

How important are tobacco prices in the propensity to start and quit smoking? An analysis of smoking histories from the Spanish National Health Survey

Angel López Nicolás*

**Dept. Economics and Business and
Centre de Recerca en Economia i Salut (CRES)**

Universitat Pompeu Fabra

(This version, December 2001)

C/Ramón Trias Fargas, 25
08005 Barcelona
angel.lopez@econ.upf.es

I am grateful to Paul Contoyannis, Andrew Jones, David Madden, Edward Norton, Simon Peters, Joan R. Villalbi, two anonymous referees and audiences at the Ministerio de Sanidad y Consumo seminar on “Precio, Impuestos y Contrabando de Tabaco en España: Evolución y Relevancia para la Salud”, the XXV Simposio de Análisis Económico and the X European Workshop on Econometrics and Health Economics for useful comments and help. Part of this research was carried out during a Ministerio de Educación y Cultura funded visit to the Department of Economics and Related Studies of the University of York, whose warm hospitality is thanked. The usual disclaimer applies.

Introduction

With a view to retrieving parameters that measure the sensitivity of participation in the consumption of tobacco to a policy instrument such as prices, this paper investigates whether movements in tobacco prices over the period 1957-1997 have had a significant effect on the propensity to i) start smoking and ii) quit smoking in the Spanish population. The method used consists in analysing the retrospective information supplied by individuals in a pool of the 1993, 1995 and 1997 editions of the Spanish National Health Survey and estimating models for the duration up to smoking (for both smokers and non-smokers), and the duration of smoking (for smokers) where prices, among other covariates such as demographics and proxies for public health policies, enter the specification for the hazard function.

The evidence for the Spanish case has an interest that goes beyond that of local policy makers and academics, for the features of the Spanish tobacco market are somewhat different with respect to the average European Union country. Firstly, while the average price of a packet cigarettes has experimented a marked increase in the last two decades, a 20-unit packet of a representative brand still costs 85% as much as the average European Union price after adjusting for purchasing power disparities [1]. Secondly, the Spanish tobacco market is characterised by the coexistence of two types of product (which account for more than 95% of tobacco sales): black cigarettes and blond cigarettes. Although these varieties are very similar in terms of nicotine and tar content, availability, advertising and presentation, there exists a big average price gap between them. Moreover, within each of the black and blond groups there is also a good deal of price/product differentiation (Tabacalera [2] presents three price ranges in the blond class in their annual report). The existence of a wide price/product spectrum in the market is not unrelated to the fact that, as reported in Delipalla and O'Donnell [3], Spain is one of the EU countries where the specific component of tobacco taxes is smallest and therefore producer prices have a greater weight in the after-tax price. Thirdly, the scope of price/product differentiation in the Spanish market is augmented by the "tax free brand" that results from the high level of cigarette smuggling activities which, as Joosens and Raw [4] suggest, turns out to be among the highest in the European Union. In these circumstances it is interesting to analyse the effect of price changes on consumption. In particular, given the ample price range for very similar

products, price increases might induce consumers to switch between brands in order to maintain the same level of consumption (in terms of cigarettes per unit of time) and outlay. Clearly, in this situation the ability of taxes to discourage participation would be severely impaired.

The National Health Surveys, where the unit of observation is the individual, contain information on i) the current status with respect to smoking, ii) retrospective information on smoking (age of starting, period elapsed since quitting etc.). This permits the specification and estimation of econometric models for the duration of the habit and the duration of the time elapsed until take up as functions of prices, demographics and public policy measures. For the Spanish population, this paper offers an alternative view to i) the studies that use aggregate data on sales [5,6], which tend to yield biased estimates for price elasticities because of the inclusion of sales to non residents and the omission of purchases of smuggled cigarettes, ii) the studies that use household expenditure data [7, 8, 9, 10], which are limited in the sense that the unit of observation is the household and that the measure of consumption is expenditure on tobacco, and iii) other studies using duration analysis [11], which abstain from considering the effect of prices and public health measures. The analysis in this paper also adds to the existing evidence from international studies on the topic [12,13,14,15,16], which provide a useful methodological guide for the current exercise.

The main findings are that prices have at most a very weak effect on the starting hazard. Thus a 10% increase in prices would increase duration up to starting by about 0.7%. In contrast, the price floor for tobacco, proxied by the average price for black tobacco cigarettes, has a significant effect on the quitting hazard, and this effect is robust across specifications and genders. The estimates suggest that increasing the price floor of cigarettes by 10% would shorten smoking spells by about 14%.

Section 2 contains descriptive evidence on price movements and the estimating sample. Section 3 presents and discusses the econometric estimates and section 4 concludes with a discussion of the fiscal policy implications of the main findings.

2. Data

2.1 Tobacco prices

The official tobacco price indices provided by the National Statistics Office go back to 1976. This short history is not sufficient for this paper's goal so I resort to the annual reports of Tabacalera, the company which has held a monopoly on the distribution of tobacco to the Spanish market for most of last century. Tabacalera offers the average price of a packet of black cigarettes and a packet of blond cigarettes separately as well as the share of each of these two over total sales starting in 1957. Figure 1 presents the evolution of the price of a packet of cigarettes (in constant 2000 euros) distinguishing between black and blond types as well as a weighted (by sales) average price of the two latter. The figure also presents the share of blond cigarettes over total cigarette sales.

Figure 1 about here

Observe that a packet of blond cigarettes was cheaper in 1997 than in the late 50's. At the start of the 70's the price of blond cigarettes descends markedly and in 1975 Tabacalera launches one of the most popular blond brands (Fortuna) coinciding with the opening of the domestic market to foreign blond brands such as Marlboro or Winston. It has been argued that advertising underwent a structural change in this period too, for women were targeted more intensely than ever before [17]. This may explain both the commonly perceived evidence that women tend to smoke blond cigarettes in a higher proportion than men and the marked rise in the share of the blond variety over total cigarette sales. Indeed, smoking prevalence among women increased from 17% in 1970 to 27% in 1990 whereas the male figure descended from 65% to 45% in the same period [17]. However, a concurrent explanatory factor for the rising share of blond cigarettes might be an increasing preference for higher value brands associated to rising incomes. Unfortunately the National Health Surveys do not provide information on the type of cigarettes consumed by individuals so it is difficult to discriminate among these two possibilities. In any case these considerations justify a separate analysis for men and women and, in the absence of information on what type of cigarettes the individual smokes, the utilisation of the price of blond and the price of black varieties in alternative

model specifications (the two series correlation coefficient is 0.6). A final point to be noted from the evidence shown in figure 1 is that over the 90's both series rise markedly as a result of the tax increases required by the fiscal harmonisation process in the European Union.

