambda - x_i'\beta \right] \right\} \Big/ \sqrt{1-\rho^2} \right) y_i^{(\lambda-1)} \frac{1}{\sigma_i} \phi\left(\left[(y_i^\lambda - 1)/\lambda - x_i'\beta \right] \Big/ \sigma_i \right) \right] \tag{13}$$

d) Box-Cox Heteroscedastic Double-hurdle model with independent error terms

$$L = \prod_0 \left[1 - \Phi(z_i'\alpha)\Phi\left((x_i'\beta + 1/\lambda)/\sigma_i\right) \right] \prod_+ \left[\Phi(z_i'\alpha) y_i^{(\lambda-1)} \frac{1}{\sigma_i} \phi\left(\left[(y_i^\lambda - 1)/\lambda - x_i'\beta \right] \Big/ \sigma_i \right) \right] \tag{14}$$

Model (c) is a general model that nests all the other three double-hurdle specifications presented and also encompasses a wide range of standard limited dependent variable models (Jones and Yen, 2000). Figure 1 summarizes the relationships between all the

models considered, showing the relevant restrictions on the likelihood function (13) implied by the nested specifications.

(Figure 1 about here)

The economic interpretation of limited dependent variable models frequently focuses on the analysis of the marginal effects of regressors on the expected value of y_i (Jones and Yen, 2000), which can be decomposed into an effect on the probability of purchase and an effect on the conditional level of expenditure[2]. The unconditional mean of y_i in the Box-Cox Double-hurdle model can be written as:

$$E(y_i) = P(y_i > 0)E(y_i \mid y_i > 0) \tag{15}$$

The conditional expectation of y_i is:

$$E(y_i \mid y_i > 0) = E(y_i \mid w_i > -z_i'\alpha, y_i^* > -x_i'\beta - \frac{1}{\lambda}) \tag{16}$$

and, assuming independence between error terms of participation and consumption[3], can be written as:

$$E(y_i \mid y_i > 0) = \left[\Phi\left(\frac{x_i'\beta + 1/\lambda}{\sigma_i} \right) \right]^{-1} \int_0^\infty \frac{y_i^\lambda}{\sigma_i} \phi\left(\frac{y_i^T - x_i'\beta}{\sigma_i} \right) dy_i \tag{17}$$

Given independence, the probability of a positive consumption level is:

$$P(y_i > 0) = \Phi(w_i'\alpha_i)\Phi\left(\frac{x_i'\beta + 1/\lambda}{\sigma_i} \right) \tag{18}$$

---

[2] This decomposition follows the approach proposed by McDonald and Moffitt (1980) for the decomposition of the unconditional mean of the dependent variable in the Tobit model.
[3] Here, for simplicity, we focus on the Independent Box-Cox Double-Hurdle model. Details on the derivation of the conditional mean for the Box-Cox Double-Hurdle model with dependent errors can be found in Jones and Yen (2000).

Marginal effects can be obtained by differentiating equations (16), (17) and (18) with respect to each explanatory variable[4]. From these marginal effects, elasticities can be derived. In particular, using equation (15), the elasticity of the conditional mean with respect to regressor x_{ij} can be written as:

$$e_j = \frac{\partial E(y_i)}{\partial x_{ij}} \frac{x_{ij}}{E(y_i)} = \frac{\partial P(y_i > 0)}{\partial x_{ij}} \frac{x_{ij}}{P(y_i > 0)} + \frac{\partial E(y_i \mid y_i > 0)}{\partial x_{ij}} \frac{x_{ij}}{E(y_i \mid y_i > 0)} \qquad (19)$$

where the two addends are the elasticity of the probability of observing a positive expenditure (e_j^P) and the elasticity of conditional consumption (e_j^{cc}). For continuous variables, the elasticities are computed at the sample means. For categorical explanatory variables, marginal effects are used to compute percentage changes in probability, conditional level and unconditional level when the value of the variable shifts from zero to one, holding all the other variables constant (Yen and Jones, 1996; Newman et al., 2003).

4. Data and Variables

The data used in the empirical analysis are taken from the 2002 Italian Household Budget Survey (IHBS), which is conducted by the Italian Central Statistics Office (ISTAT). This survey, together with Bank of Italy's Survey of Household Income and Wealth, represents the main and most comprehensive source of microdata for analysing consumption behaviours of Italian households. The ISTAT survey covers a random sample of 27499 households throughout the country and provides detailed information on family expenditures (non-durable and durable) as well as on household socio-economic and demographic characteristics. Data on non-durable consumption are collected in a diary that records household expenditures on a wide range of non-durable

---

[4] Analytical details on the derivation of conditional and unconditional marginal effects for the Box-Cox double-hurdle model can be found in Yen (1993) and Jones and Yen (2000).

goods and services over a one-week period and is subsequently expressed on a monthly basis. No price or quantity information is reported and consumption of each good or service is measured as reported expenditure. Moreover, data refer to the household rather than to individuals, providing no information on intra-household income allocation and making it necessary to account for family composition effects when analysing the demand for certain non-durable goods.

