THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS
DEPARTMENT OF ECONOMICS & RELATED STUDIES

The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data

Martin Forster
Andrew Jones

DISCUSSION PAPER 176

The role of tobacco taxes in starting and quitting smoking: duration analysis of British data

Martin Forster and Andrew M. Jones
Department of Economics and Related Studies
University of York
York YO10 5DD
United Kingdom
Tel: +44 1904 433790; Fax: +44 1904 433759
e-mail: mf8@york.ac.uk, amj1@york.ac.uk

January 2000

CENTRE FOR HEALTH ECONOMICS DISCUSSION PAPER SERIES

The Discussion Paper (DP) Series was established in 1984 to make research available quickly to a broad readership drawn from the academic community, professional health care staff and policy makers, both within the UK and internationally. Material published in DP form ranges from work in progress, to edited versions of completed project reports, from papers intended ultimately for journal publication, to those which seek to stimulate immediate debate. Hence the variation in length and complexity of DPs.

In the past, DPs have been subject to a minimum of editorial revision, it being principally seen as the responsibility of individual authors to ensure that appropriate standards were met. Since June 1996, a more formal system of internal peer-review has been introduced to consolidate the high levels of quality that have been achieved in the past. That system involves the refereeing of DPs in draft form, with any recommended revisions being verified independently by a senior member of staff.

DPs are not a substitute for publication in peer-reviewed journals. It is to be hoped that journal editors will recognise the nature of the DP series, and accept that this does not compromise the status of those papers published by the Centre which may later be submitted for consideration in more formal publications. Comments and suggestions on DPs are encouraged and should be sent directly to the authors.

The authors

Martin Forster is a lecturer in the Department of Economics and Related Studies, University of York and Professor Andrew Jones is Director of the Graduate Programme in Health Economics at the University of York.

Further Copies

Further copies of this document (at price £5.00 to cover the cost of publication, postage and packing) are available from:

The Publications Office
Centre for Health Economics
University of York
York YO10 5DD

Please make cheques payable to 'The University of York'. Details of other papers can be obtained from the same address or telephone (01904) 433648. Visit our website on www.york.ac.uk/inst/che

© Centre for Health Economics

Summary

This paper presents new evidence on the determinants of starting and quitting smoking using duration data from the British *Health and Lifestyle Survey* (HALS). Self-reported data on individual smoking histories coupled with the availability of a long time series for the tax rate on cigarettes are used to construct a longitudinal data set in which the tax rate is treated as a time varying covariate. This overcomes the problem of the lack of cross section variation in prices that has plagued previous studies of smoking in Britain. The study is the first to identify tax-price elasticities for starting and quitting in Britain. Results for age of starting smoking are reported for a split-population duration model. Results for the number of years smoked prior to quitting are reported for Cox, Weibull and gamma models. All of the models are estimated separately for males and females and extensive diagnostic tests are used to guide model specification. A sensitivity analysis is used to assess the robustness of the estimated tax elasticities for starting and quitting. Since the early 1990s successive governments have had a commitment to annual increases in the real level of tobacco taxes, to achieve health policy objectives and encourage people to stop smoking. Our estimated price elasticities directly relate to the impact of above inflation tax rises on the number of years smoked by current smokers. The estimates of the impact of tax on the probability of starting and the age of starting are not encouraging as we do not find a significant effect. However the point estimates of the elasticity of quitting are well defined for males and robust for both males and females. All of our point estimates are in the range -0.40 to -0.63. If the typical number of years smoked is 25 years this implies that the 5% real increase in tobacco duty would lead, on average, to a reduction of between 6 and 9.5 months of smoking for each smoker. Recent estimates suggest that there are around 12.1 million current smokers in the U.K. (ASH 1999). Then the potential saving in total number of years smoked across the population is substantial.

Keywords
Smoking initiation and cessation; Tobacco taxes; Duration analysis

1 Introduction

This paper presents new evidence on the determinants of starting and quitting smoking using duration data from the British *Health and Lifestyle Survey* (HALS). Self-reported data on individual smoking histories coupled with the availability of a long time series for the tax rate on cigarettes allow us to construct a longitudinal data set in which the tax rate is treated as a time varying covariate. This overcomes the problem of the lack of cross section variation in prices that has plagued previous studies of smoking in Britain. Only a handful of previous studies have used duration analysis to model the hazard rates of starting and quitting and all of these have used data from the United States. This is the first study to identify tax-price elasticities for starting and quitting in Britain, where the elasticity measures the proportionate impact on years of smoking of a proportionate change in the real tax on cigarettes.

Results for age of starting smoking are reported for a split-population duration model like that used by Douglas and Hariharan (1994). Results for the number of years smoked prior to quitting are reported for Cox, Weibull and gamma models. All of the models are estimated separately for males and females, with extensive diagnostic tests used to guide model specification. Results and specifications are compared to those in the U.S. literature. We find that, for the starting models, the split-population log-logistic duration/probit participation specification is well specified for males but not for females. For the quitting models, the gamma specification is preferred to the specifications used by Douglas (1998) and Tauras and Chaloupka (1999), although tax elasticities of quitting show little variation across models.

To assess the robustness of the estimated tax elasticities for starting and quitting we carry out sensitivity analyses. For the models of starting we assess the effect of recall bias with respect to the age at which an individual started smoking by rescaling the duration variable to measure calendar time and investigating systematic reporting bias implied by the hazard function. We find little evidence of recall bias using this method. For the models of quitting we compare different semiparametric and parametric specifications of the baseline hazard, compare discrete and continuous time specifications, and allow for the influence of unobserved heterogeneity using a mixture model. We also assess recall bias, by rescaling the duration variable for quitting to investigate quitting by calendar year, and find strong evidence of 5 and 10 year recall bias. Our models are adapted in response to this evidence. We present estimates that include measures of past smoking. The sensitivity analyses suggest that the estimated tax elasticities of quitting are robust to all of these factors.

The elasticity of age of starting with respect to the tax on cigarettes is estimated at +0.10 (p=0.001) for males and +0.05 (p=0.132) for females. The results suggest that there is no effect of tax rates on the probability of starting smoking. The estimated elasticity of the number of years smoked before quitting with respect to the tax rate ranges from -0.41 to -0.63 for males and from -0.40 to -0.55 for females. Parental smoking increases the probability of becoming a smoker and reduces the age of starting but has no significant effect on quitting. Those with more education and in higher socio-economic groups are less likely to start and, if they do, tend to start later. There is clear socio-economic gradient in success in quitting. Those with more education smoke for shorter durations.

Our results are relevant for current U.K. government public health initiatives. Recent White Papers on smoking and on public health (Department of Health, 1998b, 1999) have stressed the central role of tobacco control in the policy goals of improving population health and of reducing socio-economic inequalities in health. Since the early 1990s successive governments have had a commitment to annual increases in the real level of tobacco taxes, to achieve health policy objectives and encourage people to stop smoking. In the Budget of July 1997, the Chancellor of the Exchequer announced that in future budgets tobacco duties would increase on average by at least 5 per cent in real terms, and the most recent White Paper on smoking (Department of

Health 1998b) reaffirmed this commitment.

Our estimated price elasticities directly relate to the impact of above inflation tax rises on the number of years smoked by current smokers. The estimates of the impact of tax on the probability of starting are not encouraging as we do not find a significant effect. However the point estimates of the elasticity of quitting are well defined for males and robust for both males and females. All of our point estimates are in the range -0.40 to -0.63. If the typical number of years smoked is 25 years this implies that the 5% real increase in tobacco duty would lead, on average, to a reduction of between 6 and 9.5 months of smoking for each smoker.

2 Previous evidence on starting and quitting

Grossman (1999) provides a comprehensive review of the evidence of the influence of prices on substance use and abuse, and recent surveys of the literature on starting and quitting smoking are provided by Douglas (1998) and Tauras and Chaloupka (1999). There are two main approaches to modelling smoking behaviour: those that treat starting and quitting as binary events within a discrete choice framework, and those that use duration models.

Studies of smoking among teenagers typically report greater price responsiveness than among the population at large (see for example Lewit, Coate and Grossman (1981), Lewit and Coate (1982), Chaloupka and Grossman (1996), Chaloupka and Wechsler (1997), Evans and Farreley (1998), Harris and Chan (1999)). But these often use the level of consumption or the overall participation rate as their dependent variable, rather than the age of starting smoking. A number of retrospective studies have examined the effects of demographic variables, health and past smoking, on the propensity to start and quit smoking. These include Jones (1989, 1994), Sander (1995) Yen and Jones (1996), Hsieh (1998) and Dorsett (1999).