2.2 Smoking episodes in the National Health Surveys

The main sources of data are the 1993, 1995 and 1997 pooled editions of the National Health Survey carried out by the Ministerio de Sanidad y Consumo. These are nationally representative surveys covering many health related and socio-economic aspects among which smoking histories are well covered. The smoking episodes are constructed from the questions "Do you smoke?" (made to all respondents), "What age were you when you started smoking?" (made to current smokers and ex smokers) and "How long ago did you stop smoking"? (made to ex smokers). The 1993 edition made available by the Ministerio does not contain valid information for the last question so the quitting analysis only uses the 1995 and 1997 editions.

Also, since data for prices are only available as far back as 1957, not all individuals in the pooled sample can be used in the econometric analysis. The chosen strategy consists in selecting the individuals who were born after 1956 for the starting analysis. These individuals are assumed to be at risk of starting smoking since their birth. For the quitting analysis I use individuals born after 1947 and assume that they are at risk of quitting since they start smoking. As none of the individuals used in the quitting analysis started smoking before 1957, there are price data for all periods at risk. The transformation of the cross sectional information into a multiple-record-per-individual spell is necessary in order to use prices (which do not display geographical variation) as an explanatory variable in the duration models. Since the price data are an annual series, I expand the data on every individual by the number of years elapsed until starting (starting analysis) or since starting up to quitting (quitting analysis) and subsequently the price data are merged by calendar year.

My primary concern consists in identifying the effects of changes in the real price of cigarettes on the hazards of starting and quitting and, given that there is no

cross sectional variation in prices, it is important to control for secular changes in attitudes towards smoking. This is done by means of the inclusion of a polynomial of third order in time, cohort dummy variables designed to evaluate whether gradually younger generations behave differently, and dummy variables that capture the effect of advertising campaigns and legislation that might have changed perceptions towards smoking in the past decade. These controls have been used in Jiménez et al. [10] and respond to the following policies: Firstly, from 1984 onwards it was forbidden to advertise tobacco in some media (although the ban did not apply to light brands for some time). Also, smoking was banned in some public transport media and the prohibition to sell tobacco to under 16's was decreed. Secondly, as of 1992 the ban on smoking in public transport media was extended to flights, and the health warning campaigns were intensified. In the econometric specifications, these measures are proxied with two dummy variables which are activated for 1984 and onwards and 1992 and onwards respectively.

The other variables in the parametric specifications for the hazard functions are the following:

- i) The logarithm of real price of a 20-unit packet of cigarettes. As mentioned before, I shall use different price series: black cigarettes, blond cigarettes and a weighted average of the latter in alternative model specifications.
- ii) A dummy variable activated if the individual has obtained a university degree.
- iii) A dummy variable activated if the individual has completed secondary education.
- iv) Dummies that control for the following cohort effects:
 - 1) Born between 1967 and 1976
 - 2) Born after 1976

The effect of gender is taken into account by means of estimating separate models for men and women. Table 1 contains summary statistics for the samples used in the analyses of starting and quitting.

Table 1 about here

3. Econometric specification and estimates

.

3.1 Starting

The sample used for the analysis of starting contains individuals who “fail” (start smoking) and individuals who are not observed starting to smoke. The latter are usually considered (right) censored spells corresponding to individuals who, had the monitoring period been sufficiently long, would have been observed failing eventually. It is common practice to avoid this assumption when analysing duration up to smoking initiation on the grounds that the duration process does not apply to individuals who are inherently non-smokers. Hence the split population duration model specification, whose basic feature consists in augmenting the duration specification with a binary choice model for participation. Let $s_i = \mathbf{1}(\text{individual } i \text{ smokes})$ be a binary indicator signalling whether the individual smokes and let t_i be the time variable. In a sample of N individuals, the log likelihood function for this model is then given by

$$\text{Log}L = \sum_{i=1}^N s_i \log[P(s_i = 1)f(t_i)] + (1 - s_i) \log[1 - P(s_i = 1) + P(s_i = 1)S(t_i)]$$

Where $f(\cdot)$ and $S(\cdot)$ are the probability density function and the survival function respectively. Therefore the contribution to the likelihood of an individual who starts smoking at age t is the probability of ever starting times the density of starting at t , whereas the individuals who are not observed starting contribute with the probability of never starting plus the probability of ever starting times the probability of having survived (not started) up to the age at the survey date. A more complete discussion on this model is given in recent contributions to the

literature [15, 16], which, in conjunction with other useful methodological guides such as Morris et al. [18], I follow when performing specification tests. In this application I use a standard probit for the binary choice component of the model. The choice for the distributional assumption adopted for the duration process is suggested by the plots in figure 2, which show non parametric estimates of the survival function and the hazard function for starting. The hazard for starting is non-monotonic, which suggests the use of a log-logistic distribution. Therefore I estimate the standard log logistic and the split population log logistic models using the three different measures of prices discussed before.

Figure 2 about here

As table 2 shows, the log-likelihood values in the three alternative specifications of the standard log logistic model are very close. Moreover, none of the price coefficients is significant in these specifications. On the other hand, the values for the likelihood function for the split population log-logistic models suggest that the specification using the weighted average price measure is the one that best fits the data for both men and women, so the following specification search focuses on the models that use the average price measure. In order to discriminate among the standard log-logistic model and the split population log-logistic model, I compare the predicted survival functions with their non-parametric estimates. Also, I analyse the cumulative Cox-Snell residuals arising from the two models on the premise that a correctly fitted model will produce cumulative Cox-Snell residuals which resemble a censored sample from a standard exponential distribution and, therefore, a plot of the cumulative hazard function for these data should lie on a straight line from the origin with slope equal to one.

Table 2 about here

Figure 3,4, and 5 about here

Figures 3 and 4 present, for both men and women, the Kaplan Meier estimate of the survival function and the predicted survival functions from the standard log-logistic model (using the whole sample and the sample of starters only) and the split population log-logistic model (using the whole sample). Focusing first on the evidence for men in figure 3, note that the predicted survival function from the log-logistic model estimated on all the sample eventually reaches zero, but it does so after more than 40 years at risk, a prediction which is inconsistent with the fact that very few smokers start after the age of 20, as the Kaplan Meier estimate of the survival function for smoking men shows. If the standard log logistic model is estimated only on the sample of starters then its predicted survival function resembles very much its non parametric counterpart, but this is gained at the cost of dropping from the sample all right-censored observations, some of which might correspond to future smokers. The predicted survival function from the split population model, in contrast, adequately captures the fact that most starts take place before the age of 20. A similar argument can be made from the evidence on women in figure 4. These results provide a first piece of evidence in favour of the split population specification. The analysis from the cumulative Cox-Snell residuals in figure 5 also supports the split population specification. Indeed, the plot for the standard log-logistic specification for both men and women departs clearly from the 45° line whereas the plot for the split population specification is very close to it, especially in the case of men. The log-logistic split population model is therefore the preferred specification and table 3 presents its whole set of parameters.

