The Hazards of Starting and Quitting Smoking: Some Australian Evidence

MICHAEL P. KIDD*
Department of Economics, University
of Aberdeen, Scotland

SANDRA HOPKINS
School of Economics and Finance, Curtin University
of Technology, Western Australia, Australia

The empirical analysis employs individual level data from the Australian Health Survey combined with retrospective data on tobacco price matched to the age at which the individual started and quit smoking. Split-population hazard models are estimated for both starting and quitting smoking. The analysis suggests price plays a significant role in the decision to start smoking but not in the decision to quit. Further sensitivity analysis of different age groups and an alternative data source, questions the robustness of the significant role of price in the smoking initiation decision. From a policy perspective, the results indicate that increases in tobacco taxation can be an important instrument in reducing the incidence of smoking, but should be combined with other mechanisms such as mandating smoke-free environments and antismoking education. Our results strongly support the targeting of antismoking campaigns towards teenagers.

I Introduction

The prevalence of smoking in the Australian population has fallen rapidly in the last 55 years. In 1945, 72 per cent of men and 26 per cent of women smoked. The prevalence of male smoking declined to 30 per cent by 1989 and 25 per cent by 1998 (Quit Victoria, 2003). For women, the pattern is somewhat different with the prevalence rising over time from 26 per cent in 1945 to 35 per cent in

1978, followed by a subsequent decline to approximately 27 per cent by 1989 and 20 per cent in 1998. (Quit Victoria, 2003). After a decline in teenagers' smoking propensities in the 1980s, the start-up rates for teenagers of both genders actually increased or at least remained constant, over the 1990s (Quit Victoria, 2003).

Despite the overall fall in the prevalence of smoking, it remains a major public policy concern. Tobacco consumption is a contributory factor in more deaths than all other drugs, both licit and illicit, combined. One estimate of the economic costs of tobacco use in Australia places the combined tangible and intangible costs at a total of $12 billion with the benefits in terms of government revenue of $1.9 billion (Collins and Lapsley, 1996). Given the enormity of the economic costs of smoking, Australian public policy over the last 30 years has attempted to target both the propensity to start smoking and the propensity to quit. The addictive nature of tobacco products indicates that targeting the initial decision to start smoking is likely to be more effective. This is particularly true given that start-up rates peak in the teenage years and rapidly decline thereafter.

177

There are three basic mechanisms which the government can use to discourage people from starting to smoke and to induce them to quit once they have started: sales taxes on tobacco products which lead to increased prices; regulatory intervention which can prohibit tobacco advertising and increase the distribution of smoke-free environments; and anti-smoking education.

The current study examines the pattern and main determinants of both the age of initiation of smoking and the decision to quit. One of the determinants over which the government has considerable control through its use of selective taxes is the price of tobacco. Our major focus is on the impact of the price of smoking on the decision to start and the decision to quit smoking; and whether the impact differs by gender. The study employs individual level data from the National Health Survey (NHS) (ABS, 1990) and from the National Drug Strategy Household Survey (NDSHS) of 1998 (Australian Institute of Health and Welfare, 1999). These cross-section data sets are combined with retrospective time series data on real tobacco price matched to the age at which the individual started and quit smoking. This allows one to assess whether price plays an important role in smoking decisions. This is clearly an important issue from a public policy perspective.

The remainder of the paper is divided into five sections. The next section provides a brief overview of the existing literature, sections 3 and 4 describe the data and empirical implementation, Section 5 summarises the main econometric results and Section 6 concludes.

II The Empirical Literature

Much of the literature pertaining to starting and quitting smoking is based on the rational addiction model of Becker & Murphy (1988). This model assumes that people only take up an addictive habit after fully weighing up the costs and benefits. An addictive good is one that exhibits adjacent complementarity: an increase in the current consumption of the good increases its future consumption. A rational individual, in making a decision to start smoking, will do so if and only if at any given point in time, the discounted lifetime marginal benefit of acquiring the habit exceeds its discounted lifetime marginal cost.

The perceived marginal benefits of smoking include the current and future pleasure it gives. The pleasure includes the perception of increased maturity and acceptance by peers. The marginal costs are derived from current and future monetary outlays, present and future health costs and the effort

to smoke. The effort to smoke has increased with changes in smoking regulations and restrictions. The marginal costs of smoking tend to increase overtime with an increase in the probability of smoking related illness.

Bardsley and Olekalns (1999) use Becker's rational addiction model of aggregate consumption to examine cigarette and tobacco consumption in Australia over the period 1962/63–1995/96. They model aggregate current consumption of the addictive good, tobacco as a function of: household disposable income, the proportion of the adult population under the age of 45, the proportion of people who report workplace smoking bans, real per capita advertising expenditure in main media outlets by tobacco companies and real per capita expenditure by antismoking organisations.

Their results support the Becker model, namely; current consumption is affected by past and anticipated future consumption, and price and income are both significant. They find that most of the variation in consumption has been driven by price (including taxes), and by income and demographic effects. The model suggests that, *ceteris paribus* consumption will rise as the population ages and real incomes rise. Their results place critical importance on the role of tobacco taxes with almost all the reduction in tobacco consumption attributed to tobacco taxes, with minimal role for regulatory intervention.

The literature on the economics of tobacco has utilised time series aggregate data as well as individual survey data in an attempt to estimate the importance of price in the decision to start and quit smoking. In the current paper, we use micro level data to look at the relationship between smoking behaviour, individual characteristics and price. A finding that price is important in the decision to start and/or quit smoking would be consistent with the rational addition model. Recent US and UK research using micro level data has also focused on the impact of price on initiation of the smoking decision. Research using US data by Douglas and Hariharan (1994), Douglas (1998) and DeCicca *et al.* (2002) and UK data. Forster & Jones (2003)[1] suggests that cigarette prices are not a statistically significant determinant of smoking initiation. In contrast, US studies by Tauras and Chaloupka (1999), Tauras (1999) and Ross and Chaloupka (2003) all find

---

[1] It is important to note that the results presented in the corrigendum are rather different to those included in the original paper. In the corrected results price is no longer significant.

a statistically significant role for price. Thus the current evidence is somewhat mixed.

Our analysis follows that of Douglas and Hariharan for the US and Forster and Jones for the UK. We use duration analysis to model the hazard rate of both starting to smoke and quitting. A split population specification is also employed. The focus of the analysis is to provide further evidence on whether the price of tobacco is an important determinant of the decisions to start and quit smoking.

III The Data

Two data sources are employed in the empirical analysis: the National Health Survey, NHS (1990) and the National Drug Strategy Household Survey, NDSHS (1998). Both data sets represent a single cross-section survey providing individual level information on a random sample of the Australian population. The data sets include information on an array of demographic variables. In addition and of crucial importance, both also include retrospective information on individual smoking behaviour. The information includes smoker status, that is whether the individual is a current or ex-smoker or had never smoked, and the age at which the individual started and quit smoking if relevant. Despite the availability of the more recent 1998 survey, we decided to focus our interest on the 1990 data. The major reason for this preference is the small size of the 1998 sample.[2] Unfortunately, there are subtle differences in the questions asked across the two data sets and so results are not directly comparable.[3] However, some sensitivity analysis across data sets is carried out in order to examine the robustness of our overall results.

The 1990 NHS includes information on the age of the individual at the survey date[4]. This allows the

[2] The sample is less than one fifth the size of the 1990 data set.