In this analysis, household monthly expenditure on cigarettes and tobacco is used as the dependent variable. As already underlined, the ISTAT survey only records the value of household expenditures and so quantity and quality effects cannot be identified and analyzed. Since the survey does not provide any information on the presence of smokers in the household, inference on participation in tobacco consumption has to be made on the basis of recorded expenditure, identifying non-smoking households as those with zero tobacco expenditure. Expenditures on tobacco, as all expenditures on commonly used non durable goods, are monitored for only one week. The fact that the collection of information only covers a seven-day period, with a 33.4 percent of consuming households, introduces some room for undetected infrequency of purchases. However, for commodities like tobacco, it is unlikely that observed zero expenditures capture infrequent purchases and they probably stem from abstentions or corner solutions (Garcia and Labeaga, 1996). Using household data to model tobacco consumption, which is mainly the result of an individual decision, may generate some bias in the analysis of consumption behaviours. For this reason, it is necessary to correctly account for family size and composition effects as well as for the socio-economic characteristics of the household, since the household situation plays an active role in modifying individual preferences.

Household tobacco expenditures are assumed to be expressible as a linear combination of explanatory variables that are assumed to separately affect participation and consumption decisions. The explanatory variables selected, together with their sample statistics, are presented and fully described in Table 1. They include household income (proxied by total expenditure), alcohol consumption, the head of the household's age, and dummy variables indicating education, marital status, gender and working status of the household's head, the percentage of adult male members, the presence of children under fourteen years old, home ownership, and participation in alcohol consumption.

All expenditure variables are expressed in real terms, by deflating current values using ISTAT regional price indexes, in order to account for price variability across regions. Moreover, in order to obtain a per-equivalent adult measure of household consumption, all expenditure variables are adjusted for family size by using the modified OECD equivalence scale, which assigns a weight equal to one to the first adult in the household, 0.5 to each other adult and 0.3 to each child under fourteen years of age.

(Table 1 about here)

5. Results

In this section, estimation results are presented and discussed, with specific attention devoted firstly to the choice of the most appropriate model specification and then to the analysis of the effects of explanatory variables by calculating and decomposing elasticities.

5.1. Model Specification

One of the main objective of this paper is to test whether univariate or bivariate models are adequate for analysing tobacco consumption behaviour of Italian households.

All the double-hurdle specifications discussed in Section 3 have been estimated by maximizing the logarithm of the likelihood functions (9), (10), (13) and (14). One parameter estimation issue in double-hurdle models concerns the choice of the regressors for participation and consumptions equations. As it is known, the choice of the explanatory variables to be included in the two hurdle does not rest on any a priori theory and may be somewhat arbitrary. Given that the inclusion of the same set of regressors in each hurdle makes the parameters identification difficult, exclusion restrictions must be imposed[5]. In empirical applications the first hurdle is usually assumed to be a function of non-economic factors affecting household's smoking decision, so that economic variables can be excluded from the first equation (Newman et al, 2003). Their exclusion is motivated by the discrete random preference theory, according to which sample selection is determined exclusively by non-economic factors (Pudney, 1989; Yen, 2005a).

The foregoing arguments require, before presenting estimation results, a discussion of the explanatory variables included in the model. The independent variables considered are intended to encompass the determinants of both smoking participation and tobacco consumption decisions and their choice rests on suggestions taken from previous empirical literature (Jones, 1989; Blaylock and Blisard, 1992; Garcia and Labeaga, 1996;

---

[5] In estimating the final model we started with a specification that included all explanatory variables in both hurdles; insignificant variables were gradually dropped, with exclusion restrictions giving identification higher reliability.

Yen and Jones, 1996; Yen, 2005a) and on more specific issues connected with identification problems. Their inclusion in either participation or consumption equations is justified by economic, demographic and sociological factor, such as habit formation, information and social awareness on damages and health risks connected with smoking and restriction to smoking due to security reasons or to health guidelines for public places.

In this study, both participation and consumption decisions are postulated to be influenced by the presence of children under fourteen years old within the household (*Child014*), the percentage of adult male members (*PercMale*), age (*Age*), gender (*MaleHH*), marital status (*Single*), education level (*Highedu*) and occupation (*Whitecollar*) of the household's head.