Table 3 about here

Focusing first on the participation coefficients in table 3, note that education has a different effect across genders. In fact, having completed a university degree decreases the probability of ever smoking in men but not in women. The same can be said about having completed secondary education, although the effect is smaller in size. With the data at hand, it is impossible to discern between the alternative explanations for the effect of education on smoking. As Madden [16] argues more educated individuals may i) be better informed about the health hazards of smoking, and or ii) know better what to do

in the face of such information and or iii) have a greater preference for the future (which leads them to invest in human capital in the forms of both avoidance of health hazards and education acquisition). Some (or all) of these factors seem to be in operation in the case of men, but not in the case of women. The mechanism behind this differential behaviour might be related to the history of women's liberation in Spain. Indeed, it has been argued [17] that as of the 70's smoking was portrayed as a process of social equalisation with respect to men (a goal also achievable by means of studying). Such phenomenon could account for the lack of a negative relationship between education level and the probability of participation of women. Finally, the coefficients on the cohort dummies show that individuals born on 1977 or after are more likely to start smoking than the rest of individuals. This effect is greater in the case of women.

Focusing now on the coefficients in the duration component of the models, note first that, although significant, the price effect is very small in size. The elasticity of time elapsed until starting is 0.069 for men and 0.076 for women. At the mean starting age, (17 years), this implies that a 10% increase in prices would delay starting by roughly one month and a half. Completion of either a university degree or secondary education increases the duration up to starting for both men and women. The estimated effects suggest that, with respect to the omitted category (primary education), university graduates last 7.8% (men) or 5.6% (women) longer. The corresponding figures for secondary school graduates are 3.9% (men) and 3.1% (women). Note also that the public health measures enforced in 1984 do not seem to have had a significant impact on duration whereas those enforced as of 1992 seem to have increased the duration up to starting by around 4%. The estimates also show that younger cohorts tend to start smoking earlier. With respect to the pre 1967 cohort, individuals born between 1967 and 1976 take around 5% less to start smoking. For the post 1977 cohort the duration is 16% shorter. The top half section of figure 6 presents, for both genders, the combined effect of the time controls on the survival rate at 14 years of age implied by the model for an individual with primary education facing the time average of the average price of cigarettes. The discontinuities in the schedules correspond to the estimated cohort effects discussed above and the vertical lines on 1984 and 1992 mark the activation of

the public policy dummies. Note that for both genders the predicted survival rate descends markedly after the mid 70's. The predicted proportion of 14 year olds who have not started smoking reaches its minimum in the mid 80's, but in the last years covered in the sample it increases in a remarkable way.

Figure 6 about here

3.2 Quitting

Figure 7 presents the Kaplan Meier estimates of the survival function and the hazard function for quitting for both sexes. Their shapes would suggest the use of models that can accommodate positive duration dependence in the hazard. Thus for the analysis of quitting I first obtain estimates for the Cox proportional hazard model, the Weibull and the Gamma model and subsequently search for the preferred specification.

Figure 7 about here

Table 4 presents the values for the log-likelihood functions of these alternative specifications of the functional form using the three available measures of prices. Observe that for each of the three possible functional choices, the greatest value for the likelihood function is attained with the price series for black cigarettes. Moreover, the other two price variables are not significant in any of the specifications.

Table 4 about here

The next step consists in discriminating between the three choices of functional form using the price series for black cigarettes. The graphical methods for model comparison used in the analysis of starting (i.e. plots of the cumulative Cox-Snell residuals and comparison of predicted and non-parametric survival functions) do not offer a clear basis for model discrimination in the context of quitting, so I resort to tests for the proportional hazards assumption, the Reset test in the fashion proposed by Peters [19], and the Akaike information criterion. The results for these tests are presented in table 5. Note that the Grambsch and Therneau [20] global test for proportional hazards assumption is passed but there are three variables in the male specification for which the

test based on the re-scaled Schoenfeld residuals indicates rejection of the proportional hazards. Using a 5% significance level, the Reset overall specification tests reject the Cox models and the Gamma specification for women. In contrast, the Weibull specification (both for males and females) and the Gamma specification for males pass this test. To discriminate between these two models note that the Weibull distribution is nested within the Gamma distribution, and therefore it is possible to test the parametric restriction that collapses the latter into the former. This restriction cannot be rejected at the 5% significance level. The comparison of the two models on the basis of the Akaike information criterion further confirms the Weibull as the preferred specification. Its estimates are presented in table 6.

Tables 5,6 about here

Focusing first on the effect of prices, observe that the estimates for the elasticity of duration with respect to the price of black tobacco are -1.32 ($t\text{-value}=-2.23$) and -1.5 ($t\text{-value}=-1.50$) for men and women respectively. Thus prices appear to exert a significant and non-negligible effect on the propensity to quit: if as an example we take as reference a duration of 10 years, the estimates suggest that a 10% increase in prices would lead to a reduction of duration of 16 months for men and 18 months for women. There is also a sizeable and significant effect associated to having completed a university degree. For men this characteristic is associated to a duration 37% shorter than that of the default category (primary education) whereas for women the duration is 19% shorter. It is noteworthy that the effect of secondary education in males is also significant, non-negligible and, at 27%, larger than the university degree effect for women. Note also that the effect of secondary education is not significant in the case of women. These disparate patterns for education effects across genders are consistent with the findings for starting behaviour reported before and, again, might be a consequence of the fact that smoking has been successfully portrayed as a means to achieve social equality with respect to men in advertising. Also, in accordance with the results obtained in the starting analysis, the public health measures decreed in 1984 do not seem to have had a significant effect but those enforced since 1992 are associated to a considerable reduction in the duration of smoking: 48% and 57% for men and women respectively. The estimation results also suggest that younger cohorts tend to have shorter smoking durations. The bottom half of figure 6 presents the combined effect of

the time controls in the model for the survival rate after four years at risk (of quitting smoking) of individuals with primary education facing the mean price of black cigarettes over the sample period. Up until the end of the 60's the survival rate after four years at risk was stable for men and increasing for women. After this point the graphs show a gradual decline in the survival rate for men and a marked decline in that of women. The overall picture provided by the upper and lower halves of figure 6 would suggest the existence of two different periods over the last quarter of a century as far as attitudes towards smoking are concerned in the Spanish population. The first period runs from the mid 70's up to the end of the 90's, when both men and women started to smoke earlier but at the same time gave up earlier. The second period runs for the rest of the years covered in the sample, when the starting age gradually increased and the reduction in the duration of smoking persisted (the sudden recovery in the predicted survival rate of men in the two oldest cohorts after 1992 would seem to contradict this latter statement, but note that this might be due to the scarcity of data for these individuals, as there are few smokers born before 1977 who had been at risk just four years in 1992 or thereafter).

