THE PROBABILITY OF NON-PURCHASING TOBACCO OF A SMOKER

Daniel Miles

98-58

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 98-58
Statistics and Econometrics Series 27
October 1998

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34)91 624-9849

THE PROBABILITY OF NON-PURCHASING TOBACCO OF A SMOKER

Daniel Miles*

Abstract

This paper addresses the problem of zeroes in tobacco expenditure. Generally, tobacco demand is estimated using limited dependent variable models, i.e. Tobit or Double Hurdle Models, which take into account the zero expenditure problem under the assumption that a relatively important number of smokers declared a zero in tobacco expenditure. Clearly, if all zeroes where from non-smokers then demand estimation could be done using traditional methods over the positive expenditure observations. Based on the Spanish Expenditure Survey we estimate the conditional probability of non-expenditure by a smoker, finding that such probability is extremely small. This suggests that smokers buy quite regularly and hence it is possible to estimate the tobacco demand using only the positive observations.

Keywords: Zeroes in tobacco expenditure; limited dependent variable models; count regression models; Spanish Expenditure Survey.

*Miles, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Madrid. Spain. Research funded by Spanish Dirección General de Enseñanza Superior, referencia PB95-0292 . This research is part of my doctoral dissertation at Universidad Carlos III de Madrid. I will like to thank the comments of Miguel A. Delgado.

1. INTRODUCTION

Household budget surveys usually recover information on expenditures for a very short period of time and consequently recorded expenditures are often zero. These zero expenditures are interpreted in different ways. One one side, for goods such as cloth, a zero arises because of the infrequency of purchases, i.e. given the short duration of the survey the household decides not to purchase, although been a consumer. On the other side, for goods such as tobacco, these zero expenditures are commonly interpreted either as resulting from a non-smoker household or as a corner solution of the utility maximization problem for a smoker. In general, no information is available in the surveys so as to infer the reasons that gave up the zero expenditure in tobacco and this fact complicates its demand estimation.

Usually, tobacco demand estimation has been achieved using limited dependent variable models, e.g. Tobit or Double Hurdle (Cragg, 1971; Jones, 1989; García and Labeaga, 1996, among others). These models take into account the fact that some obervations accumulate in the same value, e.g., zero, and rely on two implicit assumptions. In first place, these models assume that there is a relative important percentage of smokers that had reported a zero expenditure in tobacco, although not being able to test this assumption[1]. In second place, the same decision process in zero is assumed for non-smokers and smokers, for whom a zero can only correspond to a corner solution (Blundell and Meghir, 1987; Pudney, 1989)[2].

[1]Clearly, if only an irrelevant percentage of zeroes correspond to smokers, then it seems feasible to estimate the demad of tobacco using only positive observation. Following Pudney (1989), the Tobit or Double Hurdle models are only ad hoc modifications of traditional regression methods for cases where some observations accumulate in some value, i.e., zero.

[2]Note that a zero in the Double Hurdle model is also interpreted as a corner solution. The first hurdle characterizes the differences between a smoker and a non smoker while the second is

In this paper we question the mechanic application of Tobit models for estimating tobacco demand by discussing both of these assumptions. First, we argue that it seems more reasonable to interpret a zero expenditure of a smoker as infrequency of purchase rather than as a corner solution. That is, smokers declaring zero in the survey consume from their stock, which implies that they should have stockpiled cigarettes the weeks before to the survey. Second, using Robin's (1993) approach we estimate the conditional probability of non-purchasing for a smoker. For a sample of the Spanish Expenditure Survey 1990-91 we find that the percentage of smokers declaring zero is extremely small. This result is in contradiction with the implicit assumption supporting the Tobit model, which states that a relative important number of smokers declare a zero expenditure, and suggests that it is feasible to estimate the tobacco demand using only the positive observations.

The data used in this paper was obtained from the Spanish Expenditure Survey 1990-91 (EPF; Cardelús et al. 1995). This survey reports information on the quantity of packs of cigarettes bought during the week of the survey as well as the frequency of purchases. We have selected a subsample where the head of the household is employed and with ages between 15 and 65 years, and considered exclusively cigarette expenditure, having a total of 10.009 observations[3].

The paper is organized as follows. In the first section we study the percentage of smokers that in the week of the survey had stockpiled cigarettes. Given that zeroes from smokers arise because of infrequency of purchase, some smokers should have accumulated tobacco during the week of the survey. We find that the percentage of smokers that stockpiled during the survey is relatively small.

Notice that if we assume that the cigarette purchase process is stationary and

interpreted as a corner solution for the smoker (Pudney, 1989).

[3]We have also eliminated those households with more members than the mean plus three standard deviations, and the same for total expenditure.

that the smokers who stockpile are uniformly distributed along the weeks of the year, then the percentage of smokers that had stockpiled during the weeks before the survey could be approximated by those who had stockpiled during the week of the survey.

In the second section we follow the approach of Robin (1993) to estimate the conditional probability of non-purchasing for a smoker. Robin proposed a method for identifying the probability of non-purchasing by a consumer. The originality of our paper arises from the fact that, in tobacco empirical literature, the probability of non-purchasing for a smokers has never been estimated before. Following this procedure we conclude that only an extremely small proportion of smokers had declared a zero during the survey. Given this last result, in the third section we estimate the cigarette Engel curve and observe that the income elasticity is lower than the one obtained using limited dependent variable models.