[3] The key smoking variables defined in the NHS (1990) survey are: (1) age commenced regular smoking-where regular is defined to be one or more cigarettes per day on average, and (2) age quit regular smoking. The NDSHS data includes the two key questions: (1) what age were you when you started smoking daily? and (2) what age were you when you last smoked daily? These subtle changes in language can lead to significant changes in results.

[4] The publicly released data set groups age into 5 year bands and thus it is not possible to precisely track calendar time. Fortunately, we were able to negotiate with the ABS allowing remote access to the unreleased data which does incorporate actual age. One advantage of the NDSHS (1998) data source is that it does include actual age. As a check on robustness we replicate our analysis using this alternative data source.

age at which the individual started or quit smoking to be translated into calendar time. We also have a separate time series data set documenting the relative price of cigarettes. Thus merging the data sets allows one to track the tobacco price level facing each individual at each year of their life.

The price index of tobacco and cigarettes comes from an unpublished ABS source. The series is called the Consumer Price Index (CPI), by capital city by expenditure class (Quarterly)-cigarettes and tobacco 13th series. The quarterly series was converted into an annual series by the ABS. The series runs from 1963 to 1999 and provides the index for each capital city as well as a weighted average of all eight capitals. Figure 1 plots the weighted average of the price index.[5] Tobacco prices have moved over time in response to a number of forces, one of which is tobacco taxes which in Australia are imposed by state and territory governments. The use of capital city data captures the differential rates of state taxes on tobacco products.

Many changes have occurred in the packaging and marketing of cigarettes over time, including changes in the strength of the tobacco in a cigarette and the number of cigarettes in a pack. The price index adopted in the current study attempts to take into account a number of these issues. The number of cigarettes in a pack is irrelevant as the CPI for tobacco and cigarettes is based on the price per cigarette. The quality of cigarettes is monitored by the ABS. According to the ABS (2003), the price of cigarettes is quality adjusted as required, with quality measured by the quantity of tobacco used in each cigarette. Information on the amount of tobacco per cigarette is obtained regularly from the manufacturers.

When we merge the two data sets, namely the ABS price index for cigarettes and tobacco and the smoking behaviour overtime of individuals surveyed in the NHS (ABS, 1990), we can match the tobacco price level facing each individual at each year of their life. The tobacco price index varies over time and by location. The price index for one of the eight capital cities is used to match to the state of residence of a given individual at the time of the survey. Unfortunately we must assume that an individual's regional location at age 18 is the same as that at the time of the survey.

[5] Bardsley and Olekalns (1999) derive a tobacco price series using published National Accounts data. Their series does not vary by state/capital city. To explore the relationship of this derived series with the unpublished (weighted average) series used in the current analysis, we calculate the simple correlation. The series are clearly extremely similar with a correlation of 0.96.

FIGURE 1
Relative Tobacco Price Index-Unpublished ABS Index-weighted Average Australian Capital City
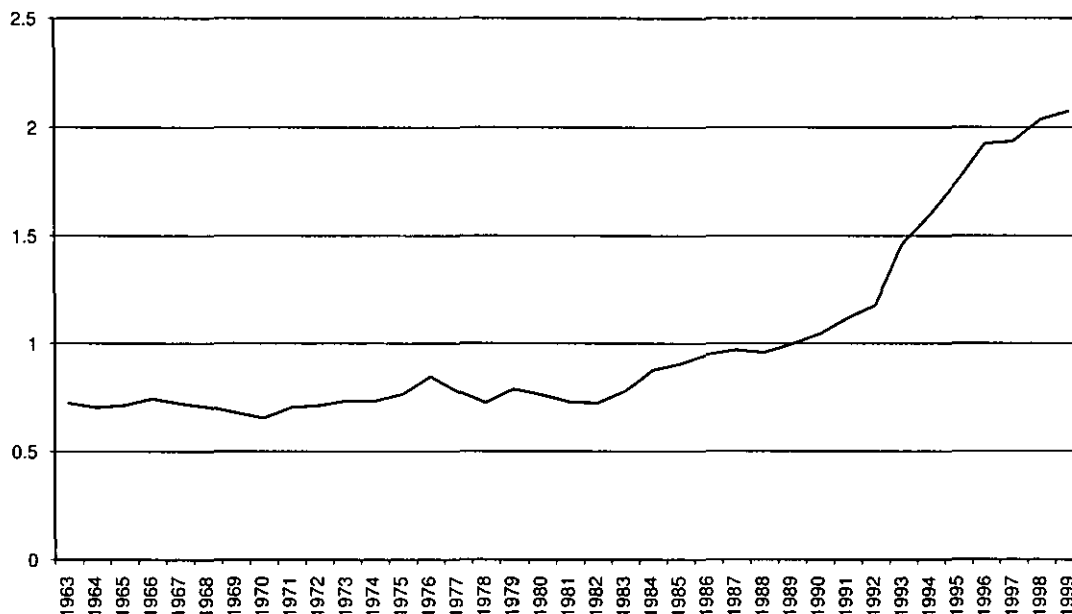
The focus of the empirical analysis is estimating the hazards of starting and quitting smoking, where the hazard is defined as the probability of exit, or the decision to start or quit smoking at a given age, conditional on not exiting previously. We assume that individuals are first exposed to the risk of starting to smoke from age 10.[6] Unfortunately, given that the price data is only available from 1963 onwards, this restricts our analysis to individuals aged up to 37 in 1990. Also the smoking questions are only asked of individuals aged 18 and over in the survey year of 1990.

The existing literature suggests completed education level as a key determinant of smoking behaviour. The empirical regularity is that the more educated are less likely to smoke and conditional on smoking, begin later. The basic theoretical rationale is that those with more education have a lower discount rate, valuing the future more highly. This in turn implies that the future health costs are more heavily weighted than the present consumption benefits. Thus, following Douglas and Hariharan (1994) we limit attention to those individuals who might reasonably

be assumed to have completed their education. Thus our sample is restricted to 27-37 year olds.

This sample is fairly restrictive and thus, despite our reservations about including younger individuals (who might not have completed their education), we also separately examine the smoking behaviour of 18-26 year olds. This is useful as it sheds light on the robustness of our results. As additional sensitivity analysis we also briefly examine results for the alternative data source, the 1998 NDSHS.

The two key duration variables in the empirical analysis are time until the individual begins smoking and for those who start, the duration of smoking, measured as the time from initiation to cessation.[7] The former variable is clearly defined for those individuals observed to start smoking. The remainder of the sample is treated as right censored at their current age. The time to quit duration variable is defined only for those individuals who have at some point started to smoke. Given this subsample, the duration variable is well defined for those observed to quit and right censored for those continuing to smoke at the survey date.

---

[6] This defines the time origin and was chosen as very few individuals start to smoke prior to age 10. Douglas (1998) selects age 11 as the time origin.

[7] Given data deficiencies multiple spells of starting and quitting cannot be separately identified.