A dummy variable indicating the presence of children is included in the model postulating that the household would not smoke or at least would attempt to moderate tobacco consumption when small children are present (Blaylock and Blisard, 1993; Kerr et al., 2004). The percentage of adult male members in the household (*PercMale*) is included to account for gender differences in tobacco consumption and to test for the presence of gender-differentiated smoking habits (Angulo et al., 2001). The education variable (equal to one if the household's head has at least a high school education, zero otherwise) has been often used in previous empirical studies (Jones, 1989; Blaylock and Blisard, 1992, 1993; Yen, 2005b; Yen and Jensen, 1996; Garcia and Labeaga, 1996; Yen and Jones, 1996), suggesting that individuals with better education may be more aware of the health risks connected with consumption of unhealthy goods such as tobacco and alcohol. Being a white-collar worker reflects the individual's social class and may help in explaining how smoking habits vary among different social groups. Age is considered to assess how age-related health problems affect smoking behaviours

and to verify the existence of a significant lifecycle pattern for both tobacco participation and consumption decisions[6].

In the participation equation, we include an additional binary variable indicating whether the household displays a high expenditure level (over the 75th percentile of the observed distribution) on alcoholic beverages (*HighAlc*), as a proxy for habit formation tendencies (Blaylock and Blisard, 1993).

Specific variables accounting for economic conditions have been introduced in consumption equation. Total household expenditure (*Income*) is included as a proxy for current income. A variable indicating whether the household lives in a home that is owned or being bought (*OwnerOcc*) is included, following the suggestions of Atkinson et al (1984) and Jones (1989), as a proxy for wealth and economic stability. Further, household alcohol expenditure (*Alcohol*) is included as a proxy to verify the presence of complementary relationships with household expenditures on alcoholic beverages. The consumption equation also includes quadratic terms of age and income to capture possible non-linear relationships with tobacco expenditure (Jones, 1989; Garcia and Labeaga, 1996).

5.2 Statistical Tests and Estimation Results

In order to correctly analyze the determinants of tobacco expenditures and to model household smoking behaviour, one first task relates to the choice of the most appropriate specification. Our selection strategy consists in testing the bivariate model with dependent error terms, which is the most general specification and encompasses all

---

[6] Jones (1989) included the individual's age and its square as explicative variables, while Yen and Jensen (1996) used both household age composition and the age of the household head, showing significant life-cycle patterns for both participation and consumption decisions.

the other bivariate and univariate models discussed in Section 3, against its nested

alternatives, by means of conventional and adjusted (Vuong, 1989) likelihood ratio

tests. However, it should be underlined that the validity of the LR tests strongly rests on

the assumption that the general model is not misspecified (Yen and Jones, 1996); in

particular homoscedasticity and normality assumptions should not be violated.

Distributional assumptions assume crucial relevance in limited dependent variable

models, since maximum-likelihood estimation will lead to inconsistent parameter

estimates when normality and homoscedasticity are not fulfilled (Maddala and Nelson,

1975; Arabmazard and Schmidt, 1982). For these reasons, preliminary tests for the

validity of the distributional assumptions are necessary. To this end LR test for

homoscedasticity and Pagan and Vella's (1989) moment base test for normality have

been carried out on both Tobit and double-hurdle specifications[7]; the results are

presented in Table 2.

(Table 2 about here)

As can be noted, all equations present severe problems of non-normality and

heteroscedasticity, with LR test values well above the relevant critical values in both

Tobit and double-hurdle models. The violation of homoscedasticity requires allowance

for heteroscedastic error terms in the univariate and bivariate specifications. Following

Yen (1993), we relax homoscedasticity assumption by specifying standard deviation σ_i^2

as a function of the continuous variables of the model, as in equation (7), and allowing

it to vary across observations. For this reason, all the models considered in the

---

[7] Details on distributional tests in censored and limited dependent variable models can be found in Bera, Jarque and Lee (1984), Pagan and Vella (1989) and Wells (2003)

remainder of the discussion account for heteroscedasticity, with a variance equation that includes only the continuous regressors that result statistically significant in generating heteroscedasticity[8]. Results of normality tests reveal that estimation of standard Tobit and double-hurdle models may lead to inconsistent results, supporting the necessity of a non-normal generalization of these models. As previously introduced in Section 3, following Yen (1993) and Yen and Jones (1996, 2000), we consider a Box-Cox transformation of the dependent variable that relaxes normality assumption on the conditional distribution of y_i and includes as special cases linear and logarithmic transformations. The results of the normality tests also can be interpreted as a strong indication of the superiority of the univariate and bivariate Box-Cox generalizations with respect to their standard counterparts.

Once the diagnostics of the model have been analyzed, we now turn to the choice of the most appropriate model. As previously shown in Figure 1, all restricted models can be obtained by placing the relevant restrictions on the likelihood function (13) and can be interpreted as special cases of the Box-Cox double-hurdle model with dependent error terms[9].