2. STOCKPILING BEHAVIOUR

It seems more reasonable to understand a zero in tobacco expenditure as a consequence of the infrequency of purchase rather than as a corner solution. First, cigarette consumption generates addiction which implies a relative inelastic demand equation with few close substitutes (Becker et al., 1993). Second, addiction implies that a smoker will try to buy a pack of cigarettes, maybe giving up other small habits. Third, the share of non-durable expenditure allocated in tobacco is relatively small, i.e. less than 3% for 1980-81, 1990-91 considering only positive expenditure. Therefore, changes in the tobacco relative price should be very important so as to push a smoker towards a corner solution, given its preferences[4].

In this paper we assume that a zero in tobacco expenditures is a consequence of the smoker purchasing process[5]. Hence, the differences in the purchasing decisions

[4]In Spain the relative price of tobacco did not change substantially during the year 1990-91.

[5]Given that the budget surveys are focused on households, here a smoker is a household which

between smokers will determine whether they declare or not a zero expenditure. On one hand there are those smokers that buy tobacco regularly so they will always declare a positive expenditure. On the other, some smokers prefer to stockpile packs of cigarettes and consume from their stock[6]. These last type of smokers will buy tobacco in large quantities in very few visits to the shop, and so, being the potential smokers declaring a zero in the survey.

Notice that, if we assume that the cigarette purchase process is stationary and that the smoker that stockpile are uniformly distributed along the weeks of the year, the percentage of smokers that had stockpiled during the weeks before the survey could be approximated by those who had stockpiled during the survey. Therefore, in this section we will be interested on those smokers that had stockpiled during the week of the survey.

For this, we first discuss the relationship between the quantity of packs and the number of purchases for those smokers that declare a positive expenditure. Second, given that the stock duration depends on the level of consumption, we analyze the mean consumption of cigarettes per smoker household. At last, we analyze the stability of the purchasing process during the year. We conclude that those smokers that had stockpiled tobacco during the survey are a small percentage of the total smokers that reported a positive expenditure. Therefore, using the reasoning of the last paragraph, it seems that the percentage of smokers that declared a zero in this survey is relatively small.

First, in the subsample used in this paper, 32% of the households declared a zero expenditure in cigarettes. That is, only three of ten households declared a zero expenditure in tobacco [7].

declares a positive expenditure in tobacco, as is common in this literature.

[6]These behaviour could be interpreted as if these smokers suffer a transpotation cost to the tobacco shop larger than the stockpiling cost and hence, they prefer to accumulate tobacco.

[7]Using the Spanish National Health Survey of 1993 (Encuesta Nacional de Salud, 1993) we find

A smoker that stockpiles is the one that goes few times to the tobacco shop and purchases large quantities of cigarettes. In the following table we present the joint frequency of the quantity and number of purchases for those smokers that declared a positive expenditure.

TABLE 1: Joint Frequency of Quantity and Number of Purchases for Positive Expenditure .

Number of Packs Week	Number of Purchases per Week								Total Packs
	1	2	3	4	5	6	7	>7	
1	4.14								4.14
2	1.03	5.75							6.78
3	0.19	4.36	4.15						8.70
4	0.12	2.99	1.16	4.70					8.97
5	0.06	1.25	0.62	1.64	3.14				6.71
6	0.03	0.41	0.53	1.31	0.94	4.33			7.55
7	0.01	0.10	0.10	0.83	0.57	1.41	9.65		12.67
8-10	2.32	0.09	0.19	0.84	0.87	2.50	4.17	6.39	17.37
11-12	0.09	0.9	0.57	0.09	0.20	0.94	1.08	3.95	7.83
13-15	0.00	0.16	0.74	0.54	0.45	0.59	3.59	5.88	11.95
16-20	0.00	0.22	0.07	0.19	0.17	0.37	0.65	3.80	5.47
>20	0.00	0.01	0.03	0.06	0.01	0.00	0.01	1.74	1.86
Total Purchases	8.0	16.2	8.2	10.2	6.4	10.1	19.2	21.7	

Notice that only 8% of these smokers went once to the tobacco shop and, from these, close to 30% bought 8 or more packs of cigarettes. That is, less than 3% of the smokers with a positive expenditure went only once to the tobacco shop buying a relatively important quantity of cigarettes.

that for a similar subsample 50% of the interviewed persons did not smoke, 33% declared never smoked while 17% been an ex-smoker. The Encuesta Nacional de Salud, 1993, was developed by the Centro de Investigaciones Sociológicas. The interviews were to individuals giving a total of 21061 adults observations.

Also, from the table above it can be observed a mode in the diagonal of the table, suggesting that most smokers bought one or two packs of cigarettes per visit to the shop. Given that more than 90% of the smokers went more than once to the tobacco, this means that most of the smokers visit regularly the tobacco shop buying small quantities of tobacco. To discuss this fact with more detail, in Table 2 we present the mean ratio of quantity of packs to the number of purchases.

TABLE2: Positive Expenditure Subsample: Distribution of the ratio between quantity and number of purchases per week.