FIGURE 2
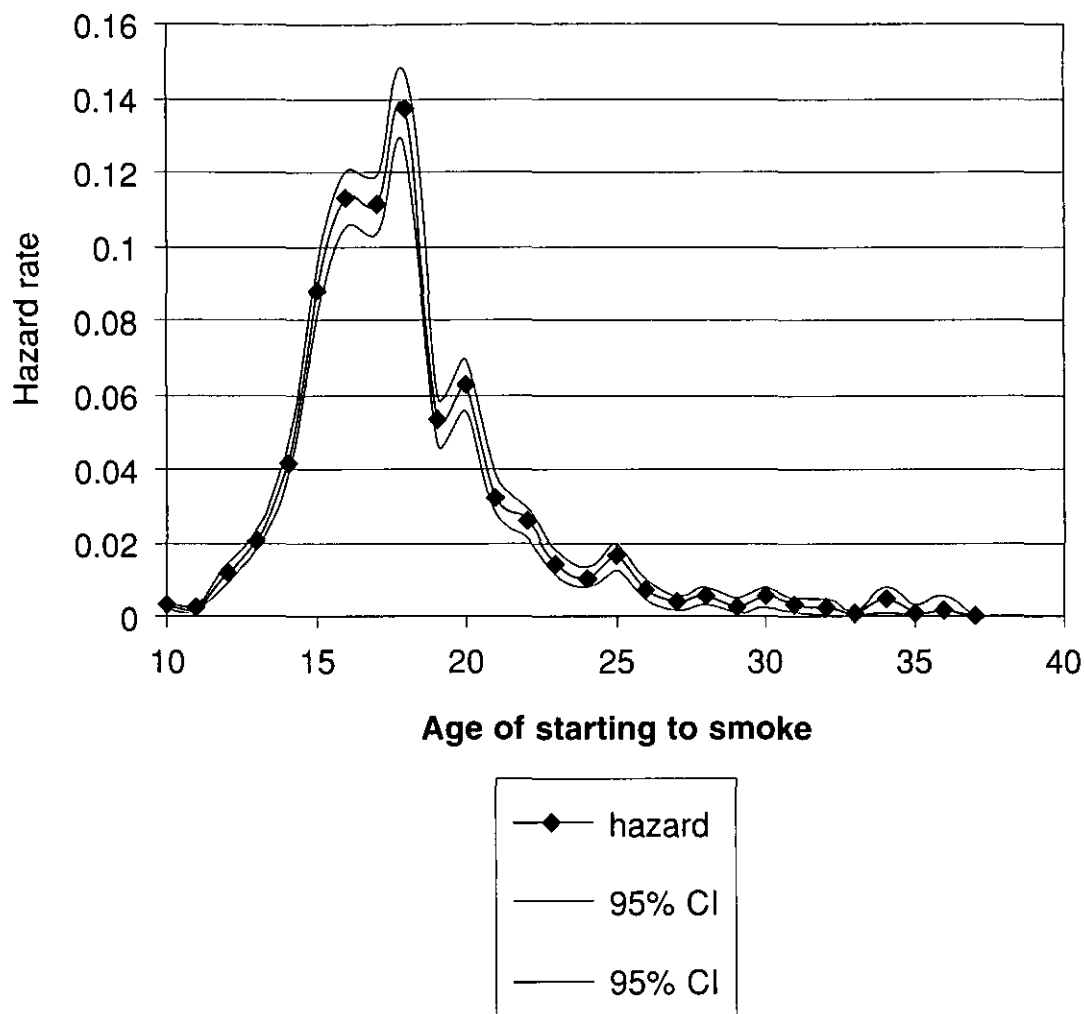*Hazard of Starting to Smoke*

Age of starting to smoke

Figure 2 illustrates the hazard rate of age of starting to smoke.[8] The pattern is similar to that previously reported in the literature, namely, the hazard rate reaches a maximum in the teenage years and declines steeply thereafter.[9] Thus, from a policy perspective, targeting anti-smoking cam-

[8] The figure also includes the 95 per cent confidence interval around the hazard.

[9] The overall pattern is extremely similar across gender, although the male graph is displaced upwards.

paigns at teenagers would appear to be particularly useful. If an individual has not started to smoke by age 20, the probability of a future start is extremely low.

Table 1 includes summary statistics on the sample used in the analysis of starting to smoke. The number of subjects is 9 402, with 5 051 starting to smoke prior to the survey date. The total number of observations measured as the number of subjects times the number of time periods is 141 022. The average number of time periods an individual remains in the

TABLE 1
Description of Variables in the Data Set

Variable	Description	Male	Female
Ever smoked percentage		58.5	49.1
Age started smoking	For sample who started prior to survey date	17.06	17.41
Ever quit smoking percentage	For sample of those observed to start prior to survey date	34.9	39.1
Age		31.95	31.87
Degree	Bachelor degree of higher	0.16	0.12
Trade	Trade certificate or apprenticeship	0.29	0.03
Diploma	Post-secondary certificate or qualification	0.15	0.32
Other-Education	Other postsecondary qualification	0.01	0.02
Australian	Equal to one if Australian born, zero otherwise	0.74	0.75
Number of observations/subjects in combined sample		4 619	4 783

analysis is 15 years. The male sample consists of 4 619 subjects, of whom 2 704 or 58.5 per cent started to smoke prior to the survey date. The female sample has 4 783 subjects of whom 2 347 or 49.1 per cent are current or ex-smokers.

As documented earlier, the age range of the sample is restricted to 27–37 year olds, the average age is 31.9 years. The age range for starting to smoke for men, in the uncensored sample is from 10 to 34, with a mean of 17.1 years. Similarly, for females the range is from 10 to 36, with a mean of 17.4. In terms of the overall sample, both censored and uncensored, 74.9 per cent are Australian born, the remainder being immigrants. Approximately 14 per cent of the total sample has the highest category of educational qualification, that is, degree or higher, with 45 per cent in the lowest category of no postsecondary qualification. Dividing the sample by gender reveals 16.4 per cent of men have a degree compared to 12 per cent of females. Interestingly within the sample of those who are observed to start smoking by the survey date, only 10.4 (6.9) per cent of men (women) have degrees.

Figure 3 illustrates the hazard for quitting smoking.[10] As suggested by previous research, the hazard has an upward trend, although it is relatively flat in the current context. The hazard suggests that once having started to smoke, escape from the addiction tends not to occur quickly. A similar overall pattern is reported by Douglas (1998). The empirical analysis of quit behaviour involves a total sample of 4 946 subjects, approximately 37 per cent of whom quit, with the remainder treated as right-censored. The

[10] The overall shape is similar across gender.

sample is divided into 2 618 men (i.e. 53 per cent) and 2 328 females. The confidence interval around the hazard is extremely large towards the upper end of the distribution. This is to be expected, as the observed number of quits in this range is small. For the restricted sample of observed quitters, the average age of starting to smoke is just under 17, with the associated age of quitting being 25.1. Thus the average number of periods spent within the analysis sample is just under 8 years. Approximately, half of this sample is male.

Given that the earliest available cigarette price data are for 1963, the upper age range of our sample is set at 37. Unfortunately, this means that our empirical analysis of quits is somewhat limited. Many of the quits will occur later in life. On the other hand the 27–37 sample age restriction captures most of the smoking initiation behaviour. This is not because most people start smoking between these years but rather prior to age 27. Given the historical cigarette price data and information on age of initiation we can capture the essence of the starting to smoke decision. Given the relative strengths of the data, our analysis focuses more on the initiation of smoking than quitting.

One final issue to be briefly discussed is the potential recall bias associated with the data. The data set is a single cross-section and thus relies on recall of dates of starting and quitting smoking. A straightforward method of examining the importance of recall bias is to plot the hazard rate of starting/ quitting smoking in terms of calendar time. This allows one to identify bunching of dates around five year periods; that is dates ending in a zero or five. Our analysis suggests that this is not a significant issue in the current context.