Before proceeding to a discussion of the policy implications of these econometric results, it is worthwhile to consider some points related to their sensitivity with respect to specification and possible measurement errors in the data. Firstly, the way the question "How long ago did you stop smoking" is posed might lead individuals to round off their answers at multiples of 5. This will result in "heaps" of reported quits at calendar years that are 5, 10, 15, 20 etc. years before the interview date. Since prices only vary by calendar year, this "systematic" heaping might affect the estimated price coefficients, so in order to assess the importance of this potential problem I have estimated the Weibull model with the same right hand side variables plus a set of 4 dummy variables activated at 5, 10, 15, and 20 years before the interview date. The estimated price elasticities for this model are -1.51 (t-value= -2.42) and -1.97 (t-value= -2.67) for men and women respectively. Thus the heaping effect in this data set seems to slightly underplay the effect of prices, but in any case the qualitative conclusion regarding the sign and significance of the price parameters is robust. This evidence is in line with the results in Forster and Jones [15], who investigate the influence of recall errors using a variety of methods (including the addition of dummy variables for the calendar years affected by heaping) in the context of the duration

models used in this exercise. There are other potentially bias-inducing specification issues, such as the fact that I have used a continuous time specification and the fact that I have considered neither the effect of unobserved heterogeneity nor the existence of habits, but the evidence obtained by Forster and Jones when evaluating the robustness of their results might be extrapolated to suggest that the estimated price effects in the type of duration models used in this paper are robust to these alternative specifications.

4. Discussion

In this paper I have estimated parametric models for the hazard of starting and quitting smoking and the results suggest a few stylised facts of interest. Firstly, the public health measures adopted in the early 90's seem to have exerted a significant effect on the propensity to start and to quit smoking. The effect on starting is small in size but smoking duration drops very substantially after the 1992 measures were enforced. This result highlights the importance of policies destined to increase the non-monetary cost of smoking (restrictions to use) and the dissemination of information on the effects of smoking on health. Secondly, prices have a very small effect on the propensity to start smoking. The results from the split population model situate the elasticity of delay with respect to prices around 0.07. Thirdly, there is a significant price effect, that of black cigarettes, that is robust across the three specifications of the quitting hazard and also across genders. Indeed, the price elasticities for the duration of the smoking spell estimated with the Weibull models are situated around -1.3 for men and -1.5 for women. It is remarkable that this effect is found for the price of black cigarettes but neither with the price of the blond variety nor the weighted price index. The fact that the most marked increase in the prices of black tobacco coincides with the last years in the data, a period where overall attitudes towards smoking might have started to change due to the arrival of evidence on the health effects of tobacco to the mass media and the implementation of restrictions and warning campaigns, could raise suspicions as to whether the estimated coefficients are really capturing the effect of prices. But note that the specifications include controls for these time effects, so the estimated coefficients for the price of black cigarettes seem to be reflecting an independent effect amidst the general trend.

This evidence is consistent with two (possibly coexisting) underlying phenomena. On one hand the population might be composed of one group with a high degree of price responsiveness which smokes black tobacco due to its low comparative cost, and another group which smokes blond tobacco and is less price responsive. Thus only the black tobacco smokers would respond to price increases. Another explanation is that the whole of the population is price responsive but smokers move across the price/brand spectrum when prices increase in order to maintain consumption (in terms of number of cigarettes per unit of time) and expenditure. Thus the only way through which price changes lead to quitting is when the price floor increases. Clearly, the key to discern among these two hypotheses is the cross elasticity between the black and the blond varieties. Although there is no direct evidence on this parameter for the Spanish case, results in the international literature suggest that smokers switch between varieties of tobacco after price rises. For instance, as reported in Chaloupka and Warner [21], Ohsfeld and Boyle [22] find a significant and positive effect of cigarette taxes on the use of smokeless tobacco in the US. Similarly, Thompson and McLead [23] and Pekurinen [24] find that Canadian and Finnish cigarette smokers would switch from manufactured to hand roll cigarettes in response to increases in the prices of manufactured cigarettes. This would lend support to the second of the hypotheses stated above, since the switch from manufactured cigarettes to either smokeless tobacco or hand rolls appears to be a much more drastic change in consumer choice than a switch between the black and blond varieties of manufactured cigarettes available in the Spanish market.

This has two clear implications for fiscal/public health policy. Firstly taxes should be geared towards the elimination of the big price differentials between varieties. As mentioned before, Spain is one of the EU countries where the specific component of tobacco taxes is smallest, which permits a wider range of variation for final consumer prices. An obvious way to restrict such range would be to increase the specific component. Moreover, the results in Delipalla and O'Donnell [3] also provide another justification for increasing the specific component if tobacco taxes are intended to discourage consumption, since they find that ad valorem cigarette taxes are undershifted to the consumers whereas specific taxes are overshifted. Finally, these results highlight the fact that the incidence of smuggling is crucial for the effectiveness of tax policy at

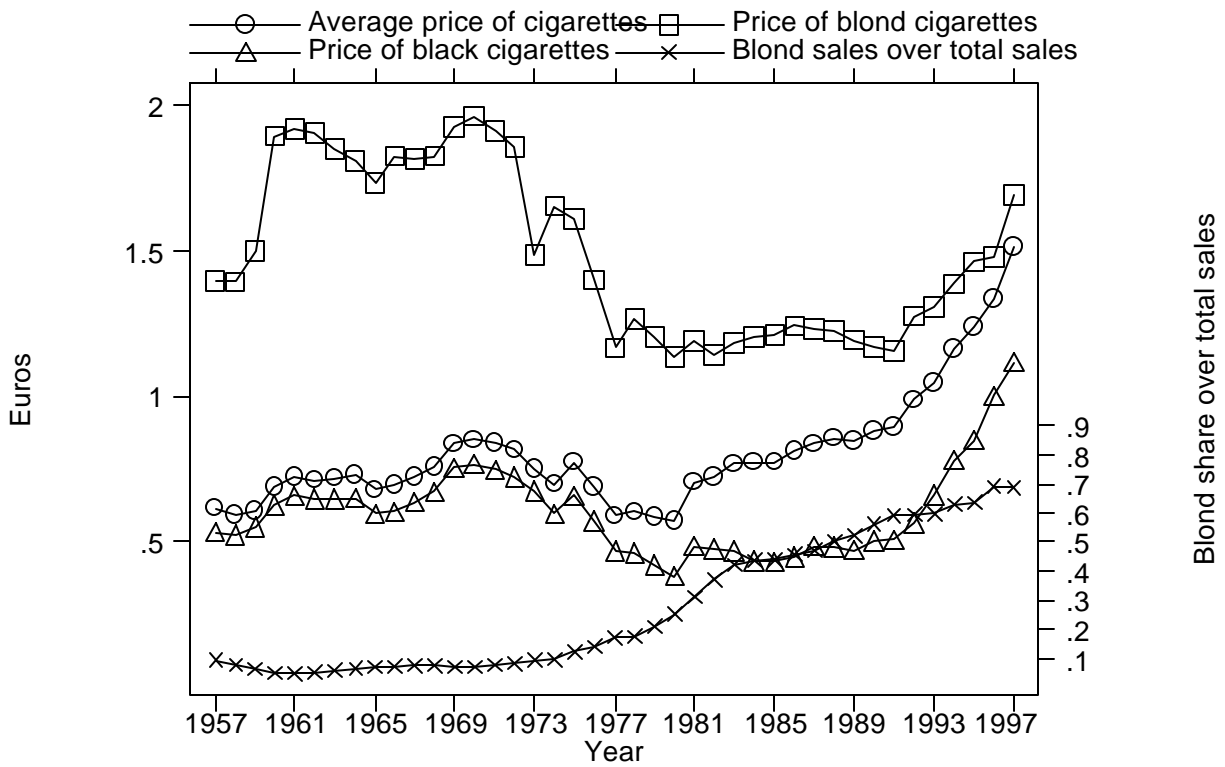
discouraging the habit, so any increase in the efforts to eliminate this low price variety from the market are likely to have an impact on the prevalence of smoking.