						% Packs/Purchases		
N. Purchases	Mean	Q_{25}	Q_{50}	Q_{75}	Máx.	1-2	3-4	≤ 5
1	4.00	1.00	1.00	10.0	12.0	64.6	3.9	31.5
2	1.96	1.00	1.50	2.00	12.5	80.7	11.5	7.8
3	1.82	1.00	1.00	2.00	8.00	79.2	10.6	10.2
4	1.51	1.00	1.25	1.75	6.25	88.4	10.8	0.8
5	1.43	1.00	1.20	1.60	4.60	86.4	13.4	0.2
6	1.38	1.00	1.17	1.50	3.33	90.6	9.40	0
7	1.32	1.00	1.00	1.57	3.14	95.2	4.80	0
8>	1.25	1.00	1.12	1.38	2.89	95.9	4.10	0
Total	1.69	1.00	1.17	1.67	12.5	87.6	7.62	4.78

Note: The mean quantity of packs per purchase is defined as the ratio between the total quantity of packs bought per week to the total number of purchases. Q_{25},Q_{50},Q_{75} represent the first, second and third quartil.

Again, only a small percentage of smokers bought a relatively important quantity of cigarettes in each visit to the shop, i.e., only 4.78% of the smokers bought, in mean, 5 or more packs of cigarettes per purchase. From the table above is clear that most smokers bought small quantities per visit to the shop, i.e. nearly 90%, bought between one or two packs of cigarettes. That is, it seems as most smokers do not stockpile cigarettes, visiting regularly the tobacco shop. This suggests that they will usually declare a positive expenditure during the survey.

Naturally, stockpiling cigarettes should be defined in terms of the level of consumption. It is not the same a stock of 10 packs for a smoker that consumes two packs a day than for another that smokes only a couple of cigarettes a day. Let Q_h be the total packs of cigarettes bought by smoker h during the week of the survey and N_h the number of times he went to the shop, therefore $(20Q_h/N_h)$ is the consumption of cigarettes in the interpurchase period, being 20 the number of cigarettes per pack. Now, if we assume that purchases are done equally spaced between the days of the week, the number of purchases per day is given by $(N_h/7)$ and, in a trivial way, the consumption per day is given by $(20Q_h/N_h)(N_h/7)$, or $(20Q_h/7)$ (see Kay, Keen and Morris, 1984) [8].

In table 3 we present the daily number of cigarettes consumed for those smokers that have declared a positive expenditure and had done between two and 7 purchases in the week.

TABLE 3: Consumption of Cigarettes per day, defined as the ratio between the number of packs bougth to the number of purchases, times 20, the number of cigarettes per pack.

Number Purchases	Mean	Q_{25}	Q_{50}	Q_{75}	Min	Max
2	11	6	9	11	6	71
3	16	9	9	17	9	68
4	17	11	14	20	11	71
5	20	14	17	23	14	66
6	24	17	20	26	17	57
7	26	20	20	31	20	63
Total	23	11	20	31	6	71

Note: Round to near integer.

[8]Note that the assumption of equally spaced purchases is restrictive only for those smokers that bought two times in a week. If some of these households bought the first and last day of the week, then it si not true that they consume their stock every 3.5 days, which is what is implied by our assumption of equally dispersed purchases.

Observe that a smoker household consumes approximately one pack of cigarettes per day (20 cigarettes), in mean, with a maximum of approximately 4 packs and a minimum of six cigarettes a day[9]. From the National Health Survey (1993; NHS), which directly asks the level of consumption of cigarettes, it is found that daily smokers consume a mean of one pack a day with a maximum of 4 packs a day. That is, the results that we obtain by our approximation are similar to those derived from a survey that explicitly asks for the level of cigarette consumption. Note that this results suggest that those smokers that by more than once a week will consume these packs during the week, and therefore, they do not stockpile tobacco.

From this same survey non habitual smokers consume, in mean, 4 cigarettes a day, and, therefore, a pack of cigarette lasts, at most, one week for these non habitual smokers. If in Table 1 we use this level of cigarette consumption for those smokers that only bought once a week we find that, at most, 4% of the total smokers declaring a positive expenditure had stockpiled tobacco during the week of the survey[10]. Again, it seems that only a small percentage of smokers had stockpiled.

At last, to observe that the purchasing process is stable through time we have constructed a time series of the weeks where the survey took place. The survey in Spain was held during 48 weeks from April 1990 to March 1991. Using these weekly information in the following graphs we present the order statistics of the "time series of weekly cross sections"

In Graph 1 we present the order statistics for the quantity of packs bought per

[9]Notice that when we are talking about a smoker we are really referring to a smoker household. This is the usual way in which the empirical literature related with tobacco defines smokers when the data is obtained from budget surveys.

[10]That is, with a consumption of 4 cigarettes a day a pack lasts one week, at most. From Table 1 we observe that 4.14% bought one pack in one visit and, given that it lasts one week, this smokers buy tabacco every week if we assume that the purchasing process and consumption are stable throught time.

week and the number of purchases done, for the whole year.

Insert Graph 1

As it can be observed from the graph, these statistics are relatively stable along the year. In graph 2 we present the same statistics for the share of non-durable expenditure allocated in cigarettes and total non-durable expenditure.

Insert Graph 2

These graphs suggest that smoker's consumption behaviour is stable through time. Note that the only important tobacco price increment along the year was in January 1991, being nearly of 5% (week from 35 to 40). Despite of it, we do not observe important changes neither on quantities purchased nor on the share of expenditure assigned to tobacco. That is, it does not seem like smokers were pushed towards a corner solution by this price increase[11].

Hence, if zeroes arise because of infrequency of purchase and if behaviour is stable through time, it is possible to approximate the percentage of smoker that declared zero with that of those who stockpiled during this survey. From the discussion above, the percentage of smokers that had accumulated in this survey is relatively small, suggesting then that the smokers that declared zero are a small percentage of total smokers.