FIGURE 3
Hazard of Quitting Smoking
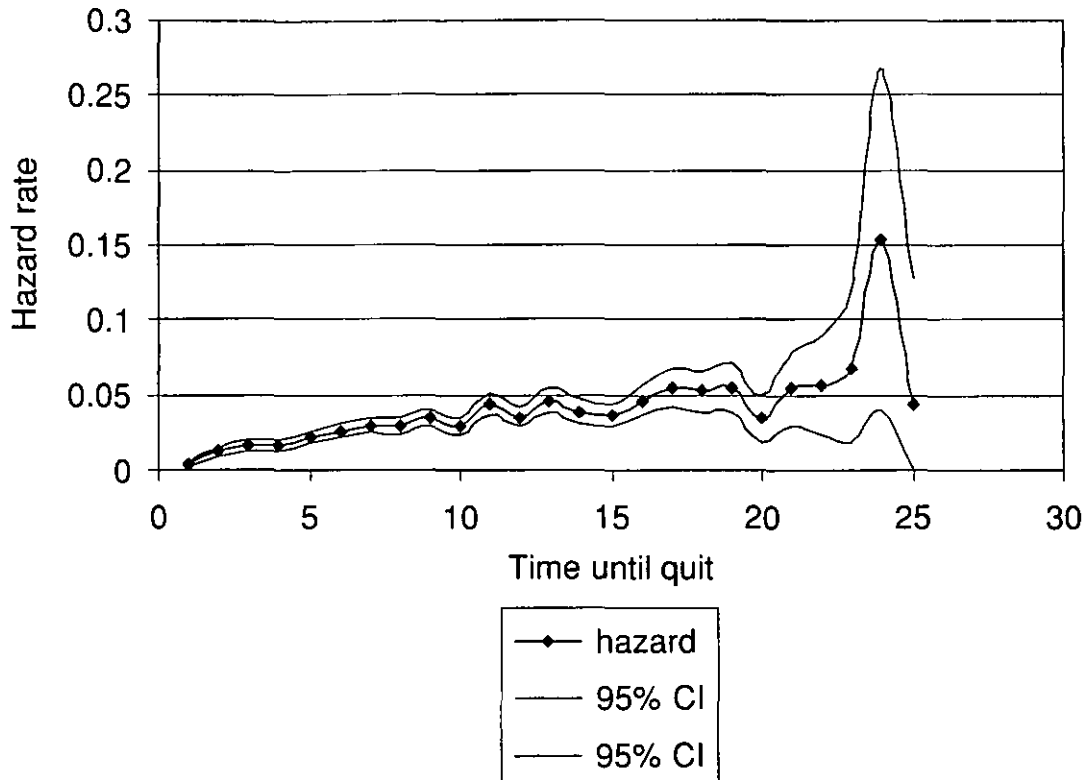
IV Empirical implementation

We now turn to a discussion of appropriate econometric methods and implementation. An examination of the hazard for the age of starting to smoke (Figure 2) clearly reveals that a non-monotonic function is required to adequately capture the shape. Within the parametric family, either the log-logistic or the log-normal appear to be obvious choices. Results generally differ little between the two specifications and thus following Douglas and Hariharan (1994) we report results for the former.

Figure 4, and in particular the function labelled 'life', represents the Kaplan–Meier survivor function associated with the age of starting to smoke.[11] The crucial aspect of the shape is that by age 37, approximately one half of the sample has not yet started

to smoke.[12] This observed pattern suggests that the standard hazard model is inappropriate as it implicitly assumes that all individuals will eventually fail (i.e. start smoking). We therefore follow Douglas and Hariharan's suggestion of estimating a split population hazard model.

Possibly the most well-known application of this method is the paper by Schmidt and Witte (1989) which examines criminal recidivism. Rather than assume all released prisoners will eventually re-offend, Schmidt and Witte treat the probability of eventual recidivism as an additional parameter to estimate. The distribution of time to the re-offence is assumed to apply only to those individuals who will eventually re-offend. The analogy with the current

[11] Please ignore the other curves for now as these are discussed later.

[12] The precise proportions are given in the summary statistics. These are 46 per cent in total and, 42 and 51 per cent for males/females, respectively.

FIGURE 4
Survivor Functions
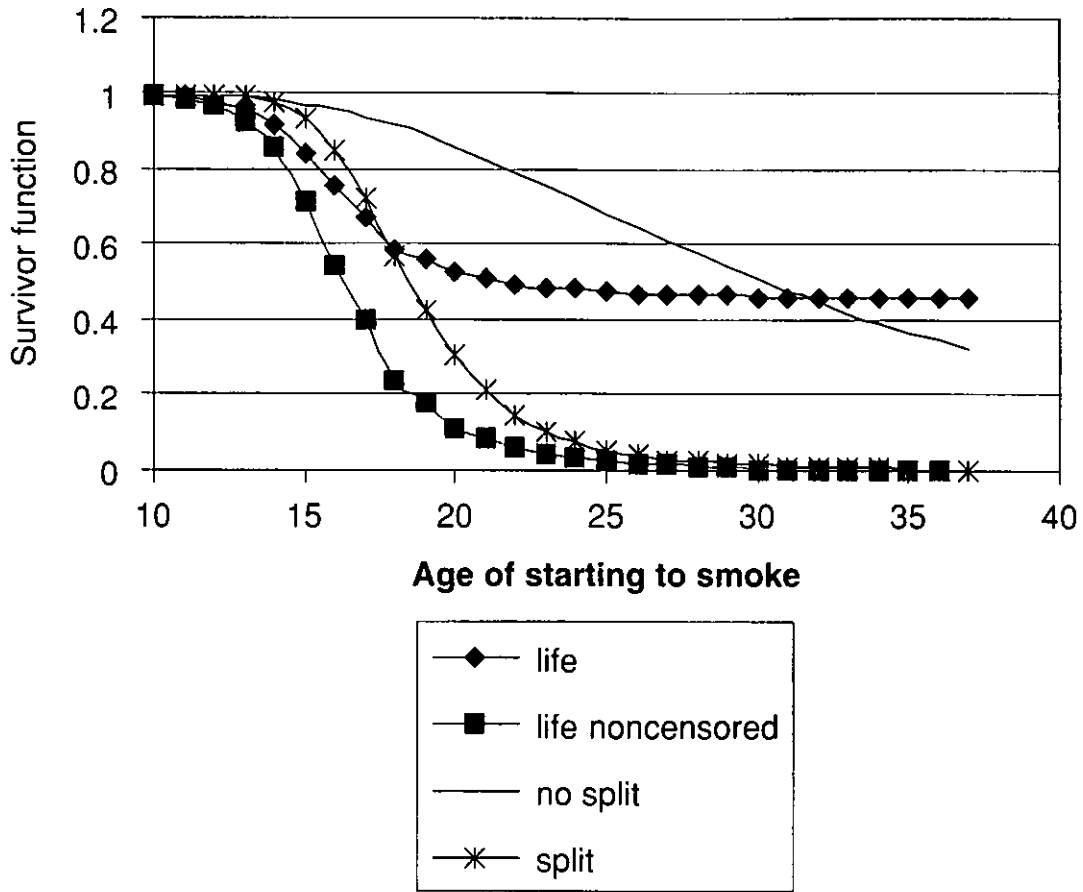
application is straightforward: time to failure is represented by the age at which smoking starts. As in the original Schmidt and Witte paper, if the individual has not exited (i.e. re-offended or in our context started to smoke) by the time of the survey they are assumed to be right-censored.