An alternative to estimating discrete choice models is to use duration analysis, as in Douglas and Hariharan (1994), Douglas (1998) and Tauras and Chaloupka (1999), all of whom use U.S. data. Douglas and Hariharan (1994) use the split population model of Schmidt and Witte (1989) to model the hazard of starting smoking, using data from the 1978 and 1979 Smoking Supplements to the U.S. National Health Interview Survey. They find evidence that those with higher lifetime educational attainment are less likely to start smoking and, if they do start, start later. Females too are less likely to start and do so at later ages. No evidence is found that higher real prices (measured at the time the individual was aged 18 and as the change in the real price between the age of 15 and 18) significantly reduce the probability of starting smoking or significantly increase the age at which individuals start.

Douglas (1998) considers the hazards of both starting and quitting using the 1987 U.S. National Health Interview Survey. He uses a split-population model based on an ordered probit, which distinguishes between those who never start smoking, those who start and will eventually quit and those who start and never quit. Price is treated as a time varying covariate in the survival function for quitting. He uses a log-logistic function for the starting hazard and a Weibull model for the quitting hazard. The price of cigarettes during the teenage years has no significant effect on the hazard of starting smoking or the probability of becoming a smoker and the number of years an individual smokes is found to have an approximately unitary elasticity with respect to the price of cigarettes.

Tauras and Chaloupka (1999) estimate Cox proportional hazard models of quitting using longitudinal data from the U.S. Monitoring the Future Surveys consisting of a representative sample of high school seniors split by gender. They estimate elasticities of the hazard function with respect to cessation to be significant at +1.12 and +1.19 for males and females respectively.

2.1 Comment

Tauras and Chaloupka (1999) express concern about using retrospective data sets in analyses of smoking behaviour. They argue that recall bias can be a serious problem in such studies, leading to errors in individuals' reported age of starting and quitting smoking which can bias parameter estimates. Further bias may occur if the individual's current state of residence is used to impute past prices of cigarettes - cross-state variation in cigarette prices occurs in the U.S. and data for individuals who move states will be subject to measurement error. Shmueli (1996) has criticised the use of measures of self-assessed health measures in retrospective analysis of quitting behaviour. This is because of problems with unobservable heterogeneity bias, which may also be a problem if measures such as previous peak consumption are used to predict quitting behaviour.

An additional problem with previous studies - both retrospective and longitudinal - is their lack of diagnostic tests to assess the fit of the models. When fitting their Cox proportional hazards models, Tauras and Chaloupka (1999) do not report whether the assumption of proportional hazards is valid for the data and parameters of their model. If the assumption is not valid, parameter estimates of quitting elasticities will be inconsistent. Douglas and Hariharan (1994) provide a graphical assessment of their split-population model but Douglas (1998) does not report the specification adequacy of his model.

This paper addresses these issues as follows. Because U.K. tobacco taxes are set at a national level, we do not encounter the problems of state-level variation that occur with U.S. data. To avoid the problems highlighted by Shmueli (1996), we estimate parsimonious models that include only independent variables that are likely to be exogenous to the individual such as gender, ethnic origin, parental smoking, education level and social class. We check the sensitivity of the parameter estimates to variation in the tax data used in the starting and quitting equations and the inclusion of measures of previous smoking behaviour. We test the models using a number of diagnostic tests from the econometric and biostatistics literature and we check the sensitivity of the quitting equations using various parametric and semi-parametric forms of the hazard function. Recall bias is tested using a rescaling of the duration variable by calendar year in a way previously used by Tunali and Pritchett (1997), which allows us to check for systematic bias in reporting by calendar year. We then adjust our models to take account of this reporting bias.

Ideally this kind of analysis should use a prospective longitudinal data set, as advised by Tauras and Chaloupka (1999). However, such high quality data is not available for the analysis of smoking behaviour over a long time horizon. By using retrospective data we can begin our period of analysis in 1920 and take account of long-run variations in cigarette taxes. This allows us to assess the effect of both price effects and the separate influence of the health scares associated with smoking that occurred during the 1960s. Our diagnostic checks and sensitivity analyses are intended to yield robust estimates of the impact of the tax rate on cigarettes as well as the effect of demographics, health scares and parental smoking behaviour, on the decisions to start and quit smoking.

3 Data and sampling

The Health and Lifestyle Survey is a study designed to record the lifestyles, personal circumstances and the physical and mental health of a large representative sample of individuals aged 18 and over living in households in England, Scotland and Wales in 1984 (Cox et al., 1987). The focus in this paper is on the first wave of the survey (hereafter HALS) and the smoking histories of participants.

HALS consisted of two home visits. In the first, an interviewer questioned the participant about self-reported health, health attitudes and health-related lifestyle such as diet, exercise, smoking and alcohol consumption. Data were also collected on demographic characteristics, employment status, qualifications and household income. In the second visit, a nurse took various measures of physiological and cognitive function and gave participants a self-completion questionnaire to collect information on personality and mental health.

The HALS sample who participated in the original home interview survey numbered 9003 individuals, a response rate of 73.5% of those initially randomly selected. Cox et al. (1987) believe that the study sample is 'a good and representative sample of the population'. The analysis in this paper uses self-reported information on individuals' smoking histories to construct duration variables, and is therefore retrospective. The following sections describe the calculation of the duration data, the price data and the other covariates used in the estimation.

3.1 Duration data

The *Health and Lifestyle Survey* contains retrospective information on whether or not an individual started smoking. It also provides information that separates non-smokers into those who have never smoked and those who class themselves as ex-smokers, allowing the analysis to be extended to distinguish between starting and quitting. These discrete choices are augmented by data on the age of starting smoking and the number of years an individual has smoked for. These can be interpreted as 'failure times' and estimated by duration analysis. In contrast to Douglas's most recent paper on starting and quitting smoking (Douglas, 1998), we model the starting and quitting processes separately, so that we can assess the fit of the models using various diagnostic tests.

In our data a smoker is defined as someone who has smoked at least one cigarette per day for a minimum of six months. For the analysis of starting, the self-reported variables FAGAGE and EXFAGAGE are used to measure the age of starting in years for those individuals who have smoked at some point in their lives.[1] The indicator variable $c = 1$ denotes a current or ex-smoker at HALS, $c = 0$ otherwise. This is a self-reported measure.[2]

Analysis of the duration of smoking is carried out on the sub-sample of individuals who had started smoking ($c = 1$). The variable SMKYRS_i is calculated for individual i as:

$$\text{SMKYRS}_i = \text{INTERVIEW}_i - \text{DOS}_i - \delta_i \text{QUIT}_i.$$

$\text{INTERVIEW}_i - \text{DOS}_i$ measures the number of days between the date of the interview[3] and the date individual i started smoking and QUIT_i measures the number of days since the individual quit for ex-smokers at the time of HALS (for whom $\delta_i = 1$).[4] Individuals who have quit smoking represent completed spells and those who are still smoking at the time of HALS represent censored spells. This duration variable is then rounded to the nearest year. Using these measures of duration, we can also identify the calendar year in which an individual started smoking and the calendar year in which they quit smoking. These can be linked to time series data on tax rates. For the HALS sample the year of starting ranges from 1909 to 1984 and for quitting it ranges from 1913 to 1985. Matching this information with annual tax data provides scope for exploiting a long time series with sufficient variability in the tax rate on cigarettes to identify elasticities of starting and quitting.

[1] These variables relate to questions 58(a) and 60(a) in the survey questionnaire which ask 'How old were you when you started to smoke cigarettes?' to current and ex-smokers.

[2] This is constructed from questions: 55(a) - 'Now, do you regularly smoke cigarettes, that is, do you regularly smoke at least one cigarette a day?' and 55(c) - 'Have you ever smoked at least on cigarette a day for as long as six months?'.

[3] The date of interview for each respondent is not provided in the HALS data set, but Cox (1999) provided us with information on estimated dates calculated from other records in the survey. All variables are converted to a common time scale using the 'elapsed time' construction in STATA release 6.

[4] QUIT_i is computed from the survey variable EXFAGAN which relates to question 60(f) - 'How long ago did you completely stop smoking cigarettes?'.