The specification tests carried out are reported in Table 3. Firstly, we tested the hypothesis of independent errors between participation and consumption equations; the issue of dependency in double-hurdle models is a problem of great relevance, but it has often been disregarded in previous empirical works (Newman et al, 2003; Moffatt,

---

[8] In principle all explanatory variables can be included in the heteroscedasticity specification; however, doing so would considerably increase the number of parameters to be estimated. So we focused our attention only on the variables that are more likely to cause heteroscedasticity and then we tested alternative specification excluding those variables that are not significantly different from zero.

[9] The Heckman sample selection model can also be obtained as a restricted specification, assuming that participation decision dominates consumption decision. Vuong specification test for non-nested models supports the inadequacy of the Heckman model. The results are not presented here, but they are available from the authors.

2005). The results of the LR test ($\chi^2_{(1)} = 0.78$ with a p-value equal to 0.377) clearly

indicates that dependency is not relevant; this result is in line with findings of Jones

(1989), Blaylock and Blisard (1993) and Garcia and Labeaga (1996) and demonstrates

that the independent Box-Cox double-hurdle model is an acceptable alternative to the

dependent model[10]. On the basis of the results of Vuong specification test for nested

models (Vuong, 1989), all the other restricted specifications are rejected, each with a *p-*

value of less than 0.0001. The interpretation of these results is twofold. Firstly, they

suggest the inadequacy of the univariate Tobit specification in modelling tobacco

consumption behaviours, given the existence of separate participation and consumption

decisions. On the other hand, the results give further support to the generalized

specification to account for non-normal and heteroscedastic error terms. Thus, the

model that best rationalizes tobacco expenditure data is the independent Box-Cox

double-hurdle model.

(Table 3 about here)

Maximum-likelihood estimates are presented in Table 4. In order to account for

differences in estimated parameters, we report the results of both standard and Box-Cox

independent double-hurdle models, even if the discussion is focused only on the latter.

Analyzing the estimated parameters, it is possible to highlight that all the

coefficients, with the exception of that of education in the consumption equation and

that of occupational status in the participation equation, are significant at the one

---

[10] Smith (2003) puts into question the relevance of the dependent double-hurdle model itself, asserting that this model contains too little statistical information to support estimation of dependency, even when dependency is truly present.

percent level. Before to show the effects of explanatory variables, we underline that the Box-Cox parameter ($\lambda = 0.1864$) is significantly different from one and zero; this implies a Box-Cox transformation that is different from both a linear transformation and a logarithm specification.

(Table 4 about here)

Turning to explanatory variables, income has a positive effect on household tobacco consumption, while income squared, which accounts for non-linearity relationships, is negative. These outcomes are in line with the findings of Garcia and Labeaga (1996) and imply that tobacco expenditure rises as household income increases, but at a decreasing rate. The estimated age effects are negative and significant in both participation and consumption equations, indicating the existence of a strong lifecycle pattern (Kerr et al., 2004; Aristei et al., 2005). As in Yen and Jones (1996) and Yen (2005b), the probability of smoking and tobacco expenditure levels decrease with the age; moreover, the estimated age squared parameter shows a positive non-linear relationship between consumption levels and age.

The estimated parameters of the variables included in both hurdles have the expected signs and indicate that having a high education and being a white collar worker reduce the probability of smoking and the level of tobacco expenditure. The presence of children under fourteen years old significantly reduces the likelihood of smoking, indicating that households tend to refrain from smoking when small children are present. However, limited to the smokers sub-sample, the presence of children under fourteen years old increases the level of tobacco expenditure. The estimated coefficients for the dummy variable indicating the gender of the household's head (*MaleHH*) reveal

the existence of differentiated behaviours, with female headed households presenting both a higher probability of smoking and higher tobacco expenditures. On the other hand, the variable indicating the percentage of adult male members within the household (*PercMale*) exerts a significant and positive effect on both participation and consumption decisions. These evidences highlight the existence of important gender differences in tobacco consumption patterns. Moreover, it is worth noticing that the marital status variable exerts opposite effects on participation and consumption. In particular, being a single reduces the likelihood to smoke but, conditional on smoking, it increases tobacco expenditure.

Home ownership seems to negatively affect tobacco consumption levels; this result, previously found by Jones (1989) and Yen (2005a), suggests that belonging to wealthier social groups may induce moderate consumption.

Finally, the positive coefficient of the dummy variable indicating a high level of alcohol expenditure (*HighAlc*) reveals that the probability of being a smoker is higher for those households that are strong drinkers. Moreover, tobacco expenditure is also found to be positively correlated with the level of alcohol expenditure, showing the existence of a complementarity relationship in the consumption of addictive and habit-generating goods.