References

1. Montes, A. y J.R. Villabi (2001) "The Price of Cigarettes in the European Union". Próxima publicación en *Tobacco Control*
2. Tabacalera (1998). Informe Anual. Madrid.
3. Delipalla, S. and O. O'Donnell (2001) "Estimating tax incidence, market power and market conduct: The European cigarette industry". *International Journal of Industrial Organisation*. Vol 19. pp. 885-908
4. Joosens, L. y M. Raw (1995) "Smuggling and Cross Border Shopping of Tobacco in Europe". *British Medical Journal* 310:1393-1397
5. Escario Gracia, J. (2000). Modelos impositivos sobre el consumo de tabaco. Tesis doctoral. Departamento de Análisis Económico. Universidad de Zaragoza.
6. Valdés B. (1993) "Cigarette Consumption in Spain: Empirical Evidence and Implications for Public Health Economics." *Applied Economics*, 25:149-156
7. Labeaga, J.M. (1999) "A Double Hurdle Rational Addiction Model with Heterogeneity: Estimating the Demand for Tobacco ". *Journal of Econometrics* 35: 45-62
8. Jiménez, S. y J.M. Labeaga (1994) "Is it Possible to Reduce Tobacco Consumption via Alcohol Taxation?". *Health Economics* 3: 231-241
9. García, J. y 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
10. Jiménez, S., J.M. Labeaga y A. López Nicolás (1998) "Participation, Heterogeneity and Dynamics in Tobacco Consumption: Evidence from Cohort Data". *Health Economics* 7:401-414
11. Ahn, N. and A. Molina (2001) "Smoking in Spain: Analysis of Initiation and Cessation." Documento de Trabajo 2001-02. FEDEA. Madrid.
12. Douglas, S. And Hariharan, G. (1994) "The Hazard of Starting Smoking: Estimates from a Split Population Duration Model". *Journal of Health Economics* 13:213-230
13. Douglas, S. (1998) "The Duration of the Smoking Habit". *Economic Inquiry* 36:49-64
14. Tauras, J.A. and Chaloupka, F.J. (1999) "Determinants of Smoking Cessation: an Analysis of Young Adult Men and Women". WP 7262. NBER. Washington.
15. Forster, M. y A. Jones (2001) "The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data.". *Journal of the Royal Statistical Society (Series A)* 164 part 3: 1-31.
16. Madden, D. (2001) "Do Tobacco Taxes Influence Starting and Quitting Smoking? Evidence from a Sample of Irish Women". Mimeo. University College Dublin.
17. Consell Asessor sobre Tabaquisme a Catalunya (2001). Informe Tècnic Sobre L'hàbit Tabaquic a Catalunya. DSSS. Generalitat de Catalunya. Barcelona.
18. Morris, C., E.C. Norton and X. Zhou (1994). "Parametric Duration Analysis of Nursing Home Usage". In Lange, N.et al. (eds.) *Case Studies in Biometry*. John Wiley and Sons. New York.

19. Peters, S. (2000). "On the Use of the Reset Test in Microeconomic Models". *Applied Economics Letters* 7:361-365
20. Grambsch, P. And Therneau, T. (1994). "Partial Hazards Tests and Diagnostics Based on Weighted Residuals". *Biometrika* 81:515-526.
21. Chaloupka, F. And K. Warner (2000). "The Economics of Smoking". In Culyer, A. and J.P. Newhouse (eds.) *Handbook of Health Economics*. North Holland.
22. Ohsfeldt, R. And R.G. Boyle and E.I. Capilouto (1999). "Effects of Tobacco Excise Taxes on the Use of Smokeless Tobacco Products". *Health Economics* 6(5):525-532.
23. Thompson, M.E. and I. McLeod (1976). "The Effects of Economic Variables upon the Demand for Cigarettes in Canada". *Mathematical Scientist* 1:121-132.
24. Pekurinen, M. (1989). "The Demand for Tobacco Products in Finland", *British Journal of Addiction* 84:1183-1192.

Figure 1



Cigarette prices in 2000 euros and blond share of sales

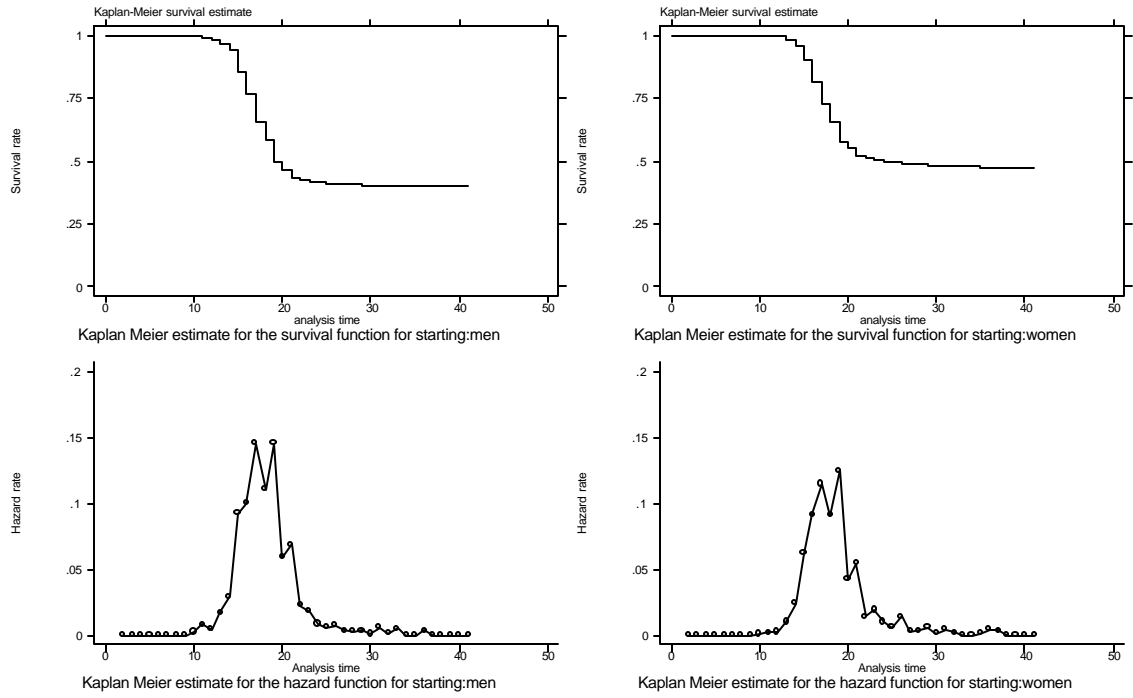
Blond share over total sales

Table 1. Descriptive statistics (sample means)