3.2 Price data

Empirical work on smoking in the United States has been able to exploit state and local variations in tobacco taxes to identify cross section variation in the price of cigarettes (see Chaloupka (1991), Douglas and Hariharan (1994) and Douglas (1998)). Tax rates in Britain are set each year by central government and the real price of cigarettes does not vary across regions to the degree that it does in the U.S.. This has prevented comparable analyses of the price elasticities of starting and quitting in Britain. In this study we solve the problem by using time series variation coupled with retrospective data on smoking durations.

We use the 'tax per cigarette', calculated using the total receipts from tobacco duty as a share of total sales volume as a proxy for the real price of cigarettes. The second edition of U.K. Smoking Statistics (Wald and Nicolaides-Bouman, 1991) provides an annual series for financial years 1920-21 to 1989-90 for the total receipts from tobacco duty for the UK.[5] Wald and Nicolaides-Bouman also present a series for total annual sales of manufactured cigarettes (numbers in millions) for the UK for the calendar years 1905 to 1987.[6] Tax receipts are divided by sales volume to give an annual series for the tax per cigarette in constant 1913-14 prices. As tobacco duty has been a high proportion of the price of cigarettes throughout the century this may be a reasonable proxy for the real price of cigarettes. However the measure will be contaminated by variations in the share of tax in the full price of cigarettes over time.[7]

The tax rate is the relevant policy instrument for the government if it wishes to use fiscal policy to influence smoking. In this sense, our estimates can be viewed as elasticities for the 'policy response', relating changes in levels of taxation to their effect on starting and quitting. To analyse the impact of the tax on cigarettes on starting and quitting smoking these data need to be mapped to the individual observations in the HALS data.

3.3 Other covariates

Because of the potential problems associated with predicting past behaviour as a function of individual characteristics that are measured at the time of the HALS survey (Shmueli, 1996), we use a parsimonious set of exogenous covariates for the starting and quitting models. This attempts, as far as possible, to use covariates that were exogenously determined prior to an individual's starting or quitting decision and therefore avoids covariates, such as health and past smoking status, that may be prone to unobservable heterogeneity bias. Summary statistics for the independent variables are presented in Table 1.

Educational status is measured by the highest qualification attained by the individual, running from the lowest (HQNONE - no qualifications) through to the highest (HQDG - higher degree qualification). Additional covariates chosen to control for other potential influences on smoking are social class (coded RGSC and included in the quitting models only), ethnic origin (coded ETH), gender, and parental smoking (PARSM).

The fact that all the variation in tax rates is attributable to variation across calendar years raises an identification problem for separating the time trend and price effects. Our solution is to use a 4th order polynomial to impose a smooth but flexible time trend and to identify price effects by variations around this trend. We therefore created a variable YEAR to measure the number of years since 1920 (the first year for which the tax data for cigarettes were available)

[5]For 1920-21 to 1975-76 this is measured in 1913-14 prices and for 1976-77 to 1989-90 it is measured in 1974 prices.

[6]Calendar year is matched with the first year of the fiscal year. That is, the 1920 volume data is matched with the 1920-21 tax data.

[7]In a study of smoking among U.S. College students, Chaloupka and Wechsler (1997) use information on state excise tax rates as an alternative measure of cigarette prices as a response to the potential endogeneity of the State level price data. Encouragingly they report that 'estimated elasticities based on the models using tax data as an instrument for price are generally very similar to those based on models using the price itself'.

Table 1: Variable definitions and sample means

		starting		quitting	
		males	females	males	females
N. obs.		3737	4861	2480	2482
N. failures		2460	2508	1176	938
variable	definition	mean			
agestrt	age of starting	-	-	15.45	18.54
start	=1 if started smoking	0.66	0.52	1.00	1.00
rgscls	social class 1/student	-	-	0.03	0.04
rgsc2	social class 2	-	-	0.20	0.20
rgsc3a	social class 3 non-manual	-	-	0.40	0.38
rgsc3	social class 3 manual	-	-	omitted regressor	
rgsc4	social class 4	-	-	0.20	0.19
rgsc5n	social class 5	-	-	0.07	0.07
hqnone	no qualification	0.45	0.51	0.63	0.67
hqcseO	highest qualification cse/O level	omitted regressor			
hqA	highest qualification A level	0.05	0.05	0.03	0.04
hqhnd	highest qualification h.n.d	0.12	0.10	0.08	0.07
hqdg	highest qualification degree	0.13	0.12	0.09	0.08
hqoth	highest qualification other	0.07	0.03	0.08	0.03
ethwheur	white/European	omitted regressor			
ethipb	Indian/Pakistani/Bangladeshi	0.02	0.14	0.01	0.00
ethbawi	black/African/West Indian	0.01	0.01	0.01	0.00
ethothnw	other non-white	0.01	0.01	0.00	0.01
parsm0	neither parent smoked	omitted regressor			
parsm1	only mother smoked	0.06	0.07	0.05	0.06
parsm2	only father smoked	0.48	0.46	0.58	0.50
parsm3	both parents smoked	0.33	0.33	0.26	0.32
year	years since 1920	35.38	36.54	41.24	45.34
lnstax/lnqtax	ln(tax per cigarette)	0.19	0.21	0.22	0.20

and included a quartic polynomial in YEAR to capture any trends in the data independent of the price effects.[8]

4 Methods

4.1 Starting

For the smokers in the sample we use the reported age of starting and the duration data can be interpreted as a complete spell. However the sample also contains individuals who have not started. In the standard duration model these observations are interpreted as incomplete spells, and it is assumed that all of these individuals will eventually 'fail' and start smoking. They are classed as 'right censored' at the time of the survey and estimation of the model has to allow for these incomplete spells.

In their analysis of U.S. data on the age of starting smoking, Douglas and Hariharan (1994) argue that standard duration analysis techniques may not be appropriate and that a split-population model should be used. The theory and logic behind the split-population duration model used to analyse starting is explained in detail in Schmidt and Witte (1989), who apply it to the study of criminal recidivism, and Douglas and Hariharan (1994) and Douglas (1998), who apply it to the study of starting smoking. In the split population duration model, the duration process applies only to those individuals who are predicted eventually to start smoking. In this paper we model eventual failure using a probit specification, that is:

$$P(\text{eventually start smoking}) = P(s=1) = \Phi(\boldsymbol{\alpha}'\mathbf{x}_i),$$

[8] Section 5.1 shows how this approach is adapted to allow for the possibility of recall bias in the starting and quitting durations.

$$P(\text{never start smoking}) \quad = P(s=0) = \quad 1 - \Phi(\boldsymbol{\alpha}'\mathbf{x}_i),$$

where $\mathbf{x}_i$ is a vector of covariates, Φ is the cumulative density function for the standard normal distribution and $\boldsymbol{\alpha}$ is a parameter vector. The probability of starting smoking at a given time t is then defined conditional upon eventually starting.

Following Douglas and Hariharan (1994), and the evidence of a plot of the Kaplan-Meier estimate of the hazard function for age of starting, we choose a log-logistic distribution to model duration. The probability density function ($f(t|s=1)$) and the survival function ($S(t|s=1)$) of the log-logistic distribution are, respectively (Greene, 1993):

$$\begin{aligned} f(t|s=1) &\equiv \frac{\lambda^{\frac{1}{\gamma}} t^{\frac{1}{\gamma}-1}}{\gamma\left(1+(\lambda t)^{\frac{1}{\gamma}}\right)^2}, \\ S(t|s=1) &\equiv \frac{1}{1+(\lambda t)^{\frac{1}{\gamma}}}, \end{aligned}$$

where $\lambda = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$ and γ is a scale parameter.

The contribution to the log-likelihood function for the split-population model then becomes, for individual i:

$$c_i \ln\left[\Phi(\boldsymbol{\alpha}'\mathbf{x}_i) f(t|s=1)\right] + (1-c_i)\ln\left[1-\Phi(\boldsymbol{\alpha}'\mathbf{x}_i) + \Phi(\boldsymbol{\alpha}'\mathbf{x}_i) S(t|s=1)\right].$$

For those who are observed smokers in the sample, $c_i = 1$ and the contribution is simply the log of the probability of being a smoker, $\Phi(\boldsymbol{\alpha}'\mathbf{x}_i)$, multiplied by the probability density function of starting at the observed starting age, $f(t|s=1)$. For those who are observed as not starting $c_i = 0$ the contribution is the probability of never starting, $1 - \Phi(\boldsymbol{\alpha}'\mathbf{x}_i)$, plus the probability of starting after the age observed at the time of the survey, $\Phi(\boldsymbol{\alpha}'\mathbf{x}_i)S(t|s=1)$.