In the next section we estimate the probability of non-purchasing for a smoker, following Robin (1993).

3. PROBABILITY OF NON-PURCHASING

In this section, using the results of Robin (1993), we estimate the percentage of smokers that declared a zero expenditure in tobacco.

[11]The Spanish inflation during the 1990 was above 5%, suggesting that there were no important tobacco relative price changes during the year.

Let N_h be the number of purchases per week by household h, where $N_h > 0$ if the household bought tobacco and $N_h = 0$ if it did not. Following Robin (1993), let ε_h be unobservable characteristics that determine whether household h is a smoker or not. That is, given a set of observable characteristics, z_h, there could be other which are unobservable (heterogeneity), ε_h, which determine whether the household is a smoker, e.g. given a realization ε, the expected value of expenditure is zero $E(N_h \mid z_h, \varepsilon_h = \varepsilon) = 0$. In the same way, let C be those unobservable characteristics of an smoker, defined by

$$C(z_h) = \{\varepsilon \mid E(N_h \mid z_h, \varepsilon_h = \varepsilon) > 0\}.$$

If a household declares a positive expenditure is a consumer with probability one. Hence, the probability of the number of purchases, given that is a smoker and had purchased, is equal to the probability of the number of purchases given that it had purchased,

$$\Pr(N_h \mid z_h, \varepsilon_h \in C(z_h), N_h > 0) = \Pr(N_h \mid z_h, N_h > 0).$$

That is, the left hand side of the last equality is identifiable using the set of positive observations of N_h.

Now, assuming a parametric distribution function of N_h, $F(n_h)$, such that its complete distribution is recoverable from the truncated one, $F(n_h \mid n_h > 0)$, the probability of not purchasing been a consumer, $\Pr(N_h = 0 \mid z_h, \varepsilon_h \in C(z_h))$, becomes identifiable (see Robin, 1993, pg. 926; Flinn and Heckman, 1982). The, as $\Pr(N_h = 0 \mid z_h)$ is identifiable from the proportion of zero expenditures on total observations, the probability of consuming

$$\Pr(\varepsilon_h \in C(z_h) \mid z_h) = \frac{\Pr(N_h > 0 \mid z_h)}{\Pr(N_h > 0 \mid z_h, \varepsilon_h \in C(z_h))}. \tag{1}$$

is identifiable.

As Robin states, the parametric assumption is a sufficient condition for identification, allowing to estimate the complete distribution using a truncated sample of observations. Therefore, the information contained in the number of purchases parametrically identifies the proportion of households with the same observable characteristics which do not purchase because they are non-consumers

Given the information in the survey, it amounts assuming a recoverable parametric density to identify the percentage of smokers that had declared zero in the subsample (Flinn and Heckman 1982; Robin, 1993). For example Robin (1993), with data on the frequency of purchase on different types of food found that from near a 7% of zeros in bread purchases, 5% corresponds to non-consumers. That is, only 2% of bread consumers declared zero.

To estimate the probability of non-purchasing it is necessary to estimate the truncated distribution function of the number of purchases. For this, we assume that the purchasing process is distributed as a Negative Binomial distribution

$$\Pr\left(N_h = n_h \mid z_h, \varepsilon_h \in C\left(z_h\right)\right) = \frac{\Gamma\left(n_h + \delta\right)}{\Gamma\left(n_h + 1\right)\Gamma\left(\delta\right)} \left(\lambda_i/\delta\right)^{n_h} \left(1 + \lambda_i/\delta\right)^{-(n_h+\delta)}.$$

Now, given that $\Pr\left(N_h = n_h \mid z_h, \varepsilon_h \in C\left(z_h\right), N_h > 0\right) = \Pr\left(N_h = n_h \mid z_h, N_h > 0\right)$, the truncated distribution is specified as

$$\Pr\left(N_h = n_h \mid z_h, N_h > 0\right) = \frac{\Gamma\left(n_h + \delta\right)}{\Gamma\left(n_h + 1\right)\Gamma\left(\delta\right)} \left(\lambda_i/\delta\right)^{n_h} \left(1 + \lambda_i/\delta\right)^{-(n_h+\delta)} \left(1 - \left(1 + \lambda_i/\delta\right)^{-\delta}\right)^{-1},$$

where we correct by $\Pr\left(N_h > 0 \mid z_h, \varepsilon_h \in C\left(z_h\right)\right) = 1 - \left(1 + \lambda_i/\delta\right)^{-\delta}$ following Groger and Carson (1991) or Creel and Loomis (1990).

We have then

$$E\left(N_h \mid z_h, N_h > 0\right) = \lambda_i \left(1 - \left(1 + \lambda_i/\delta\right)^{-\delta}\right)^{-1}$$

where, as usual, we assume that

$$\lambda_i = \exp\left(z_h'\beta\right),$$

where β are the parameters of interest (see Cameron and Trivedi, 1996 or Lawless, 1987, among others).

In Table A.1 in the appendix we present the estimation results for the number of packs of cigarettes bought during the week of the survey and the number of purchases, using a negative binomial for the whole sample and a truncated one.