Variable	Starting analysis		Quitting analysis	
	Men	Women	Men	Women
University degree*	.13	.14	.13	.17
Secondary education*	.44	.41	.36	.37
Primary education*	Omtd.	Omtd.	Omtd.	Omtd.
	Regressor	Regressor	Regressor	Regressor
Born before 1967*	Omtd.	Omtd.	Omtd.	Omtd.
	Regressor	Regressor	Regressor	Regressor
1967-1976 cohort*	.51	.49	.30	.35
1977+ cohort*	.07	.06	.06	.07
Starts / quits smoking*	.56	.48	.20	.21
Starting age	17	17.4	17	18
Sample size	7092	6913	2305	1817
Failures	3992	3386	474	395

* Starred variables are dummy variables

FIGURE 2

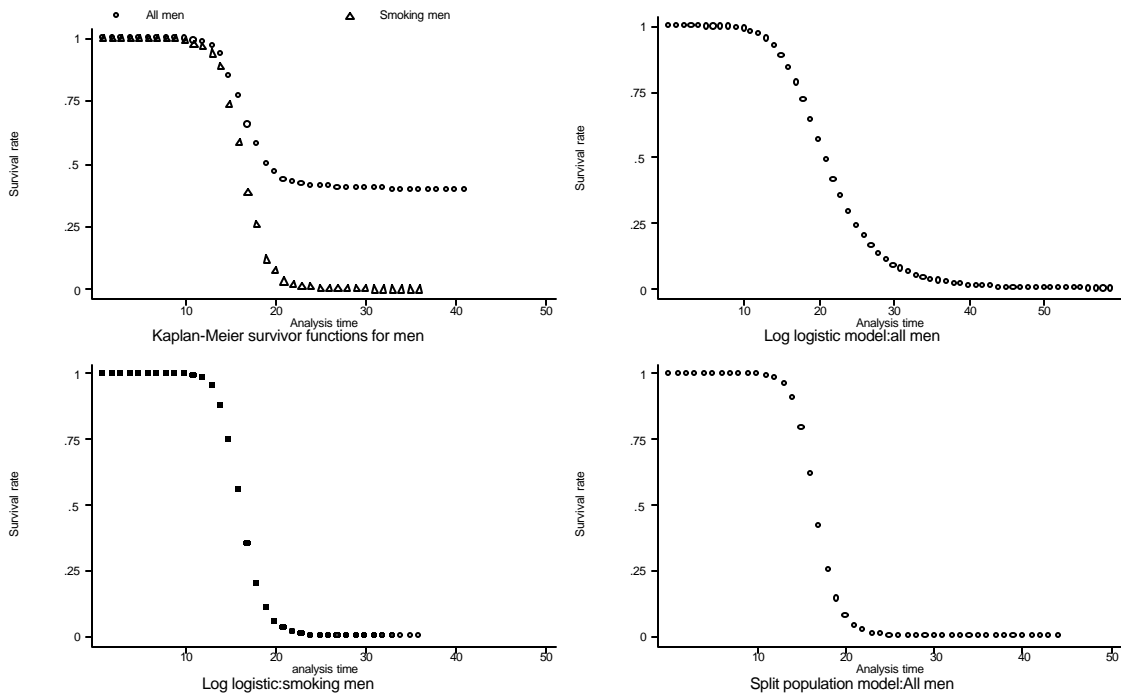


Non parametric survival and hazard functions for starting

Table 2. Log likelihood values and price coefficients for alternative specifications of starting hazard

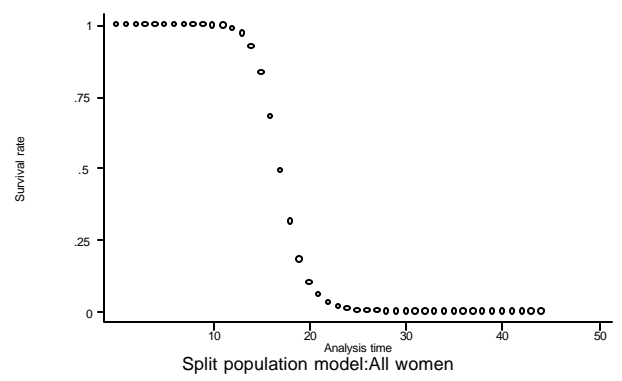
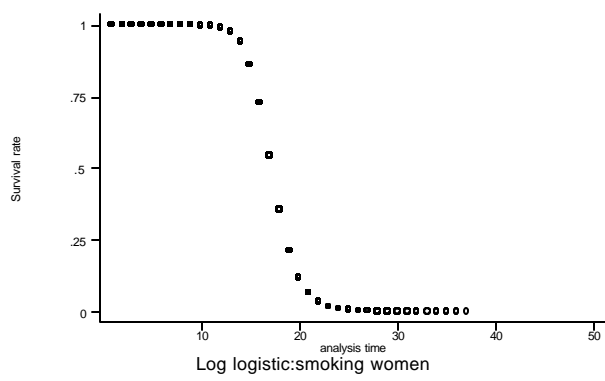
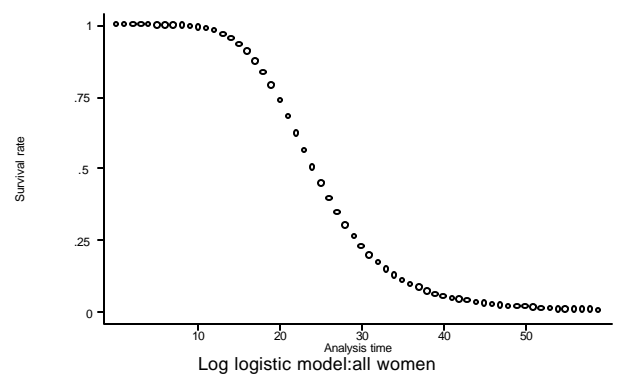
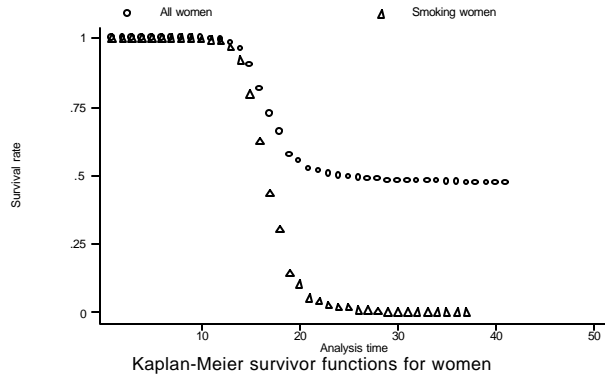
Model specification		Log Likelihood		Price coefficient / t value	
Log logistic		Men	Women	Men	Women
	Log Average Price	-3823.47	-4368.02	.016 (0.39)	.033 (.68)
	Log Black Price	-3823.26	-4367.87	-.296 (-.76)	-.041 (0.88)
	Log Blond Price	-3823.54	-4368.25	.003 (.06)	.007 (.10)
Split population		Men	Women	Men	Women
	Log Average Price	-13322.9	-12216.48	.069 (2.6)	.076 (2.6)
	Log Black Price	-13325.6	-12219.4	.03 (1.4)	.019 (.75)
	Log Blond Price	-13326	-12218.8	.031 (.9)	.053 (1.34)