Those individuals who start smoking can be linked to the level of tax in the calendar year that they started but this cannot be done for those respondents who had not started at the time of the HALS survey. Our solution is to attribute the level of tax in the calendar year that they were aged 16 (the modal age of starting). To check for robustness we compare the estimates to ones using the level of tax when the participants were 14 and 18.

4.2 Quitting

Previous models of quitting have used semi-parametric and parametric duration models to examine the effects of covariates on the years smoked prior to quitting. We take the sub-sample of individuals who had smoked at some point in their lives and estimate three models - the (semi-parametric) Cox proportional hazards model (Cox, 1972) and the (parametric) Weibull and gamma models.[9] All three models use time-varying covariates to model the effect of changes in the tax on cigarettes that occurred during the time an individual was a smoker on the duration of smoking.

In the Cox proportional hazards model, the hazard function at time t for individual i is defined as the product of an unspecified baseline hazard function, $h_0(t)$, and a proportionality factor, $\exp(\boldsymbol{\beta}'\mathbf{x}_i(t))$:

$$h_i(t;\mathbf{x}_i(t)) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i(t)), \tag{1}$$

where $\mathbf{x}_i(t)$ is a vector comprising of time variant and time invariant covariates.

Specifying the baseline hazard function $h_0(t)$ in (1) as $h_0(t) = hpt^{p-1}$ gives the Weibull proportional hazards model, which can yield a monotonic increasing, decreasing or constant

[9]The nature of the quitting data in the HALS means that the duration of smoking has to be viewed as a single spell; the model is not able to deal with multiple spells of smoking and with repeated attempts to quit.

hazard rate according to the sign of p. We estimate both continuous and discrete time versions of the Weibull model. Also we use a gamma mixture model to allow for unobservable heterogeneity. Estimation of the discrete time models exploits the fact that the data set is reshaped into longitudinal format to construct the time varying covariates. Jenkins's (1995) estimation routine and program are used to compute both discrete time Weibull models, with and without gamma heterogeneity.

Finally, the gamma model defines the hazard function as:

$$h_i(t; \mathbf{x}_i(t)) = \frac{\frac{|\kappa|}{\Gamma(\kappa^{-2})}(\kappa^{-2})^{\kappa^{-2}} \exp\left[\kappa^{-2}(\kappa z - e^{\kappa z})\right]}{1 - I\left(\kappa, \kappa \exp\left(\frac{z}{\sqrt{\kappa}}\right)\right)},$$

where $z = (\ln t - \exp(\beta' \mathbf{x}_i(t)))/\sigma$ and $I(k, a)$ is the incomplete gamma function and when $\kappa \neq 0$.

The tax rate is treated as a time varying covariate, as in Douglas (1998). For completed spells the contribution to the sample likelihood for these observations is the probability density function $f(t)$ at the time of quitting and the relevant observation on the tax rate is the value in the calendar year of quitting. For censored observations the contribution to the likelihood is the survival function. Tax enters this as a time varying covariate - using the value of the tax rate in all of the calendar years during which the respondent is a smoker.

4.3 Testing for misspecification

If the incorrect distribution is chosen to model either starting or quitting, parameter estimates may be biased. We use a number of diagnostic tests to test for misspecification in our models.

In the starting models, we compare the predicted survival functions from the split population models with those obtained by (non-parametric) Kaplan-Meier estimation of the survival function on the full sample and sub-sample of smokers. We check whether the predicted proportion of starters obtained from the split-population model is close to the actual proportion observed in the data. We also use plots of the Cox-Snell residuals for the observed failures in the sample to assess the general fit of the split population model for those who fail (for a general discussion of the Cox-Snell residuals see Klein and Moeschberger (1997)). In non-split duration models, a correctly fitted model should yield Cox-Snell residuals which resemble a (censored) sample from a uniform distribution. A plot of the non-parametric estimate of the cumulative hazard function for this data should therefore be a straight line through the origin. Finally, we use the squares of the fitted linear predictions in a RESET-type test of functional form.

For the quitting data, we use the re-scaled Schoenfeld residuals (Schoenfeld, 1982) and the 'global test' (Grambsch and Therneau, 1994) to test for non-proportionality of the hazard with respect to the covariates included in the Cox proportional hazards model. The re-scaled Schoenfeld residuals have an expected value of zero under the null hypothesis of proportional hazards. However, under the alternative of non-proportional hazards they will demonstrate time-dependency. We test the correlation of the residuals with a function of time - in our models this is 1 minus the Kaplan-Meier estimate of the survival function for the data. The global test for no time dependency is asymptotically distributed as a Chi-Squared variable with p degrees of freedom. In the Cox, Weibull and gamma models we also use the Cox-Snell residuals and RESET-like tests as additional tests for misspecification.

We discriminate between pooled models and models split by gender using likelihood ratio tests. For the quitting data we choose an appropriate model in the light of the diagnostic tests and, to guard against over-fitting, the Akaike information criterion (Akaike, 1974). Further we use a Wald test for $\kappa = 1$ in the gamma model (the Weibull model is a special case of the Gamma model when $\kappa = 1$).

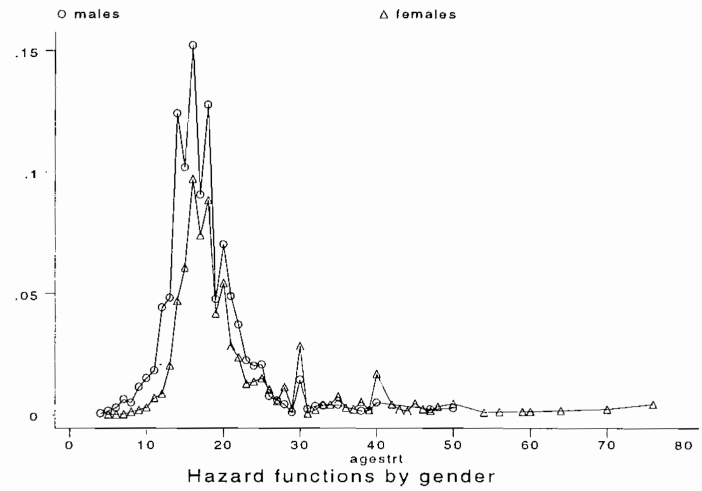

Figure 1: Kaplan-Meier hazard rates for starting smoking by gender

5 Results

5.1 Starting

Kaplan-Meier estimates of the hazard function for males and females are presented in figure 1. The non-monotone shape of the functions suggest that a log-logistic or log-normal specification might provide a suitable fit for the data, as was found by Douglas and Hariharan (1994) and Douglas (1998). LR tests of the models split by gender compared to a model estimated on the full sample indicate that the models should be analysed for male and female sub-samples separately. Results are presented in table 2.

Plots of predicted survival functions are presented in figures 2(a) and (b). These figures show that non-parametric estimation of the survival function using the full sample ('K-M - full sample', including both observed starters and observed non-starters) leads to a survival function that reaches a limit at around 0.35 for males and 0.50 for females, corresponding to the proportion of the sample who had never smoked at the time of HALS. The Kaplan-Meier estimates of the survival functions on the sub-samples who had smoked at some point in their lives ('K-M - smokers') show a much steeper descent and reaches zero around the late 20s for both groups, showing that virtually all smokers had started smoking by the time they were 30. The non-split log-logistic model's predicted hazard functions are also given in figures 2(a) and (b) ('Non-split log logistic') to show how badly they approximate the true survival function for smokers. The survival function for the full sample should reach zero, implying that all individuals in the sample will eventually 'fail' (start smoking). For males this occurs by age 80 but for females the function increases at later ages, suggesting misspecification of the model. Finally, the predicted survival functions for the split-population model for those who are observed to smoke at HALS ('split-population') are also plotted. They follow the Kaplan-Meier estimates for the subsample of smokers reasonably well for males and to a lesser extent for females.