The split population analysis is based upon an unobserved qualitative variable D_i representing whether or not the individual i will *eventually* exit, that is, start to smoke, together with a parametric distribution for the time to exit, conditional upon eventual exit. We model the probability of individual i eventually exiting via the cumulative normal distribution function $\Phi(z_i, \theta)$, where z_i is a set of time independent regressors and θ an associated vector of parameters. As noted above, the qualitative variable

D_i is not observed. However, an associated variable δ_i represents whether or not the individual has ever smoked as at the survey date, where δ_i equals one and zero, respectively. Clearly if δ_i equals one then D_i equals 1, if δ_i equals zero, the individual might eventually begin smoking ($D_i = 1$) or not ($D_i = 0$). Conditional upon eventual exit, we model the duration of time to starting to smoke via the log-logistic distribution. Define $S(t; x_i(t), \beta)$ as the survivor function at period t, $x_i(t)$ a set of explanatory variables including time varying variables and β an associated vector of parameters. The survivor function provides the probability of the spell lasting at least until period t. Finally for individual i who started smoking at time T_i define the associated density function $f(T_i; x_i(t)$ (β). Following Douglas and Hariharan (1994)

the i'th individual's contribution to the likelihood function for the split population model is given as:

$$L_i(\theta, \beta; T_i, z_i, x_i) = \delta_i\Phi(z_i, \theta)f(T_i; x_i, \beta)$$
$$+ (1 - \delta_i)[(1 - \Phi(z_i, \theta)) + \Phi(z_i,\theta)S(t; x_i, \beta)] \quad (1)$$

If individual i is a current or ex-smoker at the time of the interview ($\delta_i = 1$), then the individual's contribution to the likelihood function is equal to the product of the probability of *ever* being a smoker, $\Phi(z_i, \theta)$ multiplied by the value of the density function, $f(T_i, x_i, \beta)$ defined at the time period the individual started. If the individual i is not yet a smoker at the time of the interview ($\delta_i = 0$), the individual's contribution is equal to the probability of never being a smoker, $1 - \Phi(z_i, \theta)$ plus the product of the probability of eventually failing and the survivor function, $\Phi(z_i, \theta)S(t; x_i, \beta)$.

Explanatory variables incorporated into the qualitative dimension and the conditional duration model components of the split population specification are represented by the vector of characteristics, z_i, and x_i, respectively, where the latter may be a function of time.[13] The model specification allows one to differentiate between factors which influence the probability of whether or not an individual will ever start smoking and factors which influence the timing or the age at which smoking begins.

Following the literature, our main focus is on the role of price as a determinant of smoking behaviour. Douglas and Hariharan (1994) in their analysis define price to be the price of tobacco as at age 18, where 18 is considered to be a crucial year in terms of the hazard of starting to smoke. Thus, their price variable varies across individuals, because of age and region of residence.[14] In this setting, price is fixed for a given individual and thus can be included in both parts of the regression model, that is the qualitative dimension representing whether the individual will ever begin smoking or not, and the hazard component capturing the age smoking commences. The current study follows the more recent work by Douglas (1998) and Forster and Jones (2001; 2003) which treat price as a time varying covariate. This approach is preferred as it allows one to track the relevant price at each calendar year that the individual is exposed to the possibility of starting to smoke. Given the time varying nature of the price variable, it can only be incorporated into the

hazard part of the specification. We introduce price in logarithmic terms so that the coefficient can be interpreted directly as an elasticity. Given the crucial nature of the price variable and the fact that it is treated as a time varying covariate, it is important to control for other time effects. Following Forster and Jones, we include a set of variables capturing time in a flexible polynomial, that is a quartic in the number of years since 1963, the first calendar year of the analysis.

All the other variables are non-time varying and are thus included in both parts of the split population specification. We restrict inclusion to variables suggested by the existing literature which can be treated as exogenous to the smoking initiation decision. Previous studies point to the role of gender and race as important demographic characteristics. To this end a dummy variable representing male is included. Unfortunately, the data set does not include information on race and thus we substitute a variable capturing whether or not the individual was born in Australia. Finally, we also include a set of dummy variables capturing education. Given our restriction on sample age, an individual's education level can reasonably be considered to reflect completed educational attainment.

We now turn to a discussion of the econometric specification of the quit hazard. The shape of the hazard illustrated in Figure 3 is monotonic. Douglas (1998) adopts the weibull parametric form, whereas Forster and Jones (2001; 2003) estimate the more flexible gamma specification. The latter is attractive as it has the weibull and log-normal as special cases. To be more specific, Douglas estimates a split population model combining the start and quit qualitative dimensions into an ordered probit model and the associated duration specifications are log-logistic and weibull, respectively. Douglas argues that a split population framework is once again appropriate given the large proportion of smokers who do not quit prior to the survey date. Forster and Jones employ the equivalent argument in the context of starting to smoke but ignore it in the context of the quit hazard. We report results for the weibull specification in both a non-split and split setting and gamma results in the former.[15] In the context of the gamma results we test to see if either the weibull or log-normal hazards are appropriate.

---

[13] An identical vector of variables x, is also included in the quit hazard specification.

[14] As in the current analysis, the region of residence as at the survey date is assumed to match that at age 18.

[15] Combining a gamma hazard with a split population model is beyond the scope of the current analysis. The results across specification appear extremely robust.

TABLE 2a
Non-Split Population Hazard Results: Age at Initiation of Smoking

Duration	Male/female combined		Male		Female	
	Coefficient	T-statistic	Coefficient	T-statistic	Coefficient	T-statistic
Log-tobacco-price	0.062	0.820	0.094	0.940	0.068	0.590
yrtime1	0.155	3.570	0.087	1.580	0.228	3.070
yrtime2	−0.038	−6.560	−0.024	−3.290	−0.055	−5.710
yrtime3	0.003	8.800	0.002	4.680	0.004	7.520
yrtime4	0.000	−9.230	0.000	−4.910	0.000	−7.990
Australian	−0.064	−3.850	−0.007	−0.340	−0.140	−5.270
degree	0.463	18.770	0.461	15.220	0.475	11.470
trade	0.036	1.680	0.048	2.170	−0.005	−0.080
diploma	0.129	7.170	0.151	5.420	0.125	5.050
Other-education	−0.012	−0.180	0.032	0.330	−0.064	−0.640
male	−0.172	−11.040				
constant	2.399	10.170	2.074	6.720	2.589	6.870
LogLikelihood	−19 415.700		−10 050.300		−9 325.700	

yrtime1–4 is a quartic time trend.
Control group for all regressions: Non-Australian born, female with no postschool education.

V Econometric results

(i) Age of initiation of smoking habit

We begin by estimating a log-logistic parametric hazard specification capturing the distribution of time until initiation of smoking. The results are in Table 2a. The qualitative impact of a variable is interpreted as follows. A positive coefficient sign implies the variable increases the age of starting to smoke. The first column of estimates represents results for the combined male and female sample. *Ceteris paribus*, being male significantly decreases the age of starting to smoke. Similarly, being Australian born rather than an immigrant also significantly decreases the age of onset of smoking. The education dummies are all measured relative to the default of the lowest level of no postsecondary qualifications. The coefficient on the highest education category of degree (or higher) is positive and highly significant. This suggests *ceteris paribus*, that more educated individuals have a later starting date. This accords with *a priori* expectations. The literature interprets high levels of education as indicative of a lower individual discount rate thus placing greater emphasis on future health prospects. Finally, and of most interest, is the fact that the coefficient on the log of tobacco price is positive but insignificantly different from zero.