5.3 Elasticities

In assessing the impact of explanatory variables, the presence of parameter estimates with opposite signs in the two hurdles and the Box-Cox transformation complicate the interpretation of the estimated effects. Thus, the impact of explanatory variables can be better explored by computing elasticities.

The elasticities of probability, conditional level and unconditional level with respect to explanatory variables are calculated by using the formulas in equation (19). For statistical inference, standard errors are computed using the delta method (Su and Yen, 1996; Spanos, 1999).

(Table 5 about here)

Estimated elasticities and discrete effects for the categorical variables, along with their standard errors, are presented in Table 5. Analysing the effects of the continuous variables, income has a positive effect on the conditional level of tobacco consumption; the elasticity are both significant, with a net effect on unconditional mean equal to 0.2638, a value which is in line with the empirical findings of Yen (2005a). This finding shows that economic factors play an important role in determining smoking decision and indirectly suggest that observed zero consumption may well be the result not only of abstention but also of standard corner solution, confirming the validity of the double-hurdle specification for modelling tobacco expenditure.

The elasticities with respect to age suggest that households with older heads are less likely to smoke and, conditional on smoking, consume less tobacco than younger households. This result is consistent with previous studies, even if the estimated effect on the unconditional level is lower than that found by Yen (1999).

The level of alcohol consumption positively affects the unconditional level of tobacco consumption, revealing the existence of significant complementarities between smoking and drinking, with tobacco consumption rising as expenditures on alcoholic beverages grow.

The elasticities with respect to *PercMale* are all positive and significant, with an overall effect on the unconditional level equal to 0.0443. This result highlights how men and women exhibit different consumption patterns in relation to tobacco, with both smoking probability and conditional level increasing as the ratio of male to female members increases.

The analysis of the effects of binary variables gives further support to the evidence highlighted in the discussion of parameters estimates. In particular, it is worth noticing that education plays a negative and significant role only on the probability but not on the conditional level of smoking. The impact of education on the unconditional level of tobacco consumption is negative and significant at the 10 percent significance level (equal to -0.347) as the negative effect on probability obviously dominates. These evidences suggest that more educated individuals are more aware of the health risks associated with smoking and they are less likely to participate to tobacco consumption.

The occupational variable does not influence the probability of smoking, but white-collar workers are found to consume less tobacco conditional and unconditional on smoking.

The marital status variable exerts opposite effects on probability and conditional expenditure; more precisely, being a single reduces tobacco consumption probability by about 3 percent, but significantly increases the level of expenditure conditional on the smoking households sub-sample (the estimate effect is equal to 3.1230). However, given that the positive effect on conditional level dominates the negative effect on probability, the net effect on unconditional expenditure is clearly positive and equal to 1.3693.

Analysing the effect of household's head gender, it can be pointed out that the female-headed households are more likely to smoke and tend to consume more tobacco than the male-headed ones. The effects on probability, conditional and unconditional level are negative and significant, with an overall effect on the unconditional effect of equal to -0.7942. This result is not in contrast with the evidences connected to the effect of the percentage of male members and it is mainly connected to the structure of the female-headed households. In fact, the female-headed families in our sample are mainly single adult and single adult with adult children households and the effect exerted by this particular household structure may interact with the effect of household's head gender, making the separate identification of the two effects difficult.

Moreover, it should be noted that the categorical variable indicating the presence of children is characterized by a negative elasticity of probability (equal to -1.7%). However, conditional on smoking, the elasticity of consumption is positive and offsets the negative effect of probability.

Finally, heavy alcohol drinking households are characterized by positive and significant elasticities of probability, conditional and unconditional levels. In particular, this result suggests that the probability of smoking is higher (of about 5%) for those households displaying habit generation tendencies and less aware of the health risks connected with tobacco and alcohol consumption.

6. Concluding Remarks

In this paper a Box-Cox double-hurdle model has been applied to data on tobacco expenditure from the Italian Household Budget Survey. The test strategy to obtain the best specification that out-perform nested models was carried out by means of a

general-to-particular approach. The results obtained show that independence in the Box-Cox double-hurdle model is a non-restrictive assumption, while Tobit model, standard double-hurdle model with independence and Box-Cox Tobit model seem to be restrictive specifications, since they failed to account for non-normality in the bivariate distribution and for separate individual decisions on smoking participation and tobacco consumption.

The results based on the estimated parameters and elasticities confirm that decisions to smoke are related to income and demographic characteristics. In particular, income positively affects tobacco expenditure, indirectly showing the presence of habit-related behaviours in low income households which lead towards a lower income elasticity.