A further test of the fit of the models is given by plots of the Cox-Snell residuals. These are presented in figure 3 for males and females. Figures 3(a) and (b) plot the Cox-Snell residuals for the full sample, i.e. for models that assume all individuals will eventually start smoking. Figures

Table 2: Split population log-logistic/probit results for starting

	males		females	
N. obs.	3737		4861	
N. failures	2460		2508	
	duration	participation	duration	participation
hqnone	0.0035	0.1996	0.0438	0.3277
	(0.272)	(2.903)	(3.315)	(5.638)
hqA	0.0616	0.0308	0.1207	-0.0469
	(2.672)	(0.249)	(5.154)	(-0.427)
hqhnd	0.0411	-0.2073	0.0494	-0.1558
	(2.370)	(-2.546)	(2.764)	(-2.021)
hqdg	0.0992	-0.2427	0.1030	-0.1930
	(6.434)	(-2.958)	(6.480)	(-2.621)
hqoth	0.0215	0.1175	0.1189	0.1529
	(1.021)	(1.120)	(3.930)	(1.233)
ethipb	0.1036	-0.1354	0.1180	-1.136
	(1.435)	(-0.745)	(1.131)	(-4.913)
ethbawi	0.1018	-0.0058	0.0106	-0.6060
	(2.028)	(-0.025)	(0.146)	(-3.00)
ethothnw	0.0381	0.2699	-0.0001	-0.1013
	(0.961)	(1.030)	(-0.002)	(-0.411)
parsm1	-0.0753	0.5486	-0.1072	0.3667
	(-3.628)	(4.936)	(-4.568)	(4.021)
parsm2	-0.0517	0.4679	-0.0335	0.2372
	(-3.179)	(6.685)	(-1.875)	(4.033)
parsm3	-0.1124	0.5812	-0.1161	0.3837
	(-6.428)	(7.713)	(-6.350)	(6.054)
year	0.0154	-0.0170	-0.0235	0.2115
	(2.684)	(-0.487)	(-2.288)	(5.576)
year2/100	-0.0568	0.1470	0.1841	-0.8923
	(-1.380)	(0.600)	(2.983)	(-3.637)
year3/1000	0.0038	-0.0433	-0.0493	0.1578
	(0.378)	(-0.733)	(-3.522)	(2.736)
year4/10000	0.0004	0.0033	0.0042	-0.0096
	(0.536)	(0.731)	(4.017)	(-2.175)
lnstax	0.1032	-0.1164	0.0499	0.1254
	(3.330)	(-0.715)	(1.507)	(0.921)
cons	2.675	0.2838	2.9355	-2.1063
	(82.228)	(1.545)	(46.518)	(-10.063)
Log L.	-8750.68		-10346.10	
χ^2 (n)	264.45 (16)		250.29 (16)	
P(fail) - actual	0.6583		0.5159	
P(fail) - predicted	0.6708		0.5398	
RESET (LR test p)	0.2957		0.0013	

Robust t-statistics in parentheses.

3(c) and (d) plot the Cox-Snell residuals for the observed failures in the split-population models. Figures 3(a) and (b) suggest major misspecification of the models that assume all individuals will eventually start smoking. Figure 3(c) suggests that the male split-population model provides a fairly good fit, as the residuals lie close to the 45° line, figure 3(d) suggests that the female split-population model is misspecified. This confirms the RESET results of table 2 and the plots based on the survival functions in figures 2(a) and (b), that the split-population model seems to fit the male sub-sample better than the female sub-sample.[10]

In table 2 the first set of coefficients for each split-population model relate to duration time and can be interpreted in terms of qualitative effects on the age of starting. The second set of coefficients relate to the probability of never starting. For both men and women parental smoking (PARSM1-3) increases the probability of starting smoking and implies an earlier age at starting. Measures of educational attainment suggest that those with higher levels of education

[10]Experiments with different sets of regressors and with a log-normal duration distribution failed to improve the performance of the model for females.

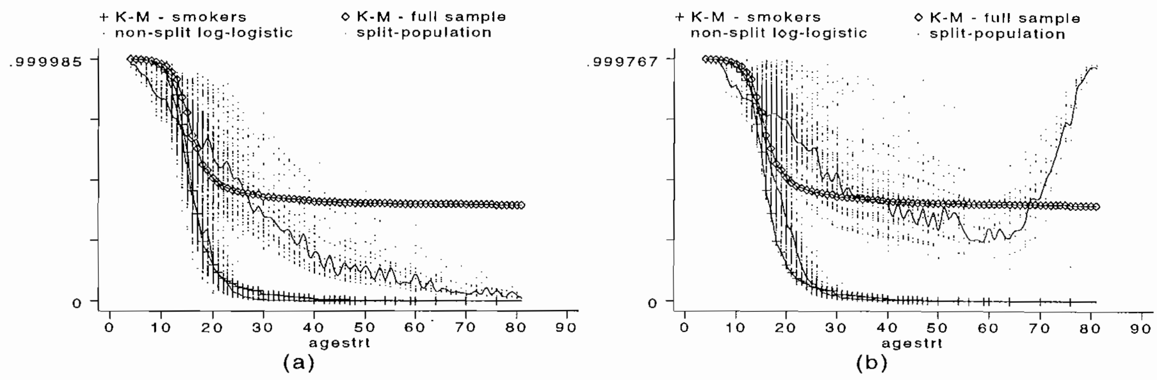

Figure 2: Predicted survival functions for (a) males and (b) females

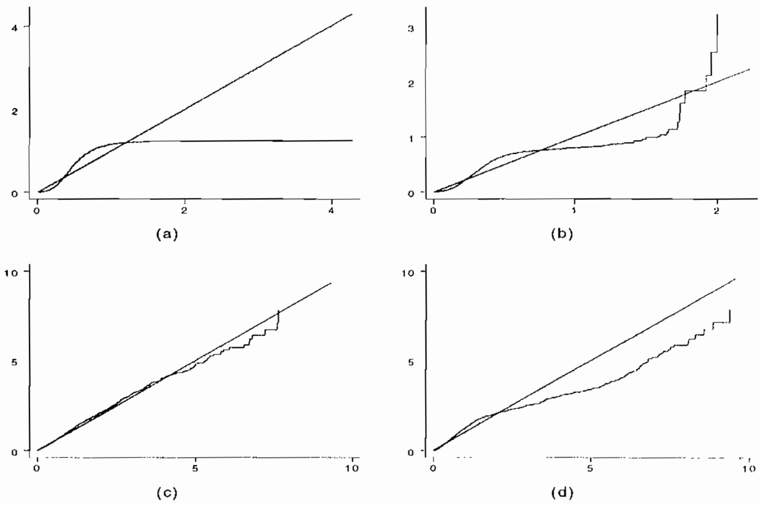

Figure 3: Cox-Snell residual plots - (a) non-split, males; (b) non-split, females; (c) split, males; (d) split, females

are less likely to start, and if they do, they start later. The evidence of a tax effect on the probability and timing of starting is weak.[11] The coefficient for the tax rate LNSTAX is not statistically significant for both participation and duration for women and for participation by

[11]The duration models are presented in accelerated failure time format and can be interpreted as regression equations for ln(failure time). The natural logarithm of the tax rate is used as a covariate and the coefficient can be interpreted directly as an elasticity.

males. There is a significant positive effect of an increased tax rate on the delay before males start. The elasticity of delay is +0.10 ($p = 0.001$).

These results are consistent with Douglas and Hariharan (1994) and Douglas (1998) but do not support the evidence cited by the recent U.K. Independent Inquiry into Inequalities in Health (Department of Health, 1998a) that 'studies in the United States and Canada indicate that young people's intention to smoke and their uptake of smoking are highly price sensitive'.

Sensitivity analysis

There is a problem of missing data on the tax rate for those individuals who had not started smoking at the time of HALS, that is, the parameter estimates in table 2 are reported using tax data for the year in which the individual started smoking for smokers and the year in which the individual was 16 (the modal age of starting) for non-smokers. To test the sensitivity of the elasticity of starting with respect to the tax on cigarettes, additional models were run using tax at 14 and 18 for non-smokers. Estimated elasticities are +0.09 and +0.11 (for the male aged 14 and 18 models) and +0.03 and +0.07 (for the female aged 14 and 18 models). None of the parameter estimates are significant in the participation equations. The same applies to the models in which the tax at ages 14, 16 and 18 is used for all individuals including those who had started smoking.