Given the statistical significance of the gender coefficient and the findings of the previous literature, we decided to split the sample by gender. The results for each group are provided in the remaining columns of Table 2a. A likelihood ratio test of the null hypothesis that all male and female slope coefficients are equal is rejected with a Chi-squared statistic of 79.3. Generally speaking, the pattern of coefficients across the two groups is qualitatively similar in terms of sign and significance. However, minor gender differences do exist; the Australian born coefficient is negative and significant (not significant) for females (males). The trade education variable is positive (negative) and significant (not significant) for men (women). As in the combined specification, log price remains insignificant in both the male and female specifications.

Figure 4 illustrates the survivor function where 'Life' represents the Kaplan-Meier plot for the combined censored and uncensored sample. The equivalent plot for the sample restricted to the non-censored subjects, that is those with an observed start date for smoking, is labelled life-non-censored. The non-split survivor function is derived from the pooled male-female hazard results reported in Table 2a. Given that the non-split model is assumed to apply to all individuals whether censored or not, the estimated survivor function is rather different to the 'Life' plot in the figure. This suggests that the specification is inappropriate.

Following Douglas and Hariharan (1994), we extend the simple hazard model to incorporate a split population specification. The results are included in Table 2b. Once again we show a set of estimates for

TABLE 2b
Split Population Hazard Results: Age at Initiation of Smoking

Duration	Male/female combined		Male		Female	
	Coefficient	T-statistic	Coefficient	T-statistic	Coefficient	T-statistic
Log-tobacco-price	0.133	2.750	0.162	2.320	0.122	1.830
yrtime1	0.332	13.380	0.259	7.310	0.411	11.180
yrtime2	−0.048	−14.360	−0.038	−7.930	−0.059	−12.140
yrtime3	0.003	15.240	0.002	8.580	0.003	12.780
yrtime4	−0.043	−14.500	−0.035	−8.200	−0.052	−12.230
Australian	−0.051	−4.200	−0.077	−4.640	−0.014	−0.790
degree	0.222	11.400	0.230	8.780	0.214	7.460
trade	0.019	1.310	0.020	1.210	0.049	1.210
diploma	0.072	5.630	0.068	3.190	0.081	5.210
Other-eduation	−0.015	−0.350	−0.044	−0.640	0.005	0.090
male	−0.052	−4.790				
constant	1.015	6.980	1.060	5.030	0.869	4.270
Probit						
Australian	0.040	1.170	−0.148	−2.850	0.213	4.640
degree	−0.584	−12.000	−0.622	−9.060	−0.563	−8.030
trade	−0.045	−1.010	−0.079	−1.470	0.085	0.720
diploma	−0.165	−4.490	−0.235	−3.610	−0.129	−2.900
Other-education	0.023	0.170	−0.096	−0.450	0.100	0.560
male	0.249	7.790				
constant	0.193	5.190	0.648	11.770	0.011	0.250
Shape						
constant	0.197	69.510	0.207	50.380	0.185	48.080
LogLikelihood	−18 474.854		−9 628.989		−8 795.094	
Chi squared	2 159.300		1 225.960		986.460	

males and females combined, together with separate results by gender. The top half of the table illustrates the duration or hazard results, that is the time until smoking initiation. These results are comparable to those reported in Table 2a but now they are assumed to refer only to those individuals who will eventually fail, that is, start to smoke. The bottom half of the table refers to the probit model capturing the tendency for an individual to ever start smoking.

The first issue is to assess the relative performance of the split versus the non-split specification. We begin by focusing on the pooled male-female specification. The split model collapses to the non-split model under the assumption that $\Phi(z_i, \theta) = 1$, implying that all individuals will eventually fail. Schmidt and Witte (1989) point out that the likelihood ratio test statistic of this restriction is not in this case chi-squared distributed. However, they suggest that a test statistic of 100 implies a resounding rejection of the null hypothesis that $\Phi(z_i, \theta) = 1$. In our case the associated test statistic is 1881 and thus we can safely reject the non-split specification.

An alternative method of evaluating the split versus the non-split model is to examine the associated survivor functions illustrated in Figure 4. As pointed out earlier, the Kaplan–Meier survivor function (Life) and the non-split survivor function for the complete sample (no split) differ widely. In the split model, the survivor function is assumed to apply only to the subset of individuals who will eventually start smoking. Of course this subset is not observable, but it can be approximated by the set of individuals who started smoking prior to the survey date. The Kaplan–Meier survivor function for the subset of subjects observed to start smoking is illustrated in Figure 4 and labelled 'Life non-censored'. The survivor function associated with the split population model drawn at the sample mean vector of characteristics for the uncensored sample (split) is also included. Clearly the overall shape is relatively close to the Kaplan–Meier profile for the non-censored sample. Thus, this once again favours the split model.

Restricting the sample to uncensored observations only, that is, individuals who are observed to begin

smoking prior to the survey date, provides a final check on the split population model. Most individuals who will ever start smoking will already have done so by age 27–37. Thus, estimating a traditional non-split hazard model on the subsample of individuals who are observed to start smoking should lead to similar results for a split-model for the combined uncensored/censored samples. As discussed briefly below, the results are in fact similar.

The first column of results in Table 2b is for the pooled sample of males and females combined. Comparing the top half of the table with Table 2a reveals similar qualitative results in terms of both sign and significance. In general, however, the size and level of significance of coefficients in the split model tend to be larger. In particular and of central importance is the price coefficient, which rises in magnitude and significance in the split model. Thus, price now plays a significant role. A rise in price leads to an increase in the age of initiation of the smoking habit.

The probit results are included in the bottom half of the table. These results model the tendency to eventually begin smoking. The percentage of the total combined sample predicted to eventually start smoking is 58.4. The percentage is 64.4 for men and 52.7 for females. These compare with 53.7, 58.5 and 49.1, respectively, for the actual proportions observed to have started by the survey date. This suggests that a very small percentage of individuals who have not started smoking by the survey date will ever start smoking. Of course, the eventual failure rates are not observed, but one informative check on results is available. Assume that if the probability of eventual failure for a given subject is greater than or equal to 0.5, they are predicted to be an eventual failure, and not otherwise. A check on results is provided by examining what proportion of those who are observed to fail are not predicted to fail given the above criterion. The model appears to work rather well with only 4.7 per cent of the sample being mislabelled, 3.8 and 7.0 per cent of males and females, respectively.

Although the probit and duration components of the model are estimated simultaneously, variables are not constrained to have concordant effects. The results in general, however, do appear to be in harmony, thus just as having a degree is predicted to delay the onset of smoking, it also decreases the probability of ever starting to smoke. Similarly, being male not only decreases the age of starting but also increases the probability of ever starting to smoke.

The final two columns of Table 2b split the sample by gender. Qualitatively, the male-female results are similar to those reported in the first column.

A likelihood ratio test that the male-female specifications might be pooled is rejected, with a likelihood ratio test statistic of 101.4. The associated critical value at 5 per cent is 26.3. A likelihood ratio test of the split population versus non-split model rejects the latter in both the male and female specifications. In general, the male and female results are qualitatively similar. However, the role of the Australian dummy differs between the genders in both components of the split population model. In the duration component, it is significant (negative) for men, but insignificant for women. In the probit, it is significant for both genders but of opposite sign. Being Australian decreases the probability of ever starting to smoke for men and increases it for women.