A substantial decline in the participation is found when the age of the household's head increases, together with an important lifecycle pattern in consumption. Estimation results highlight the presence of significant gender differences in both smoking participation and tobacco consumption patterns. High education and white collar occupation are found to reduce the likelihood of smoking and tobacco expenditure levels, suggesting that households with low educational levels and belonging to lower social classes have not yet benefit from policies concerning health consequences of smoking. Again, the data show that being a single determines a lower probability on smoking initiation, but conditional on smoking, the consumption level is higher.

Finally, complementarities between tobacco and alcoholic beverages consumption are present in both the probability and the level of smoking. This relationship is consistent with a wide body of empirical evidences and suggests that anti-smoking policies and public health strategies aimed at reducing alcohol abuse should be jointly addressed towards those households with higher levels of alcohol and tobacco expenditures.

References

Amemiya, T. (1984) Tobit Models: A Survey, *Journal of Econometrics*, **84**, 3-61.

Angulo, A.M., J.M. Gil and A. Gracia (2001) The demand for alcoholic beverages in Spain, *Agricultural Economics*, **26**, 71-83.

Arabmazard, A. and P. Schmidt (1982) An Investigation of the Robustness of the Tobit Estimator to non-normality, *Econometrica*, **50**, 1055-1063

Aristei, D., F. Perali and L. Pieroni (2005) Cohort Analysis of Alcohol Consumption: a Double-Hurdle Approach, Working Paper CHILD 09/2005.

Atkinson, A.B., J. Gomulka, and N.H. Stern (1984) Household expenditure on tobacco 1970-1980: evidence from the Family Expenditure Survey, ESRC Programme on Taxation, Incentives, and the Distribution of Income, London School of Economics, Discussion Paper No. 60.

Bask, M. and M. Melkersson (2004) Rationally addicted to drinking and smoking?, *Applied Economics*, **36**, 373 – 381.

Becker, G. S. and K. Murphy (1988) A Theory of Rational Addiction, *Journal of Political Economy*, **96**, 675-701.

Becker, G.S., M. Grossman and K.M. Murphy (1994) An Empirical Analysis of Cigarette Addiction, *American Economic Review*, **84**, 396-418.

Bera, A.K., C.M. Jarque, and L. Lee (1984) Testing the Normality Assumption in Limited Dependent Variable Models, *International Economic Review*, **25**, 563-78.

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

Blaylock, J.R. and W.N. Blisard (1993) Wine consumption by US men, *Applied Economics*, **24**, 645-651.

Blundell, R. and C. Meghir (1987) Bivariate Alternatives to the Univariate Tobit Model, *Journal of Econometrics*, **34**, 179-200.

Chaloupka, F. (1991) Rational Addictive Behavior and Cigarette Smoking, *Journal of Political Economy*, **99**, 722-42.

Cragg, J. (1971) Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrica*, **39**, 829-844.

Deaton, A. and M. Irish (1984) Statistical models for zero expenditures in household budgets, *Journal of Public Economics*, **23**, 59-80.

Garcia, J. and J.M. Labeaga (1996) Alternative Approaches to Modelling Zero Expenditure: An Application to Spanish Demand for Tobacco, *Oxford Bulletin of Economics and Statistics*, **58**, 489-506.

Heckman, J. (1979) Sample selection bias as a specification error, *Econometrica*, **47**, 153-161.

Jones, A.M. (1989) A double-hurdle model of cigarette consumption, *Journal of Applied Econometrics*, **4**, 23-39.

Jones, A.M. (1992) A note on computation of the double-hurdle model with dependence with an application to tobacco expenditure, *Bulletin of Economic Research,* **44**, 67-74.

Jones, A.M. and J.M. Labeaga (2003) Individual heterogeneity and censoring in panel data estimates of tobacco expenditure, *Journal of Applied Econometrics*, **18**, 157-177.

Jones, A.M. and S.T. Yen (2000) A Box-Cox double-hurdle model, *The Manchester School*, **68**, 203-221.

Kerr, W.C., T.K. Greenfield, J. Bond, Y. Ye, and Rehm, J. (2004) Age, period and cohort influences on beer, wine and spirits consumption trends in the US National Alcohol Surveys, *Addiction,* **99**, 1111-1120.

Labeaga, J.M. (1999) A double-hurdle rational addiction model with heterogeneity: estimating the demand for tobacco, *Journal of Econometrics*, **93**, 49-72.

Maddala, G.S. (1983) *Limited dependent and qualitative variables in econometrics*, Cambridge: Cambridge University Press.

Maddala, G.S. and F.D. Nelson (1975) Specification Errors in Limited Dependent Variable Models, *NBER Working Paper 96*.

McDonald, J. and R. Moffitt (1980) The Uses of Tobit Analysis, *The Review of Economics and Statistics,* **62**, 318-312.