Using retrospective data sets has been criticised as a way of analysing smoking durations by Tauras and Chaloupka (1999), who argue that asking individuals to recall events many years ago can lead to self-reporting bias in the dependent variable. To assess whether this is the case with these data, the duration data can be transformed according to the methods of Tunali and Pritchett (1997) so that the duration variable measures calendar time rather than the age of starting or quitting. The hazard functions are recalculated as a function of calendar year, with the data left truncated at the calendar year when the individual was four years old for the starting models (this being the earliest age at which an individual reported starting smoking in HALS and as such is used to define the criterion for entering the 'risk set' for starting smoking). Figure 4 shows that the hazard of starting smoking by calendar year using those individuals who were at risk of starting in each calendar year. There is little evidence that the hazard of starting shows systematic patterns of recall bias. It does show a peak in the hazard, for both men and women, during the years of the Second World War and convergence of the hazard function for men and women during the 1970s and early 1980s.

5.2 Quitting

The analysis of the hazard of quitting is carried out on the sub-sample of individuals who had smoked at some point in their lives. Those individuals who were current smokers at the time of HALS can be interpreted as 'incomplete spells' and defined as censored observations in the survival models. Results of various tests for misspecification and splitting the model by sex are reported in table 3 for the Cox, Weibull and gamma models.

Table 3 serves two main purposes: it allows us to choose the most suitable model for the data on the basis of the diagnostics, and it allows us to choose whether a model split by gender is preferable to a model estimated on the full sample. The LR tests indicate that, for each specification, a model split by gender is preferred, in line with the work of Tauras and Chaloupka (1999) but in contrast to Douglas (1998).

Diagnostic tests suggest that the non-proportional hazards specification of the gamma model is preferred to the proportional hazards Cox and Weibull models. The Cox models fail the re-scaled Schoenfeld residual tests on three (males) and one (females) variables, with tax variable LNQTAX failing the test for proportionality for the model estimated on males and having a p-value of 0.06 in the model estimated on females. Furthermore, the male model fails Grambsch

Table 3: Diagnostics for quitting

	Cox		Weibull No heterogeneity		Gamma	
	males	females	males	females	males	females
L.-likelihood	-7965.49	-6435.58	-2070.34	-2140.13	-2066.90	-2137.96
LR test	2892.36		81.90		71.40	
Schoenfeld i. (p)	lhqA(0.03)	regsc1s(0.03)	-	-	-	-
	lhqdg(0.04)	lnqtax(0.06)	-	-	-	-
	lnqtax(0.04)		-	-	-	-
Schoenfeld g. (p)	0.06	0.14	-	-	-	-
RESET (Wald p)	0.88	0.24	0.910	0.28	0.79	0.513
κ test	-	-	-	-	7.27	5.67
AIC	-	-	4186.68	4326.26	4181.80	4323.92
Test of heterogeneity			21.650	2.332		
			($p = 0.000$)	($p = 0.1267$)		
			21.152	2.504		
			($p = 0.000$)	($p = 0.1135$)		

Notes
LR test - model estimated on full sample against model split by gender.
Schoenfeld i. - covariates which failed test for proportionality using re-scaled Schoenfeld residuals.
Schoenfeld g. - global test for proportionality of the hazard using method of Grambsch and Therneau (1994).
κ test - test for $\kappa = 1$ in the gamma model. $\kappa = 1$ implies a Weibull model.
AIC - Akaike information criterion - lowest value suggests the preferred model for males and females.
Test of heterogeneity - LR test for Jenkins (1995) discrete time models with and without gamma heterogeneity.

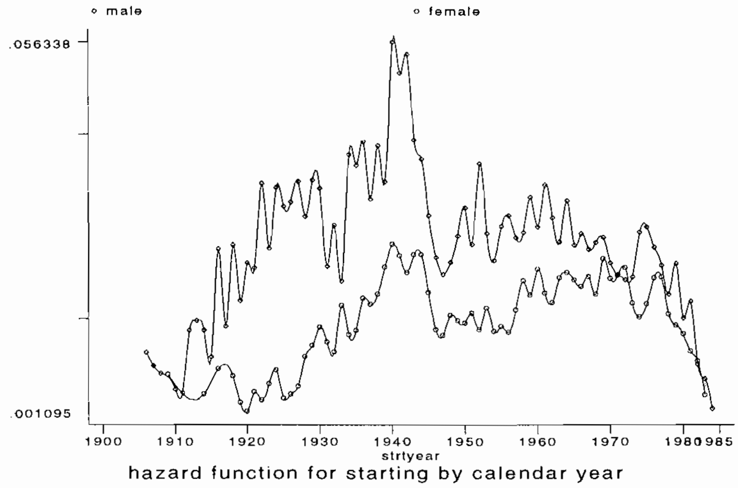

Figure 4: Hazard function for starting by calendar year

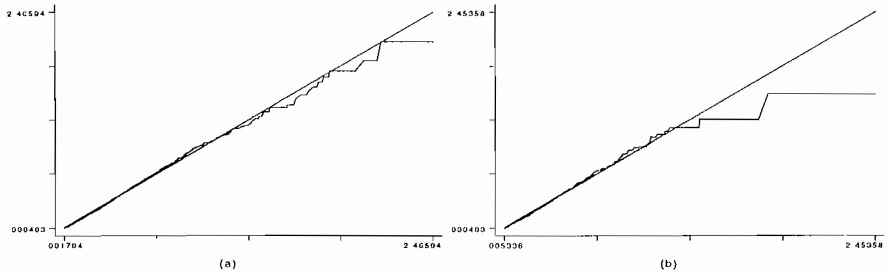

Figure 5: Cox quitting models: Cox-Snell residuals plots for (a) males and (b) females

and Therneau's global test ($p = 0.06$). Plots of the Cox-Snell residuals are shown in figures 5 to 7 and show little difference between the models. All of the plots fit the 45° line quite well, with the gamma performing slightly better for both males and females. Although the Weibull models and the gamma models pass the RESET tests, both the test of $\kappa = 1$ in the gamma models and the AIC suggest that the gamma model is preferred to the Weibull model and we concentrate on the gamma results in our discussion. Results from the gamma and Weibull models are presented in table 4.

The elasticities of years of smoking with respect to the tax rate are estimated at -0.60 ($p = 0.008$) and -0.46 ($p = 0.132$) for males and females respectively.[12] As in Douglas (1998)

[12]The models are again presented in accelerated failure time format and can be interpreted as regression

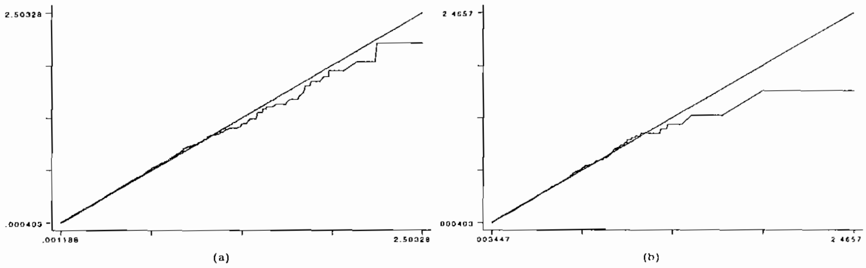

Figure 6: Weibull quitting models: Cox-Snell residuals plots for (a) males and (b) females

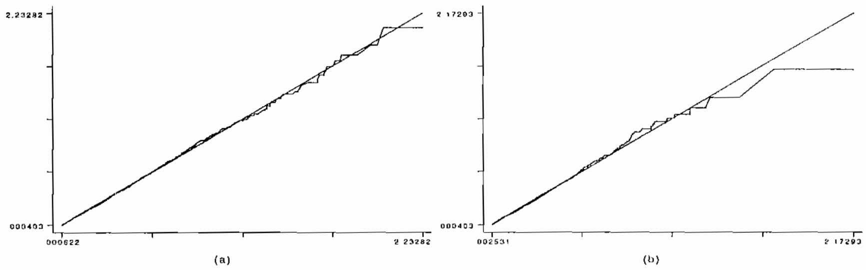

Figure 7: Gamma quitting models: Cox-Snell residuals plots for (a) males and (b) females

and Tauras and Chaloupka (1999), those with higher educational qualifications have a shorter duration of smoking, as do those in higher occupational classes. Parental smoking, which was a strong predictor of starting the habit, has little effect on quitting, with only the covariate PARSM3 - both parents smoked significantly correlated with increased smoking duration for females. Similarly, there is little effect of the ethnic origin of the individual on the duration of smoking.