Of most interest to the current study is the log-price coefficient. In this case, the male coefficient is larger than the associated female coefficient, and is significantly different from zero at the 5 per cent level. The female coefficient on the other hand is not significantly different from zero at 5 per cent but is significant at the 10 per cent level. The price coefficient can be interpreted as a price elasticity of starting to smoke. The results suggest that the male elasticity is 0.16 and 0.12 for females. Forster and Jones (2001; 2003) use a similar methodology to that adopted here, defining the price variable as a time varying covariate, but adopting the log of the tax on tobacco rather than actual price as in the current study. However, Forster and Jones cite a study by Chaloupka and Wechsler (1997) which suggests that elasticities based on tax and price produce similar results. Forster and Jones (2001; 2003) estimate much smaller tax elasticities of 0.006 for men and −0.028 for women where both are insignificant at the 10 per cent level. Thus our results contrast with Forster and Jones (2001; 2003), Douglas (1998) and Douglas and Hariharan (1994) who all report an insignificant role for price.

(ii) Sensitivity analysis

The sensitivity of the overall regression results is explored in a number of distinct ways. First as a check on the split-population results, we estimate a non-split model on the subsample of uncensored observations, that is those individuals observed to start smoking prior to the survey date, 1990. The sample of interest is restricted to those aged 27–37 and thus the vast majority of individuals who have not started smoking by that time are unlikely to ever start. This implies that the regression results for the split model on the entire sample and the non-split model estimated on the uncensored sample should be similar. Focusing on the key parameter of interest,

namely the coefficient on the log price variable, recall the split-population results for the male-female pooled model is 0.133; and 0.162 and 0.122 for the separate male and female specifications. The corresponding results for the non-split censored model are 0.02, 0.39 and 0.079. Thus overall the results are reasonably close.[16]

The next check on robustness is to examine whether the results are sensitive to the age group selected. Our original sample included those aged 27–37 in 1990. Note that this is not intrinsically restrictive given we are interested in modelling age at starting to smoke and the majority of individuals start prior to age 27. However, if we consider ages 15–20 as representative of the highest risk of starting to smoke, our age restriction implies a focus in terms of calendar time of 1968–83. Figure 1 illustrates the trend in tobacco price and suggests the index is relatively stable over this calendar period.

In an attempt to examine the sensitivity of results by age group we replicate the results for the younger age group 18–26.[17] This age group in 1990 falls within the critical 15–20 age group from 1979 onwards. This is clearly a period of faster change in the tobacco index price. As suggested earlier, the major problem associated with analysis of this younger age group is we do not observe completed education level, a key determinant of smoking behaviour. However, ignoring this issue (in other words, we assume that the observed and final completed education levels are the same for all), re-estimating the model reveals that the major parameter of interest, that is the tobacco price effect falls slightly for men from 0.16 to 0.11 and increases slightly for women from 0.12 to 0.14. The levels of statistical significance are also affected, with the male coefficient becoming insignificant (t-value of 1.25), and the female coefficient maintaining significance at the 10 per cent level. Thus overall, the results for women are similar across age groups 18–26 and 27–37, but for men price plays a significant role for those aged 27–37 but not for those aged 18–26.[18]

The switch between significance and insignificance of the price coefficient for men aged 27–37 and 18–26 is somewhat counter-intuitive given the trend in the tobacco price index. However, Douglas (1998) cites several studies which report conflicting conclusions relating to the relative price sensitivity of youth and adult smoking behaviour. Several studies suggest greater price responsiveness of adults, whereas others reach the opposite conclusion. One possible explanation for relative price insensitivity of youth is the role played by peer pressure.[19]

To further explore the sensitivity of the results, we next replicate the analysis using the alternative data source of the NDSHS (Australian Institute of Health and Welfare, 1999). Recall that although the data source is more recent its major disadvantage is the small sample frame. Given this, our focus is on examining the sensitivity of our original results rather than deriving new results. We begin by attempting to replicate the 27–37 age group in the 1990 sample using the 1998 data source. This involves selecting the 35–45 age group in 1998 and truncating their exposure at 1990.[20] In this case the price coefficient for males falls to 0.06 and once again is insignificantly different from zero at the 10 per cent level (t-statistic of 0.4 in fact). For females the estimated price coefficient rises to 0.29 and is significant at the 5 per cent level (t-statistic of 2.25). To explore the sensitivity of truncating the risk period at 1990, we allow the period to run through to 1998. Intuitively one would not expect this to have a significant impact as the number of individuals switching between being censored, that is not having started smoking at 1990 who then switch to being uncensored by 1998, is very small. The male price effect is once again totally insignificant, the female coefficient drops slightly to 0.23 with a t-statistic of 1.75, thus a very similar probability value of 7.9 per cent. Finally we use the 1998 data to also replicate results for the equivalent of the 18–26 age group sample in 1990.

[16] By way of comparison Forster and Jones (2001, 2003) report male/female results for the split model of 0.006 and −0.028. The non-split restricted sample results are −0.021 and −0.034, respectively.

[17] The total number of observations for this age group is 70 112, with 34 344 males and 35 768 females. This represents a total of 6 971 subjects, 3 401 of whom start to smoke.

[18] A likelihood ratio test of whether the model specification can be pooled across age groups is resoundingly rejected, with a test statistic of 294.1 for males and 371.8 for females.

[19] An alternative explanation for the insignificance of price for the younger age group is a loss of precision associated with the smaller sample size (i.e. 25 per cent smaller than the 27–37 age group). Also as suggested earlier, the overall fit of the model for the younger age group is somewhat compromised by the inaccuracy of the completed education variable.

[20] The total number of observations is 31 400 with 13 109 men and 18 291 women. This corresponds to a total of 2 107 subjects of whom 1082 begin to smoke. As suggested in the text, this sample is much smaller than the equivalent sample in the NHS (1990), at approximately one third the size.

TABLE 3

Hazard Results: Time Until Quit-Pooled Male-Female Specification

Duration	Weibull		Gamma		Weibull Split	
	Coefficient	T-statistic	Coefficient	T-statistic	Coefficient	T-statistic
Log-tobacco-price	0.199	0.930	0.245	1.210	0.172	0.940
yrtime1	−0.373	1.350	−0.190	−0.520	−0.305	−0.630
yrtime2	0.025	−1.030	0.012	0.350	0.021	0.480
yrtime3	−0.001	0.740	0.000	−0.240	−0.001	−0.390
yrtime4 *1000	0.009	−0.620	0.001	0.070	0.000	0.290
Australian	−0.027	−1.300	−0.042	−1.000	0.015	0.330
degree	−0.504	−9.910	−0.576	−9.050	−0.422	−6.730
trade	−0.136	−3.010	−0.147	−2.740	−0.113	−2.290
diploma	−0.261	−6.890	−0.290	−6.370	−0.201	−3.520
Other-education	−0.045	−0.150	−0.054	−0.320	−0.283	−1.610
male	0.149	3.240	0.180	4.530	0.459	5.740
constant	5.096	2.170	3.899	2.820	4.329	2.300
Probit						
Australian					0.231	1.580
degree					0.508	1.990
trade					0.044	0.160
diploma					0.281	1.690
Other-education					−0.637	−1.820
male					4.241	0.180
constant					0.232	1.220
Ancilliary parameters						
scale	1.519				1.606	28.340
k			0.315	3.190		
Loglikelihood	−4 021.516		−4 001.406		−8 016.790	

This involved selecting the sample of individuals aged 26–34 in 1998.[21] In this case there is no evidence of a significant role for price of cigarettes as a determinant of age of starting to smoke.