Following Robin (1993), from these results we can estimate the probability that a smoker household purchases tobacco as

$$\widehat{P}\left(N_h > 0 \mid z_h, \varepsilon_h \in C\left(z_h\right)\right) = 1 - \frac{1}{n}\sum_{i=1}^{n}\left(1 + \widehat{\lambda}_i/\widehat{\delta}\right)^{-\widehat{\delta}} \tag{2}$$

where $\widehat{\lambda}_i$ and $\widehat{\delta}$ are the estimated parameters from the truncated binomial density function (Grogger and Carson, 1991; Robin, 1993). With this estimation and given that $P\left(N_h > 0 \mid z_h\right)$ is identifiable from the proportion of positive to total observations, we could estimate the probability of being a nonsmoker from equation (1), presented in table 4. We have applied Robin method using both, the number of purchases and the quantity of packs bought as dependent variables in (2)

Table 4 Estimation of the Probability of Non-purchasing of Smokers .

	Probability of Purchase of a Smoker[1]	Purchase in the Sample[2]	Probability NonSmoker[3]	Zeros in the Sample[4]	Probability Smokers Zero[5]
Purchase	97.80%	67.80%	30.67%	32.20%	1.53%
Quantity	98.78%	67.80%	31.36%	32.20%	0.84%

Note: Purchase or Quantity refers to the dependent variable used for estimating the negative binomial for the application of Robin approach. 1. $\left(1 - \sum_{i=1}^{n}\left(1 + \widehat{\lambda}_i/\widehat{\delta}\right)^{-\widehat{\delta}}/n\right)$; 2. $\sum_{i=1}^{n} I\left(N_h > 0\right)/n$; 3. $\left(1 - \frac{\widehat{\Pr}(N_h>0|z_h)}{\widehat{\Pr}(N_h>0|z_h,\varepsilon_h\in C(z_h))}\right)$; 4. $\sum_{i=1}^{n} I\left(N_h = 0\right)/n$; 5. $\left[\left(\sum_{i=1}^{n} I\left(N_h = 0\right)/n\right) - \left(1 - \frac{\widehat{\Pr}(N_h>0|z_h)}{\widehat{\Pr}(N_h>0|z_h,\varepsilon_h\in C(z_h))}\right)\right]$

The first column in the table above estimates the probability that a smoker buys tobacco during the week of the survey. For the case where the number of purchases

is used as the dependent variable in Robin's approach, 97.8% of the smokers will buy during the week of the survey or 98.8% if we use the information contained in the number of packs bought. In the second column we present the proportion of positive observations to total observations in the sample. The third column is the probability of been a non-smoker, obtained as the complement of equation (1), from where it results that nearly one third of the sample is non-smoker. The fourth column states the number of zeroes in the sample. At last, the fifth column is the difference between columns four and three. That is, the proportion of zeroes that correspond to smokers or equivalently, the probability that a smoker does not purchase.

As it can be observed, less than 2% of the smokers had declared a zero during the survey. That is, an extremely small percentage of smokers declared a zero during the survey. This result is in contradiction with the common implicit assumption supporting the Tobit models, which states that a relative important percentage of smokers had declared a zero. A possible interpretation of this result could be find in the cigarette addiction and on a small transportation cost to the shop, which makes smokers buy cigarettes regularly, not stockpiling. Therefore, it looks as if smokers do not declare zero in tobacco expenditures in the surveys.

4. THE TOBACCO ENGEL CURVE

Given the result of last section, we had estimated the Engel curve for tobacco using only the positive observations, given that practically all zeros arise from non smokers.

First, in Graph 3 we present the nonparametric regression of the proportion of expenditure allocated in tobacco, evaluated at 1990 prices. Note that, given that during the week of the survey the smokers consume all the packs of cigarettes they purchase, expenditure measures consumption. This means that we do not have a measurement error problem in the Engel curve estimation. Also, given that addiction is relatively constant in time and that the budget share in tobacco expenditure is

small, we can also ignore the endogeneity problems of total expenditure[12].

Insert Graph 3

Notice that during the eighties the proportion of expenditure allocated to tobacco has increased, for a same level of total expenditure. Also, the slope of the non-parametric Engel curve has increased for 1990, indicating an important increment in the income elasticity. At last, for both years, it seems that a Working-Leser seems to be correct specification for the Engel curve of tobacco.

We have estimated the Engel curve for 1990-91, presenting the results in Table 6.

[12]The tobacco Engel curve could be thought as been derived from a quasi-linear utility function. Given that the budget allocated in tobacco is small, total expenditure can be thought as given exogenously.

Table 6 Tobacco Engel Curve

Variable	Coef.	Stdev	Variable	Coef.	Stdev
Constant	13.6969	2.6526	Estudy	10.0882	1.7829
North	-0.1809	0.0614	Week 1	4.1468	1.8728
South	0.1356	0.0612	Week 2	0.7288	1.8053
East	-0.0657	0.0690	Week 3	0.4286	1.9904
Madrid	-0.0830	0.0818	Age Partner	-0.0240	0.0032
Catalunya	0.1390	0.0898	Women No W	-2.0043	1.6480
Town	3.3570	1.5954	Food Out	0.0312	0.0149
Service	2.1120	1.4987	Log Exp	-0.7840	0.1768
Industry	0.2776	0.0987	ExpTown	-0.2178	0.1069
Construction	0.3500	0.1100	ExpServ	-0.1197	0.1011
Manual	-0.4333	0.1365	ExpStud	-0.6588	0.1180
Blue Collar	0.4283	0.1242	ExpWeek1	-0.2822	0.1257
Child less 8	-0.0752	0.0542	ExpWeek2	-0.0548	0.1214
Child 8-17	-0.1644	0.0412	ExpWeek3	-0.0316	0.1340
Child 17-25	0.1522	0.0446	ExpWomenNW	0.1307	0.1101
Size Household	0.0930	0.0377			

Note: Values multiplied by 100

First, the sign of the estimated parameters is similar to that obtained from the estimation of the univariate negative binomial model (see appendix). Second, the income elasticity, evaluated at the mean of the household characteristics is 0.33 and in the median is of 0.28[13].