Sensitivity analysis

In order to assess the robustness of the results to the influence of recall bias, the use of continuous versus discrete time specifications of the Weibull model, unobservable heterogeneity and the role

equations for ln(failure time). The natural logarithm of the tax rate is used as a covariate and the coefficient can be interpreted directly as an elasticity. Elasticities for the semi-parametric Cox models cannot be calculated in this way, since they do not assume any functional form for the dependent variable failure time.

Table 4: Accelerated failure time specifications for quitting: gamma and weibull

	males		females	
No. obs.	2480		2482	
No. failures	1176		938	
Time at risk	65420		56551	
	gamma	Weibull	gamma	Weibull
rgscls	-0.0664	-0.0652	-0.5637	-0.5012
	(-0.543)	(-0.542)	(-3.588)	(-3.424)
rgsc2	0.0400	0.0411	-0.0332	-0.0264
	(0.493)	(0.523)	(-0.303)	(-0.255)
rgsc3a	0.2082	0.2036	0.2039	0.2012
	(2.758)	(2.791)	(1.985)	(2.057)
rgsc4	0.3220	0.3108	0.1607	0.1481
	(3.694)	(3.726)	(1.394)	(1.342)
rgsc5n	0.3184	0.3120	0.4096	0.3913
	(2.640)	(2.694)	(2.469)	(2.432)
hqnone	0.1976	0.2049	0.1798	0.1975
	(2.572)	(2.828)	(1.888)	(2.181)
hqA	-0.0339	0.0119	-0.1128	-0.0901
	(-0.235)	(0.085)	(-0.655)	(-0.540)
hqhnd	-0.0200	-0.0178	-0.1601	-0.1469
	(-0.209)	(-0.195)	(-1.198)	(-1.176)
hqdg	-0.2491	-0.2108	-0.3436	-0.3163
	(-2.469)	(-2.273)	(-2.796)	(-2.665)
hqoth	0.3462	0.3531	-0.1905	-0.1825
	(3.085)	(3.243)	(-1.041)	(-1.059)
ethipb	0.8926	0.8698	-0.0212	0.0427
	(2.442)	(2.337)	(-0.033)	(0.068)
ethbawi	-0.1771	-0.2019	-0.1061	-0.1404
	(-0.711)	(-0.859)	(-0.225)	(-0.293)
ethothnw	-0.0836	-0.1101	0.3841	0.3447
	(-0.279)	(-0.410)	(0.809)	(0.755)
parsm1	0.1738	0.1545	0.0923	0.0698
	(1.350)	(1.257)	(0.603)	(0.479)
parsm2	-0.0542	-0.0498	0.2005	0.1883
	(-0.664)	(-0.673)	(1.895)	(1.906)
parsm3	0.0822	0.0551	0.2811	0.2400
	(0.901)	(0.671)	(2.467)	(2.289)
year	-0.0118	-0.0185	0.0052	0.0102
	(-0.125)	(-0.182)	(0.038)	(0.071)
year2/100	0.3281	0.2981	0.2664	0.2041
	(0.684)	(0.589)	(0.377)	(0.281)
year3/1000	-0.1107	-0.0994	-0.0948	-0.0810
	(-1.137)	(-0.983)	(-0.655)	(-0.551)
year4/10000	0.0090	0.0082	0.0072	0.0064
	(1.350)	(1.190)	(0.726)	(0.637)
lnqtax	-0.5975	-0.5285	-0.4587	-0.4138
	(-2.658)	(-2.421)	(-1.355)	(-1.235)
cons	4.1264	4.3700	4.1887	4.3277
	(6.2027)	(6.038)	(4.104)	(4.070)
$\ln(\sigma)$, $\ln(p)$	-0.2258	0.3137	0.0280	0.0798
	(-5.173)	(11.040)	(0.484)	(2.586)
κ	0.7773		0.7792	
	(9.414)		(8.403)	
Log L.	-2066.90		-2137.96	-2140.13
χ^2 (n)	274.35 (21)		249.73 (21)	266.13
RESET (Wald test p)	0.789			

Robust t-statistics in parentheses.

of measures of past cigarette smoking, we carried out an extensive sensitivity analysis of our results. Table 5 summarises the results for the estimated tax elasticity of quitting for these various models.

Assessment of recall bias was carried out in a similar manner to the starting models. For each individual we reconstructed the duration variable to measure calender year of quitting,

Table 5: Sensitivity analysis of the elasticity of quitting

		males	females
1. gamma	benchmark	-0.60	-0.46
2. gamma	adjusted for recall bias	-0.49	-0.49
3. gamma	with measures of past smoking	-0.63	-0.55
4. Weibull continuous	benchmark	-0.53	-0.41
5. Weibull continuous	adjusted for recall bias	-0.45	-0.48
6. Weibull continuous	with measures of past smoking	-0.52	-0.40
7. Weibull discrete	benchmark	-0.55	-0.44
8. Weibull discrete	with gamma heterogeneity	-0.41	-0.41
9. Weibull discrete	adjusted for recall bias	-0.46	-0.52
10. Weibull discrete	adjusted for recall bias, with gamma heterogeneity	-0.45	-0.48

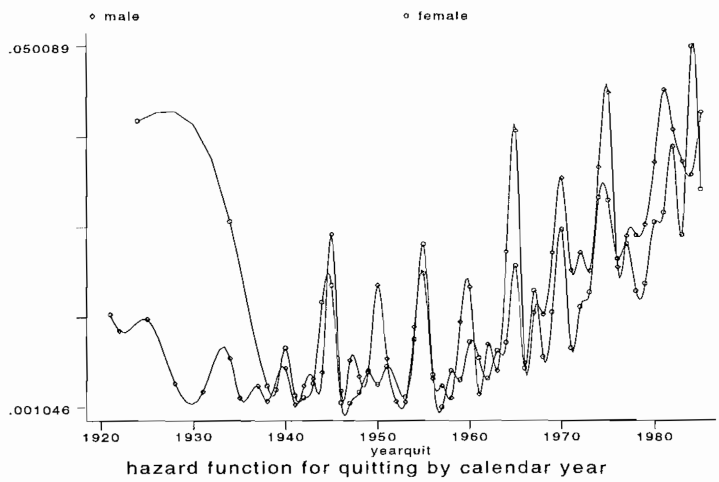

Figure 8: Hazard function for quitting by calendar year

left truncated on the calendar year of starting. Results for the hazard of quitting by calendar year are presented in figure 8. There appears to be serious recall bias for the quitting models, with peaks in the hazard being recorded at five and ten year intervals.[13] This suggests that, when questioned, respondents rounded off the 'number of years since they quit' (the variable EXFAGAN) to the nearest five or ten year mark. To control for this, we estimated quitting models with dummy variables to capture the effects of the five and ten year recall bias. These variables equal one if the calendar year in question is a multiple of 5 or 10 years prior to the year in which the respondent was interviewed. Parameter estimates for models adjusted for recall bias are reported in rows 2, 5, 9 and 10 of table 5.[14] Comparison of rows 2, 5, 9 and 10 of

[13]The time trend in the hazard function in figure 8 is informative. Prior to 1940, particularly for women, the data is too sparse to be meaningful. During the 1940s and 1950s the hazard function is relatively flat and, after the health scares of the 1960s it progressively increases.

[14]In proportional hazards models, such dummy variables have a natural interpretation as 'spikes' or mass points in the baseline hazard.

table 5 with the relevant benchmark models shows that adjusting for recall bias has little effect on estimated elasticities of quitting. In the gamma model the estimated elasticity changes from -0.60 to -0.49 for males and from -0.46 to -0.49 for females. In the Weibull models there is again a small absolute fall for males and rise for females.

Douglas (1998) includes measures of previous cigarette consumption and of age of starting to capture addiction and lagged duration dependence in his estimates of the hazard of quitting. We have avoided these variables in our preferred models because of a concern with unobservable heterogeneity bias. However, for comparison the models are re-estimated including the age of starting smoking along with a measure of previous peak consumption.[15] Estimated quitting elasticities are reported in rows 3 and 6 of table 5 and appear robust to these specifications; all remain negative and significant at the 5% level for males and negative but not significant for females. Estimates show little change in the value of the elasticity of quitting for males and a slight increase in the absolute value for the female gamma model (0.09) although there is little change for the female Weibull model.