Moffatt, P.G. (2005) Hurdle models of loan default, *Journal of the Operational Research Society*, **56**, 1063-1071.

Newman, C., M. Henchion, and A. Matthews (2003) A double-hurdle model of Irish household expenditure on prepared meals, *Applied Economics*, **35**, 1053-1061.

Pagan, A. and P. Vella (1989) Diagnostic tests for models based on individual data: A survey, *Journal of Applied Econometrics*, **4**, S29-S59.

Pudney, S. (1989) *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes*, New York: Basil Blackwell.

Smith, M.D. (2003) On dependency in Double-Hurdle models. *Statistical Papers*, **44**, 581-595.

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

Su, S-J. and S.T. Yen (1996) Microeconometric Models of Infrequently Purchased Goods: An Application to Household Pork Consumption, *Empirical Economics,* **21**, 513-533.

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

Tobin, J. (1958) Estimation of relationships for limited dependent variables, *Econometrica*, **26**, 24-36.

Vuong, Q.H. (1989) Likelihood ratio test for model selection and non-nested hypotheses, *Econometrica*, **57**, 307–333.

Wells, C. (2003) Retesting Fair's (1978) Model on Infidelity, *Journal of Applied Econometrics*, **18**, 237-239.

Wooldridge, J.M. (2002) Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: MIT Press.

Yen, S.T. (1993) Working wives and food away from home: the Box-Cox double hurdle model, *American Journal of Agricultural Economics*, **75**, 884-895

Yen, S.T. (1995) Alternative transformations in a class of limited dependent variable models: alcohol consumption by US women. *Applied Economics Letters*, **2**, 258–262.

Yen, S.T. (1999) Gaussian versus count-data hurdle models: cigarette consumption by women in the US, *Applied Economics Letters*, **6**, 73-76.

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

Yen, S.T. (2005b) Zero observations and gender differences in cigarette consumption, *Applied Economics*, **37**, 1839–1849

Yen, S.T. and A.M. Jones (1996) Individual Cigarette Consumption and Addiction: a Flexible Limited Dependent Variable Approach, *Health Economics*, **5**, 105-117.

Yen, S.T. and H.H. Jensen (1996) Determinants of household expenditures on alcohol, *The Journal of Consumer Affairs*, **30**, 48-67.

Tables

Table 1 – Variable definitions and descriptive statistics

		Sample statistics			
Variable	Definition	*Full sample (n=27499)*		*Consuming (n=9184)*	
		Mean	*SD*	*Mean*	*SD*
DEPENDENT VARIABLE					
TOBACCO	Per- adult household tobacco consumption	8.053	15.294	24.11264	17.696
EXPLANATORY VARIABLES (CONTINUOUS)					
AGE	Age of the household's head	56.057	15.818	51.730	13.634
AGESQR	Age of the household's head squared	3392.56	1822.87	2861.844	1482.522
PERCMALE	Percentage of adult male members in the household	0.456	0.272	0.510	0.237
INCOME	Proxied by per-equivalent adult household total expenditure and scaled by 100	11.166	8.580	11.946	9.019
INCOMESQR	Income squared	198.306	614.281	224.038	698.944
ALCOHOL	Per- adult household alcohol consumption	8.310	15.485	10.291	16.997
EXPLANATORY VARIABLES (BINARY)					
MALEHH	Equals 1 if the household's head is male	0.746		0.828	
HIGHEDU	Equals 1 if the household's head has at least a high school education, zero otherwise	0.286		0.294	
WHITECOLLAR	Equals 1 if the household's head is in a white collar occupation, zero otherwise	0.186		0.208	
OWNEROCC	Equals 1 if the household owns its home, zero otherwise	0.743		0.702	
SINGLE	Equals 1 for a single adult household without children, zero otherwise	0.290		0.207	
CHILD014	Equals 1 if any child aged 0-14 is present in the household, zero otherwise	0.243		0.290	
HIGHALC	Equals 1 if the household consumes a high level (over the 75th percentile of the observed distribution) of alcoholic beverages, zero otherwise	0.555		0.658	

Table 2 – Diagnostic tests

Model	Homoscedasticity	Normality
Tobit	90.82 (2) [0.000]	134.603 (2) [0.000]
Double-hurdle	825.2 (3) [0.000]	138.122 (2) [0.000]

Note: the degrees of freedom of each χ^2 statistic are reported in round brackets while the p-value of each test is in squared brackets.