In summary, tobacco price appears to be a significant determinant of smoking for 27–37 year olds in 1990. This holds for both males and females. However, sensitivity analysis across age group and data source sheds doubt on the robustness of this finding. The lack of a significant role accords with the recent UK empirical literature and some of the more recent US studies.

(iii) The quit decision

We now turn, relatively briefly, to a discussion of the quit hazard results which are contained in

[21] The total number of observations is 33 602, corresponding to 2 331 subjects of whom 1 198 begin to smoke. Once again, this is a relatively small sample, approximately one third of the equivalent sample size in the NHS (1990).

Table 3. The first two columns of results are for the pooled male-female non-split model with weibull and gamma hazard specifications, respectively. Initially, we tested whether male-female specifications could legitimately be pooled. The null hypothesis of equal slope coefficients was not rejected at the 5 per cent level. In the weibull specification, the likelihood ratio test statistic was 17.1 and 14.5 in the gamma specification, compared to a critical value of 18.3 at the 5 per cent level. Given the test results and in order to conserve space, we include only the pooled results. However, the separate male-female results will be briefly mentioned.

The ancillary parameter k for the gamma specification provides a direct test of the nested log-normal ($k = 0$) and weibull ($k = 1$) specifications against the more general alternative of gamma. A Wald test of the restriction $k = 0$ leads to a test statistic of 10.2, the critical value at 1 per cent significance is 6.63 and thus the log-normal can be rejected. The equivalent test of the weibull, leads to a test statistic of

48.2 and thus once again rejection.[22] Although the gamma specification is preferred in a statistical sense, qualitatively the results are similar to those of the weibull. In both cases the likelihood ratio test of all slope coefficients simultaneously equal to zero is resoundingly rejected, with a test statistic of 199.1 for the weibull and 224.9 for the gamma specification. Despite this result, few of the regressors are individually significant.

In terms of interpretation, a positive estimated coefficient should be interpreted as the associated variable increasing the time to quit. The overall pattern of results is similar to the age of starting to smoke results. Thus, being male significantly increases the time to quit. Higher educational attainment decreases the time to quit. Once again of greatest interest is the log price coefficient. In both specifications, the coefficient is insignificant with a probability value of 0.294 in the weibull and 0.224 in the gamma specification.[23]

The final column of Table 3 contains the results from estimating the split population weibull specification over the pooled male and female samples. The results provide further sensitivity analysis of the impact of the price of tobacco on quit behaviour. Initially we estimated separate models for males and females but did not reject the null hypothesis of equal slope coefficients at conventional significance levels.[24] Formally the split population specification is preferred to the non-split model, the likelihood ratio test statistic is 116.8. However, more importantly from a sensitivity perspective, the results are very similar to the non-split model. The log-price tobacco coefficient is insignificant with a probability value of 0.35.

Of most interest to the current study is the role of price in the timing of the quit decision. The results are consistent across gender and specification, it appears that price plays no significant role in the time to quit decision.[25] This result contrasts with that

---

[22] We also carried out equivalent tests in the context of the male-female specifications. The log-normal restriction was rejected at the 5 per cent level for both males and females with a test statistic of 12.7 and 3.92, respectively. The weibull specification is also rejected for both males and females with a test statistic of 17.5 and 37.9, respectively.

[23] In the separate male-female specifications the prob. values for the weibull are 0.151 and 0.917, and 0.191 and 0.588 for the gamma specification.

[24] The likelihood ratio statistic was 4.32 for the weibull and 12.56 for the gamma specification.

[25] Note that although significance does not vary across specification, the estimated coefficient is much larger in the weibull compared to the gamma specification for females.

reported by Forster and Jones (2001; 2003). In their analysis, they report a negative and significant impact of the log of tax on the timing of the quit decision for men, and negative but insignificant for women. Douglas (1998) reports no significant impact of price on either the decision to start or quit smoking.

· VI Conclusions

The empirical analysis employed a sample of individuals aged 27–37 drawn from the NHS (1990) data set. In this restricted sample, we find evidence that the price of tobacco plays a significant role in an individual's decision to start smoking but not in their decision to quit. However, further sensitivity analysis of different age groups and an alternative data source calls into question the robustness of the role of price in the smoking initiation decision. This questionable robustness of the price variable in our analysis and in a number of overseas studies means that the government cannot rely solely on price to combat or reduce smoking propensities. The smoking rates of teenagers continue to be a public policy issue of considerable concern. The social costs of smoking are sufficiently high to warrant ongoing exploration of the role of price in the decision to smoke. We have taken an important step in this paper in analysing this issue using available Australian data sets. Econometric exploration of future surveys of smoking behaviour may help to clarify the role of price.

REFERENCES

Australian Bureau of Statistics (1990), *National Health Survey 1989–90*. Australian Bureau of Statistics, Canberra.

Australian Bureau of Statistics (2003), *Australian Consumer Price Index: Concepts, Sources and Methods 2003*, cat no. 6461.0. Australian Bureau of Statistics, Canberra.

Australia Institute of Health and Welfare (1999), *National Drug Strategy Household Survey (1998)*, computer file. The Australian National University, Canberra.

Bardsley, P. and Olekalns, N. (1999), 'Cigarette and Tobacco Consumption: Have Anti-Smoking Policies Made a Difference?', *Economic Record* 75, 225–40.

Becker, G.S., and Murphy, K.M. (1988), 'A Theory of Rational Addiction', *Journal of Political Economy*, August 96, 675–700.

Chaloupka, F.J. and Warner, K.E. (2000), 'The Economics of Smoking', in: J.P. Newhouse and A.J. Culyer (eds.) *Handbook of Health Economics*, Vol. 1B, North Holland, Amsterdam.

Collins, D.J. and Lapsley, H.M. (1996), 'The Social Costs of Drug Abuse in Australia 1988 and 1992'. Monograph

Series no. 30. Australian Government Publishing Service, *Canberra*.

DeCicca, P., Kenkel, D., and Mathios, A. (2002), 'Putting Out the Fires: Will Higher Taxes Reduce the Onset of Youth Smoking?', *Journal of Political Economy* **110**, 144–69.

Douglas, S. (1998), 'The Duration of the Smoking Habit', *Economic Inquiry* **23**, 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**, 213–30.

Forster, M. and Jones, A. (2001), 'The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data', *Journal of the Royal Statistical Society, Series A, Part 3.* **164**, 517–47.

Forster, M. and Jones, A. (2003), 'The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data', *Journal of the Royal Statistical Society, Series A*. Corrigendum. forthcoming.

Quit Victoria (2003), 'Smoking Rates'. Available [online]: http://www.quit.org.au/.

Ross, H. and Chaloupka, F.J. (2003), 'The Effect of Cigarette Prices on Youth Smoking', *Health Economics* **12**, 217–46.

Schmidt, P. and Witte, A.D. (1989), 'Predicting Criminal Recidivism using "Split Population" Survival Time Models', *Journal of Econometrics* **40**, 141–59.

Tauras, J.A. (1999), 'The Transition to Smoking Cessation: Evidence from Multiple Failure Duration Analysis', Working paper no. 7262, NBER, Cambridge, MA.

Tauras, J.A. and Chaloupka, F.J. (1999), 'Determinants of Smoking Cessation: An Analysis of Young Adult Men and Women', Working paper no. 7262, NBER, Cambridge, MA.