Row 8 of table 5 reports estimated elasticities when controlling for unobserved gamma heterogeneity using Jenkins's (1995) estimator. The estimated elasticity changes from -0.55 to -0.41 for males and from -0.44 to -0.41 for females. Table 3 shows that the unobservable heterogeneity parameter is significant in the discrete time Weibull/gamma mixture model estimated on the male sub-sample, but not on the female sub-sample. Row 10 of table 5 shows that, after adjusting for recall bias, the absolute value of the elasticity of quitting falls by only 0.01 for males.[16]

6 Policy implications

The goal of reducing the U.K. death rate from cancer in people under 75 by at least one fifth by the year 2010 is one of the four broad targets of the recent public health White Paper, 'Our Healthier Nation' (Department of Health, 1999). Many of these cancer deaths are attributed to smoking and this target has been linked to the Government's stated policy of reducing socio-economic inequalities in health. In July 1999, the Secretary of State for Health told the House of Commons:

> '... tobacco smoking is the principal cause of the inequalities in health among adults in this country. Seventy per cent. of deaths of working-class people, over and above what one would expect among middle class people, are the result of smoking' (Hansard, 1999)

This view is informed by the Independent Inquiry into Inequalities in Health (Department of Health, 1998a). Their review of the evidence lead to the conclusion that 'smoking is an important component of differences in mortality between social classes.'

In December 1998 the Government published a White Paper, 'Smoking Kills' (Department of Health, 1998b), to define their strategy towards smoking. The White Paper reaffirms the use of above inflation increases in tobacco taxes for health policy. Since the early 1990s successive governments have had a commitment to annual increases in the real level of tobacco taxes,

[15] Previous peak consumption is constructed from the two variables FAGMAX and EXFAGMAX which relate to question 56(b) - 'What is the maximum number of cigarettes you have regularly smoked in a day? and 50(c) - 'What was the maximum number of cigarettes you ever regularly smoked in a day?'.

[16] We also compared estimates from proportional hazards versions of the Weibull models with those of the Cox models. There was little variation. For males, the Cox models yield a proportional shift in the baseline hazard of 2.05 (compared to 2.06 for the Weibull model), and for females the estimates are 1.53 (compared to 1.55). When adjusted for recall bias the estimates are identical to three decimal places for males and are 1.65 compared to 1.67 for females. Elasticities are a little higher for the discrete time Weibull models compared to the continuous time Weibull models.

to achieve health policy objectives and encourage people to stop smoking. In the Budget of July 1997, the Chancellor of the Exchequer announced that, in future, tobacco duties would be increased on average by at least 5 per cent in real terms. This commitment was reiterated in the 1998 and 1999 Budgets. For example, in his March 1999 statement the Chancellor of the Exchequer announced:

> ' ..duty on tobacco will rise by the normal escalator of 5 per cent above inflation....a policy on cigarettes which successive British Governments have adopted for good and urgent health reasons.' (H.M. Treasury, 1999)

What are the implications of our results for this policy? Our estimated price elasticities directly relate to the impact of above inflation tax rises on the number of years smoked by current smokers (and indirectly on the hazards of starting and quitting). The estimates of the impact of tax on the probability of starting are not encouraging as we do not find significant effect. However the point estimates of the elasticity of quitting are well defined for males and robust for both males and females. Point estimates are in the range -0.4 to -0.63. If the typical number of years smoked is 25 years this implies that the 5% real increase in tobacco duty would lead, on average, to a reduction of between 6 and 9.5 months of smoking for each smoker. Recent estimates suggest that there are around 12.1 million current smokers in the U.K. (ASH 1999). Then the potential saving in total number of years smoked across the population is substantial.

Acknowledgements
Data from the *Health and Lifestyle Survey* were supplied by the ESRC Data Archive. Neither the original collectors of the data nor the Archive bear any responsibility for the analysis or interpretations presented here. Thanks to Brian Cox for supplying us with the estimated dates of interview for the HALS sample. Thanks to Paul Contoyannis and participants in the York Seminars in Health Econometrics for comments. Thanks to Javier Bezos for the titlesec LaTeXpackage and to Stephen Jenkins for making available the pgmhaz program for STATA. Any errors are the sole responsibility of the authors.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction and automatic control*, AC-19:716–723.

ASH (1999). Basic Facts No.1: Smoking Statistics, Action on Smoking and Health. http://www.ash.org.uk/papers/basic01.html.

Chaloupka, F. J. (1991). Rational addictive behaviour and cigarette smoking. *Journal of Political Economy*, 99:722–742.

Chaloupka, F. J. and Grossman, M. (1996). Price, tobacco control policies and youth smoking. NBER Working paper 5740.

Chaloupka, F. J. and Wechsler, H. (1997). Price, tobacco control policies and smoking among young adults. *Journal of Health Economics*, 16:359–373.

Cox, B. D. (1999). Personal communication.

Cox, B. D., Blaxter, M., Fenner, J., Golding, J., and Gore, M. (1987). *The Health and Lifestyle Survey*. Health Promotion Research Trust, London, first edition.

Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 74:187–220.

Department of Health (1998a). *Independent Inquiry into Inequalities in Health*. The Stationery Office, Stationery Office, London.

Department of Health (1998b). *Smoking Kills*. The Stationery Office, Stationery Office, London.

Department of Health (1999). *Our Healthier Nation*. The Stationery Office, Stationery Office, London.

Dorsett, R. (1999). An econometric analysis of smoking prevalence among lone mothers. *Journal of Health Economics*, 18:429–441.

Douglas, S. (1998). The duration of the smoking habit. *Economic Inquiry*, XXXVI:49–64.

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.

Evans, W. N. and Farrelly, M. C. (1998). The compensating behaviour of smokers: taxes, tar and nicotine. *RAND Journal of Economics*, 29:578–595.

Grambsch, P. and Therneau, T. (1994). Partial hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526.

Greene, W. (1993). *Econometric analysis*. Macmillan, New York, second edition.

Grossman, M. (1999). The economics of substance use and abuse. In *Paper presented at the Taipai International Conference on Health Economics, 25-27 March 1999*.

Hansard (1999). Comment by Frank Dobson, Secretary of State for Health. House of Commons Debates for 2 July 1999 (pt 3).

Harris, J. E. and Chan, S. W. (1999). The continuum of addiction: cigarette smoking in relation to price among Americans 15-29. *Health Economics*, 8:81–86.

H.M.Treasury (1999). Budget speech 1999. http://www.hm-treasury.gov.uk/budget/1999/speech.html.

Hsieh, C.-R. (1998). Health risk and the decision to quit smoking. *Applied Economics*, 30:795–804.

Jenkins, S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57:129–138.

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

Jones, A. M. (1994). Health, addiction, social interaction and the decision to quit smoking. *Journal of Health Economics*, 13:93–110.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis.* Springer-Verlag, New York, first edition.

Lewit, E. and Coate, D. (1982). The potential for using excise taxes to reduce smoking. *Journal of Health Economics*, 1:121–145.

Lewit, E. M., Coate, D., and Grossman, M. (1981). The effect of government regulation on teenage smoking. *Journal of Law and Economics*, 545-570:545–570.

Sander, W. (1995). Schooling and quitting smoking. *Review of Economics and Statistics*, 129:191–198.

Schmidt, P. and Witte, A. D. (1989). Predicting criminal recidivism using 'split population survival time models'. *Journal of Econometrics*, 40:141–159.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69:239–241.

Shmueli, A. (1996). Smoking cessation and health: a comment. *Journal of Health Economics*, 15:751–754.

Tauras, J. A. and Chaloupka, F. J. (1999). Determinants of smoking cessation: an analysis of young adult men and women. NBER Working paper 7262.

Tunali, I. and Pritchett, J. (1997). Cox regression with alternative concepts of waiting time: the New Orleans yellow fever epidemic of 1853. *Journal of Applied Econometrics*, 12:1–25.

Wald, N. and Nicolaides-Bouman, A. (1991). *U.K. Smoking Statistics.* Oxford University Press, Oxford, second edition.

Yen, S. and Jones, A. M. (1996). Individual cigarette consumption and addiction: a flexible limited dependent variable approach. *Health Economics*, 5:105–117.