Figure 3



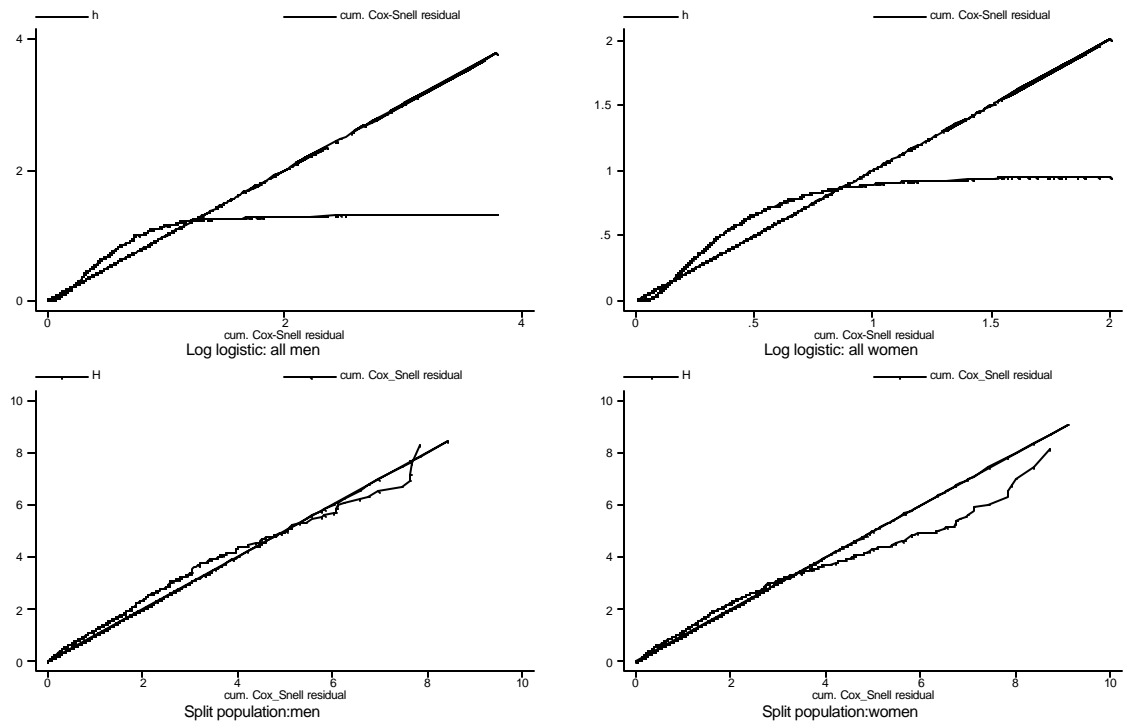
Comparison of survival functions for starting:men

Figure 4



Comparison of survival functions for starting:women

Figure 5



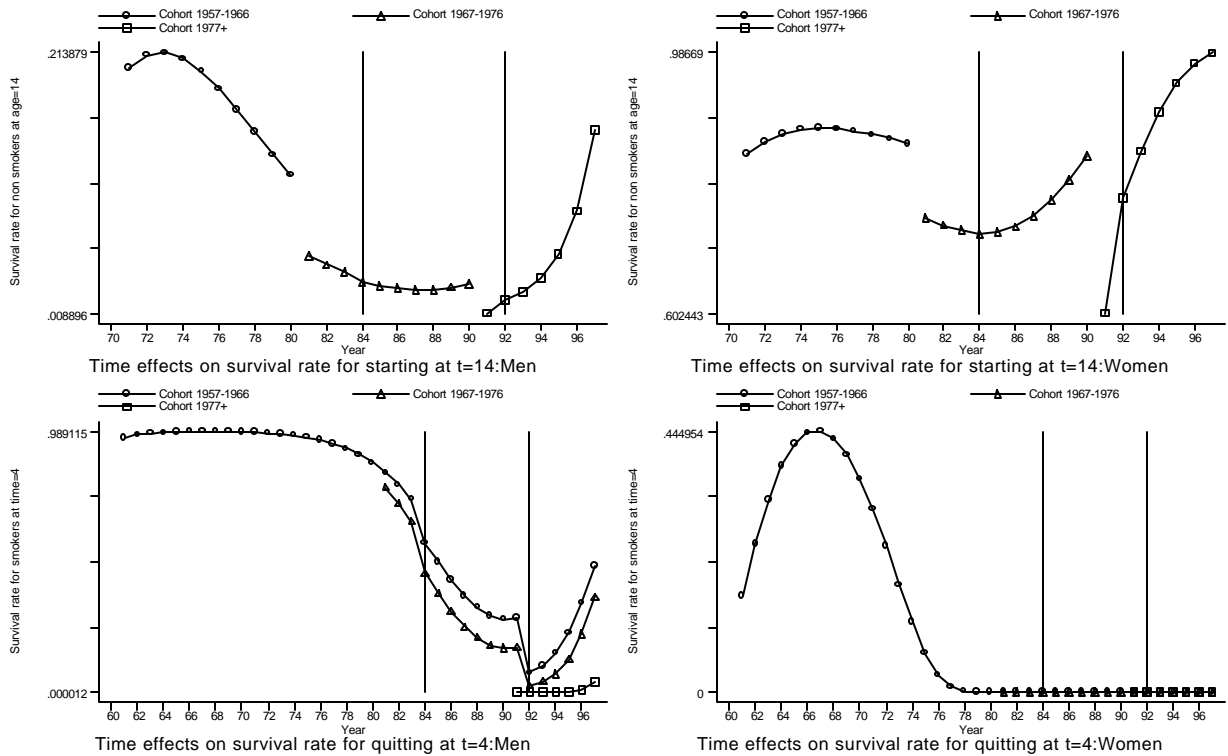
Cumulative Cox-Snell residuals

Table 3. Parameter estimates for log-logistic split population model for the duration up to starting