[13]If we compare these elasticities with that obtained using limited dependent variable models, i.e., using Spanish data, García and Labeaga (1996) find an elasticity of 0.72 for 1980-81 data, where the one obtained from the Engel curve is much smaller.

CONCLUDING REMARKS

Zero expenditures introduces several complications when estimating demand equations. If all zeroes correspond to non-consumers, then demand equation can be consistently estimated using only the positive observations. But, if the zeroes are reported either by non-consumers as well as by consumers, as is the case of tobacco, care should be taken in the way the zero expenditure is treated.

Usually, the treatment of the zero expenditure in tobacco demand is undertaken using limited dependent variable models. Implicitly, this models assume that there is a relatively important number of smokers that declare a zero expenditure in tobacco. In this paper we are concerned with discussing whether smoker household tend to declare zero in tobacco expenditure in the budget surveys. We find that only a extremely small percentage of smokers declared a zero expenditure in the survey. Hence, it is feasible to estimate its demand equation using only positive observations, allowing a direct interpretation of the estimated demand parameters.

REFERENCES

[1] Becker, G., Grossman, M. and Murphy, K. (1994) "An Empirical Analysis of Cigarette Addiction", *The American Economic Review,* 84(3) , 396-418.

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

[3] Cardelús, M., Arévalo, R. and Ruiz Castillo, J. (1995) "La Encuesta de Presupuestos Familiares de 1990-91", Documento de Trabajo 95-07(05), Universidad Carlos III de Madrid.

[4] Cragg, J. (1971) "Some Statitical Models for Limited Dependent Variables with Application to the Demand of Durable Goods", *Econometrica,* 39, 829-44.

[5] Creel, M. and Loomis, J. (1990) "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California", *Journal of Agricultural Economics,* 72, 434-41.

[6] Garcia, J. and Labeaga, J. (1996) "Alternative Approaches to Modelling Zero Expenditure: An Application to Spanish Demand for Tabacco", *Oxford Bulletin of Economics and Statistics,* 58(3), 489-506..

[7] Grogger, J. y Carson (1991) "Models for Truncated Counts", *Journal of Applied Econometrics,*6, 225-38.

[8] Jones, A. (1989) " A Note on Computation of the Double-Hurdle Model with Dependence with an Application to Tobacco Expenditure", *Bulletin of Economic Research* 44(1) , 67-74.

[9] Kay, J., Keen, M. and Morris, C (1984) "Estimation Consumption from Expenditure Data", *Journal of Public Economics,* 23, 161-81.

[10] Lawless, J. (1987) "Negative Binomial and Mixed Poisson Regression", *The Canadian Journal of Statistics,* 15(3), 209-25.

[11] Mullahy, J. (1997) "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior", *The Review of Economics and Statistics,* 586-93.

[12] Pudney, S. (1989) "*Modelling Individual Choice: the Econometrics of Corners, Kinks and Holes",* Oxford: Basil Blackwell.

[13] Robin, JM. (1993) "Ecomometric Analysis of the Short-run Fluctuations of Household's Purchases", *Review of Economic Studies,* 60, 923-34.

[14] Winkelmann, R. (1997) *Econometric Analysis of Count Data,* Second Edition, Springer.