Table 3 – Specification tests

Model	Test type	Test value
Box-Cox dependent double-hurdle vs. Box-Cox independent double-hurdle	LR	0.78 (1) [0.3771]
Box-Cox independent double-hurdle vs. Box-Cox Tobit	Vuong	21.074*
Box-Cox independent double-hurdle vs. Independent double-hurdle	Vuong	38.428*
Independent double-hurdle vs. Tobit	Vuong	12.311*

Note: the degrees of freedom of the χ^2 statistic of the LR test are reported in round brackets while the corresponding p-value is in squared brackets. In the Vuong tests, the asterisk indicates that the null hypothesis of model equivalence is rejected at the 1% significance level.

Table 4 – Heteroscedastic double-hurdle estimates

Variable	Heteroscedastic Double-Hurdle Model			Box-Cox Heteroscedastic Double-Hurdle Model		
	Participation	Consumption	Het.	Participation	Consumption	Het.
AGE	-0.0032 (5.60)	-0.9120 (7.08)	–	-0.0101 (16.42)	-0.0560 (9.02)	–
AGESQR	–	0.0067 (5.66)	–	–	0.0004 (7.84)	–
MALEHH	-0.0453 (4.48)	-2.2069 (2.26)	–	-0.0704 (6.52)	-0.0884 (2.13)	–
HIGHEDU	-0.0354 (4.43)	0.0886 (0.14)	–	-0.0327 (3.91)	-0.0297 (0.98)	–
WHITECOLLAR	-0.0035 (0.39)	-2.5902 (3.48)	–	-0.0182 (1.78)	-0.1666 (4.86)	–
PERCMALE	0.1764 (8.49)	3.2854 (2.59)	–	0.2461 (11.93)	0.1646 (2.66)	–
SINGLE	-0.1071 (8.57)	23.5328 (33.75)	–	-0.0804 (8.28)	1.1662 (30.94)	–
CHILD014	-0.0451 (4.36)	1.6787 (2.36)	–	-0.0435 (3.87)	0.1237 (3.89)	–
OWNEROCC	–	-4.6780 (8.07)	–	–	-0.2355 (8.46)	–
INCOME	–	12.8827 (8.05)	0.2489 (15.00)	–	0.6113 (14.34)	0.0476 (2.77)
INCOMESQR	–	-3.1280 (6.22)	-0.0078 (6.55)	–	-0.0819 (8.75)	0.0047 (1.97)
ALCOHOL	–	0.0994 (2.84)	0.005 (5.49)	–	0.0077 (9.40)	–
HIGHALC	0.1014 (8.51)	–	–	0.1309 (11.67)	–	–
CONSTANT	0.0261 (1.21)	27.1526 (7.57)	–	0.6647 (16.70)	4.5645 (26.40)	–
Box-Cox parameter (λ)					0.1864 (17.50)	
Log Likelihood		-50881.02			-31097.36	

Note: the absolute value of each t-statistic is reported in round brackets.

*Table 5 – Elasticities with respect to continuous variables
and effects of binary variables*

Variables	Probability	Conditional level	Unconditional level
Continuous variables			
AGE	-0.2551***	-0.0748***	-0.3299***
	(0.0155)	(0.0040)	(0.0134)
PERCMALE	0.0339***	0.0104***	0.0443***
	(0.0042)	(0.0022)	(0.0011)
INCOME	–	0.2638***	0.2638***
		(0.0184)	(0.0184)
ALCOHOL	–	0.0025***	0.0025***
		(0.0003)	(0.0003)
Discrete variables			
MALEHH	-0.0272***	-0.8969***	-0.7942***
	(0.0104)	(0.0561)	(0.0426)
HIGHEDU	-0.0127***	-0.2952	-0.3470*
	(0.0058)	(0.2460)	(0.2015)
WHITECOLLAR	-0.0071	-1.4820***	-0.5296**
	(0.0088)	(0.4940)	(0.2493)
SINGLE	-0.0313***	3.1230***	1.3693***
	(0.0094)	(0.0919)	(0.0437)
CHILD014	-0.0169**	1.2554***	0.7603***
	(0.0085)	(0.0523)	(0.0335)
OWNEROCC	0.0462***	-2.1037***	-0.5305***
	(0.0067)	(0.0915)	(0.0286)
HIGHALC	0.0506***	2.0469***	1.0921***
	(0.0062)	(0.1279)	(0.0341)

Notes: Asymptotic standard errors of estimated elasticities and discrete effects are reported in round brackets. Asterisks indicate levels of significance: *** = 0.01, ** = 0.05 and * = 0.10.

Figures

Figure 1 – Nested Models

Box-Cox Double-Hurdle
model with dependence

$\rho = 0$

Box-Cox Double-Hurdle
model with independence

$\lambda = 0$ $\alpha_j = 0 \ \forall \ j \neq 0$

Standard Double-Hurdle
model with independence   Box-Cox Tobit model

$\alpha_j = 0 \ \forall \ j \neq 0$ $\rho = 0$

Standard Tobit model