Parameter	Log-logistic Split population model			
	Men		Women	
	Participation	Duration	Participation	Duration
Log of average price		.069 (2.68)		.076 (2.60)
University degree*	-0.551 (-10.68)	.078 (11.47)	-0.007 (-0.16)	.056 (8.14)
Secondary education*	-0.255 (-6.48)	.039 (9.13)	.055 (1.48)	.031 (6.53)
1967-1976 cohort	-0.272 (-7.20)	-0.052 (-6.36)	.002 (0.01)	-0.048 (-5.29)
1977+ cohort	1.322 (0.65)	-0.162 (-8.71)	1.987 (0.45)	-0.162 (-6.90)
1984 and after*		-0.008 (-0.93)		.001 (0.11)
1992 and after*		.045 (4.11)		.034 (2.88)
Trend		1.336 (20.04)		1.461 (22.95)
trend^2		-1.682 (-20.35)		-1.840 (-23.48)
trend^3		0.070 (20.60)		.077 (23.74)
Constant	.744 (22.77)	-32.900 (-18.13)	.126 (0.40)	-36.120 (-20.44)
Gamma		.075 (65.85)		0.076 (59.22)
Log L		-13322.90		-12216.48
N		7092		6913
Failures		3992		3386

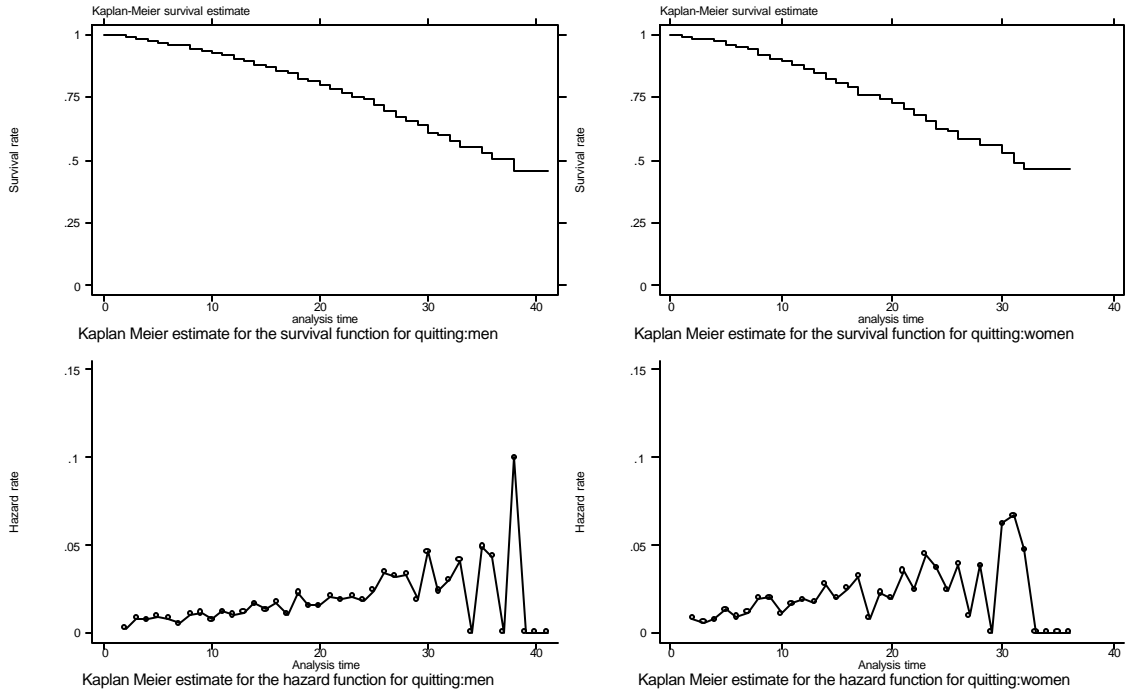
* Starred variables are dummy variables
t-statistics in brackets

Figure 6



Effects of time controls in the starting and quitting models

Figure 7



Non parametric survival and hazard functions for quitting

Table 4. Log likelihood values and price coefficients for alternative specifications of the duration up to quitting

Model specification		Log Likelihood		Price coefficient / t value	
Cox PH		Men	Women	Men	Women
	Log Average Price	-3178.15	-2592.51	1.73 (0.47)	0.53 -(0.49)
	Log Black Price	-3175.68	-2590.36	8.1 (2.24)	8.84 (2.09)
	Log Blond Price	-3178.25	-2591.84	0.88 -(0.08)	0.1 -(1.25)
Weibull		Men	Women	Men	Women
	Log Average Price	-1186.77	-1005.45	-0.35 (0.48)	0.38 (0.44)
	Log Black Price	-1184.29	-1003.2	-1.32 -(2.23)	-1.5 -(2.11)
	Log Blond Price	-1186.88	-1004.85	0.84 (0.09)	1.41 (1.18)
Gamma					
	Log Average Price	-1185.86	-1008.56	-0.43 (0.57)	0.40 (0.43)
	Log Black Price	-1183.72	-1006.78	-1.31 -(2.55)	-1.56 -(1.90)
	Log Blond Price	-1187.14	-1008.06	0.01 (0.01)	1.46 (1.09)

Table 5. Specification tests for parametric models for the the duration up to quitting

	Cox Model		Weibull model		Gamma model	
	Men	Women	Men	Women	Men	Women
Log Likelihood	-3175.68	-2590.36	-1184.29	-1003.2	-1183.72	-1006.43
Reset test (p)	0.0003	0.0016	0.18	0.06	0.08	0.03
Akaike information criterion			2392.58	2030.4	2393.44	2039.56
Test for Gamma=Weibull (p)					0.9	0.1
Schoenfeld residuals						
Global test for PH (p)	0.28	0.11				
Individual test for PH (rejections) (p)						
Trend	0.02					
Trend2	0.03					
Trend3	0.003					

Table6. Estimates for the Weibull model

Parameter	Weibull model	
	Men	Women
Log of black price	-1.322 (-2.23)	-1.504 (-2.11)
University degree*	-0.372 (-4.55)	-0.197 (-2.18)
Secondary education*	-0.277 (-4.24)	-0.107 (1.39)
1967-1976 cohort	-0.207 (-2.22)	-0.067 (-0.65)
1977+ cohort	-0.932 (-5.57)	-0.698 (-3.69)
1984 and after*	-0.266 (-1.38)	0.183 (0.77)
1992 and after*	-0.487 (-3.96)	-0.572 (-3.84)
trend	8.396 (3.73)	10.120 (3.08)
trend^2	-1.088 (-3.74)	-1.299 (-3.15)
trend^3	0.046 (3.73)	0.054 (3.20)
Constant	-202.600 (-3.65)	-248.660 (-2.92)
Ln p	0.461 (7.82)	0.390 (6.00)
Log L	-1184.29	-1003.20
N	2305	1817
Failures	474	395