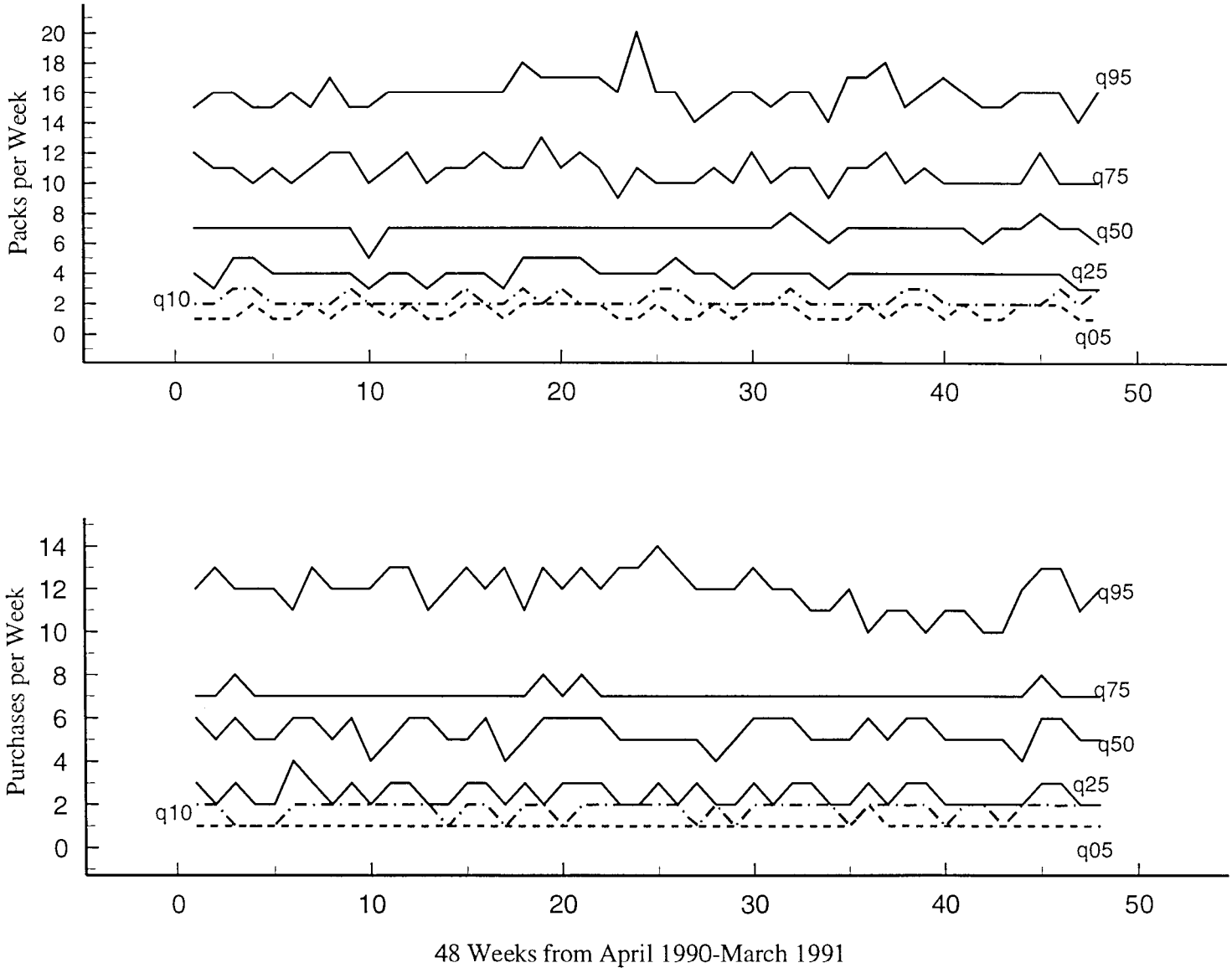
Table A1 Negative Binomial

	Number of Purchases					
	Total Sample		Positive Sample			
	Negative Binomial				Negative Truncated Binomial	
Variable	Coef.	ST	Coef.	ST	Coef.	ST
Constant	.7940	.2405	1.342	.1662	1.290	.1803
Size Household	.2027	.0171	.1179	.0123	.1267	.0131
Child less 8 Years	-.1657	.0253	-.1018	.0181	-.1104	.0195
Child 8-17 Years	-.1892	.0202	-.1105	.0145	-1191	.0155
Larger 25	.0936	.0192	.0575	.0143	.0607	.0151
Age Partner	-.0130	.0014	-.0055	.0010	-.0060	.0010
Town	.0700	.0220	.0614	.0157	.0659	.0170
Madrid	.0819	.0504	.0328	.0359	.0359	.0386
Catalunya	-.2047	.0435	-.0814	.0303	-.0884	.0331
South	.1531	.0275	.0703	.0197	.0757	.0212
North	-.0607	.0307	-.0391	.0219	-.0423	.0238
East	-.0399	.0345	-.0294	.0244	-.0316	.0265
Service	.2854	.0511	.1320	.0362	.1446	.0397
Industry	.2140	.0557	.0972	.0395	.1064	.0433
Construction	.2409	.0588	.1039	.0417	.1135	.0457
Manual	-.2752	.0659	-.1543	.0459	-.1671	.0498
Blue Collar	.3918	.0607	.2144	.0422	.2317	.0457
Study	.1690	.0364	.0970	.0259	.1062	.0283
Women not W	-.0593	.0257	-.0489	.0182	-.0533	.0197
Alcohol	.8E-04	.1E-04	.3E-04	.1E-04	.4E-04	.1E-04
Outside Food	.0483	.5E-02	.0069	.0035	.0071	.0037
LogIncome	-.0110	.0230	-.0014	.0159	-.0021	.0172
Alfa	.8396	.0191	5.757	.1868	4.441	.1693

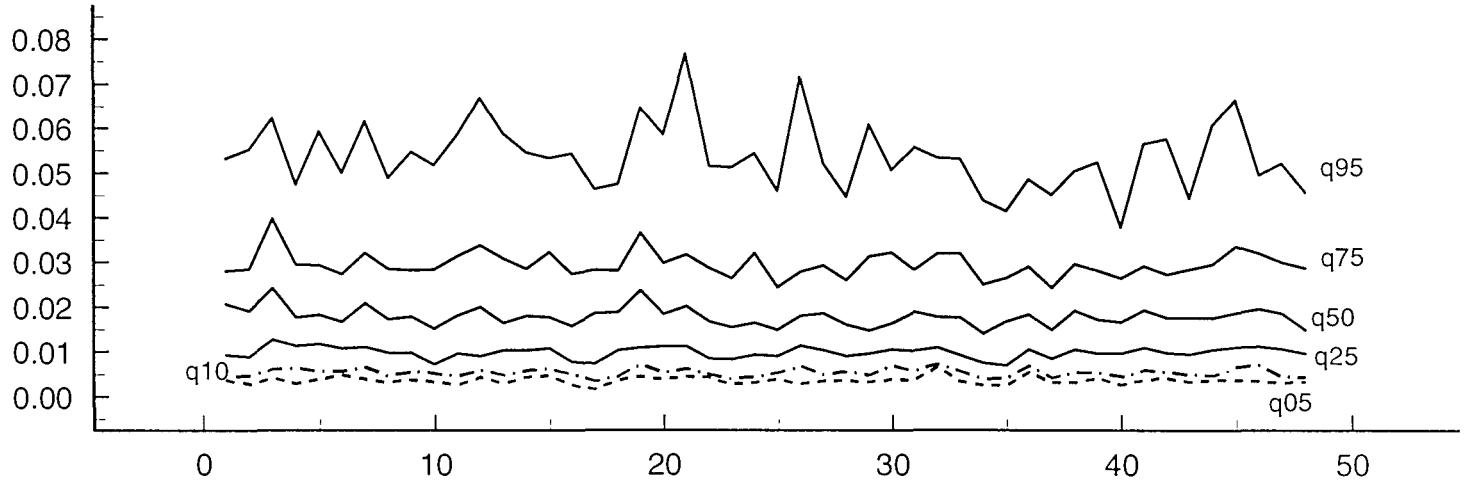
Table A1 continue

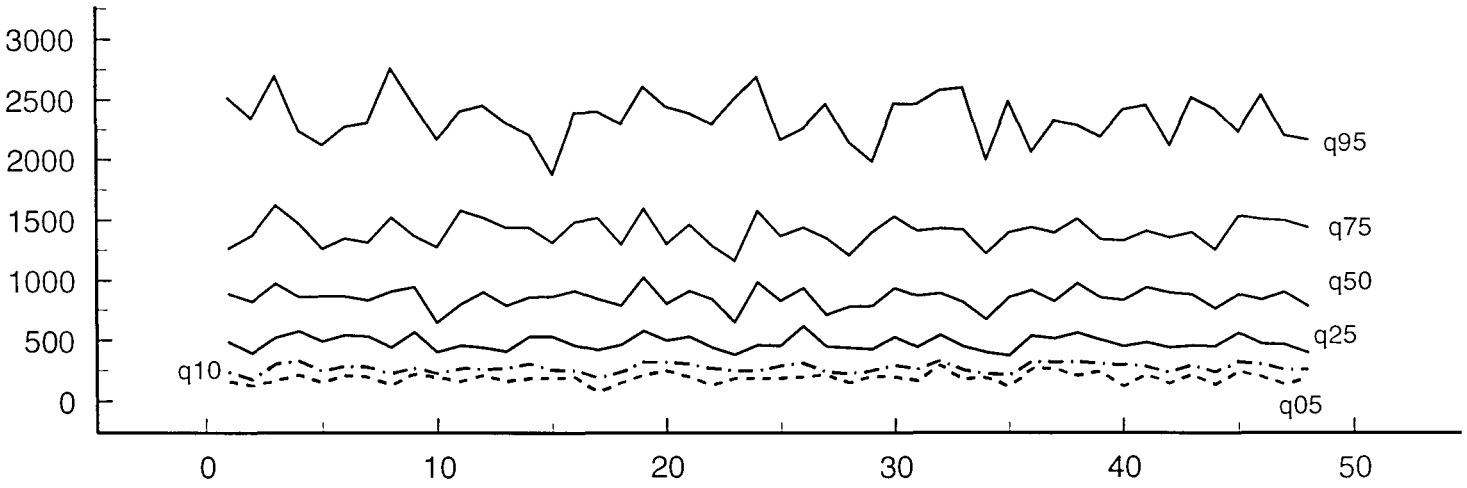
	Number of Packs					
	Total Sample		Positive Sample			
	Negative Binomial				Negative Truncated Binomial	
Variable	Coef.	ST	Coef.	ST	Coef.	ST
Constant	1.039	.2446	1.637	.1695	1.606	.1768
Size Household	.1728	.0172	.0886	.0123	.0921	.0127
Child less 8 Years	-.1252	.0254	-.0642	.0181	-.0668	.0188
Child 8-17 Years	-.1519	.0203	-.0760	.0146	-.0790	.0152
Larger 25	.0585	.0191	.0219	.0142	.0225	.0147
Age Partner	-.0123	.0014	-.0046	.0010	-.0049	.0010
Town	.0661	.0224	.0520	.0158	.0541	.0165
Madrid	.1054	.0513	.0657	.0369	.0680	.0382
Catalunya	-.1482	.0454	-.0256	.0320	-.0267	.0334
South	.1465	.0277	.0645	.0197	.0669	.0205
North	-.0979	.0310	-.0758	.0221	-.0791	.0232
East	-.0123	.0345	-.0048	.0241	-.0049	.0251
Service	.2192	.0510	.0652	.0353	.0684	.0370
Industry	.1926	.0560	.0717	.0389	.0751	.0408
Construction	.2126	.0590	.0753	.0412	.0787	.0431
Manual	-.2707	.0664	-.1516	.0464	-.1576	.0483
Blue Collar	.3678	.0609	.1896	.0423	.1970	.0441
Estudy	.0941	.0371	.0237	.0264	.0248	.0276
Women not W	-.0292	.0260	-.0256	.0183	-.0267	.0191
Alcohol	.1E-03	.1E-04	.5E-04	.1E-04	.5E-04	.1E-04
Outside Food	.0511	.0052	.0079	.0035	.0081	.0036
LogIncome	.0156	.0235	.0205	.0163	.0213	.0171
Alfa	.6669	.0143	4.284	.1021	3.784	.1057

Graph 1 Number of Packs and Frequency of Purchases per Week



Graph 2 Expenditure Share and Monetary Expenditure in Tobacco.





Semana de la Encuesta Abril 1990-Marzo 1991

Graph 3 Nonparamtric Tobacco Engel curve.